Biometrical Journal **52** (2020) 61, zzz–zzz / DOI: 10.1002/bimj.200100000

Supplementary material for "Sample size and power considerations for cluster randomized trials with count outcomes subject to right truncation"

Fan Li[*,1,2,3] and **Guangyu Tong** [1,3]

[1] Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA
[2] Center for Methods in Implementation and Prevention Science, Yale University, New Haven, CT, USA
[3] Yale Center for Analytical Sciences, New Haven, CT, USA

Appendix A: Derivation of equation (3) and (4)

Based on the conditional model (1) in the main text, we find that given $X_i = 1$ (intervention arm) and random effect $\phi_i = \phi$, the following conditional expectation is

$$\begin{aligned}\lambda_{(1)} =& \mathrm{E}(Y_{ij}|X_i = 1, \phi_i = \phi, 0 \leq Y_{ij} \leq T)\\ =& \sum_{y=0}^{T} yP(Y_{ij} = y|\lambda_{ij}, 0 \leq Y_{ij} \leq T) = \sum_{y=0}^{T} y\frac{\lambda_{ij}^y}{y!Q_T(\lambda_{ij})} = \sum_{y=1}^{T} \lambda_{ij}\frac{\lambda_{ij}^{y-1}}{(y-1)!Q_T(\lambda_{ij})}\\ =& \lambda_{ij}\sum_{t=0}^{T-1}\frac{\lambda_{ij}^t}{t!Q_T(\lambda_{ij})} = \lambda_{ij}\frac{Q_{T-1}(\lambda_{ij})}{Q_T(\lambda_{ij})} = \exp(\beta_0 + \beta_1 + \phi)\frac{Q_{T-1}(\exp(\beta_0 + \beta_1 + \phi))}{Q_T(\exp(\beta_0 + \beta_1 + \phi))},\end{aligned}$$

where $\lambda_{ij} = \exp(\beta_0 + \beta_1 + \phi)$ in the intervention arm by model assumption. For mathematical rigor, we allow $T = \infty$ to indicate no truncation. Similarly, given $X_i = 0$ and random effect $\psi_i = \psi$, we find the conditional expectation

$$\lambda_{(0)} = \mathrm{E}(Y_{ij}|X_i = 0, \psi_i = \psi, 0 \leq Y_{ij} \leq T) = \lambda_{ij}\frac{Q_{T-1}(\lambda_{ij})}{Q_T(\lambda_{ij})} = \exp(\beta_0 + \psi)\frac{Q_{T-1}(\exp(\beta_0 + \psi))}{Q_T(\exp(\beta_0 + \psi))},$$

where $\lambda_{ij} = \exp(\beta_0 + \psi)$ in the control arm by model assumption. Because $Q_t(\lambda) = \sum_{k=0}^{t} \lambda^k/k!$ is a non-decreasing function of t, we have $Q_T(\lambda) \geq Q_{T-1}(\lambda)$ for any $\lambda > 0$. Therefore

$$\lambda_{(1)} \leq \exp(\beta_0 + \beta_1 + \phi) \;, \lambda_{(0)} \leq \exp(\beta_0 + \psi),$$

where the equality holds only when $T = \infty$ (i.e., no truncation). In other words, the truncated means under a finite T are smaller than the untruncated means.

Appendix B: Conversion formulas

We provide a set of conversion formulas to connect the marginal model and conditional model parameters, when the outcome is subject to right truncation. Given the conditional model parameters and the truncation

*Corresponding author: Fan Li, Department of Biostatistics, Yale School of Public Health, e-mail: fan.f.li@yale.edu

point T, the marginal mean for the intervention and control clusters can be obtained as

$$\begin{aligned}\mu_{(1)} &= \int_{-\infty}^{\infty} \exp(\beta_0+\beta_1+\phi)\frac{Q_{T-1}(\exp(\beta_0+\beta_1+\phi))}{Q_T(\exp(\beta_0+\beta_1+\phi))}f(\phi)d\phi \\ &= \exp(\beta_0+\beta_1)\mathrm{E}_\phi\left\{\frac{\exp(\phi)Q_{T-1}(\exp(\beta_0+\beta_1+\phi))}{Q_T(\exp(\beta_0+\beta_1+\phi))}\right\} \qquad (B.1)\\ \mu_{(0)} &= \int_{-\infty}^{\infty} \exp(\beta_0+\psi)\frac{Q_{T-1}(\exp(\beta_0+\psi))}{Q_T(\exp(\beta_0+\psi))}g(\psi)d\psi \\ &= \exp(\beta_0)\mathrm{E}_\psi\left\{\frac{\exp(\psi)Q_{T-1}(\exp(\beta_0+\psi))}{Q_T(\exp(\beta_0+\psi))}\right\}, \qquad (B.2)\end{aligned}$$

where $f(\phi)$ and $g(\psi)$ are densities of $\mathrm{N}(0,\sigma_1^2)$ and $\mathrm{N}(0,\sigma_0^2)$. In the simulation study, our data generating process follows the conditional model (1) in the main text subject to right truncation. To apply the proposed sample size formulas within the marginal modeling framework, we have determined the marginal quantities through the following set of conversion formulas. The marginal RR is taken as $\Delta = \log\{\mu_{(1)}/\mu_{(0)}\}$, following equations (B.1) and (B.2). To compute the arm-specific marginal variance $\tau_{(0)}$ and $\tau_{(1)}$, we first obtain the conditional expectation of the second moment

$$\begin{aligned}\mathrm{E}(Y_{ij}^2|\lambda_{ij},0\le Y_{ij}\le T) &= \sum_{y=0}^{T} y^2\mathrm{P}(Y_{ij}=y|\lambda_{ij},0\le Y_{ij}\le T) = \sum_{y=0}^{T} y^2\frac{\lambda_{ij}^y}{y!Q_T(\lambda_{ij})} \\ &= \sum_{y=1}^{T}\lambda_{ij}(y-1)\frac{\lambda_{ij}^{y-1}}{(y-1)!Q_T(\lambda_{ij})} + \sum_{y=1}^{T}\lambda_{ij}\frac{\lambda_{ij}^{y-1}}{(y-1)!Q_T(\lambda_{ij})} \\ &= \sum_{y=2}^{T}\lambda_{ij}^2\frac{\lambda_{ij}^{y-2}}{(y-2)!Q_T(\lambda_{ij})} + \sum_{t=0}^{T-1}\lambda_{ij}\frac{\lambda_{ij}^{t}}{t!Q_T(\lambda_{ij})} \\ &= \lambda_{ij}^2\sum_{t=0}^{T-2}\frac{\lambda_{ij}^{t}}{t!Q_T(\lambda_{ij})} + \lambda_{ij}\sum_{t=0}^{T-1}\frac{\lambda_{ij}^{t}}{t!Q_T(\lambda_{ij})} \\ &= \lambda_{ij}^2\frac{Q_{T-2}(\lambda_{ij})}{Q_T(\lambda_{ij})} + \lambda_{ij}\frac{Q_{T-1}(\lambda_{ij})}{Q_T(\lambda_{ij})}.\end{aligned}$$

Marginalizing over the random effects, we obtain

$$\begin{aligned}\tau_{(1)} =& \mathrm{Var}(Y_{ij}|X_i=1,0\le Y_{ij}\le T) \\ =& \exp(2(\beta_0+\beta_1))\mathrm{E}_\phi\left\{\frac{\exp(2\phi)Q_{T-2}(\exp(\beta_0+\beta_1+\phi))}{Q_T(\exp(\beta_0+\beta_1+\phi))}\right\} \\ &+ \exp(\beta_0+\beta_1)\mathrm{E}_\phi\left\{\frac{\exp(\phi)Q_{T-1}(\exp(\beta_0+\beta_1+\phi))}{Q_T(\exp(\beta_0+\beta_1+\phi))}\right\} - \mu_{(1)}^2 \\ \tau_{(0)} =& \mathrm{Var}(Y_{ij}|X_i=0,0\le Y_{ij}\le T) \\ =& \exp(2\beta_0)\mathrm{E}_\psi\left\{\frac{\exp(2\psi)Q_{T-2}(\exp(\beta_0+\psi))}{Q_T(\exp(\beta_0+\psi))}\right\} + \exp(\beta_0)\mathrm{E}_\psi\left\{\frac{\exp(\psi)Q_{T-1}(\exp(\beta_0+\psi))}{Q_T(\exp(\beta_0+\psi))}\right\} - \mu_{(0)}^2.\end{aligned}$$

Finally, to obtain the arm-specific marginal ICC $\rho_{(0)}$ and $\rho_{(1)}$, we observe the conditional cross-moment

$$\mathrm{E}(Y_{ij}Y_{ik}|\lambda_{ij},\lambda_{ik},0\le Y_{ij}\le T,0\le Y_{ik}\le T) = \left\{\lambda_{ij}\frac{Q_{T-1}(\lambda_{ij})}{Q_T(\lambda_{ij})}\right\}^2,$$

due to conditional independence. Therefore, marginalizing over the random effects, we compute the arm-specific marginal cross-moment as

$$\varpi_{(1)} = \mathrm{E}(Y_{ij}Y_{ik}|X_i = 1, 0 \le Y_{ij} \le T) = \exp(2(\beta_0+\beta_1))\mathrm{E}_\phi\left[\frac{\exp(2\phi)\left\{Q_{T-1}(\exp(\beta_0+\beta_1+\phi))\right\}^2}{\left\{Q_T^2(\exp(\beta_0+\beta_1+\phi))\right\}^2}\right]$$

$$\varpi_{(0)} = \mathrm{E}(Y_{ij}Y_{ik}|X_i = 0, 0 \le Y_{ij} \le T) = \exp(2\beta_0)\mathrm{E}_\psi\left[\frac{\exp(2\psi)\left\{Q_{T-1}(\exp(\beta_0+\psi))\right\}^2}{\left\{Q_T(\exp(\beta_0+\psi))\right\}^2}\right].$$

This then gives the arm-specific ICC as $\rho_{(1)} = \{\varpi_{(1)} - \mu_{(1)}^2\}/\tau_{(1)}$ and $\rho_{(0)} = \{\varpi_{(0)} - \mu_{(0)}^2\}/\tau_{(0)}$. The expectations in all conversion formulas are respect to the Gaussian distribution of the random effects, and can be approximated either through numerical quadrature or Monte Carlo integration.

Appendix C: Derivation of variance expression under the independence working correlation

Because the independence GEE is written as

$$\sum_{i=1}^N \mathbf{D}_{1i}'\mathbf{V}_{1i}^{-1}(\mathbf{Y}_i-\boldsymbol{\mu}_i) = \sum_{i=1}^N \begin{pmatrix} 1 \\ X_i \end{pmatrix}\sum_{j=1}^{m_i}(Y_{ij}-\mu_{ij}) = \sum_{i=1}^N \begin{pmatrix} 1 \\ X_i \end{pmatrix}\sum_{j=1}^{m_i}(Y_{ij}-\exp(\gamma_0+\gamma_1X_i)) = \mathbf{0},$$

we can see that the solution $(\widehat{\gamma}_0, \widehat{\gamma}_1)'$ must satisfy

$$0 = \sum_{i=1}^N\sum_{j=1}^{m_i}(Y_{ij}-\exp(\widehat{\gamma}_0+\widehat{\gamma}_1X_i)) = \sum_{i=1}^N\sum_{j=1}^{m_i}X_i(Y_{ij}-\exp(\widehat{\gamma}_0+\widehat{\gamma}_1)) + \sum_{i=1}^N\sum_{j=1}^{m_i}(1-X_i)(Y_{ij}-\exp(\widehat{\gamma}_0))$$

$$0 = \sum_{i=1}^N\sum_{j=1}^{m_i}X_i(Y_{ij}-\exp(\widehat{\gamma}_0+\widehat{\gamma}_1X_i)) = \sum_{i=1}^N\sum_{j=1}^{m_i}X_i(Y_{ij}-\exp(\widehat{\gamma}_0+\widehat{\gamma}_1)). \quad \text{(C.1)}$$

The above system of equations directly imply

$$\exp(\hat{\gamma}_0) = \frac{\sum_{i=1}^N\sum_{j=1}^{m_i}(1-X_i)Y_{ij}}{\sum_{i=1}^N(1-X_i)m_i}. \quad \text{(C.2)}$$

Substituting this in (C.1), we have

$$\exp(\hat{\gamma}_0+\hat{\gamma}_1) = \frac{\sum_{i=1}^N\sum_{j=1}^{m_i}X_iY_{ij}}{\sum_{i=1}^N X_im_i}. \quad \text{(C.3)}$$

Combing equations (C.2) and (C.3) gives equation (10) in the main text.

We next provide detailed derivations of the variance expression. Because the independence working correlation differs from the true correlation matrix, the derivation of σ_{ind}^2 should be based on the sandwich variance. One should also note that the variance function of this independence GEE approach is misspecified under right truncation, but the sandwich variance remains robust to such misspecification. By definition, we have the true (inverse) of the sandwich variance $\mathbf{\Omega}_1^{-1}\mathbf{\Omega}_0\mathbf{\Omega}_1^{-1}$, where $\mathbf{\Omega}_1 = \sum_{i=1}^N \mathbf{D}_{1i}'\mathbf{V}_{1i}^{-1}\mathbf{D}_{1i}$

and $\mathbf{\Omega}_0 = \sum_{i=1}^N \mathbf{D}_{1i}'\mathbf{V}_{1i}^{-1}\text{Cov}(\mathbf{Y}_i)\mathbf{V}_{1i}^{-1}\mathbf{D}_{1i}$, where $\mathbf{V}_{1i} = \mathbf{A}_i^{1/2}\mathbf{R}_i\mathbf{A}_i^{1/2}$, with $\mathbf{R}_i = \mathbf{I}$, the identity matrix. First,

$$\begin{aligned}
\mathbf{\Omega}_1 &= \sum_{i=1}^N \mathbf{D}_{1i}'\mathbf{V}_{1i}^{-1}\mathbf{D}_{1i} = \sum_{i=1}^N \mathbf{D}_{1i}'\mathbf{A}_i^{-1/2}\mathbf{R}_i^{-1}\mathbf{A}_i^{-1/2}\mathbf{D}_{1i} \\
&= \sum_{i=1}^N \left\{ \begin{pmatrix} 1 \\ X_i \end{pmatrix} \otimes \mathbf{1}_{m_i}' \right\} \left\{ \left((1-X_i)\mu_{(0)} + X_i\mu_{(1)}\right)\mathbf{R}_i^{-1}\right\} \left\{ \begin{pmatrix} 1 & X_i \end{pmatrix} \otimes \mathbf{1}_{m_i}\right\} \\
&= \sum_{i=1}^N \begin{pmatrix} (1-X_i)\mu_{(0)} + X_i\mu_{(1)} & X_i\mu_{(1)} \\ X_i\mu_{(1)} & X_i\mu_{(1)} \end{pmatrix} \otimes \left(\mathbf{1}_{m_i}'\mathbf{R}_i^{-1}\mathbf{1}_{m_i}\right) \\
&= \sum_{i=1}^N \left(\mathbf{1}_{m_i}'\mathbf{R}_i^{-1}\mathbf{1}_{m_i}\right) \begin{pmatrix} (1-X_i)\mu_{(0)} + X_i\mu_{(1)} & X_i\mu_{(1)} \\ X_i\mu_{(1)} & X_i\mu_{(1)} \end{pmatrix},
\end{aligned}$$

where $\mu_{(0)} = \exp(\gamma_0)$, $\mu_{(1)} = \exp(\gamma_0 + \gamma_1)$, $\mathbf{1}_s$ is the $s \times 1$ vector of one's, and the last identity is because $\left(\mathbf{1}_{m_i}'\mathbf{R}_i^{-1}\mathbf{1}_{m_i}\right)$ is a scalar. Because the treatment is randomized at the cluster level, we have $N^{-1}\sum_{i=1}^N X_i \xrightarrow{p} \pi$. We then have

$$N^{-1}\mathbf{\Omega}_1 \xrightarrow{p} \left\{\text{E}\left(\mathbf{1}_{m_i}'\mathbf{R}_i^{-1}\mathbf{1}_{m_i}\right)\right\} \begin{pmatrix} (1-\pi)\mu_{(0)} + \pi\mu_{(1)} & \pi\mu_{(1)} \\ \pi\mu_{(1)} & \pi\mu_{(1)} \end{pmatrix}.$$

Define

$$\mathbf{M} = \begin{pmatrix} (1-\pi)\mu_{(0)} + \pi\mu_{(1)} & \pi\mu_{(1)} \\ \pi\mu_{(1)} & \pi\mu_{(1)} \end{pmatrix},$$

then by the Continuous Mapping Theorem and 2×2 matrix inversion, we obtain

$$\begin{aligned}
N\mathbf{\Omega}_1^{-1} &\xrightarrow{p} \left\{\text{E}\left(\mathbf{1}_{m_i}'\mathbf{R}_i^{-1}\mathbf{1}_{m_i}\right)\right\}^{-1}\mathbf{M}^{-1} \\
&= \frac{\left\{\text{E}\left(\mathbf{1}_{m_i}'\mathbf{R}_i^{-1}\mathbf{1}_{m_i}\right)\right\}^{-1}}{(1-\pi)\mu_{(0)}} \begin{pmatrix} 1 & -1 \\ -1 & 1 + \{(1-\pi)/\pi\}(\mu_{(0)}/\mu_{(1)}) \end{pmatrix},
\end{aligned}$$

