Supporting Information for "Power analyses for stepped wedge designs with multivariate continuous outcomes" by Davis-Plourde, Taljaard, and Li

Contents

Web Appendix A

Likelihood function

Specific details on the EM algorithm for our MLMM are provided below. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\sigma}^\top)^\top$, where $\boldsymbol{\beta}$ is the vector of all fixed effects and $\boldsymbol{\sigma}$ is the vector of all variance components (unique components that make up $\boldsymbol{\Sigma}_b$, $\boldsymbol{\Sigma}_s$, and $\boldsymbol{\Sigma}_\epsilon$), denote our set of parameters we wish to estimate. Also let $\boldsymbol{D}_{ij}$ denote the design matrix for the fixed effects, $\boldsymbol{\beta}$, for cluster i at period j. We can express the fully observed likelihood of our MLMM using

$$f(\boldsymbol{Y}, \boldsymbol{b}, \boldsymbol{s}|\boldsymbol{\theta}) = \prod_{i=1}^{I}\prod_{j=1}^{T}\prod_{k=1}^{N_{ij}} f(\boldsymbol{Y}_{ijk}|\boldsymbol{b}_i, \boldsymbol{s}_{ij}; \boldsymbol{\theta}) f(\boldsymbol{b}_i|\boldsymbol{\theta}) f(\boldsymbol{s}_{ij}|\boldsymbol{\theta}),$$

where $f(\boldsymbol{Y}_{ijk}|\boldsymbol{b}_i, \boldsymbol{s}_{ij}; \boldsymbol{\theta})$, $f(\boldsymbol{b}_i|\boldsymbol{\theta})$ and $f(\boldsymbol{s}_{ij}|\boldsymbol{\theta})$ are the conditional multivariate normal density of the outcome, multivariate normal density for the random cluster effects and multivariate normal density for the random cluster-by-time interactions given by

$$\begin{aligned}
f(\boldsymbol{Y}_{ijk}|\boldsymbol{b}_i, \boldsymbol{s}_{ij}; \boldsymbol{\theta}) &= (2\pi)^{-L/2}|\boldsymbol{\Sigma}_\epsilon|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij})^\top \boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij})\right\}\\
f(\boldsymbol{b}_i|\boldsymbol{\theta}) &= (2\pi)^{-L/2}|\boldsymbol{\Sigma}_b|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{b}_i^\top \boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}_i\right\}\\
f(\boldsymbol{s}_{ij}|\boldsymbol{\theta}) &= (2\pi)^{-L/2}|\boldsymbol{\Sigma}_s|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{s}_{ij}^\top \boldsymbol{\Sigma}_s^{-1}\boldsymbol{s}_{ij}\right\}.
\end{aligned}$$

Ignoring the normalization constant, we can write the log-likelihood as

$$\begin{aligned}
\ell(\boldsymbol{Y}, \boldsymbol{b}, \boldsymbol{s}|\boldsymbol{\theta}) \approx & -\frac{\sum_{i=1}^{I}\sum_{j=1}^{T} N_{ij}}{2}\log|\boldsymbol{\Sigma}_\epsilon| - \frac{IT}{2}\log|\boldsymbol{\Sigma}_s| - \frac{I}{2}\log|\boldsymbol{\Sigma}_b| - \frac{1}{2}\sum_{i=1}^{I}\boldsymbol{b}_i^\top\boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}_i - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{T}\boldsymbol{s}_{ij}^\top\boldsymbol{\Sigma}_s^{-1}\boldsymbol{s}_{ij}\\
& -\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij})^\top\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij}).
\end{aligned}$$

To generate the score functions for the maximization step we take the partial derivative of our log-likelihood with respect to a particular parameter, set the expression equal to zero, and solve for that parameter.

Score functions for EM algorithm

Score function for fixed effects, $\boldsymbol{\beta}$

To compute the score function for $\boldsymbol{\beta}$ we only need to consider the parts of the log-likelihood that contain this parameter, specifically, $-\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij})^\top\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij})$. Before we take the partial derivative, we can further simplify this expression by removing any terms that do not involve $\boldsymbol{\beta}$ leaving us with

$$\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}} \boldsymbol{\beta}^\top \boldsymbol{D}_{ij}^\top \boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{Y}_{ijk} - \boldsymbol{b}_i - \boldsymbol{s}_{ij}) - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}} \boldsymbol{\beta}^\top \boldsymbol{D}_{ij}^\top \boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{D}_{ij}\boldsymbol{\beta}.$$

Taking the partial derivative with respect to $\boldsymbol{\beta}$ gives us

$$\frac{\partial}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}} \boldsymbol{D}_{ij}^\top \boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{Y}_{ijk} - \boldsymbol{b}_i - \boldsymbol{s}_{ij}) - \sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}} \boldsymbol{D}_{ij}^\top \boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{D}_{ij}\boldsymbol{\beta}.$$

To derive the expression of the score function we set the partial derivative equal to zero and solve for $\boldsymbol{\beta}$ giving us

$$\begin{aligned} 0 &= \sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}} \boldsymbol{D}_{ij}^\top \boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{Y}_{ijk} - \boldsymbol{b}_i - \boldsymbol{s}_{ij}) - \sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}} \boldsymbol{D}_{ij}^\top \boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{D}_{ij}\boldsymbol{\beta} \\ S(\boldsymbol{\beta}) &= \left(\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}} \boldsymbol{D}_{ij}^\top \boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{D}_{ij}\right)^{-1} \sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}} \boldsymbol{D}_{ij}^\top \boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{Y}_{ijk} - \boldsymbol{b}_i - \boldsymbol{s}_{ij}). \end{aligned}$$

Score function for covariance components of random effects, $\boldsymbol{\Sigma}_b$ and $\boldsymbol{\Sigma}_s$

To generate the score function for $\boldsymbol{\Sigma}_b$ we only need to consider $-\frac{I}{2}\log|\boldsymbol{\Sigma}_b| - \frac{1}{2}\sum_{i=1}^{I}\boldsymbol{b}_i^\top\boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}_i = \frac{I}{2}\log|\boldsymbol{\Sigma}_b^{-1}| - \frac{1}{2}\sum_{i=1}^{I}\boldsymbol{b}_i^\top\boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}_i$ of the log-likelihood. Taking the partial derivative with respect to $\boldsymbol{\Sigma}_b^{-1}$ we have

$$\frac{\partial}{\partial\boldsymbol{\Sigma}_b^{-1}} = \frac{I}{2}\boldsymbol{\Sigma}_b - \frac{1}{2}\sum_{i=1}^{I}\boldsymbol{b}_i\boldsymbol{b}_i^\top.$$

Again setting the partial derivative equal to zero and solving for $\boldsymbol{\Sigma}_b$ gives us our score function

$$\begin{aligned} 0 &= \frac{I}{2}\boldsymbol{\Sigma}_b - \frac{1}{2}\sum_{i=1}^{I}\boldsymbol{b}_i\boldsymbol{b}_i^\top \\ S(\boldsymbol{\Sigma}_b) &= \frac{1}{I}\sum_{i=1}^{I}\boldsymbol{b}_i\boldsymbol{b}_i^\top. \end{aligned}$$

To generate the score function for $\boldsymbol{\Sigma}_s$ we only need to consider $-\frac{IT}{2}\log|\boldsymbol{\Sigma}_s| - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{T}\boldsymbol{s}_{ij}^\top\boldsymbol{\Sigma}_s^{-1}\boldsymbol{s}_{ij} = \frac{IT}{2}\log|\boldsymbol{\Sigma}_s^{-1}| - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{T}\boldsymbol{s}_{ij}^\top\boldsymbol{\Sigma}_s^{-1}\boldsymbol{s}_{ij}$ of the log-likelihood. Taking the partial derivative with respect to $\boldsymbol{\Sigma}_s^{-1}$ we have

$$\frac{\partial}{\partial\boldsymbol{\Sigma}_s^{-1}} = \frac{IT}{2}\boldsymbol{\Sigma}_s - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{T}\boldsymbol{s}_{ij}\boldsymbol{s}_{ij}^\top.$$

Setting the partial equal to zero and solving for $\boldsymbol{\Sigma}_s$ gives us our score function

$$0 = \frac{IT}{2}\boldsymbol{\Sigma}_s - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{T}\boldsymbol{s}_{ij}\boldsymbol{s}_{ij}^\top$$

$$S(\boldsymbol{\Sigma}_s) = \frac{1}{IT}\sum_{i=1}^{I}\sum_{j=1}^{T}\boldsymbol{s}_{ij}\boldsymbol{s}_{ij}^\top.$$

Score function for covariance components of error term, $\boldsymbol{\Sigma}_\epsilon$

To generate the score function of $\boldsymbol{\Sigma}_\epsilon$ we only need to consider $-(\sum_{i=1}^{I}\sum_{j=1}^{T}N_{ij}/2)\log|\boldsymbol{\Sigma}_\epsilon| - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij})^\top\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij}) = (\sum_{i=1}^{I}\sum_{j=1}^{T}N_{ij}/2)\log|\boldsymbol{\Sigma}_\epsilon^{-1}| - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij})^\top\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij})$ of the log-likelihood. Taking the partial derivative with respect to $\boldsymbol{\Sigma}_\epsilon^{-1}$ we have

$$\frac{\partial}{\partial\boldsymbol{\Sigma}_\epsilon^{-1}} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{T}N_{ij}}{2}\boldsymbol{\Sigma}_\epsilon - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij})(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij})^\top$$

Setting the partial derivative equal to zero and solving for $\boldsymbol{\Sigma}_\epsilon$ gives us our score function

$$\begin{aligned}
0 &= \frac{\sum_{i=1}^{I}\sum_{j=1}^{T}N_{ij}}{2}\boldsymbol{\Sigma}_\epsilon - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij})(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij})^\top \\
S(\boldsymbol{\Sigma}_\epsilon) &= \frac{1}{\sum_{i=1}^{I}\sum_{j=1}^{T}N_{ij}}\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij})(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{b}_i - \boldsymbol{s}_{ij})^\top.
\end{aligned}$$

This is broken down into the following components

$$\begin{aligned}
S(\boldsymbol{\Sigma}_\epsilon) = \frac{1}{\sum_{i=1}^{I}\sum_{j=1}^{T}N_{ij}}\Bigg\{&\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta})(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta})^\top + \sum_{i=1}^{I}\sum_{j=1}^{T}N_{ij}\boldsymbol{b}_i\boldsymbol{b}_i^\top + \sum_{i=1}^{I}\sum_{j=1}^{T}N_{ij}\boldsymbol{s}_{ij}\boldsymbol{s}_{ij}^\top \\
&- \sum_{i=1}^{I}\boldsymbol{b}_i\left(\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta})^\top\right) - \sum_{i=1}^{I}\left(\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta})\right)\boldsymbol{b}_i^\top \\
&- \sum_{i=1}^{I}\sum_{j=1}^{T}\boldsymbol{s}_{ij}\left(\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta})^\top\right) - \sum_{i=1}^{I}\sum_{j=1}^{T}\left(\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta})\right)\boldsymbol{s}_{ij}^\top \\
&+ \sum_{i=1}^{I}\boldsymbol{b}_i\left(\sum_{j=1}^{T}N_{ij}\boldsymbol{s}_{ij}^\top\right) + \sum_{i=1}^{I}\left(\sum_{j=1}^{T}N_{ij}\boldsymbol{s}_{ij}\right)\boldsymbol{b}_i^\top\Bigg\}.
\end{aligned}$$

Expected values for EM algorithm

We need to derive expressions for $\mathbb{E}\,\boldsymbol{b}_i$, $\mathbb{E}\,\boldsymbol{s}_{ij}$, $\mathbb{E}\,\boldsymbol{b}_i\boldsymbol{b}_i^\top$, $\mathbb{E}\,\boldsymbol{s}_{ij}\boldsymbol{s}_{ij}^\top$, and for the expected value of all crossproducts of $\boldsymbol{b}_i$ and $\boldsymbol{s}_{ij}$. To achieve this we re-parameterize our MLMM to the equivalent expression

$$\boldsymbol{Y}_{ijk} = \boldsymbol{D}_{ij}^\top\boldsymbol{\beta} + \boldsymbol{M}_{ij}^\top\boldsymbol{\phi}_i + \boldsymbol{\epsilon}_{ijk},$$

where $\boldsymbol{\phi}_i = (\boldsymbol{b}_i^\top, \boldsymbol{s}_{ij}^\top)^\top$ is a vector of all random effects for cluster i and follows a multivariate normal distribution characterized by $\mathcal{N}(0_{T+1}, \boldsymbol{\Sigma}_\phi)$ with $\boldsymbol{\Sigma}_\phi = \text{diag}\left(\boldsymbol{\Sigma}_b, \boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_s\right)$ where $\boldsymbol{I}_T$ denotes a $T \times T$ identity matrix. Further, $\boldsymbol{M}_{ij}$ is the design matrix for the random effects, $\boldsymbol{\phi}_i$, for cluster i at time j. This gives us the following likelihoods and likelihood function

$$
\begin{aligned}
f(\boldsymbol{Y}_{ijk}|\boldsymbol{\phi}_i, \boldsymbol{\theta}) &= (2\pi)^{-L/2}|\boldsymbol{\Sigma}_\epsilon|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{M}_{ij}^\top\boldsymbol{\phi}_i)^\top \boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{M}_{ij}^\top\boldsymbol{\phi}_i)\right\} \\
f(\boldsymbol{\phi}_i|\boldsymbol{\theta}) &= (2\pi)^{-L/2}|\boldsymbol{\Sigma}_\phi|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{\phi}_i^\top \boldsymbol{\Sigma}_\phi^{-1}\boldsymbol{\phi}_i\right\} \\
f(\boldsymbol{Y}, \boldsymbol{\phi}|\boldsymbol{\theta}) &= \prod_{i=1}^{I}\prod_{j=1}^{T}\prod_{k=1}^{N_{ij}} f(\boldsymbol{Y}_{ijk}|\boldsymbol{\phi}_i, \boldsymbol{\theta}) f(\boldsymbol{\phi}_i|\boldsymbol{\theta}).
\end{aligned}
$$

Ignoring the normalization constant, we can write the log-likelihood as

$$
\begin{aligned}
\ell(\boldsymbol{Y}, \boldsymbol{\phi}|\boldsymbol{\theta}) \approx& -\frac{I}{2}\log|\boldsymbol{\Sigma}_b| - \frac{I}{2}\log|\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_s| - \frac{\sum_{i=1}^{I}\sum_{j=1}^{T} N_{ij}}{2}\log|\boldsymbol{\Sigma}_\epsilon| \\
&- \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{M}_{ij}^\top\boldsymbol{\phi}_i)^\top \boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{M}_{ij}^\top\boldsymbol{\phi}_i) - \frac{1}{2}\sum_{i=1}^{I}\boldsymbol{b}_i^\top\boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}_i - \frac{1}{2}\sum_{i=1}^{I}\boldsymbol{s}_i^\top(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_s^{-1})\boldsymbol{s}_i
\end{aligned}
$$