Next, we observe that

$$\begin{aligned}
\mathbf{\Omega}_0 &= \sum_{i=1}^N \mathbf{D}_{1i}'\mathbf{V}_{1i}^{-1}\text{Cov}(\mathbf{Y}_i)\mathbf{V}_{1i}^{-1}\mathbf{D}_{1i} \\
&= \sum_{i=1}^N \left\{ \begin{pmatrix} 1 \\ X_i \end{pmatrix} \otimes \mathbf{1}_{m_i}' \right\} \mathbf{R}_i^{-1} \left\{ \left((1-X_i)\tau_{(0)} + X_i\tau_{(1)}\right)\widetilde{\mathbf{R}}_i(\boldsymbol{\rho})\right\} \mathbf{R}_i^{-1} \left\{ \begin{pmatrix} 1 & X_i \end{pmatrix} \otimes \mathbf{1}_{m_i}\right\} \\
&= \sum_{i=1}^N (1-X_i)\tau_{(0)} \left\{\mathbf{1}_{m_i}'\mathbf{R}_i^{-1}\widetilde{\mathbf{R}}_i(\rho_{(0)})\mathbf{R}_i^{-1}\mathbf{1}_{m_i}\right\} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \\
&\quad \sum_{i=1}^N X_i\tau_{(1)} \left\{\mathbf{1}_{m_i}'\mathbf{R}_i^{-1}\widetilde{\mathbf{R}}_i(\rho_{(1)})\mathbf{R}_i^{-1}\mathbf{1}_{m_i}\right\} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},
\end{aligned}$$

where we generically denote the true marginal variance of the truncated outcome in each arm by $\tau_{(0)}$ and $\tau_{(1)}$. Now define

$$\mathbf{G}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \ \mathbf{G}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \tag{C.4}$$

asymptotically, we have

$$N^{-1}\boldsymbol{\Omega}_0 \xrightarrow{p} \left\{\mathrm{E}\left(\mathbf{1}'_{m_i}\mathbf{R}_i^{-1}\widetilde{\mathbf{R}}_i(\rho_{(0)})\mathbf{R}_i^{-1}\mathbf{1}_{m_i}\right)\right\}(1-\pi)\tau_{(0)}\mathbf{G}_0 + \left\{\mathrm{E}\left(\mathbf{1}'_{m_i}\mathbf{R}_i^{-1}\widetilde{\mathbf{R}}_i(\rho_{(1)})\mathbf{R}_i^{-1}\mathbf{1}_{m_i}\right)\right\}\pi\tau_{(1)}\mathbf{G}_1.$$

Because $\mathbf{R}_i = \mathbf{I}$, we have

$$\begin{aligned}
&\mathbf{1}'_{m_i}\mathbf{R}_i^{-1}\mathbf{1}_{m_i} = m_i\\
&\mathbf{1}'_{m_i}\mathbf{R}_i^{-1}\widetilde{\mathbf{R}}_i(\rho_{(0)})\mathbf{R}_i^{-1}\mathbf{1}_{m_i} = m_i\{1+(m_i-1)\rho_{(0)}\}\\
&\mathbf{1}'_{m_i}\mathbf{R}_i^{-1}\widetilde{\mathbf{R}}_i(\rho_{(1)})\mathbf{R}_i^{-1}\mathbf{1}_{m_i} = m_i\{1+(m_i-1)\rho_{(1)}\}
\end{aligned}$$

Now we multiply out the sandwich variance $N\boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Omega}_1^{-1}$, and the probability limit of the $(2,2)$-th element is obtained as

$$\begin{aligned}
&\left[N\boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Omega}_1^{-1}\right]_{(2,2)}\\
&\xrightarrow{p}\sigma^2_{\text{ind}} = \frac{\kappa^2_{(0)}}{(1-\pi)}\frac{\mathrm{E}\left[m_i\{1+(m_i-1)\rho_{(0)}\}\right]}{\bar{m}^2} + \frac{\kappa^2_{(1)}}{\pi}\frac{\mathrm{E}\left[m_i\{1+(m_i-1)\rho_{(1)}\}\right]}{\bar{m}^2},
\end{aligned}$$

where $\bar{m} = \mathrm{E}(m_i)$ is the mean cluster size, $\kappa_{(0)} = \sqrt{\tau_{(0)}}/\mu_{(0)}$ and $\kappa_{(1)} = \sqrt{\tau_{(1)}}/\mu_{(1)}$ are the coefficient of variation (CV) of the truncated counts in the control and intervention arms, respectively.

Finally, to see why we could use CV of the cluster sizes to further approximate the variance, we write $\mathrm{E}(m_i^2) = \mathrm{Var}(m_i) + \bar{m}^2 = (1+\eta^2)\bar{m}^2$, where η is the CV of cluster sizes. Therefore

$$\begin{aligned}
\mathrm{E}\left[m_i\{1+(m_i-1)\rho_{(0)}\}\right] =& \bar{m}[1+\{(1+\eta^2)\bar{m}-1\}\rho_{(0)}],\\
\mathrm{E}\left[m_i\{1+(m_i-1)\rho_{(1)}\}\right] =& \bar{m}[1+\{(1+\eta^2)\bar{m}-1\}\rho_{(1)}].
\end{aligned}$$

This leads to variance expression (15) in the main text

$$\sigma^2_{\text{ind}} = \frac{\kappa^2_{(0)}}{(1-\pi)\bar{m}}\left[1+\{(1+\eta^2)\bar{m}-1\}\rho_{(0)}\right] + \frac{\kappa^2_{(1)}}{\pi\bar{m}}\left[1+\{(1+\eta^2)\bar{m}-1\}\rho_{(1)}\right].$$

Appendix D: Derivation of variance expression under the arm-specific exchangeable working correlation

When using the SEE approach, both the variance functions and correlation models are correctly specified, and we can derive the asymptotic variance based on the model-based variance $\boldsymbol{\Omega}_1^{-1}$, where $\boldsymbol{\Omega}_1 = \sum_{i=1}^N \mathbf{D}'_{1i}\mathbf{V}_{1i}^{-1}\mathbf{D}_{1i}$, $\mathbf{V}_{1i} = \widetilde{\mathbf{A}}_i^{1/2}(\boldsymbol{\tau})\widetilde{\mathbf{R}}_i(\boldsymbol{\rho})\widetilde{\mathbf{A}}_i^{1/2}(\boldsymbol{\tau})$, with $\widetilde{\mathbf{R}}_i(\boldsymbol{\rho})$ the arm-specific exchangeable correlation matrix, defined in equation (11) of the main text. Observe that

$$\begin{aligned}
\boldsymbol{\Omega}_1 &= \sum_{i=1}^N \mathbf{D}'_{1i}\mathbf{V}_{1i}^{-1}\mathbf{D}_{1i} = \sum_{i=1}^N \mathbf{D}'_{1i}\widetilde{\mathbf{A}}_i^{-1/2}(\boldsymbol{\tau})\widetilde{\mathbf{R}}_i^{-1}(\boldsymbol{\rho})\widetilde{\mathbf{A}}_i^{-1/2}(\boldsymbol{\tau})\mathbf{D}_{1i}\\
&= \sum_{i=1}^N \left\{\begin{pmatrix} 1 \\ X_i\end{pmatrix}\otimes\mathbf{1}'_{m_i}\right\}\left\{\left((1-X_i)\frac{\mu^2_{(0)}}{\tau_{(0)}} + X_i\frac{\mu^2_{(1)}}{\tau_{(1)}}\right)\widetilde{\mathbf{R}}_i^{-1}(\boldsymbol{\rho})\right\}\left\{\begin{pmatrix} 1 & X_i\end{pmatrix}\otimes\mathbf{1}_{m_i}\right\}\\
&= \sum_{i=1}^N \frac{(1-X_i)\mu^2_{(0)}}{\tau_{(0)}}\left\{\mathbf{1}'_{m_i}\widetilde{\mathbf{R}}_i^{-1}(\rho_{(0)})\mathbf{1}_{m_i}\right\}\begin{pmatrix} 1 & 0\\ 0 & 0\end{pmatrix} +\\
&\quad \sum_{i=1}^N \frac{X_i\mu^2_{(1)}}{\tau_{(1)}}\left\{\mathbf{1}'_{m_i}\widetilde{\mathbf{R}}_i^{-1}(\rho_{(1)})\mathbf{1}_{m_i}\right\}\begin{pmatrix} 1 & 1\\ 1 & 1\end{pmatrix},
\end{aligned}$$

Then it follows that

$$N^{-1}\boldsymbol{\Omega}_1 \xrightarrow{p} \left\{\mathrm{E}\left(\mathbf{1}'_{m_i}\widetilde{\mathbf{R}}_i^{-1}(\rho_{(0)})\mathbf{1}_{m_i}\right)\right\}\frac{(1-\pi)\mu_{(0)}^2}{\tau_{(0)}}\mathbf{G}_0+$$
$$\left\{\mathrm{E}\left(\mathbf{1}'_{m_i}\widetilde{\mathbf{R}}_i^{-1}(\rho_{(1)})\mathbf{1}_{m_i}\right)\right\}\frac{\pi\mu_{(1)}^2}{\tau_{(1)}}\mathbf{G}_1,$$

where $\mathbf{G}_0$ and $\mathbf{G}_1$ are two constant matrices defined in equation (C.4). Notice

$$\begin{aligned}
\widetilde{\mathbf{R}}_i^{-1}(\rho_{(0)}) &= \frac{1}{1-\rho_{(0)}}\mathbf{I}_{m_i} - \frac{\rho_{(0)}}{(1-\rho_{(0)})\{1+(m_i-1)\rho_{(0)}\}}\mathbf{J}_{m_i}\\
\widetilde{\mathbf{R}}_i^{-1}(\rho_{(1)}) &= \frac{1}{1-\rho_{(1)}}\mathbf{I}_{m_i} - \frac{\rho_{(1)}}{(1-\rho_{(1)})\{1+(m_i-1)\rho_{(0)}\}}\mathbf{J}_{m_i}\\
\mathbf{1}'_{m_i}\widetilde{\mathbf{R}}_i^{-1}(\rho_{(0)})\mathbf{1}_{m_i} &= \frac{m_i}{1+(m_i-1)\rho_{(0)}}\\
\mathbf{1}'_{m_i}\widetilde{\mathbf{R}}_i^{-1}(\rho_{(1)})\mathbf{1}_{m_i} &= \frac{m_i}{1+(m_i-1)\rho_{(1)}},
\end{aligned}$$

where $\mathbf{I}_s$ is the $s \times s$ identity matrix and $\mathbf{J}_s = \mathbf{1}_s\mathbf{1}'_s$ is the $s \times s$ matrix of ones. Explicitly writing out the model-based variance allows us to derive the following expression for the asymptotic variance of the SEE estimator $\widehat{\gamma}_1$. To be specific, we have

$$\begin{aligned}
&\left[N\boldsymbol{\Omega}_1^{-1}\right]_{(2,2)}\\
&\xrightarrow{p}\sigma^2_{\text{aexch}} = \frac{\kappa^2_{(0)}}{(1-\pi)}\left\{\mathrm{E}\left(\frac{m_i}{1+(m_i-1)\rho_{(0)}}\right)\right\}^{-1} + \frac{\kappa^2_{(1)}}{\pi}\left\{\mathrm{E}\left(\frac{m_i}{1+(m_i-1)\rho_{(1)}}\right)\right\}^{-1}.
\end{aligned}$$

We further approximate the variance expression σ^2_{aexch} using the first two moments of the cluster size distribution. Recall that

$$\mathrm{E}\left(\frac{m_i}{1+(m_i-1)\rho_{(0)}}\right) = \frac{1}{\rho_{(0)}}\mathrm{E}\left\{\frac{m_i}{m_i+(1-\rho_{(0)})/\rho_{(0)}}\right\}.$$

And we invoke the Taylor series results of van Breukelen et al. (2007) and Candel and van Breukelen (2010), which states that for any random variable U with mean u_n and CV v_n, the second-order approximation of the following expectation is

$$\mathrm{E}\left(\frac{U}{U+\alpha}\right) \approx \left(\frac{u_n}{u_n+\alpha}\right)\left\{1 - v_n^2\left(\frac{u_n}{u_n+\alpha}\right)\left(\frac{\alpha}{u_n+\alpha}\right)\right\}$$

for any $\alpha \geq 0$. We then substitute U with m_i and α with $(1-\rho_{(0)})/\rho_{(0)}$ to get

$$\mathrm{E}\left\{\frac{m_i}{m_i+(1-\rho_{(0)})/\rho_{(0)}}\right\} \approx \left(\frac{\bar{m}}{\bar{m}+(1-\rho_{(0)})/\rho_{(0)}}\right)\left[1-\eta^2\frac{\bar{m}(1-\rho_{(0)})/\rho_{(0)}}{\{\bar{m}+(1-\rho_{(0)})/\rho_{(0)}\}^2}\right].$$

Replicating the same approximation for the expression involving $\rho_{(1)}$, we obtain

$$\begin{aligned}
\sigma^2_{\text{aexch}} \approx &\frac{\kappa^2_{(0)}\{1+(\bar{m}-1)\rho_{(0)}\}}{(1-\pi)\bar{m}}\left[1-\frac{\eta^2\bar{m}\rho_{(0)}(1-\rho_{(0)})}{\{1+(\bar{m}-1)\rho_{(0)}\}^2}\right]^{-1}\\
&+\frac{\kappa^2_{(1)}\{1+(\bar{m}-1)\rho_{(1)}\}}{\pi\bar{m}}\left[1-\frac{\eta^2\bar{m}\rho_{(1)}(1-\rho_{(1)})}{\{1+(\bar{m}-1)\rho_{(1)}\}^2}\right]^{-1}.
\end{aligned}$$

Appendix E: Derivation of variance expression under the independence working correlation assuming outcome is missing completely at random

In this Section, we incorporate missing outcomes into sample size calculation. We first introduce a missing data indicator O_{ij} which equals to 1 if the outcome of the jth patient in cluster i is observed (i.e. the participant has complete follow-up information until the end of the study). If the outcome of this patient is missing either due to non-response or drop-out, we set $O_{ij} = 0$. We assume the outcome is missing completely at random (MCAR). This simple condition allows us to present closed-form variance expressions for purpose of design calculation, and ensures the validity of complete-case GEE and SEE analyses. Here, a complete case is contributed by a participant who has continued through the end of the study. Because participants with partial follow-up information is ignored in the complete-case analysis, the following sample size calculation approach is at most conservative.

Define $\mathbf{\Gamma}_i = \text{diag}(O_{i1}, \ldots, O_{im_i})$, based on the complete-case analysis, we can express

$$\begin{aligned}
N^{-1}\mathbf{\Omega}_1 =& \frac{1}{N}\sum_{i=1}^{N}\left\{\begin{pmatrix} 1 \\ X_i \end{pmatrix} \otimes \mathbf{1}'_{m_i}\mathbf{\Gamma}_i\right\}\left\{\left((1-X_i)\mu_{(0)} + X_i\mu_{(1)}\right)\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\right\}\left\{\begin{pmatrix} 1 & X_i \end{pmatrix} \otimes \mathbf{\Gamma}_i\mathbf{1}_{m_i}\right\} \\
=& \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{1}'_{m_i}\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\mathbf{1}_{m_i}\right)\begin{pmatrix} (1-X_i)\mu_{(0)} + X_i\mu_{(1)} & X_i\mu_{(1)} \\ X_i\mu_{(1)} & X_i\mu_{(1)} \end{pmatrix} \\
\xrightarrow{p}& \left\{\text{E}\left(\mathbf{1}'_{m_i}\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\mathbf{1}_{m_i}\right)\right\}\begin{pmatrix} (1-\pi)\mu_{(0)} + \pi\mu_{(1)} & \pi\mu_{(1)} \\ \pi\mu_{(1)} & \pi\mu_{(1)} \end{pmatrix}.
\end{aligned}$$

Introducing O_{ij} in the similar fashion, we obtain

$$\begin{aligned}
N^{-1}\mathbf{\Omega}_0 =& \frac{1}{N}\sum_{i=1}^{N}(1-X_i)\tau_{(0)}\left\{\mathbf{1}'_{m_i}\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\widetilde{\mathbf{R}}_i(\rho_{(0)})\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\mathbf{1}_{m_i}\right\}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \\
& \frac{1}{N}\sum_{i=1}^{N}X_i\tau_{(1)}\left\{\mathbf{1}'_{m_i}\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\widetilde{\mathbf{R}}_i(\rho_{(1)})\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\mathbf{1}_{m_i}\right\}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\
\xrightarrow{p}& \left\{\text{E}\left(\mathbf{1}'_{m_i}\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\widetilde{\mathbf{R}}_i(\rho_{(0)})\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\mathbf{1}_{m_i}\right)\right\}(1-\pi)\tau_{(0)}\mathbf{G}_0 + \\
& \left\{E\left(\mathbf{1}'_{m_i}\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\widetilde{\mathbf{R}}_i(\rho_{(1)})\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\mathbf{1}_{m_i}\right)\right\}\pi\tau_{(1)}\mathbf{G}_1.
\end{aligned}$$

Because the working correlation matrix assumes independence, $R_i = I$, we have

$$\begin{aligned}
&\mathbf{1}'_{m_i}\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\mathbf{1}_{m_i} = \sum_{j=1}^{m_i}O_{ij} = O_{i+} \\
&\mathbf{1}'_{m_i}\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\widetilde{\mathbf{R}}_i(\rho_{(0)})\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\mathbf{1}_{m_i} = O_{i+}\{1 + (O_{i+} - 1)\rho_{(0)}\} \\
&\mathbf{1}'_{m_i}\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\widetilde{\mathbf{R}}_i(\rho_{(1)})\mathbf{\Gamma}_i\mathbf{R}_i^{-1}\mathbf{\Gamma}_i\mathbf{1}_{m_i} = O_{i+}\{1 + (O_{i+} - 1)\rho_{(1)}\}
\end{aligned}$$

Now we multiply out the sandwich variance $N\mathbf{\Omega}_1^{-1}\mathbf{\Omega}_0\mathbf{\Omega}_1^{-1}$, and the probability limit of the $(2,2)$-th element is obtained as

$$\begin{aligned}
&\left[N\mathbf{\Omega}_1^{-1}\mathbf{\Omega}_0\mathbf{\Omega}_1^{-1}\right]_{(2,2)} \\
\xrightarrow{p}& \widetilde{\sigma}^2_{\text{ind}} = \frac{\kappa^2_{(0)}}{(1-\pi)}\frac{\text{E}\left[O_{i+}\{1 + (O_{i+} - 1)\rho_{(0)}\}\right]}{\{\text{E}(O_{i+})\}^2} + \frac{\kappa^2_{(1)}}{\pi}\frac{\text{E}\left[O_{i+}\{1 + (O_{i+} - 1)\rho_{(1)}\}\right]}{\{\text{E}(O_{i+})\}^2},
\end{aligned}$$

where we use $\widetilde{\sigma}^2_{\text{ind}}$ to denote the asymptotic variance in the presence of missing outcomes. Notice that this expression is identical to expression (13) in Section 3.1, except we replace the total sample size in each cluster m_i with the number of complete cases (or referred to as the *effective cluster size*) $O_{i+} \leq m_i$. The following insights can be obtained from this expression:

- In general, the previously-derived formula for σ^2_{ind} is still applicable to outcome MCAR, once we substitute $\bar{m}$ (mean cluster size) and η (CV of cluster sizes) with the mean effective cluster size and the CV of the effective cluster sizes defined based on complete cases;

- For further approximations, assuming $\zeta = \text{E}(O_{ij})$ as the common marginal probability that the outcome of each individual is observed (in other words, $1-\zeta$ is the marginal missingness proportion), and $\xi = \text{E}(O_{ij}O_{ij'})$ as the common joint probability that any two individuals have their outcomes observed. This allows us to define a common ICC of the missingness indicator (Turner et al., 2020),

$$\begin{aligned}\varrho =&\text{Corr}(1-O_{ij}, 1-O_{ij'})\\ =&\frac{\text{E}\{(1-O_{ij})(1-O_{ij'})\}-\text{E}(1-O_{ij})\text{E}(1-O_{ij'})}{\sqrt{\text{E}(1-O_{ij})\{1-\text{E}(1-O_{ij})\}}\sqrt{\text{E}(1-O_{ij'})\{1-\text{E}(1-O_{ij'})\}}} = \frac{\xi-\zeta^2}{\zeta(1-\zeta)}\\ =&\text{Corr}(O_{ij}, O_{ij'}).\end{aligned}$$

In other words, $\xi = \zeta\{\zeta + \varrho(1-\zeta)\}$. In this case, because the outcome is missing completely at random and does not depend on the cluster size, $\text{E}(O_{i+}) = \bar{m}\zeta$, $\text{E}(O^2_{i+}) = \bar{m}[\zeta + \{(1+\eta^2)\bar{m}-1\}\xi]$, and

$$\text{E}\left[O_{i+}\{1+(O_{i+}-1)\rho_{(0)}\}\right] = \bar{m}[\zeta + \{(1+\eta^2)\bar{m}-1\}\xi\rho_{(0)}].$$

Likewise, we have $\text{E}\left[O_{i+}\{1+(O_{i+}-1)\rho_{(1)}\}\right] = \bar{m}[\zeta + \{(1+\eta^2)\bar{m}-1\}\xi\rho_{(1)}]$. This leads to an explicit expression of the asymptotic variance

$$\begin{aligned}\widetilde{\sigma}^2_{\text{ind}} =&\frac{\kappa^2_{(0)}}{(1-\pi)\bar{m}}\left[\frac{1}{\zeta} + \{(1+\eta^2)\bar{m}-1\}\frac{\xi}{\zeta^2}\rho_{(0)}\right] + \frac{\kappa^2_{(1)}}{\pi\bar{m}}\left[\frac{1}{\zeta} + \{(1+\eta^2)\bar{m}-1\}\frac{\xi}{\zeta^2}\rho_{(1)}\right]\\ =&\frac{\kappa^2_{(0)}}{(1-\pi)\bar{m}}\left[\frac{1}{\zeta} + \{(1+\eta^2)\bar{m}-1\}\left\{1+\frac{1-\zeta}{\zeta}\varrho\right\}\rho_{(0)}\right] +\\ &\frac{\kappa^2_{(1)}}{\pi\bar{m}}\left[\frac{1}{\zeta} + \{(1+\eta^2)\bar{m}-1\}\left\{1+\frac{1-\zeta}{\zeta}\varrho\right\}\rho_{(1)}\right].\end{aligned}$$

Compared to σ^2_{ind}, this new expression now requires additional assumptions on the marginal missingness proportion $(1-\zeta)$, and the the ICC of the missingness indicator. Further because $\zeta \leq 1$, we have $\widetilde{\sigma}^2_{\text{ind}} \geq \sigma^2_{\text{ind}}$, which is expected because missing outcome will reduce the effective sample size and inflated the variance. Furthermore, the variance is an increasing function in the missingness proportion $(1-\zeta)$ and the ICC of missingness indicator ϱ. More interestingly,

$$\begin{aligned}\widetilde{\sigma}^2_{\text{ind}} =&\frac{1}{\zeta}\frac{\kappa^2_{(0)}}{(1-\pi)\bar{m}}\left[1 + \{(1+\eta^2)\bar{m}-1\}\{\zeta+(1-\zeta)\varrho\}\rho_{(0)}\right] +\\ &\frac{1}{\zeta}\frac{\kappa^2_{(1)}}{\pi\bar{m}}\left[1 + \{(1+\eta^2)\bar{m}-1\}\{\zeta+(1-\zeta)\varrho\}\rho_{(1)}\right] \leq \frac{\sigma^2_{\text{ind}}}{\zeta},\end{aligned}$$

because $\zeta + (1-\zeta)\varrho \leq 1$. This result indicates that simply inflates the variance (and hence the sample size) by ζ^{-1} (inverse of one minus of the marginal missingness probability) leads to an upper bound and will at most be conservative.

- If we assume that the missingness of any two observations is independent of each other so that $\varrho = 0$ and $\xi = \zeta^2$, the variance expression simplifies to

$$\widetilde{\sigma}^2_{\text{ind}} = \frac{\kappa^2_{(0)}}{(1-\pi)\bar{m}}\left[\frac{1}{\zeta} + \{(1+\eta^2)\bar{m} - 1\}\rho_{(0)}\right] + \frac{\kappa^2_{(1)}}{\pi\bar{m}}\left[\frac{1}{\zeta} + \{(1+\eta^2)\bar{m} - 1\}\rho_{(1)}\right].$$

Appendix F: Derivation of variance expression under the arm-specific exchangeable working correlation assuming outcome is missing completely at random

When using the SEE approach, both the variance functions and correlation models are correctly specified, and we can derive the asymptotic variance based on the model-based variance $\mathbf{\Omega}_1^{-1}$, where $\mathbf{\Omega}_1 = \sum_{i=1}^N \mathbf{D}'_{1i}\mathbf{V}_{1i}^{-1}\mathbf{D}_{1i}$, $\mathbf{V}_{1i} = \widetilde{\mathbf{A}}_i^{1/2}(\boldsymbol{\tau})\widetilde{\mathbf{R}}_i(\boldsymbol{\rho})\widetilde{\mathbf{A}}_i^{1/2}(\boldsymbol{\tau})$, with $\widetilde{\mathbf{R}}_i(\boldsymbol{\rho})$ the arm-specific exchangeable correlation matrix, defined in equation (11) of the main text. In the presence of missing outcome data, we can write

$$\begin{aligned}
N^{-1}\mathbf{\Omega}_1 =& \frac{1}{N}\sum_{i=1}^N \left\{\begin{pmatrix} 1 \\ X_i \end{pmatrix} \otimes \mathbf{1}'_{m_i}\mathbf{\Gamma}_i\right\}\left\{\left((1-X_i)\frac{\mu^2_{(0)}}{\tau_{(0)}} + X_i\frac{\mu^2_{(1)}}{\tau_{(1)}}\right)\left(\mathbf{\Gamma}_i\widetilde{\mathbf{R}}_i(\boldsymbol{\rho})\mathbf{\Gamma}_i\right)^-\right\}\left\{\begin{pmatrix} 1 & X_i \end{pmatrix} \otimes \mathbf{\Gamma}_i\mathbf{1}_{m_i}\right\} \\
\xrightarrow{p}& \left\{\mathrm{E}\left(\mathbf{1}'_{m_i}\mathbf{\Gamma}_i\left(\mathbf{\Gamma}_i\widetilde{\mathbf{R}}_i(\rho_{(0)})\mathbf{\Gamma}_i\right)^-\mathbf{\Gamma}_i\mathbf{1}_{m_i}\right)\right\}\frac{(1-\pi)\mu^2_{(0)}}{\tau_{(0)}}\mathbf{G}_0 + \\
& \left\{\mathrm{E}\left(\mathbf{1}'_{m_i}\mathbf{\Gamma}_i\left(\mathbf{\Gamma}_i\widetilde{\mathbf{R}}_i(\rho_{(1)})\mathbf{\Gamma}_i\right)^-\mathbf{\Gamma}_i\mathbf{1}_{m_i}\right)\right\}\frac{\pi\mu^2_{(1)}}{\tau_{(1)}}\mathbf{G}_1,
\end{aligned}$$

where define $\mathbf{A}^-$ as the g-inverse of the rank-deficient matrix $\mathbf{A}$ (More specifically, we require this g-inverse to be the one obtained by inverting the full-rank subset of $\mathbf{A}$, and then replacing the remaining elements with zeros). Notice

$$\begin{aligned}
\mathbf{1}'_{m_i}\mathbf{\Gamma}_i\left(\mathbf{\Gamma}_i\widetilde{\mathbf{R}}_i(\rho_{(0)})\mathbf{\Gamma}_i\right)^-\mathbf{\Gamma}_i\mathbf{1}_{m_i} &= \frac{O_{i+}}{1-\rho_{(0)}} - \frac{O^2_{i+}\rho_{(0)}}{(1-\rho_{(0)})\{1+(O_{i+}-1)\rho_{(0)}\}} = \frac{O_{i+}}{1+(O_{i+}-1)\rho_{(0)}} \\
\mathbf{1}'_{m_i}\mathbf{\Gamma}_i\left(\mathbf{\Gamma}_i\widetilde{\mathbf{R}}_i(\rho_{(1)})\mathbf{\Gamma}_i\right)^-\mathbf{\Gamma}_i\mathbf{1}_{m_i} &= \frac{O_{i+}}{1+(O_{i+}-1)\rho_{(1)}}.
\end{aligned}$$

Then we have

$$\begin{aligned}
&\left[N\mathbf{\Omega}_1^{-1}\right]_{(2,2)} \\
&\xrightarrow{p}\widetilde{\sigma}^2_{\text{aexch}} = \frac{\kappa^2_{(0)}}{(1-\pi)}\left\{\mathrm{E}\left(\frac{O_{i+}}{1+(O_{i+}-1)\rho_{(0)}}\right)\right\}^{-1} + \frac{\kappa^2_{(1)}}{\pi}\left\{\mathrm{E}\left(\frac{O_{i+}}{1+(O_{i+}-1)\rho_{(1)}}\right)\right\}^{-1},
\end{aligned}$$

where we use $\widetilde{\sigma}^2_{\text{aexch}}$ to denote the asymptotic variance in the presence of missing outcomes. The following insights can be obtained from this expression:

- In general, the previously-derived formula for σ^2_{aexch} is still applicable to outcome MCAR, once we substitute $\bar{m}$ (mean cluster size) and η (CV of cluster sizes) with the mean effective cluster size and the CV of the effective cluster sizes defined based on complete cases;

- For further approximations, assuming $\zeta = \mathrm{E}(O_{ij})$ as the common marginal probability that the outcome of each individual is observed (in other words, $1-\zeta$ is the marginal missingness proportion), and $\xi = \mathrm{E}(O_{ij}O_{ij'})$ as the common joint probability that any two individuals have their outcomes observed. This allows us to define a common ICC of the missingness indicator,

$$\varrho = \mathrm{Corr}(1 - O_{ij}, 1 - O_{ij'}) = \mathrm{Corr}(O_{ij}, O_{ij'}) = \frac{\xi - \zeta^2}{\zeta(1-\zeta)}.$$

In other words, $\xi = \zeta\{\zeta + \varrho(1-\zeta)\}$. Observe that

$$\mathrm{E}\left(\frac{O_{i+}}{1 + (O_{i+} - 1)\rho_{(0)}}\right) = \frac{1}{\rho_{(0)}}\mathrm{E}\left\{\frac{O_{i+}}{O_{i+} + (1-\rho_{(0)})/\rho_{(0)}}\right\}.$$

Invoking the Taylor series results of van Breukelen et al. (2007) and Candel and van Breukelen (2010), which states that for any random variable U with mean u_n and CV v_n, the second-order approximation of the following expectation is

$$\mathrm{E}\left(\frac{U}{U+\alpha}\right) \approx \left(\frac{u_n}{u_n+\alpha}\right)\left\{1 - v_n^2\left(\frac{u_n}{u_n+\alpha}\right)\left(\frac{\alpha}{u_n+\alpha}\right)\right\}$$

for any $\alpha \geq 0$. We then substitute U with $O_{i+} = \sum_{i=1}^{m_i} O_{ij}$ and α with $(1-\rho_{(0)})/\rho_{(0)}$ to provide an approximate variance expression. Recall that O_{i+} an over-dispersed count due to the ICC ϱ; this indicates that the marginal mean $u_n = \mathrm{E}(O_{i+}) = \bar{m}\zeta$,

$$\begin{aligned}\mathrm{E}(O_{i+}^2) =& \bar{m}[\zeta + \{(1+\eta^2)\bar{m} - 1\}\xi]\\ \mathrm{Var}(O_{i+}) =& \mathrm{E}(O_{i+}^2) - u_n^2 = \bar{m}[\zeta + \{(1+\eta^2)\bar{m} - 1\}\xi - \bar{m}\zeta^2]\end{aligned}$$

and the squared CV for O_{ij} is obtained as

$$v_n^2 = \frac{\zeta + \{(1+\eta^2)\bar{m} - 1\}\xi - \bar{m}\zeta^2}{\bar{m}\zeta^2} = \eta^2 + \frac{(1-\zeta)\left[1 + \{(1+\eta^2)\bar{m} - 1\}\varrho\right]}{\bar{m}\zeta} \geq \eta^2. \qquad \text{(F.1)}$$

This expressions allow us to approximate

$$\mathrm{E}\left\{\frac{O_{i+}}{O_{i+} + (1-\rho_{(0)})/\rho_{(0)}}\right\} \approx \left(\frac{u_n}{u_n + (1-\rho_{(0)})/\rho_{(0)}}\right)\left[1 - v_n^2\frac{u_n(1-\rho_{(0)})/\rho_{(0)}}{\{u_n + (1-\rho_{(0)})/\rho_{(0)}\}^2}\right].$$

Replicating the same approximation for the expression involving $\rho_{(1)}$, we obtain

$$\begin{aligned}\widetilde{\sigma}^2_{\text{aexch}} \approx& \frac{\kappa^2_{(0)}\{1 + (\bar{m}\zeta - 1)\rho_{(0)}\}}{(1-\pi)\bar{m}\zeta}\left[1 - \frac{v_n^2\bar{m}\zeta\rho_{(0)}(1-\rho_{(0)})}{\{1 + (\bar{m}\zeta - 1)\rho_{(0)}\}^2}\right]^{-1}\\ &+ \frac{\kappa^2_{(1)}\{1 + (\bar{m}\zeta - 1)\rho_{(1)}\}}{\pi\bar{m}\zeta}\left[1 - \frac{v_n^2\bar{m}\zeta\rho_{(1)}(1-\rho_{(1)})}{\{1 + (\bar{m}\zeta - 1)\rho_{(1)}\}^2}\right]^{-1}.\end{aligned}$$

Although this variance expression is rather complicated, it only requires additional assumptions on the marginal missingness proportion $(1-\zeta)$, and the the ICC of the missingness, compared to σ^2_{aexch}. We now investigate whether $\widetilde{\sigma}^2_{\text{aexch}} \leq \sigma^2_{\text{aexch}}/\zeta$ as in Web Appendix E. Although such a relationship is not immediate, when the mean cluster size $\bar{m}$ approaches infinity, we can show that $\sigma^2_{\text{aexch}}/\zeta$ becomes a valid upper bound for $\widetilde{\sigma}^2_{\text{aexch}}$. To see this, we define

$$\begin{aligned}
q_0 =& \frac{\kappa^2_{(0)}\{1+(\bar{m}\zeta-1)\rho_{(0)}\}}{(1-\pi)\bar{m}\zeta}\left[1-\frac{\upsilon_n^2\bar{m}\zeta\rho_{(0)}(1-\rho_{(0)})}{\{1+(\bar{m}\zeta-1)\rho_{(0)}\}^2}\right]^{-1} \\
&\times\left[\frac{\kappa^2_{(0)}\{1+(\bar{m}-1)\rho_{(0)}\}}{(1-\pi)\bar{m}\zeta}\right]^{-1}\left[1-\frac{\eta^2\bar{m}\rho_{(0)}(1-\rho_{(0)})}{\{1+(\bar{m}-1)\rho_{(0)}\}^2}\right] \\
=& \frac{\{1+(\bar{m}\zeta-1)\rho_{(0)}\}^3}{\{1+(\bar{m}-1)\rho_{(0)}\}^3}\times\frac{\{1+(\bar{m}-1)\rho_{(0)}\}^2-\eta^2\bar{m}\rho_{(0)}(1-\rho_{(0)})}{\{1+(\bar{m}\zeta-1)\rho_{(0)}\}^2-\upsilon_n^2\bar{m}\zeta\rho_{(0)}(1-\rho_{(0)})} \\
=& \frac{\{1+(\bar{m}\zeta-1)\rho_{(0)}\}^3}{\{1+(\bar{m}-1)\rho_{(0)}\}^3} \\
&\times\frac{\{1+(\bar{m}-1)\rho_{(0)}\}^2-\eta^2\bar{m}\rho_{(0)}(1-\rho_{(0)})}{\{1+(\bar{m}\zeta-1)\rho_{(0)}\}^2-\eta^2\bar{m}\zeta\rho_{(0)}(1-\rho_{(0)})-(1-\zeta)\left[1+\{(1+\eta^2)\bar{m}-1\}\varrho\right]\rho_{(0)}(1-\rho_{(0)})}
\end{aligned}$$

Based equation (F.1), we have $\lim_{\bar{m}\to\infty} q_0 = \zeta \leq 1$. Likewise, we can show that

$$\begin{aligned}
\lim_{\bar{m}\to\infty} q_1 = \lim_{\bar{m}\to\infty} & \frac{\kappa^2_{(1)}\{1+(\bar{m}\zeta-1)\rho_{(1)}\}}{\pi\bar{m}\zeta}\left[1-\frac{\upsilon_n^2\bar{m}\zeta\rho_{(1)}(1-\rho_{(1)})}{\{1+(\bar{m}\zeta-1)\rho_{(1)}\}^2}\right]^{-1} \\
&\times\left[\frac{\kappa^2_{(1)}\{1+(\bar{m}-1)\rho_{(1)}\}}{\pi\bar{m}\zeta}\right]^{-1}\left[1-\frac{\eta^2\bar{m}\rho_{(1)}(1-\rho_{(1)})}{\{1+(\bar{m}-1)\rho_{(1)}\}^2}\right] \leq 1.
\end{aligned}$$

This leads to $\lim_{\bar{m}\to\infty}\widetilde{\sigma}^2_{\text{aexch}}/\{\sigma^2_{\text{aexch}}/\zeta\} \leq 1$, and indicates when $\bar{m}$ is large, sample size based on $\sigma^2_{\text{aexch}}/\zeta$ is at most conservative.