Under this equivalent MLMM expression our score function for $\boldsymbol{\Sigma}_\epsilon$ becomes

$$
\begin{aligned}
S(\boldsymbol{\Sigma}_\epsilon) =& \frac{1}{\sum_{i=1}^{I}\sum_{j=1}^{T} N_{ij}}\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{M}_{ij}^\top\boldsymbol{\phi}_i)(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta} - \boldsymbol{M}_{ij}^\top\boldsymbol{\phi}_i)^\top \\
=& \frac{1}{\sum_{i=1}^{I}\sum_{j=1}^{T} N_{ij}}\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta})(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta})^\top - \frac{1}{\sum_{i=1}^{I}\sum_{j=1}^{T} N_{ij}}\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N}\boldsymbol{M}_{ij}^\top\boldsymbol{\phi}_i(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta})^\top \\
&- \frac{1}{\sum_{i=1}^{I}\sum_{j=1}^{T} N_{ij}}\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}\boldsymbol{\beta})\boldsymbol{\phi}_i^\top\boldsymbol{M}_{ij} + \frac{1}{\sum_{i=1}^{I}\sum_{j=1}^{T} N_{ij}}\sum_{i=1}^{I}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}\boldsymbol{M}_{ij}^\top\boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top\boldsymbol{M}_{ij}.
\end{aligned}
$$

Further, the posterior distribution of the random effects for cluster i is $f(\boldsymbol{\phi}_i|\boldsymbol{Y}_i;\boldsymbol{\theta}) = f(\boldsymbol{Y}_i, \boldsymbol{\phi}_i|\boldsymbol{\theta})/f(\boldsymbol{Y}_i|\boldsymbol{\theta})$, which is proportional to

$$
\exp\left\{-\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}\boldsymbol{\phi}_i^\top\boldsymbol{M}_{ij}\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}^\top\boldsymbol{\beta}) - \frac{1}{2}\boldsymbol{\phi}_i^\top\left(\boldsymbol{\Sigma}_\phi^{-1} + N_{ij}\sum_{j=1}^{T}\boldsymbol{M}_{ij}\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{M}_{ij}^\top\right)\boldsymbol{\phi}_i\right\},
$$

where $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i11}^\top, \ldots, \boldsymbol{Y}_{iTN_{iT}}^\top)^\top$ is a vector of all outcomes measured in cluster i. Now we can easily generate the expected value of $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top$ through the realization that $f(\boldsymbol{\phi}_i|\boldsymbol{Y}_i;\boldsymbol{\theta})$ is proportional to a multivariate normal distribution with mean $\mathbb{E}(\boldsymbol{\phi}_i|\boldsymbol{Y}_i, \boldsymbol{\theta})$ and covariance $\mathrm{V}(\boldsymbol{\phi}_i|\boldsymbol{Y}_i, \boldsymbol{\theta})$ given by

$$
\mathbb{E}(\boldsymbol{\phi}_i|\boldsymbol{Y}_i, \boldsymbol{\theta}) = \left(\boldsymbol{\Sigma}_\phi^{-1} + N_{ij}\sum_{j=1}^{T}\boldsymbol{M}_{ij}\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{M}_{ij}^\top\right)^{-1}\sum_{j=1}^{T}\sum_{k=1}^{N_{ij}}\boldsymbol{M}_{ij}\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{Y}_{ijk} - \boldsymbol{D}_{ij}^\top\boldsymbol{\beta})
$$

$$\mathrm{V}(\boldsymbol{\phi}_i|\boldsymbol{Y}_i,\boldsymbol{\theta}) = \left(\boldsymbol{\Sigma}_\phi^{-1} + N_{ij}\sum_{j=1}^{T}\boldsymbol{M}_{ij}\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{M}_{ij}^\top\right)^{-1}$$

$$\mathbb{E}(\boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top|\boldsymbol{Y}_i,\boldsymbol{\theta}) = \mathrm{V}(\boldsymbol{\phi}_i|\boldsymbol{Y}_i,\boldsymbol{\theta}) + \mathbb{E}(\boldsymbol{\phi}_i|\boldsymbol{Y}_i,\boldsymbol{\theta})\,\mathbb{E}(\boldsymbol{\phi}_i|\boldsymbol{Y}_i,\boldsymbol{\theta})^\top.$$

Estimation via EM algorithm

The EM algorithm is an iterative process for maximizing the likelihood of a model and includes two steps: (1) generating the expected values of the missing data (i.e. random effects) given the current parameter estimates and (2) using those expected values to generate updated parameter estimates using score functions. The first step fittingly refers to the expectation stage and the second step to the maximization stage since the score function for each parameter produces a value that maximizes the likelihood. The EM algorithm iterates between these two steps until convergence is met.

More specifically, to start the EM algorithm we need to provide initial parameter values, denoted by $\boldsymbol{\beta}^0$ and $\boldsymbol{\sigma}^0$. In our algorithm we use individually fitted LMMs, one for each outcome, to initialize all parameters. Next, the expected values of $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top$ are computed using the expressions derived above and then used in the score functions in place of $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top$ to generate updated parameter estimates, $\boldsymbol{\beta}^1$ and $\boldsymbol{\sigma}^1$. These new parameter estimates are then used to generate new expected values which are then used to produce a new set of updated parameter values. This process continues until the likelihood converges, usually defined as a negligible change in the likelihood (i.e. 10^{-5}).

Web Appendix B

Derivation of $\widetilde{\boldsymbol{V}}_i$

MLMM under Cluster-Period Means Approach

We can simplify our approach using cluster-period means (Li et al., 2021; Davis-Plourde et al., 2021). Let $\overline{\boldsymbol{Y}}_{ij} = (\overline{Y}_{ij1},\ldots,\overline{Y}_{ijL})^\top$ where $\overline{Y}_{ijl} = \frac{1}{N}\sum_{k=1}^{N} Y_{ijkl}$. Also let $\overline{\boldsymbol{\epsilon}}_{ij} = (\overline{\epsilon}_{ij1},\ldots,\overline{\epsilon}_{ijL})^\top$ with $\overline{\epsilon}_{ijl} = \frac{1}{N}\sum_{k=1}^{N}\epsilon_{ijkl}$. Our MLMM is equivalent to

$$\overline{\boldsymbol{Y}}_{ij} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_j + X_{ij}\boldsymbol{\delta} + \boldsymbol{b}_i + \boldsymbol{s}_{ij} + \overline{\boldsymbol{\epsilon}}_{ij}.$$

Within-Period Block of $\widetilde{V}_i$

The variance of $\overline{Y}_{ijl}$ is

$$\begin{aligned}
\text{var}(\overline{Y}_{ijl}) &= \text{cov}\left(\frac{1}{N}\sum_{k=1}^{N} Y_{ijkl}, \frac{1}{N}\sum_{k=1}^{N} Y_{ijkl}\right) \\
&= \frac{1}{N^2}\left(N\text{var}(Y_{ijkl}) + N(N-1)\text{cov}(Y_{ijkl}, Y_{ijk'l})\right) \\
&= \frac{1}{N}\left(\sigma_{bl}^2 + \sigma_{sl}^2 + \sigma_{\epsilon l}^2 + (N-1)(\sigma_{bl}^2 + \sigma_{sl}^2)\right) \\
&= \sigma_{bl}^2 + \sigma_{sl}^2 + \frac{\sigma_{\epsilon l}^2}{N}.
\end{aligned}$$

The covariance between outcomes in the same period is

$$\begin{aligned}
\text{cov}(\overline{Y}_{ijl}, \overline{Y}_{ijl'}) &= \text{cov}\left(\frac{1}{N}\sum_{k=1}^{N} Y_{ijkl}, \frac{1}{N}\sum_{k=1}^{N} Y_{ijkl'}\right) \\
&= \frac{1}{N^2}\left(N\text{cov}(Y_{ijkl}, Y_{ijkl'}) + N(N-1)\text{cov}(Y_{ijkl}, Y_{ijk'l'})\right) \\
&= \frac{1}{N}\left(\sigma_{bll'} + \sigma_{sll'} + \sigma_{\epsilon ll'} + (N-1)(\sigma_{bll'} + \sigma_{sll'})\right) \\
&= \sigma_{bll'} + \sigma_{sll'} + \frac{\sigma_{\epsilon ll'}}{N}.
\end{aligned}$$

Therefore, the within-period block is $\mathbf{\Sigma}_b + \mathbf{\Sigma}_s + \dfrac{1}{N}\mathbf{\Sigma}_\epsilon$.

Between-Period Block of $\widetilde{V}_i$

Recall that subjects are sampled cross-sectionally, therefore, $k \neq k'$. The within outcome covariance between periods is

$$\begin{aligned}
\text{cov}(\overline{Y}_{ijl}, \overline{Y}_{ij'l}) &= \text{cov}\left(\frac{1}{N}\sum_{k=1}^{N} Y_{ijkl}, \frac{1}{N}\sum_{k'=1}^{N} Y_{ij'k'l}\right) \\
&= \frac{1}{N^2}\left(N^2\text{cov}(Y_{ijkl}, Y_{ij'k'l})\right) \\
&= \sigma_{bl}^2.
\end{aligned}$$

The between outcome covariance between periods is

$$\begin{aligned}
\text{cov}(\overline{Y}_{ijl}, \overline{Y}_{ij'l'}) &= \text{cov}\left(\frac{1}{N}\sum_{k=1}^{N} Y_{ijkl}, \frac{1}{N}\sum_{k'=1}^{N} Y_{ij'k'l'}\right) \\
&= \frac{1}{N^2}\left(N^2\text{cov}(Y_{ijkl}, Y_{ij'k'l'})\right) \\
&= \sigma_{bll'}.
\end{aligned}$$

Therefore, the between-period block is $\boldsymbol{\Sigma}_b$.

Combining Within-Period and Between-Period Blocks to obtain $\widetilde{\boldsymbol{V}}_i$

$\widetilde{\boldsymbol{V}}_i$ is block exchangeable, therefore, we can combine the within-period and between-period blocks using $\widetilde{\boldsymbol{V}}_i = \boldsymbol{I}_T \otimes (\boldsymbol{WP} - \boldsymbol{BP}) + \boldsymbol{J}_T \otimes \boldsymbol{BP}$, where $\boldsymbol{WP}$ is the within-period block, $\boldsymbol{BP}$ is the between-period block, $\boldsymbol{I}_T$ is a $T \times T$ identity matrix, and $\boldsymbol{J}_T$ is a $T \times T$ matrix of ones. This gives us

$$\begin{aligned}
\widetilde{\boldsymbol{V}}_i &= \boldsymbol{I}_T \otimes \left(\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon - \boldsymbol{\Sigma}_b\right) + \boldsymbol{J}_T \otimes \boldsymbol{\Sigma}_b \\
&= \boldsymbol{I}_T \otimes \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right) + \boldsymbol{J}_T \otimes \boldsymbol{\Sigma}_b. \qquad (1)
\end{aligned}$$

.

Derivation of $\widetilde{\boldsymbol{V}}_i^{-1}$

Section 2.1 of Leiva (2007) states that given a block exchangeable matrix of the form, $\boldsymbol{A} = \boldsymbol{I}_u \otimes (\boldsymbol{B} - \boldsymbol{C}) + \boldsymbol{J}_u \otimes \boldsymbol{C}$, if $\boldsymbol{B} - \boldsymbol{C}$ and $\boldsymbol{B} + (u-1)\boldsymbol{C}$ are non-singular matrices, then

$$\boldsymbol{A}^{-1} = \boldsymbol{I}_u \otimes (\boldsymbol{B} - \boldsymbol{C})^{-1} + \boldsymbol{J}_u \otimes \frac{1}{u}\left[\{\boldsymbol{B} + (u-1)\boldsymbol{C}\}^{-1} - (\boldsymbol{B} - \boldsymbol{C})^{-1}\right]. \qquad (2)$$

This means that

$$\begin{aligned}
\widetilde{\boldsymbol{V}}_i^{-1} &= \boldsymbol{I}_T \otimes (\boldsymbol{WP} - \boldsymbol{BP})^{-1} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\{\boldsymbol{WP} + (T-1)\boldsymbol{BP}\}^{-1} - (\boldsymbol{WP} - \boldsymbol{BP})^{-1}\right] \\
&= \boldsymbol{I}_T \otimes \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\left(\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon + (T-1)\boldsymbol{\Sigma}_b\right)^{-1} - \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\right] \\
&= \boldsymbol{I}_T \otimes \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} - \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\right]. \qquad (3)
\end{aligned}$$

Derivation of covariance expression, $\boldsymbol{\Omega}_\delta$

Let $\boldsymbol{Z}_i = (\boldsymbol{I}_T, \boldsymbol{X}_i) \otimes \boldsymbol{I}_L$ where $\boldsymbol{X}_i$ is the randomization schedule for cluster i. The covariance matrix for the model estimators are $(\sum_{i=1}^{I} \boldsymbol{Z}_i^\top \widetilde{\boldsymbol{V}}_i^{-1} \boldsymbol{Z}_i)^{-1}$ where

$$\sum_{i=1}^{I} \boldsymbol{Z}_i^\top \widetilde{\boldsymbol{V}}_i^{-1} \boldsymbol{Z}_i = \sum_{i=1}^{I}\left[\begin{pmatrix} \boldsymbol{I}_T \\ \boldsymbol{X}_i^\top \end{pmatrix} \otimes \boldsymbol{I}_L\right] \widetilde{\boldsymbol{V}}_i^{-1} \left[\begin{pmatrix} \boldsymbol{I}_T & \boldsymbol{X}_i \end{pmatrix} \otimes \boldsymbol{I}_L\right] = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix},$$

where $\boldsymbol{\Omega}_{11}$ is a $TL \times TL$ matrix, $\boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}_{21}^\top$ is a $TL \times L$ matrix, and $\boldsymbol{\Omega}_{22}$ is a $L \times L$ matrix. Block matrix inversion gives us $\boldsymbol{\Omega}_\delta = \boldsymbol{\Omega}^{-1}_{22} = (\boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12})^{-1}$.

We can rewrite $\widetilde{\boldsymbol{V}}_i$ as

$$\widetilde{\boldsymbol{V}}_i^{-1} = \boldsymbol{I}_T \otimes \boldsymbol{A} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right],$$

where $\boldsymbol{A} = \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}$ and $\boldsymbol{B} = \left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}$. Using this expression we can generate our $\boldsymbol{\Omega}$s.