- Finally, if we assume that the missingness of any two observations is independent of each other so that $\varrho = 0$ and $\upsilon_n^2 = \eta^2 + (1-\zeta)/(\bar{m}\zeta)$, the variance expression becomes

$$\begin{aligned}
\widetilde{\sigma}^2_{\text{aexch}} \approx & \frac{\kappa^2_{(0)}\{1+(\bar{m}\zeta-1)\rho_{(0)}\}}{(1-\pi)\bar{m}\zeta}\left[1-\left\{\eta^2+\frac{1-\zeta}{\bar{m}\zeta}\right\}\frac{\bar{m}\zeta\rho_{(0)}(1-\rho_{(0)})}{\{1+(\bar{m}\zeta-1)\rho_{(0)}\}^2}\right]^{-1} \\
&+\frac{\kappa^2_{(1)}\{1+(\bar{m}\zeta-1)\rho_{(1)}\}}{\pi\bar{m}\zeta}\left[1-\left\{\eta^2+\frac{1-\zeta}{\bar{m}\zeta}\right\}\frac{\bar{m}\zeta\rho_{(1)}(1-\rho_{(1)})}{\{1+(\bar{m}\zeta-1)\rho_{(1)}\}^2}\right]^{-1} \\
=& \frac{\kappa^2_{(0)}\{1+(\bar{m}\zeta-1)\rho_{(0)}\}}{(1-\pi)\bar{m}\zeta}\left[1-\frac{\eta^2\bar{m}\zeta\rho_{(0)}(1-\rho_{(0)})}{\{1+(\bar{m}\zeta-1)\rho_{(0)}\}^2}-\frac{(1-\zeta)\rho_{(0)}(1-\rho_{(0)})}{\{1+(\bar{m}\zeta-1)\rho_{(0)}\}^2}\right]^{-1} \\
&+\frac{\kappa^2_{(1)}\{1+(\bar{m}\zeta-1)\rho_{(1)}\}}{\pi\bar{m}\zeta}\left[1-\frac{\eta^2\bar{m}\zeta\rho_{(1)}(1-\rho_{(1)})}{\{1+(\bar{m}\zeta-1)\rho_{(1)}\}^2}-\frac{(1-\zeta)\rho_{(1)}(1-\rho_{(1)})}{\{1+(\bar{m}\zeta-1)\rho_{(1)}\}^2}\right]^{-1}.
\end{aligned}$$

Appendix G: Web Tables referenced in the main text

Table 1 True values induced from the conditional model for power calculation when $(\exp(\beta_0), \exp(\beta_1)) = (1.25, 0.55)$ and under different right truncation points.

(σ_0^2, σ_1^2)	Parameters	$T = \infty$	$T = 4$	$T = 2$	$T = 1$
$(0.05, 0.05)$	$(\mu_{(0)}, \exp(\Delta))$	(1.282,0.550)	(1.232,0.568)	(0.929,0.655)	(0.555,0.736)
	$(\tau_{(0)}, \tau_{(1)})$	(1.367,0.730)	(1.165,0.708)	(0.588,0.493)	(0.247,0.242)
	$(\kappa_{(0)}, \kappa_{(1)})$	(0.912,1.212)	(0.876,1.202)	(0.825,1.153)	(0.896,1.203)
	$(\rho_{(0)}, \rho_{(1)})$	(0.062,0.035)	(0.053,0.034)	(0.028,0.024)	(0.012,0.012)
$(0.05, 0.10)$	$(\mu_{(0)}, \exp(\Delta))$	(1.282,0.564)	(1.232,0.581)	(0.929,0.661)	(0.555,0.953)
	$(\tau_{(0)}, \tau_{(1)})$	(1.366,0.778)	(1.165,0.743)	(0.589,0.500)	(0.247,0.249)
	$(\kappa_{(0)}, \kappa_{(1)})$	(0.912,1.220)	(0.876,1.206)	(0.825,1.151)	(0.896,0.944)
	$(\rho_{(0)}, \rho_{(1)})$	(0.062,0.071)	(0.053,0.067)	(0.028,0.046)	(0.012,0.024)
$(0.05, 0.20)$	$(\mu_{(0)}, \exp(\Delta))$	(1.282,0.593)	(1.232,0.605)	(0.929,0.672)	(0.555,0.741)
	$(\tau_{(0)}, \tau_{(1)})$	(1.366,0.887)	(1.165,0.816)	(0.588,0.513)	(0.247,0.242)
	$(\kappa_{(0)}, \kappa_{(1)})$	(0.912,1.240)	(0.876,1.212)	(0.825,1.147)	(0.896,1.196)
	$(\rho_{(0)}, \rho_{(1)})$	(0.062,0.144)	(0.053,0.132)	(0.028,0.087)	(0.012,0.044)
$(0.10, 0.10)$	$(\mu_{(0)}, \exp(\Delta))$	(1.315,0.550)	(1.249,0.573)	(0.931,0.660)	(0.554,0.410)
	$(\tau_{(0)}, \tau_{(1)})$	(1.496,0.778)	(1.216,0.744)	(0.594,0.500)	(0.247,0.242)
	$(\kappa_{(0)}, \kappa_{(1)})$	(0.931,1.220)	(0.883,1.206)	(0.828,1.151)	(0.897,1.201)
	$(\rho_{(0)}, \rho_{(1)})$	(0.121,0.071)	(0.100,0.067)	(0.053,0.046)	(0.024,0.023)
$(0.10, 0.20)$	$(\mu_{(0)}, \exp(\Delta))$	(1.315,0.578)	(1.249,0.597)	(0.931,0.671)	(0.554,0.742)
	$(\tau_{(0)}, \tau_{(1)})$	(1.496,0.887)	(1.216,0.816)	(0.594,0.513)	(0.247,0.242)
	$(\kappa_{(0)}, \kappa_{(1)})$	(0.931,1.240)	(0.883,1.212)	(0.828,1.147)	(0.897,1.196)
	$(\rho_{(0)}, \rho_{(1)})$	(0.121,0.144)	(0.100,0.132)	(0.053,0.087)	(0.024,0.044)
$(0.20, 0.20)$	$(\mu_{(0)}, \exp(\Delta))$	(1.382,0.760)	(1.279,0.583)	(0.933,0.670)	(0.553,0.744)
	$(\tau_{(0)}, \tau_{(1)})$	(1.805,0.887)	(1.311,0.816)	(0.604,0.513)	(0.247,0.242)
	$(\kappa_{(0)}, \kappa_{(1)})$	(0.972,1.240)	(0.895,1.212)	(0.833,1.147)	(0.899,1.196)
	$(\rho_{(0)}, \rho_{(1)})$	(0.234,0.144)	(0.181,0.132)	(0.099,0.087)	(0.045,0.044)

Table 2 Empirical type I error rates (%) for GEE and SEE analyses when the mean cluster size $\bar{m} = 25$, cluster size CV $\eta = 0.3$, $(\exp(\beta_0), \exp(\beta_1)) = (1.25, 0.55)$. Empirical type I error rate between 4.5% and 5.5% are considered close to nominal according to the margin of error under a binomial model with 10000 replications, and are highlighted in bold font.

		Independence (GEE)				Arm-specific exchangeable (SEE)			
(σ_0^2, σ_1^2)	Variance	$T=\infty$	$T=4$	$T=2$	$T=1$	$T=\infty$	$T=4$	$T=2$	$T=1$
	N	12	12	14	22	12	12	14	22
	MB	16.8	14.9	9.8	6.9	7.1	6.9	6.9	5.8
	LZ	7.3	7.2	7.0	5.8	7.3	7.1	7.1	5.9
$(0.05, 0.05)$	MD	3.4	3.5	3.8	3.8	3.5	3.4	3.8	3.8
	KC	**5.1**	**5.1**	**5.2**	**4.9**	**5.1**	**4.9**	**5.1**	**4.9**
	FG	4.3	4.3	4.4	4.3	4.2	3.9	4.4	4.4
	AVG	4.1	4.2	**4.5**	4.3	4.2	4.0	**4.5**	4.4
	N	16	16	18	26	16	16	18	26
	MB	24.0	21.3	13.4	9.0	6.9	6.7	6.4	5.8
$(0.05, 0.10)$	LZ	7.0	6.9	6.2	5.7	6.8	6.7	6.3	5.8
	MD	4.1	4.0	3.9	4.1	4.2	3.9	3.7	4.1
	KC	**5.4**	**5.2**	**5.0**	**4.8**	**5.4**	**5.1**	**5.0**	**4.9**
	FG	**4.6**	**4.6**	4.4	4.4	**4.9**	**4.6**	4.3	**4.6**
	AVG	**4.7**	**4.5**	4.4	4.4	**4.7**	**4.5**	4.3	**4.5**
	N	24	24	26	32	24	22	24	32
	MB	34.9	30.2	19.9	12.3	6.7	6.9	6.0	5.9
$(0.05, 0.20)$	LZ	6.9	6.2	6.0	6.0	6.7	6.8	5.9	5.9
	MD	**4.7**	4.4	4.1	**4.5**	**4.8**	**4.7**	4.1	**4.5**
	KC	5.8	**5.1**	**5.1**	**5.2**	5.6	**5.5**	**4.9**	**5.2**
	FG	**5.0**	**4.7**	**4.5**	**4.8**	**5.4**	**5.2**	**4.7**	**5.0**
	AVG	**5.2**	**4.8**	**4.6**	**4.9**	**5.2**	**5.0**	**4.5**	**4.9**
	N	16	18	20	28	16	18	20	28
	MB	28.4	25.3	17.9	11.1	6.7	6.8	6.8	6.0
$(0.10, 0.10)$	LZ	7.0	6.8	6.9	6.1	6.6	6.6	6.9	6.1
	MD	4.1	4.0	4.3	4.3	3.9	3.9	4.2	**4.5**
	KC	**5.5**	**5.2**	**5.4**	**5.2**	**5.3**	**5.3**	**5.4**	**5.2**
	FG	**4.8**	**4.6**	**4.9**	**4.7**	**4.6**	**4.5**	**4.8**	**4.8**
	AVG	**4.8**	**4.6**	**4.9**	**4.7**	**4.7**	**4.5**	**4.7**	**4.8**
	N	26	26	26	34	24	24	26	34
	MB	38.5	34.1	23.8	14.0	6.5	6.1	6.0	5.9
$(0.10, 0.20)$	LZ	6.8	6.1	6.1	5.8	6.3	6.0	6.0	5.9
	MD	**4.7**	4.4	4.0	4.4	**4.8**	4.3	4.1	**4.5**
	KC	5.7	**5.1**	**5.0**	**5.1**	5.6	**5.0**	**5.0**	**5.2**
	FG	**5.0**	**4.7**	**4.5**	**4.7**	**5.2**	**4.7**	**4.6**	**4.9**
	AVG	**5.1**	**4.8**	**4.6**	**4.7**	**5.1**	**4.6**	**4.5**	**4.9**
	N	28	28	30	38	26	26	28	38
	MB	44.5	40.0	28.0	17.6	6.6	6.2	6.3	6.0
$(0.20, 0.20)$	LZ	7.0	6.6	6.1	6.1	6.5	6.0	6.3	6.1
	MD	**5.0**	**4.6**	**4.5**	**4.5**	4.4	4.1	**4.5**	**4.6**
	KC	5.9	**5.5**	**5.1**	**5.2**	**5.3**	**5.0**	**5.5**	**5.3**
	FG	**5.4**	**5.1**	**4.9**	**4.8**	**4.8**	**4.5**	**4.9**	**5.0**
	AVG	**5.5**	**5.0**	**4.9**	**4.8**	**4.8**	**4.5**	**4.9**	**5.0**

Table 3 Empirical type I error rates (%) for GEE and SEE analyses when the mean cluster size $\bar{m} = 25$, cluster size CV $\eta = 0.6$, $(\exp(\beta_0), \exp(\beta_1)) = (1.25, 0.55)$. Empirical type I error rate between 4.5% and 5.5% are considered close to nominal according to the margin of error under a binomial model with 10000 replications, and are highlighted in bold font.

		Independence (GEE)				Arm-specific exchangeable (SEE)			
(σ_0^2, σ_1^2)	Variance	$T=\infty$	$T=4$	$T=2$	$T=1$	$T=\infty$	$T=4$	$T=2$	$T=1$
$(0.05, 0.05)$	N	14	14	16	24	12	12	16	24
	MB	20.2	18.6	11.2	8.0	9.0	8.6	7.3	7.3
	LZ	9.5	9.1	7.4	7.0	9.4	9.1	7.6	7.5
	MD	**4.6**	**4.5**	3.9	4.2	3.9	3.9	3.5	4.4
	KC	6.9	6.7	**5.3**	**5.5**	6.4	6.4	**5.5**	5.9
	FG	5.7	5.7	**4.7**	**4.9**	**5.0**	**5.2**	**4.5**	**5.1**
	AVG	5.6	5.6	**4.6**	**4.9**	**5.0**	**5.0**	**4.5**	**5.0**
$(0.05, 0.10)$	N	18	18	20	28	16	16	18	28
	MB	28.3	25.3	16.4	10.6	7.7	7.5	7.9	7.2
	LZ	8.9	8.8	7.3	7.0	7.8	7.6	8.1	7.4
	MD	**5.0**	**4.7**	4.0	**4.8**	4.3	4.2	**4.9**	**5.2**
	KC	6.7	6.5	**5.5**	5.9	6.1	5.8	6.3	6.2
	FG	5.6	**5.3**	**4.8**	**5.2**	**5.4**	**5.1**	5.6	5.6
	AVG	5.7	**5.5**	**4.8**	**5.3**	**5.2**	**4.9**	5.6	5.7
$(0.05, 0.20)$	N	30	28	28	36	24	24	26	34
	MB	38.6	34.7	24.0	14.0	8.0	7.4	6.5	6.1
	LZ	8.1	7.8	7.2	6.2	7.3	6.7	6.5	6.3
	MD	**5.5**	**5.1**	**4.8**	4.4	**5.4**	**4.7**	**4.6**	**4.5**
	KC	6.7	6.2	5.8	**5.3**	6.3	5.6	**5.4**	**5.4**
	FG	5.9	**5.3**	**5.0**	**4.8**	6.0	**5.4**	**5.1**	**5.1**
	AVG	6.0	5.6	**5.1**	**4.9**	5.8	**5.2**	**5.0**	**5.1**
$(0.10, 0.10)$	N	20	20	22	30	18	18	20	30
	MB	33.7	30.5	20.1	12.1	8.2	7.8	6.8	6.6
	LZ	8.0	7.8	7.6	6.6	7.8	7.7	7.1	6.6
	MD	**4.8**	**4.5**	**4.5**	**4.6**	**4.9**	**4.7**	4.2	**4.7**
	KC	6.3	6.0	5.8	**5.5**	6.3	6.1	5.7	5.7
	FG	**5.5**	**5.4**	**5.2**	**5.1**	**5.5**	**5.3**	**4.8**	**5.1**
	AVG	**5.4**	**5.3**	**5.1**	**5.1**	5.6	**5.4**	**4.9**	**5.2**
$(0.10, 0.20)$	N	30	30	30	38	26	26	28	36
	MB	42.1	38.0	26.9	16.5	6.9	6.5	6.4	6.1
	LZ	7.7	7.3	7.0	6.4	6.7	6.5	6.3	6.2
	MD	**5.2**	**4.9**	**4.5**	**4.6**	**5.0**	**4.5**	**4.7**	**4.6**
	KC	6.5	5.9	5.7	**5.3**	5.7	**5.5**	**5.4**	**5.3**
	FG	5.8	**5.4**	**5.0**	**5.0**	**5.3**	**5.0**	**5.1**	**5.0**
	AVG	5.9	**5.4**	**5.0**	**5.0**	**5.2**	**5.0**	**5.1**	**5.0**
$(0.20, 0.20)$	N	34	32	34	42	28	26	30	40
	MB	48.1	43.7	31.2	19.9	7.4	6.3	6.3	6.2
	LZ	7.4	7.0	6.5	6.6	6.7	6.1	6.0	6.3
	MD	**5.0**	**4.8**	4.4	**4.8**	**5.0**	4.4	**4.6**	**4.9**
	KC	6.2	5.7	**5.5**	5.7	5.8	**5.2**	**5.2**	5.6
	FG	**5.5**	**5.2**	**5.0**	**5.3**	**5.4**	**4.8**	**4.8**	**5.2**
	AVG	**5.5**	**5.2**	**5.0**	**5.2**	**5.4**	**4.8**	**4.8**	**5.2**

Table 4 Empirical type I error rates (%) for GEE and SEE analyses when the mean cluster size $\bar{m} = 25$, cluster size CV $\eta = 0.9$, $(\exp(\beta_0), \exp(\beta_1)) = (1.25, 0.55)$. Empirical type I error rate between 4.5% and 5.5% are considered close to nominal according to the margin of error under a binomial model with 10000 replications, and are highlighted in bold font.