Derivation of $\boldsymbol{\Omega}_{11}$ and $\boldsymbol{\Omega}_{11}^{-1}$

$$\begin{aligned}
\boldsymbol{\Omega}_{11} &= \sum_{i=1}^{I}(\boldsymbol{I}_T \otimes \boldsymbol{I}_L)\widetilde{\boldsymbol{V}}_i^{-1}(\boldsymbol{I}_T \otimes \boldsymbol{I}_L) \\
&= \sum_{i=1}^{I}\left(\boldsymbol{I}_T \otimes \boldsymbol{A} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right]\right) \\
&= I\left(\boldsymbol{I}_T \otimes \boldsymbol{A} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right]\right) \\
&= I\widetilde{\boldsymbol{V}}_i^{-1}.
\end{aligned}$$

Therefore, $\boldsymbol{\Omega}_{11}^{-1} = \frac{1}{I}\widetilde{\boldsymbol{V}}_i = \frac{1}{I}\left[\boldsymbol{I}_T \otimes \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right) + \boldsymbol{J}_T \otimes \boldsymbol{\Sigma}_b\right]$.

Derivation of $\boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}_{21}^{\top}$

$$\begin{aligned}
\boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}_{21}^{\top} &= \sum_{i=1}^{I}(\boldsymbol{I}_T \otimes \boldsymbol{I}_L)\widetilde{\boldsymbol{V}}_i^{-1}(\boldsymbol{X}_i \otimes \boldsymbol{I}_L) \\
&= \sum_{i=1}^{I}\boldsymbol{X}_i \otimes \boldsymbol{A} + \sum_{i=1}^{I}\sum_{j=1}^{T}X_{ij}\boldsymbol{1}_T \otimes \frac{1}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right] \\
&= \sum_{i=1}^{I}\boldsymbol{X}_i \otimes \boldsymbol{A} + U\boldsymbol{1}_T \otimes \frac{1}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right] \\
&= \sum_{i=1}^{I}\boldsymbol{X}_i \otimes \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} + U\boldsymbol{1}_T \otimes \frac{1}{T}\left[\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} - \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\right],
\end{aligned}$$

where $U = \sum_{j=1}^{T} X_{ij}$.

Derivation of $\boldsymbol{\Omega}_{22}$

$$\begin{aligned}
\boldsymbol{\Omega}_{22} &= \sum_{i=1}^{I}(\boldsymbol{X}_i^{\top} \otimes \boldsymbol{I}_L)\widetilde{\boldsymbol{V}}_i^{-1}(\boldsymbol{X}_i \otimes \boldsymbol{I}_L) \\
&= \sum_{i=1}^{I}\sum_{j=1}^{T}X_{ij}^2 \otimes \boldsymbol{A} + \sum_{i=1}^{I}\left(\sum_{j=1}^{T}X_{ij}\right)^2 \otimes \frac{1}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right] \\
&= U \otimes \boldsymbol{A} + V \otimes \frac{1}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right]
\end{aligned}$$

$$= U\boldsymbol{A} + \frac{V}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right]$$
$$= U\left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} + \frac{V}{T}\left[\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} - \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\right],$$

where $V = \sum_{i=1}^{I}\left(\sum_{j=1}^{T} X_{ij}\right)^2$.

Derivation of $\boldsymbol{\Omega}_\delta = \boldsymbol{\Omega}_{22}^{-1}$

We have the following matrices for our block inversion formula, $\boldsymbol{\Omega}_\delta = \boldsymbol{\Omega}_{22}^{-1} = (\boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12})^{-1}$,

$$\boldsymbol{\Omega}_{22} = U\left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} + \frac{V}{T}\left[\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} - \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\right]$$
$$\boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}_{21}^\top = \sum_{i=1}^{I}\boldsymbol{X}_i \otimes \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} + U\boldsymbol{1}_T \otimes \frac{1}{T}\left[\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} - \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\right]$$
$$\boldsymbol{\Omega}_{11}^{-1} = \frac{1}{I}\left[\boldsymbol{I}_T \otimes \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right) + \boldsymbol{J}_T \otimes \boldsymbol{\Sigma}_b\right].$$

This gives us

$$\begin{aligned}
\boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12} &= \frac{1}{I}\Bigg(\left(W - \frac{U^2}{T}\right)\left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} + \frac{U^2}{T}\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} \\
&\quad + U^2\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\boldsymbol{\Sigma}_b\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\Bigg) \\
\boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12} &= \frac{1}{IT}\left[(ITU - TW + U^2 - IV)\left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} - (U^2 - IV)\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\right] \\
\boldsymbol{\Omega}_\delta &= IT\left[(ITU - TW + U^2 - IV)\left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} + (IV - U^2)\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\right]^{-1},
\end{aligned}$$

where $W = \sum_{j=1}^{T}\left(\sum_{i=1}^{I} X_{ij}\right)^2$.

We can map the variance component parameters to the set of unique ICC parameters by observing $\sigma_{bl}^2 = \sigma_{yl}^2\rho_1^l$, $\sigma_{bll'} = \sigma_{yl}\sigma_{yl'}\rho_1^{ll'}$, $\sigma_{sl}^2 = \sigma_{yl}^2(\rho_0^l - \rho_1^l)$, $\sigma_{sll'} = \sigma_{yl}\sigma_{yl'}(\rho_0^{ll'} - \rho_1^{ll'})$, $\sigma_{\epsilon l}^2 = \sigma_{yl}^2(1 - \rho_0^l)$ and $\sigma_{\epsilon ll'} = \sigma_{yl}\sigma_{yl'}(\rho_2^{ll'} - \rho_0^{ll'})$. We can further let $\boldsymbol{\Gamma}_0$, $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$ denote the within-period ICC matrix, between-period ICC matrix and intra-subject ICC matrix across L endpoints, defined as

$$\boldsymbol{\Gamma}_0 = \begin{pmatrix} \rho_0^1 & \rho_0^{12} & \dots & \rho_0^{1L} \\ \rho_0^{12} & \rho_0^2 & \dots & \rho_0^{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_0^{1L} & \rho_0^{2L} & \dots & \rho_0^L \end{pmatrix}, \quad \boldsymbol{\Gamma}_1 = \begin{pmatrix} \rho_1^1 & \rho_1^{12} & \dots & \rho_1^{1L} \\ \rho_1^{12} & \rho_1^2 & \dots & \rho_1^{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_1^{1L} & \rho_1^{2L} & \dots & \rho_1^L \end{pmatrix}, \quad \boldsymbol{\Gamma}_2 = \begin{pmatrix} 1 & \rho_2^{12} & \dots & \rho_2^{1L} \\ \rho_2^{12} & 1 & \dots & \rho_2^{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_2^{1L} & \rho_2^{2L} & \dots & 1 \end{pmatrix}.$$

Defining the diagonal matrix of outcome variances as $\mathbf{\Lambda}_y = \text{diag}(\sigma^2_{y1}, \dots, \sigma^2_{yL})$, we can further rewrite the covariance matrix of the intervention effect estimators in terms of the ICCs through the realization that $\mathbf{\Sigma}_b = \mathbf{\Lambda}_y^{1/2}\mathbf{\Gamma}_1\mathbf{\Lambda}_y^{1/2}$, $\mathbf{\Sigma}_s = \mathbf{\Lambda}_y^{1/2}\left(\mathbf{\Gamma}_0 - \mathbf{\Gamma}_1\right)\mathbf{\Lambda}_y^{1/2}$, and $\mathbf{\Sigma}_\epsilon = \mathbf{\Lambda}_y^{1/2}\left(\mathbf{\Gamma}_2 - \mathbf{\Gamma}_0\right)\mathbf{\Lambda}_y^{1/2}$ giving us

$$
\begin{aligned}
\mathbf{\Omega}_\delta = IT\Bigg[& \Big(ITU - TW + U^2 - IV\Big)\left(\mathbf{\Lambda}_y^{1/2}\left(\mathbf{\Gamma}_0 - \mathbf{\Gamma}_1\right)\mathbf{\Lambda}_y^{1/2} + \frac{1}{N}\mathbf{\Lambda}_y^{1/2}\left(\mathbf{\Gamma}_2 - \mathbf{\Gamma}_0\right)\mathbf{\Lambda}_y^{1/2}\right)^{-1} \\
& - \Big(U^2 - IV\Big)\left(T\mathbf{\Lambda}_y^{1/2}\mathbf{\Gamma}_1\mathbf{\Lambda}_y^{1/2} + \mathbf{\Lambda}_y^{1/2}\left(\mathbf{\Gamma}_0 - \mathbf{\Gamma}_1\right)\mathbf{\Lambda}_y^{1/2} + \frac{1}{N}\mathbf{\Lambda}_y^{1/2}\left(\mathbf{\Gamma}_2 - \mathbf{\Gamma}_0\right)\mathbf{\Lambda}_y^{1/2}\right)^{-1}\Bigg]^{-1} \\
= \frac{IT}{N}\mathbf{\Lambda}_y^{1/2}\Bigg[& \Big(ITU - TW + U^2 - IV\Big)\left(N\left(\mathbf{\Gamma}_0 - \mathbf{\Gamma}_1\right) + \mathbf{\Gamma}_2 - \mathbf{\Gamma}_0\right)^{-1} \\
& - \Big(U^2 - IV\Big)\left(TN\mathbf{\Gamma}_1 + N\left(\mathbf{\Gamma}_0 - \mathbf{\Gamma}_1\right) + \mathbf{\Gamma}_2 - \mathbf{\Gamma}_0\right)^{-1}\Bigg]^{-1}\mathbf{\Lambda}_y^{1/2} \\
= \frac{IT}{N}\mathbf{\Lambda}_y^{1/2}\Bigg[& \Big(ITU - TW + U^2 - IV\Big)\left(\mathbf{\Gamma}_2 - N\mathbf{\Gamma}_1 + (N-1)\mathbf{\Gamma}_0\right)^{-1} \\
& - \Big(U^2 - IV\Big)\left(\mathbf{\Gamma}_2 + (T-1)N\mathbf{\Gamma}_1 + (N-1)\mathbf{\Gamma}_0\right)^{-1}\Bigg]^{-1}\mathbf{\Lambda}_y^{1/2}.
\end{aligned}
$$

Connection to Hooper and Girling model: $\mathbf{\Sigma}_b$, $\mathbf{\Sigma}_s$, and $\mathbf{\Sigma}_\epsilon$ as scalars

If we only have one outcome, y, then $\mathbf{\Sigma}_\epsilon$, $\mathbf{\Sigma}_s$, and $\mathbf{\Sigma}_b$ are scalars, i.e. $\mathbf{\Sigma}_\epsilon = \sigma^2_\epsilon$, $\mathbf{\Sigma}_b = \sigma^2_s$, and $\mathbf{\Sigma}_b = \sigma^2_b$, then we would have

$$
\begin{aligned}
\text{var}(\hat{\delta}) &= IT\left[(ITU - TW + U^2 - IV)\left(\sigma^2_s + \frac{1}{N}\sigma^2_\epsilon\right)^{-1} - (U^2 - IV)\left(T\sigma^2_b + \sigma^2_s + \frac{1}{N}\sigma^2_\epsilon\right)^{-1}\right]^{-1} \\
&= \frac{IT}{N}\left[\frac{(ITU - TW + U^2 - IV)(TN\sigma^2_b + N\sigma^2_s + \sigma^2_\epsilon) - (U^2 - IV)(N\sigma^2_s + \sigma^2_\epsilon)}{(TN\sigma^2_b + N\sigma^2_s + \sigma^2_\epsilon)(N\sigma^2_s + \sigma^2_\epsilon)}\right]^{-1} \\
&= \frac{(IT/N)(TN\sigma^2_b + N\sigma^2_s + \sigma^2_\epsilon)(N\sigma^2_s + \sigma^2_\epsilon)}{(ITU - TW + U^2 - IV)(TN\sigma^2_b + N\sigma^2_s + \sigma^2_\epsilon) - (U^2 - IV)(N\sigma^2_s + \sigma^2_\epsilon)}.
\end{aligned}
$$

The total variance of the outcome, y, is $\sigma^2_y = \sigma^2_b + \sigma^2_s + \sigma^2_\epsilon$. Under the LMM we have the following ICCs and eigenvalues

$$
\begin{aligned}
\rho_0 &= \frac{\sigma^2_b + \sigma^2_s}{\sigma^2_y} \\
\rho_1 &= \frac{\sigma^2_b}{\sigma^2_y} \\
\lambda_2 &= 1 + (N-1)\rho_0 - N\rho_1 \\
&= \frac{N\sigma^2_s + \sigma^2_\epsilon}{\sigma^2_y} \\
\lambda_3 &= 1 + (N-1)\rho_0 + (T-1)N\rho_1
\end{aligned}
$$

$$= \frac{TN\sigma_b^2 + N\sigma_s^2 + \sigma_\epsilon^2}{\sigma_y^2}.$$

Using these expressions we can re-write our variance formula as

$$\mathrm{var}(\hat{\delta}) = \frac{(\sigma_y^2/N)IT\lambda_2\lambda_3}{(ITU - TW + U^2 - IV)\lambda_3 - (U^2 - IV)\lambda_2},$$

which is equivalent to the variance expression from Hooper and Girling (Hooper et al., 2016; Girling and Hemming, 2016).

Web Appendix C

Common intervention effects: derivation of variance expression, var$(\widehat{\delta'})$

Let $\boldsymbol{Z}_i = (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_\epsilon, \boldsymbol{X}_i \otimes \boldsymbol{\sigma}_\epsilon)$ where $\boldsymbol{X}_i$ is the randomization schedule for cluster i, $\boldsymbol{\sigma}_\epsilon = (\sigma_{\epsilon 1}, \ldots, \sigma_{\epsilon L})^\top$, and $\boldsymbol{\Lambda}_\epsilon = \mathrm{diag}(\boldsymbol{\sigma}_\epsilon)$. Also let $\boldsymbol{\Sigma}_b = \boldsymbol{\Lambda}_\epsilon \boldsymbol{\Sigma}_b' \boldsymbol{\Lambda}_\epsilon$ and $\boldsymbol{\Sigma}_s = \boldsymbol{\Lambda}_\epsilon \boldsymbol{\Sigma}_s' \boldsymbol{\Lambda}_\epsilon$ denote the scaled random effects. The covariance matrix for the model estimators are $\left(\sum_{i=1}^I \boldsymbol{Z}_i^\top \widetilde{\boldsymbol{V}}_i^{-1} \boldsymbol{Z}_i\right)^{-1}$ where

$$\sum_{i=1}^I \boldsymbol{Z}_i^\top \widetilde{\boldsymbol{V}}_i^{-1} \boldsymbol{Z}_i = \sum_{i=1}^I \left[\begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_\epsilon \\ \boldsymbol{X}_i^\top \otimes \boldsymbol{\sigma}_\epsilon^\top \end{pmatrix}\right] \widetilde{\boldsymbol{V}}_i^{-1} \left[\begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_\epsilon & \boldsymbol{X}_i \otimes \boldsymbol{\sigma}_\epsilon \end{pmatrix}\right] = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$$

where $\widetilde{\boldsymbol{V}}_i^{-1}$ is the same as previously derived, Ω_{11} is a $2T \times 2T$ matrix, $\Omega_{12} = \Omega_{21}^\top$ is a $2T$ vector, and Ω_{22} is a scalar. Block matrix inversion gives us $\Omega_{22}^{-1} = (\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1} = \mathrm{var}(\widehat{\delta'})$.