		Independence (GEE)				Arm-specific exchangeable (SEE)			
(σ_0^2, σ_1^2)	Variance	$T=\infty$	$T=4$	$T=2$	$T=1$	$T=\infty$	$T=4$	$T=2$	$T=1$
(0.05, 0.05)	N	16	16	18	26	14	14	18	26
	MB	23.7	21.6	14.1	9.6	10.1	10.1	9.4	8.7
	LZ	11.1	10.9	10.0	8.3	10.8	10.9	10.4	9.2
	MD	**5.1**	**5.1**	**4.8**	**4.9**	**4.8**	**5.0**	**4.7**	**4.9**
	KC	7.7	7.4	7.2	6.5	7.7	7.5	7.4	7.1
	FG	6.4	6.1	6.0	5.6	6.1	6.1	5.8	6.0
	AVG	6.3	6.0	5.8	5.6	6.2	6.0	5.9	6.0
(0.05, 0.10)	N	20	20	24	32	18	18	22	30
	MB	31.7	29.1	20.3	12.4	9.8	9.3	8.3	7.7
	LZ	10.6	10.4	8.9	7.6	9.8	9.6	8.5	8.4
	MD	**5.5**	5.6	**4.7**	**4.6**	5.6	**5.3**	**5.1**	**5.0**
	KC	7.8	7.5	6.6	6.0	7.8	7.5	6.8	6.6
	FG	6.4	6.3	**5.5**	**5.2**	6.8	6.6	6.1	5.8
	AVG	6.6	6.6	5.6	**5.2**	6.6	6.3	5.9	5.7
(0.05, 0.20)	N	36	34	34	42	26	26	28	38
	MB	44.1	40.4	28.7	17.0	9.0	8.4	8.1	6.8
	LZ	9.2	8.9	8.3	7.2	8.1	7.7	8.0	6.9
	MD	5.9	5.7	**5.2**	**4.9**	5.8	**5.4**	5.7	**5.0**
	KC	7.6	7.2	6.7	6.1	6.9	6.5	6.8	6.0
	FG	6.4	6.2	5.7	**5.4**	6.4	6.1	6.4	5.6
	AVG	6.7	6.4	5.9	**5.5**	6.3	5.9	6.2	**5.5**
(0.10, 0.10)	N	24	24	26	34	18	18	22	32
	MB	38.8	35.6	25.4	14.2	9.6	9.1	7.9	7.7
	LZ	9.9	9.7	9.0	7.3	9.5	9.2	8.2	8.1
	MD	**5.4**	**5.3**	**5.1**	**4.8**	5.8	5.6	**4.8**	**5.3**
	KC	7.4	7.3	6.9	5.9	7.6	7.4	6.6	6.6
	FG	6.3	6.3	5.9	**5.3**	6.3	6.3	5.6	5.8
	AVG	6.4	6.3	5.9	**5.3**	6.5	6.5	5.7	5.9
(0.10, 0.20)	N	38	36	38	44	28	28	30	40
	MB	46.4	43.6	31.1	19.8	8.6	7.8	7.3	6.9
	LZ	8.0	8.1	7.3	7.1	7.8	7.5	7.0	6.9
	MD	**5.3**	**5.3**	**4.8**	**5.0**	5.9	5.6	**5.1**	**5.1**
	KC	6.6	6.5	6.0	6.0	6.7	6.5	6.0	6.0
	FG	5.9	5.7	**5.2**	**5.4**	6.3	6.0	5.6	5.6
	AVG	6.0	5.8	**5.3**	**5.4**	6.3	6.0	**5.5**	5.6
(0.20, 0.20)	N	42	40	42	50	28	28	34	44
	MB	52.6	48.1	37.4	24.2	8.8	7.8	7.4	6.6
	LZ	8.2	7.8	7.4	6.9	7.7	7.3	6.8	6.5
	MD	5.6	**5.2**	**5.0**	**5.1**	5.8	**5.5**	**5.1**	**5.0**
	KC	6.9	6.4	6.3	6.0	6.8	6.4	5.9	5.7
	FG	6.2	5.7	5.6	**5.5**	6.3	5.9	**5.5**	**5.3**
	AVG	6.2	5.7	5.6	**5.5**	6.3	5.9	**5.5**	**5.3**

Table 5 Difference between empirical power and predicted power for GEE and SEE analyses when the mean cluster size $\bar{m} = 25$, cluster size CV $\eta = 0.3$, $(\exp(\beta_0), \exp(\beta_1)) = (1.25, 0.55)$. Based on 80% nominal power, difference within 0.8% are considered close to nominal according to the margin of error under a binomial model with 10000 replications, and are highlighted in bold font.

		Independence (GEE)				Arm-specific exchangeable (SEE)			
(σ_0^2, σ_1^2)	Variance	$T = \infty$	$T = 4$	$T = 2$	$T = 1$	$T = \infty$	$T = 4$	$T = 2$	$T = 1$
$(0.05, 0.05)$	N	12	12	14	22	12	12	14	22
	MB	12.2	12.4	12.0	7.3	4.3	4.1	4.0	2.9
	LZ	5.5	5.1	4.8	3.2	4.5	4.2	4.3	3.1
	MD	-4.6	-5.2	-5.2	-2.9	-4.7	-5.3	-5.1	-2.6
	KC	0.9	**0.6**	**0.2**	**0.7**	**0.6**	**0.1**	**0.4**	**0.3**
	FG	**-0.7**	-1.2	-1.9	-1.0	-2.2	-2.6	-1.7	**-0.8**
	AVG	-1.7	-2.0	-2.4	-1.1	-2.0	-2.4	-2.4	-1.0
$(0.05, 0.10)$	N	16	16	18	26	16	16	18	26
	MB	13.4	13.5	13.5	10.0	3.2	3.4	4.2	3.0
	LZ	4.1	3.9	4.7	3.5	3.4	3.5	4.6	3.2
	MD	-3.1	-3.5	-2.1	-1.3	-2.6	-2.9	-1.6	-0.9
	KC	1.0	**0.7**	1.8	1.1	**0.8**	0.9	1.6	1.3
	FG	**-0.5**	-1.0	**-0.1**	**0.0**	**-0.5**	**-0.4**	**0.7**	**0.6**
	AVG	**-0.8**	-1.4	**-0.2**	**0.0**	-0.9	-1.0	**0.2**	**0.1**
$(0.05, 0.20)$	N	24	24	26	32	24	22	24	32
	MB	17.5	16.0	13.4	12.7	2.9	3.4	4.0	3.6
	LZ	3.8	3.2	1.0	3.5	3.3	3.9	4.2	3.6
	MD	-0.9	-1.6	-4.7	**-0.5**	-1.0	-1.4	**-0.8**	**0.0**
	KC	1.7	**0.8**	-1.6	1.6	1.1	1.4	1.9	1.8
	FG	**0.6**	**-0.4**	-3.1	**0.4**	**0.6**	**0.8**	1.3	1.5
	AVG	**0.5**	**-0.4**	-3.0	**0.5**	**0.1**	**-0.2**	**0.5**	1.0
$(0.10, 0.10)$	N	16	18	20	28	16	16	20	28
	MB	17.1	12.9	13.0	9.9	4.1	3.5	2.9	2.6
	LZ	5.1	3.5	3.4	2.8	4.2	3.8	3.2	2.7
	MD	-3.3	-2.6	-2.7	-1.4	-2.8	-3.9	-2.3	-1.0
	KC	1.4	**0.7**	**0.7**	**0.8**	1.1	**0.3**	**0.8**	1.0
	FG	**-0.2**	**-0.4**	**-0.7**	**-0.3**	**-0.8**	-1.5	**-0.3**	**0.1**
	AVG	-0.9	**-0.8**	-0.9	**-0.3**	**-0.8**	-1.8	**-0.6**	**-0.2**
$(0.10, 0.20)$	N	26	26	26	34	24	24	26	34
	MB	16.7	15.4	16.4	12.8	3.3	2.6	3.5	2.6
	LZ	3.9	3.1	4.0	2.9	3.6	3.1	3.6	2.7
	MD	**-0.5**	-1.3	-1.3	-0.9	**-0.8**	-1.5	**-0.4**	**-0.4**
	KC	1.7	1.1	1.3	1.1	1.5	**0.8**	1.7	1.3
	FG	**0.7**	**0.0**	**0.1**	**0.0**	**0.8**	**0.1**	1.0	**0.7**
	AVG	**0.6**	**-0.1**	**0.1**	**0.0**	**0.4**	**-0.3**	**0.6**	**0.5**
$(0.20, 0.20)$	N	28	28	30	38	26	26	28	38
	MB	16.6	15.3	15.1	13.3	3.6	2.6	1.5	1.5
	LZ	3.4	1.9	2.1	1.9	3.9	2.7	2.1	1.6
	MD	-1.0	-2.4	-2.2	-1.5	**-0.2**	-1.4	-2.0	-1.1
	KC	1.3	**-0.2**	**0.0**	**0.4**	1.8	0.9	**0.1**	**0.5**
	FG	**0.5**	-0.9	-0.9	**-0.4**	0.9	**-0.1**	**-0.7**	**0.0**
	AVG	**0.1**	-1.3	-0.9	**-0.5**	**0.8**	**-0.3**	-0.9	**-0.2**

Table 6 Difference between empirical power and predicted power for GEE and SEE analyses when the mean cluster size $\bar{m} = 25$, cluster size CV $\eta = 0.6$, $(\exp(\beta_0), \exp(\beta_1)) = (1.25, 0.55)$. Based on 80% nominal power, difference within 0.8% are considered close to nominal according to the margin of error under a binomial model with 10000 replications, and are highlighted in bold font.

		Independence (GEE)				Arm-specific exchangeable (SEE)			
(σ_0^2, σ_1^2)	Variance	$T = \infty$	$T = 4$	$T = 2$	$T = 1$	$T = \infty$	$T = 4$	$T = 2$	$T = 1$
	N	14	14	16	24	12	12	16	24
$(0.05, 0.05)$	MB	11.0	11.1	11.3	7.7	4.8	4.6	3.1	2.6
	LZ	5.5	5.3	5.0	3.2	5.7	5.4	3.8	2.9
	MD	-4.1	-4.6	-5.5	-3.7	-6.2	-6.8	-4.6	-3.3
	KC	1.5	1.1	**0.6**	**0.4**	1.0	**0.5**	**0.4**	**0.4**
	FG	**-0.3**	**-0.8**	-1.9	-1.4	-2.9	-3.1	-1.9	-1.1
	AVG	-1.3	-1.7	-2.5	-1.7	-2.6	-3.1	-2.1	-1.3
	N	18	18	20	28	16	16	18	28
$(0.05, 0.10)$	MB	13.9	14.0	14.1	10.0	4.3	4.1	3.8	2.5
	LZ	4.8	4.8	5.3	3.1	4.9	4.8	4.5	3.0
	MD	-2.8	-3.0	-3.3	-2.6	-1.7	-2.3	-2.5	-1.8
	KC	1.6	1.2	1.7	**0.4**	2.1	1.9	1.2	0.9
	FG	**-0.1**	**-0.5**	**-0.7**	-1.1	**0.6**	**0.5**	**-0.1**	**0.1**
	AVG	**-0.5**	**-0.8**	-0.9	-1.1	**0.1**	**-0.1**	**-0.6**	**-0.4**
	N	30	28	28	36	24	24	26	34
$(0.05, 0.20)$	MB	15.9	16.5	16.7	13.2	3.1	3.0	3.8	3.9
	LZ	4.3	4.1	4.7	3.7	3.7	3.5	4.4	4.1
	MD	**-0.2**	-1.3	-1.7	**-0.6**	**-0.6**	**-0.7**	**-0.4**	0.9
	KC	2.2	1.7	1.7	1.7	1.7	1.4	2.2	2.8
	FG	1.0	**0.2**	**-0.2**	**0.3**	1.2	1.0	1.5	2.3
	AVG	1.1	**0.3**	**-0.1**	**0.5**	**0.6**	**0.4**	**0.8**	1.9
	N	20	20	22	30	18	18	20	30
$(0.10, 0.10)$	MB	14.7	14.2	14.2	11.3	3.1	2.8	3.3	2.3
	LZ	5.7	4.9	4.6	3.0	3.9	3.1	4.1	2.8
	MD	-1.4	-2.3	-3.6	-2.8	-1.6	-2.3	-2.1	-1.3
	KC	2.6	1.7	0.9	**0.0**	1.3	**0.8**	1.5	0.9
	FG	1.3	**0.2**	**-0.8**	-1.3	**0.0**	**-0.5**	**0.2**	**0.0**
	AVG	**0.7**	**-0.3**	-1.1	-1.3	**0.0**	**-0.6**	**-0.1**	**-0.2**
	N	30	30	30	38	26	26	28	36
$(0.10, 0.20)$	MB	17.8	16.3	16.8	13.4	3.2	2.5	2.3	3.2
	LZ	5.1	3.9	4.5	3.3	3.7	3.3	3.0	3.7
	MD	**0.0**	-1.3	-1.0	-1.0	**-0.5**	-1.1	-1.1	**0.4**
	KC	2.8	1.8	2.1	1.5	1.6	1.1	1.0	2.1
	FG	1.5	**0.4**	**0.5**	**0.1**	**0.8**	**0.3**	**0.4**	1.6
	AVG	1.4	**0.2**	**0.5**	**0.2**	**0.5**	**-0.1**	**-0.1**	1.3
	N	34	32	34	42	28	26	30	40
$(0.20, 0.20)$	MB	16.0	17.0	16.8	14.2	2.0	2.1	2.0	2.0
	LZ	2.8	3.5	3.4	2.4	2.9	2.6	2.7	2.5
	MD	-1.8	-1.6	-1.5	-2.1	**-0.8**	-1.6	-0.9	**-0.4**
	KC	**0.6**	1.3	1.1	**0.3**	1.2	**0.5**	**0.8**	1.2
	FG	**-0.2**	**0.4**	**-0.2**	-0.9	**0.2**	**-0.4**	**0.1**	**0.6**
	AVG	**-0.5**	**-0.2**	**-0.2**	-1.0	**0.1**	**-0.6**	**0.0**	**0.4**

Table 7 Difference between empirical power and predicted power for GEE and SEE analyses when the mean cluster size $\bar{m} = 25$, cluster size CV $\eta = 0.9$, $(\exp(\beta_0), \exp(\beta_1)) = (1.25, 0.55)$. Based on 80% nominal power, difference within 0.8% are considered close to nominal according to the margin of error under a binomial model with 10000 replications, and are highlighted in bold font.

		Independence (GEE)				Arm-specific exchangeable (SEE)			
(σ_0^2, σ_1^2)	Variance	$T=\infty$	$T=4$	$T=2$	$T=1$	$T=\infty$	$T=4$	$T=2$	$T=1$
	N	16	16	18	26	14	14	18	26
$(0.05, 0.05)$	MB	12.1	12.0	11.8	8.6	4.0	4.1	3.2	2.8
	LZ	7.0	6.8	5.4	4.0	5.1	5.2	4.1	3.5
	MD	-4.3	-5.1	-5.8	-4.3	-4.3	-5.2	-3.7	-3.5
	KC	2.9	2.3	1.0	**0.4**	1.3	1.3	1.1	**0.8**
	FG	**0.8**	**-0.1**	-1.4	-1.7	-2.2	-2.4	-1.2	-1.0
	AVG	**-0.6**	-1.2	-2.3	-1.9	-1.3	-2.1	-1.2	-1.3
	N	20	20	24	32	18	18	22	30
$(0.05, 0.10)$	MB	17.1	17.0	13.9	11.2	4.2	4.5	4.9	4.0
	LZ	8.6	8.2	6.0	4.3	5.1	5.2	5.7	4.8
	MD	-1.4	-2.0	-2.7	-2.5	-0.9	-1.0	**0.0**	**-0.2**
	KC	4.7	4.2	2.5	1.4	2.6	2.5	3.5	2.6
	FG	2.7	2.0	**0.1**	**-0.6**	1.2	1.2	2.3	1.9
	AVG	2.0	1.1	**0.1**	**-0.4**	1.1	**0.8**	1.9	1.3
	N	36	34	34	42	26	26	28	38
$(0.05, 0.20)$	MB	17.6	17.8	17.4	14.4	4.1	3.4	4.5	4.4
	LZ	5.9	6.2	6.1	4.3	5.0	4.6	5.5	5.3
	MD	1.1	**0.4**	**-0.3**	-1.5	**0.6**	**0.3**	1.3	2.1
	KC	3.7	3.6	3.4	1.7	3.0	2.6	3.8	3.8
	FG	2.4	1.8	1.2	**0.0**	2.4	2.1	3.3	3.4
	AVG	2.4	1.9	1.5	**0.2**	1.8	1.5	2.7	3.1
	N	24	24	26	34	18	18	22	32
$(0.10, 0.10)$	MB	15.7	15.1	15.2	12.5	4.4	4.1	4.9	4.1
	LZ	7.3	6.4	5.8	3.9	5.6	5.4	5.9	4.8
	MD	**-0.7**	-1.9	-2.8	-2.5	-1.1	-2.0	**-0.3**	**0.0**
	KC	3.9	2.8	1.9	0.9	2.7	2.5	3.5	2.8
	FG	2.5	1.2	**-0.1**	**-0.6**	**0.3**	**0.1**	1.6	1.4
	AVG	1.5	**0.4**	**-0.6**	**-0.7**	**0.7**	**0.3**	1.6	1.4
	N	38	36	38	44	28	28	30	40
$(0.10, 0.20)$	MB	17.8	18.1	16.5	15.1	2.4	2.4	3.5	4.7
	LZ	6.0	5.8	5.4	4.4	3.2	3.1	4.4	5.2
	MD	1.1	**0.0**	**-0.4**	-1.3	**-0.5**	**-0.7**	**0.5**	2.0
	KC	3.6	3.0	2.9	1.7	1.6	1.4	2.6	3.8
	FG	2.6	1.6	1.4	**0.2**	**0.8**	**0.5**	2.0	3.2
	AVG	2.5	1.4	1.2	**0.0**	**0.5**	**0.2**	1.6	2.9
	N	42	40	42	50	28	28	34	44
$(0.20, 0.20)$	MB	17.4	17.9	17.3	15.6	2.1	1.4	2.7	3.4
	LZ	5.5	5.4	3.4	3.2	3.1	2.2	3.3	3.6
	MD	**0.5**	**0.0**	-2.5	-1.6	**-0.6**	-1.9	**0.1**	**0.8**
	KC	3.2	2.8	**0.7**	1.0	1.2	**0.4**	1.7	2.3
	FG	2.4	1.8	**-0.7**	**-0.1**	**0.2**	**-0.7**	1.1	1.6
	AVG	1.9	1.4	-1.0	**-0.2**	**0.3**	**-0.8**	0.9	1.5

Table 8 True values induced from the conditional model for power calculation when $(\exp(\beta_0), \exp(\beta_1)) = (2.70, 0.60)$ and under different right truncation points.