Recall we can rewrite $\widetilde{\boldsymbol{V}}_i^{-1} = \boldsymbol{I}_T \otimes \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} - \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\right]$ as

$$\widetilde{\boldsymbol{V}}_i^{-1} = \boldsymbol{I}_T \otimes \boldsymbol{A} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right],$$

where $\boldsymbol{A} = \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}$ and $\boldsymbol{B} = \left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}$. Also recall that

$$\begin{aligned} \widetilde{\boldsymbol{V}}_i &= \boldsymbol{I}_T \otimes \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right) + \boldsymbol{J}_T \otimes \boldsymbol{\Sigma}_b \\ &= \boldsymbol{I}_T \otimes (\boldsymbol{C} - \boldsymbol{D}) + \boldsymbol{J}_T \otimes \boldsymbol{D} \\ \widetilde{\boldsymbol{V}}_i^{-1} &= \boldsymbol{I}_T \otimes (\boldsymbol{C} - \boldsymbol{D})^{-1} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\{\boldsymbol{C} + (T-1)\boldsymbol{D}\}^{-1} - (\boldsymbol{C} - \boldsymbol{D})^{-1}\right], \end{aligned}$$

where $\boldsymbol{C} = \boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon$ and $\boldsymbol{D} = \boldsymbol{\Sigma}_b$.

Derivation of Ω_{11} and Ω_{11}^{-1}

$$
\begin{aligned}
\Omega_{11} &= \sum_{i=1}^{I} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_\epsilon) \widetilde{\boldsymbol{V}}_i^{-1} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_\epsilon) \\
&= \sum_{i=1}^{I} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_\epsilon) \left(\boldsymbol{I}_T \otimes (\boldsymbol{C} - \boldsymbol{D})^{-1} + \boldsymbol{J}_T \otimes \frac{1}{T} \left[\{\boldsymbol{C} + (T-1)\boldsymbol{D}\}^{-1} - (\boldsymbol{C} - \boldsymbol{D})^{-1} \right] \right) (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_\epsilon) \\
&= I \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_\epsilon (\boldsymbol{C} - \boldsymbol{D})^{-1} \boldsymbol{\Lambda}_\epsilon + \boldsymbol{J}_T \otimes \boldsymbol{\Lambda}_\epsilon \frac{1}{T} \left[\{\boldsymbol{C} + (T-1)\boldsymbol{D}\}^{-1} - (\boldsymbol{C} - \boldsymbol{D})^{-1} \right] \boldsymbol{\Lambda}_\epsilon \right) \\
&= I \left(\boldsymbol{I}_T \otimes (\boldsymbol{\Lambda}_\epsilon^{-1} \boldsymbol{C} \boldsymbol{\Lambda}_\epsilon^{-1} - \boldsymbol{\Lambda}_\epsilon^{-1} \boldsymbol{D} \boldsymbol{\Lambda}_\epsilon^{-1})^{-1} + \boldsymbol{J}_T \otimes \frac{1}{T} \left[\{\boldsymbol{\Lambda}_\epsilon^{-1} \boldsymbol{C} \boldsymbol{\Lambda}_\epsilon^{-1} + (T-1) \boldsymbol{\Lambda}_\epsilon^{-1} \boldsymbol{D} \boldsymbol{\Lambda}_\epsilon^{-1}\}^{-1} - (\boldsymbol{\Lambda}_\epsilon^{-1} \boldsymbol{C} \boldsymbol{\Lambda}_\epsilon^{-1} - \boldsymbol{\Lambda}_\epsilon^{-1} \boldsymbol{D} \boldsymbol{\Lambda}_\epsilon^{-1})^{-1} \right] \right).
\end{aligned}
$$

This is of the form $\boldsymbol{I}_T \otimes (\boldsymbol{C}' - \boldsymbol{D}')^{-1} + \boldsymbol{J}_T \otimes \frac{1}{T} \left[\{\boldsymbol{C}' + (T-1)\boldsymbol{D}'\}^{-1} - (\boldsymbol{C}' - \boldsymbol{D}')^{-1} \right]$ where $\boldsymbol{C}' = \boldsymbol{\Lambda}_\epsilon^{-1} \boldsymbol{C} \boldsymbol{\Lambda}_\epsilon^{-1}$ and $\boldsymbol{D}' = \boldsymbol{\Lambda}_\epsilon^{-1} \boldsymbol{D} \boldsymbol{\Lambda}_\epsilon^{-1}$. Therefore

$$
\begin{aligned}
\Omega_{11}^{-1} &= \frac{1}{I} \left(\boldsymbol{I}_T \otimes (\boldsymbol{C}' - \boldsymbol{D}') + \boldsymbol{J}_T \otimes \boldsymbol{D}' \right) \\
&= \frac{1}{I} \left(\boldsymbol{I}_T \otimes (\boldsymbol{\Lambda}_\epsilon^{-1} \boldsymbol{C} \boldsymbol{\Lambda}_\epsilon^{-1} - \boldsymbol{\Lambda}_\epsilon^{-1} \boldsymbol{D} \boldsymbol{\Lambda}_\epsilon^{-1}) + \boldsymbol{J}_T \otimes \boldsymbol{\Lambda}_\epsilon^{-1} \boldsymbol{D} \boldsymbol{\Lambda}_\epsilon^{-1} \right) \\
&= \frac{1}{I} \left(\boldsymbol{I}_T \otimes \left[\boldsymbol{\Lambda}_\epsilon^{-1} \left\{ \boldsymbol{\Sigma}_s + \frac{1}{N} \boldsymbol{\Sigma}_\epsilon \right\} \boldsymbol{\Lambda}_\epsilon^{-1} \right] + \boldsymbol{J}_T \otimes \boldsymbol{\Lambda}_\epsilon^{-1} \boldsymbol{\Sigma}_b \boldsymbol{\Lambda}_\epsilon^{-1} \right) \\
&= \frac{1}{I} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_\epsilon^{-1} \boldsymbol{A}^{-1} \boldsymbol{\Lambda}_\epsilon^{-1} + \boldsymbol{J}_T \otimes \frac{1}{T} \boldsymbol{\Lambda}_\epsilon^{-1} \left(\boldsymbol{B}^{-1} - \boldsymbol{A}^{-1} \right) \boldsymbol{\Lambda}_\epsilon^{-1} \right)
\end{aligned}
$$

Derivation of Ω_{12} and Ω_{21}

$$
\begin{aligned}
\Omega_{12} = \Omega_{21}^\top &= \sum_{i=1}^{I} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_\epsilon) \widetilde{\boldsymbol{V}}_i^{-1} (\boldsymbol{X}_i \otimes \boldsymbol{\sigma}_\epsilon) \\
&= \sum_{i=1}^{I} \left(\boldsymbol{X}_i \otimes \boldsymbol{\Lambda}_\epsilon \boldsymbol{A} \boldsymbol{\sigma}_\epsilon + \sum_{j=1}^{T} X_{ij} \boldsymbol{1}_T \otimes \frac{1}{T} \boldsymbol{\Lambda}_\epsilon \left[\boldsymbol{B} - \boldsymbol{A} \right] \boldsymbol{\sigma}_\epsilon \right) \\
&= \sum_{i=1}^{I} \boldsymbol{X}_i \otimes \boldsymbol{\Lambda}_\epsilon \boldsymbol{A} \boldsymbol{\sigma}_\epsilon + U \boldsymbol{1}_T \otimes \frac{1}{T} \boldsymbol{\Lambda}_\epsilon \left[\boldsymbol{B} - \boldsymbol{A} \right] \boldsymbol{\sigma}_\epsilon,
\end{aligned}
$$

where $U = \sum_{i=1}^{I} \sum_{j=1}^{T} X_{ij}$.