(σ_0^2, σ_1^2)	Parameters	$T = \infty$	$T = 6$	$T = 3$	$T = 1$
$(0.05, 0.05)$	$(\mu_{(0)}, \exp(\Delta))$	(2.769,0.600)	(2.623,0.627)	(1.859,0.740)	(0.728,0.848)
	$(\tau_{(0)}, \tau_{(1)})$	(3.162,1.802)	(2.438,1.715)	(0.944,0.993)	(0.198,0.236)
	$(\kappa_{(0)}, \kappa_{(1)})$	(0.642,0.808)	(0.595,0.796)	(0.523,0.724)	(0.612,0.788)
	$(\rho_{(0)}, \rho_{(1)})$	(0.124,0.078)	(0.099,0.075)	(0.043,0.045)	(0.010,0.012)
$(0.05, 0.10)$	$(\mu_{(0)}, \exp(\Delta))$	(2.769,0.615)	(2.623,0.640)	(1.859,0.742)	(0.728,0.846)
	$(\tau_{(0)}, \tau_{(1)})$	(3.162,2.007)	(2.438,1.851)	(0.944,1.014)	(0.198,0.237)
	$(\kappa_{(0)}, \kappa_{(1)})$	(0.642,0.832)	(0.595,0.811)	(0.523,0.730)	(0.612,0.648)
	$(\rho_{(0)}, \rho_{(1)})$	(0.124,0.152)	(0.099,0.141)	(0.043,0.085)	(0.010,0.023)
$(0.05, 0.20)$	$(\mu_{(0)}, \exp(\Delta))$	(2.769,0.646)	(2.623,0.662)	(1.859,0.745)	(0.728,0.843)
	$(\tau_{(0)}, \tau_{(1)})$	(3.162,2.499)	(2.438,2.109)	(0.944,1.052)	(0.198,0.237)
	$(\kappa_{(0)}, \kappa_{(1)})$	(0.642,0.883)	(0.595,0.837)	(0.523,0.740)	(0.612,0.794)
	$(\rho_{(0)}, \rho_{(1)})$	(0.124,0.284)	(0.099,0.249)	(0.043,0.152)	(0.010,0.044)
$(0.10, 0.10)$	$(\mu_{(0)}, \exp(\Delta))$	(2.840,0.600)	(2.642,0.635)	(1.852,0.745)	(0.725,0.849)
	$(\tau_{(0)}, \tau_{(1)})$	(3.688,2.007)	(2.596,1.851)	(0.965,1.014)	(0.199,0.237)
	$(\kappa_{(0)}, \kappa_{(1)})$	(0.676,0.832)	(0.610,0.811)	(0.531,0.730)	(0.615,0.790)
	$(\rho_{(0)}, \rho_{(1)})$	(0.230,0.152)	(0.177,0.141)	(0.082,0.085)	(0.019,0.023)
$(0.10, 0.20)$	$(\mu_{(0)}, \exp(\Delta))$	(2.840,0.630)	(2.642,0.657)	(1.852,0.748)	(0.725,0.845)
	$(\tau_{(0)}, \tau_{(1)})$	(3.688,2.499)	(2.596,2.109)	(0.965,1.052)	(0.199,0.237)
	$(\kappa_{(0)}, \kappa_{(1)})$	(0.676,0.883)	(0.610,0.837)	(0.531,0.740)	(0.615,0.794)
	$(\rho_{(0)}, \rho_{(1)})$	(0.230,0.284)	(0.177,0.249)	(0.082,0.152)	(0.019,0.044)
$(0.20, 0.20)$	$(\mu_{(0)}, \exp(\Delta))$	(2.986,0.600)	(2.673,0.649)	(1.839,0.753)	(0.721,0.850)
	$(\tau_{(0)}, \tau_{(1)})$	(4.960,2.499)	(2.869,2.109)	(1.004,1.052)	(0.201,0.237)
	$(\kappa_{(0)}, \kappa_{(1)})$	(0.746,0.883)	(0.634,0.837)	(0.545,0.740)	(0.622,0.794)
	$(\rho_{(0)}, \rho_{(1)})$	(0.398,0.284)	(0.292,0.249)	(0.148,0.152)	(0.038,0.044)

Table 9 Empirical type I error rates (%) for GEE and SEE analyses when the mean cluster size $\bar{m} = 50$, cluster size CV $\eta = 0$, $(\exp(\beta_0), \exp(\beta_1)) = (2.70, 0.60)$. Empirical type I error rate between 4.5% and 5.5% are considered close to nominal according to the margin of error under a binomial model with 10000 replications, and are highlighted in bold font.

		Independence (GEE)				Arm-specific exchangeable (SEE)			
(σ_0^2, σ_1^2)	Variance	$T = \infty$	$T = 6$	$T = 3$	$T = 1$	$T = \infty$	$T = 6$	$T = 3$	$T = 1$
	N	10	10	12	20	10	10	12	20
$(0.05, 0.05)$	MB	38.2	34.2	20.9	8.4	7.3	7.0	7.0	6.1
	LZ	7.3	7.0	7.0	6.1	7.3	7.0	7.0	6.1
	MD	3.5	3.2	3.7	4.0	3.5	3.2	3.7	4.0
	KC	**5.1**	**4.9**	**5.1**	**5.0**	**5.1**	**4.9**	**5.1**	**5.0**
	FG	4.2	4.1	**4.5**	**4.5**	4.4	4.1	4.3	**4.6**
	AVG	4.1	4.0	4.3	4.4	4.1	4.0	4.3	4.4
	N	14	14	16	24	16	14	16	24
$(0.05, 0.10)$	MB	49.7	44.7	29.0	11.3	7.0	6.6	6.7	6.2
	LZ	7.0	6.6	6.7	6.2	7.0	6.6	6.7	6.2
	MD	4.3	3.9	3.8	4.3	4.3	3.9	3.8	4.3
	KC	**5.5**	**5.0**	**5.1**	**5.2**	**5.5**	**5.0**	**5.1**	**5.2**
	FG	**4.9**	4.4	4.4	**4.8**	**5.1**	**4.6**	**4.5**	**4.8**
	AVG	**4.9**	4.4	4.4	**4.8**	**4.9**	4.4	4.4	**4.8**
	N	28	24	22	30	28	24	22	30
$(0.05, 0.20)$	MB	59.7	52.9	38.6	16.7	6.7	6.4	6.2	5.8
	LZ	6.7	6.4	6.2	5.8	6.7	6.4	6.2	5.8
	MD	**5.1**	**4.6**	4.2	4.3	**5.1**	**4.6**	4.2	4.3
	KC	5.8	**5.4**	**5.1**	**5.1**	5.8	**5.4**	**5.1**	**5.1**
	FG	**5.3**	**4.9**	**4.5**	**4.7**	5.6	**5.2**	**4.7**	**4.8**
	AVG	**5.4**	**5.0**	**4.6**	**4.7**	**5.4**	**5.0**	**4.6**	**4.7**
	N	16	16	18	28	16	16	18	28
$(0.10, 0.10)$	MB	54.3	49.5	35.0	14.3	7.2	6.9	6.1	**5.4**
	LZ	7.2	6.9	6.1	**5.4**	7.2	6.9	6.1	**5.4**
	MD	4.3	4.1	3.7	3.8	4.3	4.1	3.7	3.8
	KC	5.6	**5.4**	**4.8**	**4.5**	5.6	**5.4**	**4.8**	**4.5**
	FG	**5.1**	**4.7**	4.2	4.1	**5.0**	**4.7**	4.2	4.2
	AVG	**5.0**	**4.6**	4.2	4.1	**5.0**	**4.6**	4.2	4.1
	N	28	26	26	34	28	26	26	34
$(0.10, 0.20)$	MB	62.0	55.7	41.9	18.2	6.2	5.7	**5.3**	**5.3**
	LZ	6.2	5.7	**5.3**	**5.3**	6.2	5.7	**5.3**	**5.3**
	MD	**4.7**	4.0	3.8	4.1	**4.7**	4.0	3.8	4.1
	KC	**5.3**	**4.6**	**4.5**	**4.7**	**5.3**	**4.6**	**4.5**	**4.7**
	FG	**4.9**	4.2	4.1	4.4	**5.1**	4.4	4.2	4.4
	AVG	**5.0**	4.3	4.1	4.4	**5.0**	4.3	4.1	4.4
	N	30	28	30	40	30	28	30	40
$(0.20, 0.20)$	MB	65.3	60.8	47.5	23.0	6.6	5.7	**5.5**	**5.4**
	LZ	6.6	5.7	**5.5**	**5.4**	6.6	5.7	**5.5**	**5.4**
	MD	**5.1**	4.0	4.1	4.2	**5.1**	4.0	4.1	4.2
	KC	5.9	**4.8**	**4.8**	**4.8**	5.9	**4.8**	**4.8**	**4.8**
	FG	**5.5**	4.4	4.4	**4.5**	**5.5**	4.4	4.3	**4.5**
	AVG	**5.5**	4.4	4.3	**4.5**	**5.5**	4.4	4.3	**4.5**

Table 10 Empirical type I error rates (%) for GEE and SEE analyses when the mean cluster size $\bar{m} = 50$, cluster size CV $\eta = 0.3$, $(\exp(\beta_0), \exp(\beta_1)) = (2.70, 0.60)$. Empirical type I error rate between 4.5% and 5.5% are considered close to nominal according to the margin of error under a binomial model with 10000 replications, and are highlighted in bold font.

		Independence (GEE)				Arm-specific exchangeable (SEE)			
(σ_0^2, σ_1^2)	Variance	$T = \infty$	$T = 6$	$T = 3$	$T = 1$	$T = \infty$	$T = 6$	$T = 3$	$T = 1$
	N	12	12	12	22	10	10	12	22
$(0.05, 0.05)$	MB	41.6	37.2	22.1	9.1	8.0	7.7	7.2	6.5
	LZ	7.8	7.7	7.4	6.5	8.0	7.6	7.2	6.6
	MD	3.8	3.6	3.9	4.2	3.5	3.4	3.7	4.4
	KC	**5.5**	**5.5**	**5.5**	**5.3**	**5.5**	**5.3**	**5.3**	**5.4**
	FG	**4.6**	**4.6**	**4.8**	**4.7**	**4.5**	4.2	4.3	**4.8**
	AVG	**4.5**	**4.5**	**4.6**	**4.7**	4.4	4.2	4.4	**4.8**
	N	16	16	16	26	14	14	16	24
$(0.05, 0.10)$	MB	50.1	45.1	29.5	12.1	8.0	7.4	6.7	6.2
	LZ	7.4	7.0	7.0	6.0	7.6	7.2	6.6	6.3
	MD	4.4	4.3	3.9	4.1	4.4	4.0	3.9	4.4
	KC	5.7	5.6	**5.3**	**5.0**	5.9	5.7	**5.2**	**5.4**
	FG	**4.9**	**4.7**	**4.5**	4.3	**5.1**	**5.0**	**4.6**	**4.9**
	AVG	**5.0**	**4.8**	**4.6**	4.4	**5.0**	**4.9**	**4.6**	**4.8**
	N	30	26	24	32	28	24	24	30
$(0.05, 0.20)$	MB	60.4	55.0	39.4	17.3	7.6	6.2	5.8	5.7
	LZ	7.2	6.8	6.1	6.1	7.3	5.9	5.7	5.8
	MD	**5.5**	**4.9**	4.0	**4.6**	**5.4**	4.0	3.9	4.3
	KC	6.3	5.7	**4.9**	**5.4**	6.3	**4.9**	**4.8**	**5.1**
	FG	5.8	**5.1**	4.3	**4.9**	6.1	**4.7**	**4.5**	**4.8**
	AVG	5.9	**5.2**	4.4	**5.0**	5.8	4.4	4.4	**4.7**
	N	18	18	20	28	18	16	18	28
$(0.10, 0.10)$	MB	56.0	51.1	36.3	15.1	7.0	6.6	6.1	6.2
	LZ	7.2	6.8	6.9	6.4	6.7	6.5	6.1	6.3
	MD	4.3	4.2	**4.6**	**4.7**	4.3	3.8	3.7	**4.7**
	KC	5.7	**5.4**	5.8	5.6	5.6	**5.0**	**4.9**	**5.5**
	FG	**5.0**	**4.8**	**5.2**	**5.1**	**5.0**	4.3	4.2	**5.0**
	AVG	**5.0**	**4.8**	**5.1**	**5.1**	**5.0**	4.3	4.2	**5.1**
	N	30	28	28	34	28	26	26	34
$(0.10, 0.20)$	MB	63.1	58.1	44.4	19.8	7.1	6.3	6.0	5.8
	LZ	7.1	6.4	6.3	5.9	6.8	6.2	5.8	5.8
	MD	**5.2**	**4.5**	**4.5**	4.4	**5.1**	4.3	4.3	4.4
	KC	6.1	**5.5**	**5.2**	**5.1**	5.8	**5.2**	**4.9**	**5.0**
	FG	**5.5**	**4.9**	**4.9**	**4.7**	**5.5**	**4.9**	**4.6**	**4.8**
	AVG	5.6	**5.1**	**4.9**	**4.7**	**5.5**	**4.7**	**4.5**	**4.7**
	N	34	32	32	42	30	30	30	40
$(0.20, 0.20)$	MB	67.1	61.4	49.0	24.4	6.9	6.1	5.6	**5.5**
	LZ	6.5	6.0	5.9	**5.5**	6.6	5.7	**5.5**	5.6
	MD	**5.0**	**4.5**	4.2	4.2	**4.9**	4.4	3.9	4.2
	KC	5.6	**5.1**	**5.0**	**4.8**	5.8	**5.1**	**4.6**	**4.8**
	FG	**5.3**	**4.8**	**4.7**	**4.5**	**5.4**	**4.7**	4.2	**4.5**
	AVG	**5.3**	**4.9**	**4.6**	**4.5**	**5.4**	**4.7**	4.3	**4.5**

Table 11 Empirical type I error rates (%) for GEE and SEE analyses when the mean cluster size $\bar{m} = 50$, cluster size CV $\eta = 0.6$, $(\exp(\beta_0), \exp(\beta_1)) = (2.70, 0.60)$. Empirical type I error rate between 4.5% and 5.5% are considered close to nominal according to the margin of error under a binomial model with 10000 replications, and are highlighted in bold font.

		Independence (GEE)				Arm-specific exchangeable (SEE)			
(σ_0^2, σ_1^2)	Variance	$T=\infty$	$T=6$	$T=3$	$T=1$	$T=\infty$	$T=6$	$T=3$	$T=1$
$(0.05, 0.05)$	N	12	12	14	22	10	10	12	22
	MB	43.9	39.6	25.9	10.2	8.7	8.7	8.0	6.8
	LZ	9.9	9.4	9.1	7.0	8.7	8.7	8.1	7.2
	MD	**4.5**	4.3	4.3	3.9	3.6	3.6	3.8	4.2
	KC	6.8	6.6	6.3	**5.4**	6.0	5.8	5.8	5.7
	FG	**5.4**	**5.3**	**5.5**	**4.8**	**4.7**	**4.6**	**4.6**	**4.9**
	AVG	**5.4**	**5.3**	**5.2**	**4.8**	**4.6**	**4.7**	**4.7**	**4.9**
$(0.05, 0.10)$	N	18	18	20	28	16	16	16	26
	MB	53.5	49.6	34.4	14.4	8.2	7.7	7.6	6.5
	LZ	9.0	8.7	8.2	7.1	7.4	7.1	7.4	6.5
	MD	**5.1**	**4.8**	**4.7**	**4.7**	4.4	4.4	4.3	4.4
	KC	6.7	6.5	6.3	5.9	5.9	5.7	5.7	**5.4**
	FG	5.7	**5.5**	**5.4**	**5.2**	**5.1**	**5.1**	**5.0**	**4.9**
	AVG	5.8	5.7	**5.4**	**5.2**	**5.1**	**5.0**	**4.9**	**4.9**
$(0.05, 0.20)$	N	36	30	30	36	28	24	24	32
	MB	63.7	58.8	44.2	19.6	8.2	6.8	6.8	6.6
	LZ	7.9	7.6	7.3	6.5	7.1	6.0	6.3	6.7
	MD	5.7	**5.3**	**4.8**	**4.7**	**5.2**	4.4	4.4	**4.8**
	KC	6.6	6.3	6.0	5.6	6.0	**5.2**	**5.2**	5.7
	FG	5.9	5.6	**5.1**	**5.1**	5.8	**4.9**	**4.9**	**5.5**
	AVG	6.2	5.8	**5.3**	**5.1**	5.6	**4.7**	**4.8**	**5.3**
$(0.10, 0.10)$	N	22	20	22	32	18	16	20	30
	MB	60.6	54.9	40.5	17.2	8.1	7.5	7.1	6.2
	LZ	8.2	8.3	7.0	6.6	7.1	7.0	7.0	6.4
	MD	**4.8**	**4.8**	4.2	4.4	**4.5**	4.1	4.3	4.4
	KC	6.4	6.3	5.6	**5.5**	5.7	**5.5**	**5.5**	**5.4**
	FG	**5.5**	5.6	**4.8**	**4.9**	**5.0**	**4.8**	**4.9**	**4.8**
	AVG	**5.4**	**5.5**	**4.7**	**5.0**	**5.1**	**4.8**	**4.9**	**4.9**
$(0.10, 0.20)$	N	38	34	32	40	28	26	26	36
	MB	65.2	60.6	49.2	23.0	7.9	6.1	6.2	6.2
	LZ	7.8	7.0	6.9	6.2	6.8	**5.5**	5.8	6.4
	MD	5.6	**5.0**	**4.8**	**4.6**	**5.1**	4.0	3.9	**4.9**
	KC	6.6	5.9	5.7	**5.2**	6.0	**4.7**	**4.8**	5.6
	FG	6.0	**5.4**	**5.1**	**4.9**	5.6	4.3	**4.5**	**5.3**
	AVG	6.1	**5.5**	**5.2**	**4.9**	**5.5**	4.2	4.4	**5.2**
$(0.20, 0.20)$	N	40	38	40	48	32	30	32	42
	MB	69.8	64.4	53.3	27.9	7.4	6.3	6.5	5.7
	LZ	6.9	7.1	6.3	6.3	6.5	5.7	5.9	5.7
	MD	**5.2**	**4.9**	4.4	**4.9**	**4.8**	4.2	**4.6**	**4.6**
	KC	6.0	5.9	**5.3**	**5.5**	5.7	**5.0**	**5.2**	**5.1**
	FG	**5.5**	**5.5**	**5.0**	**5.2**	**5.1**	**4.5**	**4.9**	**4.8**
	AVG	5.6	**5.4**	**5.0**	**5.2**	**5.2**	**4.5**	**4.9**	**4.8**

Table 12 Empirical type I error rates (%) for GEE and SEE analyses when the mean cluster size $\bar{m} = 50$, cluster size CV $\eta = 0.9$, $(\exp(\beta_0), \exp(\beta_1)) = (2.70, 0.60)$. Empirical type I error rate between 4.5% and 5.5% are considered close to nominal according to the margin of error under a binomial model with 10000 replications, and are highlighted in bold font.

		Independence (GEE)				Arm-specific exchangeable (SEE)			
(σ_0^2, σ_1^2)	Variance	$T = \infty$	$T = 6$	$T = 3$	$T = 1$	$T = \infty$	$T = 6$	$T = 3$	$T = 1$
	N	16	16	16	26	12	12	14	24
$(0.05, 0.05)$	MB	50.7	47.0	30.9	12.4	10.4	10.4	9.6	8.5
	LZ	12.1	11.8	11.5	8.3	10.0	9.7	9.8	9.1
	MD	**5.1**	**5.0**	**5.0**	**4.8**	**4.9**	**4.9**	**4.6**	**5.2**
	KC	7.9	7.9	7.7	6.4	7.2	7.4	7.2	7.0
	FG	6.3	6.4	6.4	**5.5**	6.0	5.9	5.8	6.0
	AVG	6.2	6.2	6.2	**5.5**	6.0	6.0	5.6	6.0
	N	24	22	24	32	16	16	18	30
$(0.05, 0.10)$	MB	59.7	55.3	40.0	16.9	9.7	9.2	8.8	7.6
	LZ	10.6	10.5	9.3	7.8	8.3	8.2	8.5	8.0
	MD	5.6	5.6	**5.2**	**4.7**	**4.9**	**4.7**	**5.0**	**5.1**
	KC	8.0	7.9	7.1	6.1	6.4	6.2	6.6	6.6
	FG	6.5	6.6	6.0	**5.3**	**5.5**	**5.4**	5.8	5.9
	AVG	6.7	6.7	6.1	**5.3**	**5.5**	**5.4**	5.8	5.8
	N	46	40	36	44	28	26	24	36
$(0.05, 0.20)$	MB	67.9	63.5	49.9	24.1	9.2	7.5	8.0	7.0
	LZ	8.7	8.2	8.5	7.2	7.6	6.6	7.4	7.2
	MD	6.3	**5.4**	**5.4**	**5.2**	5.8	**4.7**	**5.0**	**5.3**
	KC	7.3	6.8	6.9	6.0	6.6	5.7	6.2	6.2
	FG	6.5	5.8	5.8	**5.4**	6.3	**5.3**	5.8	5.8
	AVG	6.8	6.0	6.1	5.6	6.2	**5.1**	5.7	5.8
	N	28	26	28	38	18	18	20	34
$(0.10, 0.10)$	MB	64.3	59.9	45.7	20.1	9.6	8.7	8.3	6.9
	LZ	9.6	9.1	8.8	6.7	8.1	7.7	7.6	7.0
	MD	5.6	**4.9**	**4.9**	4.2	**4.9**	**4.7**	**4.7**	**4.9**
	KC	7.5	6.8	6.8	**5.3**	6.3	6.2	6.0	5.9
	FG	6.5	5.7	5.9	**4.8**	**5.5**	**5.3**	**5.3**	**5.3**
	AVG	6.4	5.8	5.8	**4.8**	5.6	**5.5**	**5.3**	**5.3**
	N	48	44	42	48	30	28	28	40
$(0.10, 0.20)$	MB	69.9	66.6	53.6	27.1	8.6	7.4	7.4	6.5
	LZ	8.4	7.7	8.1	7.2	7.1	6.6	6.8	6.4
	MD	5.7	**5.3**	**5.2**	**5.0**	**5.4**	**4.7**	**4.7**	**4.9**
	KC	6.9	6.4	6.7	6.0	6.2	**5.5**	5.6	5.6
	FG	6.1	5.7	5.8	**5.4**	5.8	**5.1**	**5.1**	**5.3**
	AVG	6.3	5.8	5.8	**5.5**	5.8	**5.1**	**5.1**	**5.2**
	N	52	48	50	60	32	30	32	48
$(0.20, 0.20)$	MB	73.5	68.8	59.7	32.5	8.5	7.2	6.7	6.4
	LZ	7.8	7.6	7.2	6.2	6.6	6.3	6.1	6.2
	MD	5.6	**5.4**	**5.1**	**4.7**	**5.1**	**4.8**	**4.6**	**5.1**
	KC	6.6	6.5	6.1	**5.5**	5.8	5.7	**5.3**	5.7
	FG	6.0	5.9	**5.5**	**5.1**	**5.4**	**5.1**	**4.9**	**5.5**
	AVG	6.1	5.9	5.6	**5.0**	**5.5**	**5.2**	**5.0**	**5.5**

Table 13 Difference between empirical power and predicted power for GEE and SEE analyses when the mean cluster size $\bar{m} = 50$, cluster size CV $\eta = 0$, $(\exp(\beta_0), \exp(\beta_1)) = (2.70, 0.60)$. Based on 80% nominal power, difference within 0.8% are considered close to nominal according to the margin of error under a binomial model with 10000 replications, and are highlighted in bold font.

		Independence (GEE)				Arm-specific exchangeable (SEE)			
(σ_0^2, σ_1^2)	Variance	$T=\infty$	$T=6$	$T=3$	$T=1$	$T=\infty$	$T=6$	$T=3$	$T=1$
$(0.05, 0.05)$	N	10	10	12	20	10	10	12	20
	MB	16.8	16.4	14.4	10.0	5.5	4.6	4.2	3.2
	LZ	5.5	4.6	4.2	3.2	5.5	4.6	4.2	3.2
	MD	-6.5	-7.6	-6.3	-3.0	-6.5	-7.6	-6.3	-3.0
	KC	**0.5**	**-0.5**	**-0.2**	**0.5**	**0.5**	**-0.5**	**-0.2**	**0.5**
	FG	-1.8	-3.1	-3.0	-1.2	-2.7	-3.4	-2.2	**-0.8**
	AVG	-2.8	-4.1	-3.2	-1.2	-2.8	-4.1	-3.2	-1.2
$(0.05, 0.10)$	N	14	14	16	24	14	14	16	24
	MB	18.6	17.8	15.6	12.5	4.4	4.0	4.3	4.0
	LZ	4.4	4.0	4.3	4.0	4.4	4.0	4.3	4.0
	MD	-4.0	-5.0	-2.4	**-0.7**	-4.0	-5.0	-2.4	**-0.7**
	KC	**0.5**	**-0.1**	1.5	1.8	**0.5**	**-0.1**	1.5	1.8
	FG	-1.5	-2.5	**-0.6**	**0.4**	-0.9	-1.5	**0.6**	1.0
	AVG	-1.8	-2.7	**-0.4**	**0.6**	-1.8	-2.7	**-0.4**	**0.6**
$(0.05, 0.20)$	N	28	24	22	30	28	24	22	30
	MB	16.4	17.5	19.1	14.8	1.8	3.4	4.5	3.8
	LZ	1.8	3.4	4.5	3.8	1.8	3.4	4.5	3.8
	MD	-1.6	-1.1	-1.2	**-0.1**	-1.6	-1.1	-1.2	**-0.1**
	KC	**0.2**	1.2	1.6	2.1	**0.2**	1.2	1.6	2.1
	FG	**-0.7**	**-0.1**	**-0.1**	**0.7**	**-0.1**	**0.7**	1.2	1.6
	AVG	**-0.6**	**0.0**	**0.3**	1.1	**-0.6**	**0.0**	**0.3**	1.1
$(0.10, 0.10)$	N	16	16	18	28	16	16	18	28
	MB	18.9	17.6	16.6	12.0	4.5	3.0	3.4	2.2
	LZ	4.5	3.0	3.4	2.2	4.5	3.0	3.4	2.2
	MD	-2.6	-4.0	-3.3	-1.9	-2.6	-4.0	-3.3	-1.9
	KC	1.0	**-0.3**	**0.4**	**0.3**	1.0	**-0.3**	**0.4**	**0.3**
	FG	**0.0**	-1.8	-1.4	-0.9	**-0.4**	-1.6	-1.0	**-0.5**
	AVG	**-0.5**	-2.0	-1.5	**-0.8**	**-0.5**	-2.0	-1.5	**-0.8**
$(0.10, 0.20)$	N	28	26	26	34	28	26	26	34
	MB	18.5	17.7	16.8	14.3	2.6	2.9	3.7	2.8
	LZ	2.6	2.9	3.7	2.8	2.6	2.9	3.7	2.8
	MD	-1.1	-1.7	**-0.5**	**-0.5**	-1.1	-1.7	**-0.5**	**-0.5**
	KC	**0.7**	**0.7**	1.8	1.3	**0.7**	**0.7**	1.8	1.3
	FG	**-0.2**	**-0.6**	**0.5**	**0.2**	**0.1**	**0.0**	1.1	**0.8**
	AVG	**-0.2**	**-0.6**	**0.6**	**0.5**	**-0.2**	**-0.6**	**0.6**	**0.5**
$(0.20, 0.20)$	N	30	28	30	40	30	28	30	40
	MB	18.9	19.2	18.2	15.0	3.8	1.8	2.6	2.3
	LZ	3.8	1.8	2.6	2.3	3.8	1.8	2.6	2.3
	MD	**0.1**	-2.2	-1.3	**-0.6**	**0.1**	-2.2	-1.3	**-0.6**
	KC	2.2	**-0.1**	**0.7**	0.9	2.2	**-0.1**	**0.7**	0.9
	FG	1.3	-0.9	**-0.3**	**0.1**	1.1	-0.9	**-0.1**	**0.2**
	AVG	1.0	-1.1	**-0.3**	**0.2**	1.0	-1.1	**-0.3**	**0.2**

Table 14 Difference between empirical power and predicted power for GEE and SEE analyses when the mean cluster size $\bar{m} = 50$, cluster size CV $\eta = 0.3$, $(\exp(\beta_0), \exp(\beta_1)) = (2.70, 0.60)$. Based on 80% nominal power, difference within 0.8% are considered close to nominal according to the margin of error under a binomial model with 10000 replications, and are highlighted in bold font.

		Independence (GEE)				Arm-specific exchangeable (SEE)			
(σ_0^2, σ_1^2)	Variance	$T=\infty$	$T=6$	$T=3$	$T=1$	$T=\infty$	$T=6$	$T=3$	$T=1$
$(0.05, 0.05)$	N	12	12	12	22	10	10	12	22
	MB	11.9	11.5	16.3	8.9	6.4	5.4	4.9	2.6
	LZ	4.8	4.0	6.0	2.8	6.3	5.7	5.3	2.7
	MD	-3.7	-4.8	-5.5	-2.3	-5.5	-6.8	-4.9	-2.3
	KC	1.4	**0.3**	1.1	**0.2**	1.3	**0.2**	**0.6**	**0.3**
	FG	**-0.2**	-1.2	-2.0	-1.1	-2.3	-2.7	-1.4	**-0.7**
	AVG	-1.1	-1.9	-2.2	-0.9	-1.9	-3.2	-2.1	-1.0
$(0.05, 0.10)$	N	16	16	16	26	14	14	16	24
	MB	15.7	15.0	18.1	11.7	4.1	3.8	4.9	3.9
	LZ	4.6	4.0	5.8	3.7	4.4	4.0	5.1	4.2
	MD	-2.4	-3.4	-2.8	-1.1	-3.5	-4.3	-2.2	-1.2
	KC	1.3	**0.7**	1.8	1.4	0.9	**0.4**	1.8	1.6
	FG	**-0.4**	-1.4	**-0.7**	**0.0**	**-0.4**	**-0.8**	1.0	**0.8**
	AVG	**-0.5**	-1.4	**-0.6**	**0.1**	-1.1	-1.8	**-0.1**	**0.3**
$(0.05, 0.20)$	N	30	26	24	32	28	24	24	30
	MB	16.6	17.3	18.4	14.6	2.2	3.2	4.8	4.1
	LZ	1.9	3.6	5.5	4.4	2.1	3.2	5.1	4.4
	MD	-1.5	-1.4	**-0.3**	**0.3**	-1.2	-1.4	**0.8**	**0.6**
	KC	**0.3**	1.3	3.0	2.3	**0.5**	1.1	3.0	2.7
	FG	**-0.8**	**-0.2**	0.9	1.0	**0.1**	**0.5**	2.7	2.2
	AVG	**-0.7**	**0.0**	1.5	1.3	**-0.3**	**-0.3**	1.8	1.8
$(0.10, 0.10)$	N	18	18	20	28	18	16	18	28
	MB	16.8	15.7	15.0	13.1	3.4	3.3	3.9	2.5
	LZ	4.8	3.5	3.7	2.7	3.6	3.8	4.2	2.6
	MD	-1.5	-3.0	-2.2	-1.7	-1.4	-3.9	-2.5	-1.1
	KC	2.0	**0.5**	0.9	**0.5**	1.2	**0.4**	1.0	**0.8**
	FG	1.0	**-0.8**	**-0.6**	**-0.6**	**0.1**	-1.5	**-0.2**	**0.1**
	AVG	**0.3**	-1.2	**-0.8**	**-0.5**	**0.0**	-1.8	**-0.6**	**-0.1**
$(0.10, 0.20)$	N	30	28	28	34	28	26	26	34
	MB	18.8	17.6	16.6	16.1	2.5	2.7	3.7	3.1
	LZ	2.7	2.8	3.6	3.7	2.4	2.7	3.9	3.2
	MD	-1.0	-1.6	**-0.5**	**-0.3**	-1.4	-1.5	**-0.1**	**-0.1**
	KC	1.0	0.9	1.7	1.8	**0.6**	**0.8**	2.2	1.6
	FG	**0.2**	**-0.3**	**0.4**	**0.5**	**0.0**	**0.0**	1.5	1.1
	AVG	**0.1**	**-0.4**	**0.6**	**0.7**	**-0.4**	**-0.5**	1.0	**0.8**
$(0.20, 0.20)$	N	34	32	32	42	30	30	30	40
	MB	17.0	16.9	18.6	15.5	2.9	1.3	2.3	1.8
	LZ	3.1	1.6	2.3	2.1	3.2	1.6	2.6	2.0
	MD	**-0.2**	-2.2	-1.7	**-0.8**	**-0.3**	-1.8	-1.3	**-0.8**
	KC	1.5	**-0.1**	**0.4**	**0.8**	1.5	**0.0**	**0.6**	**0.6**
	FG	0.9	-0.9	**-0.7**	**0.0**	0.9	**-0.8**	**-0.2**	**0.1**
	AVG	**0.6**	-1.1	-0.7	**0.0**	**0.7**	-1.0	**-0.3**	**-0.1**

Table 15 Difference between empirical power and predicted power for GEE and SEE analyses when the mean cluster size $\bar{m} = 50$, cluster size CV $\eta = 0.6$, $(\exp(\beta_0), \exp(\beta_1)) = (2.70, 0.60)$. Based on 80% nominal power, difference within 0.8% are considered close to nominal according to the margin of error under a binomial model with 10000 replications, and are highlighted in bold font.

		Independence (GEE)				Arm-specific exchangeable (SEE)			
(σ_0^2, σ_1^2)	Variance	$T = \infty$	$T = 6$	$T = 3$	$T = 1$	$T = \infty$	$T = 6$	$T = 3$	$T = 1$
	N	12	12	14	24	10	10	12	22
$(0.05, 0.05)$	MB	18.2	17.4	15.6	11.5	5.6	5.1	5.6	3.6
	LZ	9.3	8.0	6.2	4.6	6.7	6.1	6.4	3.8
	MD	-3.4	-4.5	-4.9	-3.6	-6.5	-7.4	-4.7	-2.4
	KC	4.3	3.0	1.7	0.9	1.0	**0.1**	1.6	1.2
	FG	1.7	**0.3**	-1.3	-1.1	-3.8	-4.0	**-0.7**	**-0.6**
	AVG	**0.5**	-0.9	-1.6	-1.2	-2.7	-3.7	-1.3	**-0.7**
	N	18	18	20	28	16	16	16	26
$(0.05, 0.10)$	MB	18.1	17.2	15.6	12.8	3.0	2.7	5.3	3.5
	LZ	7.0	6.2	6.3	3.8	3.8	3.8	6.0	4.4
	MD	-1.0	-1.8	-1.6	-2.2	-2.1	-2.5	-2.3	**-0.6**
	KC	3.4	2.7	3.0	1.2	1.1	0.9	2.5	2.0
	FG	1.7	**0.7**	**0.4**	**-0.6**	**0.0**	**-0.2**	1.3	1.2
	AVG	1.2	**0.6**	**0.6**	**-0.3**	**-0.4**	-0.9	**0.1**	**0.7**
	N	36	30	30	36	28	24	24	32
$(0.05, 0.20)$	MB	17.2	19.3	16.9	16.4	2.0	2.3	3.8	3.9
	LZ	3.9	4.7	6.1	5.3	2.2	3.2	5.0	4.6
	MD	**-0.1**	**-0.7**	**0.3**	**0.3**	-1.4	-2.0	**-0.2**	**0.8**
	KC	2.0	2.0	3.6	2.9	**0.5**	**0.8**	2.5	2.8
	FG	**0.7**	**0.4**	1.2	1.3	**0.1**	**0.2**	2.0	2.3
	AVG	**0.8**	**0.6**	1.7	1.7	**-0.4**	**-0.5**	1.3	1.7
	N	22	20	22	32	18	16	20	30
$(0.10, 0.10)$	MB	16.3	18.7	17.9	13.2	2.9	3.1	2.7	2.8
	LZ	5.3	5.8	5.2	3.2	3.7	3.9	3.3	3.3
	MD	**-0.8**	-2.7	-2.6	-1.7	-1.6	-3.7	-2.0	**-0.5**
	KC	2.7	2.2	1.4	0.9	1.3	**0.4**	**0.8**	1.4
	FG	1.5	**0.5**	**-0.4**	**-0.4**	**-0.1**	-1.3	**-0.2**	**0.7**
	AVG	**0.8**	**-0.1**	**-0.5**	**-0.3**	**0.0**	-1.5	**-0.4**	**0.5**
	N	38	34	32	40	28	26	26	36
$(0.10, 0.20)$	MB	17.3	17.7	18.7	16.2	2.4	1.8	3.1	3.1
	LZ	3.6	4.3	5.1	4.1	2.5	2.4	4.0	3.6
	MD	**0.1**	**-0.7**	**-0.6**	**-0.3**	-1.3	-2.0	**-0.8**	**0.6**
	KC	1.9	1.8	2.5	2.0	**0.8**	**0.4**	1.6	2.2
	FG	1.1	**0.5**	**0.6**	**0.7**	**0.1**	**-0.3**	1.1	1.7
	AVG	1.1	**0.6**	0.9	0.9	**-0.2**	**-0.8**	**0.6**	1.4
	N	40	38	40	48	32	30	32	42
$(0.20, 0.20)$	MB	18.6	18.1	17.7	16.3	2.1	1.7	1.9	1.5
	LZ	4.1	2.8	3.0	2.8	2.8	2.0	2.1	1.8
	MD	**0.7**	-1.6	-1.2	-1.0	**-0.3**	-1.6	-1.5	**-0.8**
	KC	2.2	0.9	1.0	1.1	1.5	**0.3**	**0.4**	**0.6**
	FG	1.8	**-0.1**	**0.0**	**0.1**	**0.6**	**-0.6**	**-0.3**	**0.1**
	AVG	1.6	**-0.4**	**-0.1**	**0.1**	**0.7**	**-0.6**	**-0.6**	**-0.1**

Table 16 Difference between empirical power and predicted power for GEE and SEE analyses when the mean cluster size $\bar{m} = 50$, cluster size CV $\eta = 0.9$, $(\exp(\beta_0), \exp(\beta_1)) = (2.70, 0.60)$. Based on 80% nominal power, difference within 0.8% are considered close to nominal according to the margin of error under a binomial model with 10000 replications, and are highlighted in bold font.

		Independence (GEE)				Arm-specific exchangeable (SEE)			
(σ_0^2, σ_1^2)	Variance	$T = \infty$	$T = 6$	$T = 3$	$T = 1$	$T = \infty$	$T = 6$	$T = 3$	$T = 1$
$(0.05, 0.05)$	N	16	16	16	26	12	12	14	24
	MB	15.1	14.3	18.2	11.3	3.2	2.7	4.0	4.3
	LZ	8.4	7.3	9.7	4.7	4.1	3.7	4.8	5.2
	MD	-2.0	-3.2	-3.5	-3.1	-4.3	-5.5	-3.7	-1.9
	KC	4.4	3.3	4.2	1.3	**0.2**	**-0.4**	1.3	2.1
	FG	2.6	0.9	**0.5**	-1.0	-2.9	-3.2	-1.2	**0.4**
	AVG	1.5	**0.1**	**0.3**	**-0.6**	-1.9	-2.8	-1.1	**0.2**
$(0.05, 0.10)$	N	24	22	24	32	16	16	18	30
	MB	16.1	18.6	17.2	14.7	2.4	2.0	3.9	4.1
	LZ	6.9	8.1	7.9	5.9	3.8	3.6	5.3	5.0
	MD	**-0.3**	-1.0	-1.1	-1.1	-2.9	-3.4	-1.4	**0.6**
	KC	3.9	4.1	3.9	2.8	**0.8**	**0.5**	2.3	3.1
	FG	2.0	1.8	1.0	**0.5**	**-0.6**	**-0.7**	1.3	2.3
	AVG	1.7	1.6	1.3	**0.8**	-1.0	-1.3	**0.5**	1.9
$(0.05, 0.20)$	N	46	40	36	44	28	26	24	36
	MB	17.9	18.1	19.4	16.5	1.8	1.9	3.3	4.0
	LZ	4.4	5.7	7.4	6.2	2.1	3.4	4.3	5.0
	MD	**0.4**	**0.6**	**0.8**	1.2	-1.5	**-0.6**	-1.0	1.6
	KC	2.7	3.4	4.5	4.1	**0.4**	1.4	1.8	3.5
	FG	1.5	1.9	2.3	2.2	**-0.2**	1.1	1.2	3.0
	AVG	1.5	2.1	2.7	2.8	**-0.5**	**0.4**	**0.3**	2.6
$(0.10, 0.10)$	N	28	26	28	38	18	18	20	34
	MB	16.5	18.1	17.8	14.5	2.2	1.6	2.1	3.7
	LZ	6.2	6.7	5.9	4.7	3.5	2.4	3.2	4.3
	MD	**-0.1**	-1.1	-2.2	**-0.9**	-2.5	-3.6	-3.1	**0.8**
	KC	3.5	3.3	2.6	2.1	1.0	**-0.1**	**0.2**	2.7
	FG	2.1	1.7	**0.3**	**0.6**	-1.0	-1.8	-1.2	1.9
	AVG	1.7	1.3	**0.2**	**0.6**	**-0.7**	-1.8	-1.4	1.9
$(0.10, 0.20)$	N	48	44	42	48	30	28	28	40
	MB	18.5	17.7	17.7	17.2	1.4	**0.5**	1.9	4.3
	LZ	4.7	4.8	4.9	4.9	1.6	1.2	2.7	5.1
	MD	**0.7**	**-0.1**	-1.1	**-0.4**	-1.8	-2.9	-1.3	1.9
	KC	2.9	2.8	2.1	2.5	**-0.1**	**-0.8**	**0.8**	3.7
	FG	1.8	1.2	**0.3**	**0.8**	**-0.7**	-1.4	**0.2**	3.2
	AVG	1.8	1.2	**0.5**	1.1	-0.9	-1.8	**-0.2**	2.8
$(0.20, 0.20)$	N	52	48	50	60	32	30	32	48
	MB	18.8	19.2	18.9	17.1	2.0	**0.2**	1.0	2.3
	LZ	5.7	3.3	3.5	3.4	2.6	**0.5**	2.0	3.0
	MD	1.8	-1.6	-1.5	**-0.5**	**-0.5**	-3.5	-1.7	**0.3**
	KC	3.8	1.2	1.2	1.7	1.1	-1.5	**0.4**	1.7
	FG	3.1	**0.2**	**-0.3**	**0.7**	**0.2**	-2.4	**-0.5**	1.1
	AVG	2.9	**-0.2**	**-0.2**	**0.7**	**0.3**	-2.3	**-0.7**	1.1

Table 17 Marginalized baseline event rate $\exp(\gamma_0)$, marginal RR $\exp(\Delta)$ and asymptotic variance of the intervention effect under the independence or arm-specific exchangeable working correlation when the cluster sizes are all equal (CV $\eta = 0$). We use σ^2 to generically denote the asymptotic variance under either working correlation structure because $\sigma^2 = \sigma^2_{\text{ind}} = \sigma^2_{\text{aexch}}$ under equal cluster sizes. This table corresponds to Table 3 in the main text.

$(\exp(\beta_0), \exp(\beta_1))$	(σ_0^2, σ_1^2)	N	$\bar{m}$		$T=\infty$	$T=6$	$T=5$	$T=4$	$T=3$	$T=2$	$T=1$
(1.25, 0.70)	(0.05, 0.05)	30	15	$\exp(\gamma_0)$	1.28	1.28	1.27	1.23	1.14	0.93	0.55
				$\exp(\Delta)$	0.70	0.70	0.71	0.72	0.74	0.79	0.84
				σ^2	0.46	0.45	0.44	0.42	0.37	0.31	0.30
	(0.10, 0.10)	30	45	$\exp(\gamma_0)$	1.31	1.31	1.29	1.25	1.15	0.93	0.55
				$\exp(\Delta)$	0.70	0.70	0.71	0.72	0.75	0.79	0.84
				σ^2	0.50	0.49	0.46	0.42	0.34	0.25	0.18
	(0.20, 0.20)	60	35	$\exp(\gamma_0)$	1.38	1.36	1.34	1.28	1.16	0.93	0.55
				$\exp(\Delta)$	0.70	0.71	0.71	0.73	0.76	0.80	0.85
				σ^2	0.99	0.91	0.84	0.74	0.59	0.42	0.28
	(0.30, 0.30)	90	40	$\exp(\gamma_0)$	1.45	1.41	1.38	1.31	1.17	0.93	0.55
				$\exp(\Delta)$	0.70	0.71	0.72	0.74	0.76	0.80	0.85
				σ^2	1.48	1.28	1.16	0.99	0.77	0.54	0.34
	(0.40, 0.40)	110	40	$\exp(\gamma_0)$	1.53	1.46	1.41	1.33	1.18	0.94	0.55
				$\exp(\Delta)$	0.70	0.72	0.73	0.75	0.77	0.81	0.85
				σ^2	2.05	1.62	1.44	1.21	0.94	0.65	0.41
(2.70, 0.70)	(0.05, 0.05)	25	10	$\exp(\gamma_0)$	2.77	2.62	2.48	2.23	1.86	1.35	0.73
				$\exp(\Delta)$	0.70	0.73	0.75	0.78	0.82	0.86	0.90
				σ^2	0.38	0.32	0.28	0.24	0.20	0.17	0.20
	(0.10, 0.10)	30	30	$\exp(\gamma_0)$	2.84	2.64	2.48	2.23	1.85	1.35	0.73
				$\exp(\Delta)$	0.70	0.73	0.76	0.79	0.82	0.86	0.90
				σ^2	0.48	0.36	0.30	0.23	0.17	0.12	0.10
	(0.20, 0.20)	55	20	$\exp(\gamma_0)$	2.99	2.67	2.49	2.22	1.84	1.34	0.72
				$\exp(\Delta)$	0.70	0.75	0.77	0.79	0.83	0.86	0.90
				σ^2	0.97	0.62	0.52	0.40	0.29	0.20	0.16
	(0.30, 0.30)	80	30	$\exp(\gamma_0)$	3.14	2.70	2.49	2.21	1.83	1.33	0.72
				$\exp(\Delta)$	0.70	0.76	0.78	0.80	0.83	0.87	0.90
				σ^2	1.45	0.80	0.66	0.51	0.36	0.24	0.17
	(0.40, 0.40)	110	25	$\exp(\gamma_0)$	3.30	2.72	2.50	2.21	1.82	1.32	0.71
				$\exp(\Delta)$	0.70	0.77	0.79	0.81	0.84	0.87	0.90
				σ^2	2.03	0.98	0.80	0.62	0.45	0.31	0.22

Table 18 Marginalized baseline event rate $\exp(\gamma_0)$, marginal RR e^{Δ} and asymptotic variance of the intervention effect under the independence (σ^2_{ind}) and arm-specific exchangeable (σ^2_{aexch}) working correlation when the CV of cluster sizes is $\eta = 0.6$. This table corresponds to Table 4 and 5 in the main text.

$(\exp(\beta_0), \exp(\beta_1))$	(σ_0^2, σ_1^2)	N	$\bar{m}$		$T=\infty$	$T=6$	$T=5$	$T=4$	$T=3$	$T=2$	$T=1$
(1.25, 0.70)	(0.05, 0.05)	30	15	$\exp(\gamma_0)$	1.28	1.28	1.27	1.23	1.14	0.93	0.55
				$\exp(\Delta)$	0.70	0.70	0.71	0.72	0.74	0.79	0.84
				σ^2_{ind}	0.53	0.52	0.51	0.48	0.42	0.34	0.32
				σ^2_{aexch}	0.50	0.50	0.48	0.45	0.40	0.34	0.32
	(0.10, 0.10)	30	45	$\exp(\gamma_0)$	1.31	1.31	1.29	1.25	1.15	0.93	0.55
				$\exp(\Delta)$	0.70	0.70	0.71	0.72	0.75	0.79	0.84
				σ^2_{ind}	0.65	0.63	0.60	0.54	0.44	0.31	0.21
				σ^2_{aexch}	0.53	0.51	0.49	0.44	0.36	0.27	0.19
	(0.20, 0.20)	60	35	$\exp(\gamma_0)$	1.38	1.36	1.34	1.28	1.16	0.93	0.55
				$\exp(\Delta)$	0.70	0.71	0.71	0.73	0.76	0.80	0.85
				σ^2_{ind}	1.30	1.20	1.11	0.97	0.77	0.53	0.35
				σ^2_{aexch}	1.02	0.94	0.88	0.77	0.62	0.44	0.31
	(0.30, 0.30)	90	40	$\exp(\gamma_0)$	1.45	1.41	1.38	1.31	1.17	0.93	0.55
				$\exp(\Delta)$	0.70	0.71	0.72	0.74	0.76	0.80	0.85
				σ^2_{ind}	1.99	1.70	1.54	1.31	1.02	0.70	0.44
				σ^2_{aexch}	1.51	1.30	1.18	1.01	0.80	0.56	0.37
	(0.40, 0.40)	110	40	$\exp(\gamma_0)$	1.53	1.46	1.41	1.33	1.18	0.94	0.55
				$\exp(\Delta)$	0.70	0.72	0.73	0.75	0.77	0.81	0.85
				σ^2_{ind}	2.76	2.17	1.93	1.62	1.25	0.86	0.53
				σ^2_{aexch}	2.07	1.64	1.47	1.24	0.96	0.68	0.44
(2.70, 0.70)	(0.05, 0.05)	25	10	$\exp(\gamma_0)$	2.77	2.62	2.48	2.23	1.86	1.35	0.73
				$\exp(\Delta)$	0.70	0.73	0.75	0.78	0.82	0.86	0.90
				σ^2_{ind}	0.45	0.38	0.33	0.28	0.22	0.19	0.21
				σ^2_{aexch}	0.42	0.35	0.31	0.26	0.21	0.18	0.21
	(0.10, 0.10)	30	30	$\exp(\gamma_0)$	2.84	2.64	2.48	2.23	1.85	1.35	0.73
				$\exp(\Delta)$	0.70	0.73	0.76	0.79	0.82	0.86	0.90
				σ^2_{ind}	0.63	0.47	0.39	0.30	0.21	0.14	0.11
				σ^2_{aexch}	0.50	0.37	0.31	0.24	0.18	0.13	0.11
	(0.20, 0.20)	55	20	$\exp(\gamma_0)$	2.99	2.67	2.49	2.22	1.84	1.34	0.72
				$\exp(\Delta)$	0.70	0.75	0.77	0.79	0.83	0.86	0.90
				σ^2_{ind}	1.29	0.82	0.68	0.52	0.37	0.25	0.19
				σ^2_{aexch}	0.99	0.65	0.54	0.42	0.31	0.22	0.18
	(0.30, 0.30)	80	30	$\exp(\gamma_0)$	3.14	2.70	2.49	2.21	1.83	1.33	0.72
				$\exp(\Delta)$	0.70	0.76	0.78	0.80	0.83	0.87	0.90
				σ^2_{ind}	1.95	1.07	0.88	0.67	0.48	0.32	0.21
				σ^2_{aexch}	1.47	0.82	0.67	0.52	0.38	0.26	0.19
	(0.40, 0.40)	110	25	$\exp(\gamma_0)$	3.30	2.72	2.50	2.21	1.82	1.32	0.71
				$\exp(\Delta)$	0.70	0.77	0.79	0.81	0.84	0.87	0.90
				σ^2_{ind}	2.73	1.31	1.07	0.83	0.60	0.40	0.27
				σ^2_{aexch}	2.05	1.00	0.82	0.64	0.47	0.33	0.24

References

Candel, M. J. and van Breukelen, G. J. (2010). Sample size adjustments for varying cluster sizes in cluster randomized trials with binary outcomes analyzed with second-order PQL mixed logistic regression. *Statistics in Medicine* **29,** 1488–1501.

Jung, S. H. and Ahn, C. (2003). Sample size estimation for GEE method for comparing slopes in repeated measurements data. *Statistics in Medicine* **22,** 1305–1315.

Li, D., Zhang, S., and Cao, J. (2019). Sample size calculation for clinical trials with correlated count measurements based on the negative binomial distribution. *Statistics in Medicine* **38,** 5413–5427.

Turner, E. L., Yao, L., Li, F., and Prague, M. (2020). Properties and pitfalls of weighting as an alternative to multilevel multiple imputation in cluster randomized trials with missing binary outcomes under covariate-dependent missingness. *Statistical Methods in Medical Research* **29,** 1338–1353.

van Breukelen, G. J. P., Candel, M. J. J. M., and Berger, M. P. F. B. (2007). Relative efficiency of unequal versus equal cluster sizes in cluster randomized and multicentre trials. *Statistics in Medicine* **26,** 2589–2603.