Derivation of Ω_{22}

$$
\begin{aligned}
\Omega_{22} &= \sum_{i=1}^{I} (\boldsymbol{X}_i^\top \otimes \boldsymbol{\sigma}_\epsilon^\top) \widetilde{\boldsymbol{V}}_i^{-1} (\boldsymbol{X}_i \otimes \boldsymbol{\sigma}_\epsilon) \\
&= \sum_{i=1}^{I} (\boldsymbol{X}_i^\top \otimes \boldsymbol{\sigma}_\epsilon^\top) \left(\boldsymbol{I}_T \otimes \boldsymbol{A} + \boldsymbol{J}_T \otimes \frac{1}{T} \left[\boldsymbol{B} - \boldsymbol{A} \right] \right) (\boldsymbol{X}_i \otimes \boldsymbol{\sigma}_\epsilon) \\
&= U \boldsymbol{\sigma}_\epsilon^\top \boldsymbol{A} \boldsymbol{\sigma}_\epsilon + \frac{V}{T} \boldsymbol{\sigma}_\epsilon^\top \left[\boldsymbol{B} - \boldsymbol{A} \right] \boldsymbol{\sigma}_\epsilon,
\end{aligned}
$$

where $V = \sum_{i=1}^{I} \left(\sum_{j=1}^{T} X_{ij} \right)^2$.

Derivation of var($\widehat{\delta^l}$)

We have the following matrices for our block inversion formula, $\mathrm{var}(\widehat{\delta^l}) = \Omega_{22}^{-1} = (\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1}$,

$$\begin{aligned}
\Omega_{22} &= U\boldsymbol{\sigma}_\epsilon^\top \boldsymbol{A}\boldsymbol{\sigma}_\epsilon + \frac{V}{T}\boldsymbol{\sigma}_\epsilon^\top[\boldsymbol{B}-\boldsymbol{A}]\boldsymbol{\sigma}_\epsilon \\
\Omega_{12} &= \Omega_{21}^\top = \sum_{i=1}^{I}\boldsymbol{X}_i \otimes \boldsymbol{\Lambda}_\epsilon\boldsymbol{A}\boldsymbol{\sigma}_\epsilon + U\boldsymbol{1}_T \otimes \frac{1}{T}\boldsymbol{\Lambda}_\epsilon[\boldsymbol{B}-\boldsymbol{A}]\boldsymbol{\sigma}_\epsilon \\
\Omega_{11}^{-1} &= \frac{1}{I}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_\epsilon^{-1}\boldsymbol{A}^{-1}\boldsymbol{\Lambda}_\epsilon^{-1} + \boldsymbol{J}_T \otimes \frac{1}{T}\boldsymbol{\Lambda}_\epsilon^{-1}\left(\boldsymbol{B}^{-1}-\boldsymbol{A}^{-1}\right)\boldsymbol{\Lambda}_\epsilon^{-1}\right).
\end{aligned}$$

This gives us

$$\begin{aligned}
\Omega_{21}\Omega_{11}^{-1}\Omega_{12} &= \sum_{i=1}^{I}\boldsymbol{X}_i^\top\sum_{i=1}^{I}\boldsymbol{X}_i \otimes \frac{1}{I}\boldsymbol{1}_L^\top\boldsymbol{\Lambda}_\epsilon\boldsymbol{A}\boldsymbol{\sigma}_\epsilon + U\sum_{i=1}^{I}\boldsymbol{X}_i^\top\boldsymbol{1}_T \otimes \frac{1}{IT}\boldsymbol{1}_L^\top\boldsymbol{\Lambda}_\epsilon[\boldsymbol{B}-\boldsymbol{A}]\boldsymbol{\sigma}_\epsilon \\
&= \frac{W}{I}\boldsymbol{\sigma}_\epsilon^\top\boldsymbol{A}\boldsymbol{\sigma}_\epsilon + \frac{U^2}{IT}\boldsymbol{\sigma}_\epsilon^\top[\boldsymbol{B}-\boldsymbol{A}]\boldsymbol{\sigma}_\epsilon \\
\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12} &= U\boldsymbol{\sigma}_\epsilon^\top\boldsymbol{A}\boldsymbol{\sigma}_\epsilon + \frac{V}{T}\boldsymbol{\sigma}_\epsilon^\top[\boldsymbol{B}-\boldsymbol{A}]\boldsymbol{\sigma}_\epsilon - \left(\frac{W}{I}\boldsymbol{\sigma}_\epsilon^\top\boldsymbol{A}\boldsymbol{\sigma}_\epsilon + \frac{U^2}{IT}\boldsymbol{\sigma}_\epsilon^\top[\boldsymbol{B}-\boldsymbol{A}]\boldsymbol{\sigma}_\epsilon\right) \\
&= \frac{1}{IT}\left((ITU - TW + U^2 - IV)\boldsymbol{\sigma}_\epsilon^\top\boldsymbol{A}\boldsymbol{\sigma}_\epsilon + (IV - U^2)\boldsymbol{\sigma}_\epsilon^\top\boldsymbol{B}\boldsymbol{\sigma}_\epsilon\right) \\
\mathrm{var}(\widehat{\delta^l}) &= \left[\frac{1}{IT}\left((ITU - TW + U^2 - IV)\boldsymbol{\sigma}_\epsilon^\top\boldsymbol{A}\boldsymbol{\sigma}_\epsilon + (IV - U^2)\boldsymbol{\sigma}_\epsilon^\top\boldsymbol{B}\boldsymbol{\sigma}_\epsilon\right)\right]^{-1} \\
&= IT\left[(ITU - TW + U^2 - IV)\boldsymbol{\sigma}_\epsilon^\top\left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\boldsymbol{\sigma}_\epsilon - (U^2 - IV)\boldsymbol{\sigma}_\epsilon^\top\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\boldsymbol{\sigma}_\epsilon\right]^{-1},
\end{aligned}$$

where $W = \sum_{j=1}^{T}\left(\sum_{i=1}^{I}X_{ij}\right)^2$. We can rewrite this expression in terms of the ICCs again using $\sigma_{\epsilon l}^2 = \sigma_{yl}^2(1-\rho_0^l)$, $\boldsymbol{\Sigma}_b = \boldsymbol{\Lambda}_y^{1/2}\boldsymbol{\Gamma}_1\boldsymbol{\Lambda}_y^{1/2}$, $\boldsymbol{\Sigma}_s = \boldsymbol{\Lambda}_y^{1/2}(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1)\boldsymbol{\Lambda}_y^{1/2}$, and $\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Lambda}_y^{1/2}(\boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_0)\boldsymbol{\Lambda}_y^{1/2}$ which gives us

$$\begin{aligned}
\mathrm{var}(\widehat{\delta^l}) = IT\Bigg[&(ITU - TW + U^2 - IV)\boldsymbol{\omega}^\top\left(\boldsymbol{\Lambda}_y^{1/2}(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1)\boldsymbol{\Lambda}_y^{1/2} + \frac{1}{N}\boldsymbol{\Lambda}_y^{1/2}(\boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_0)\boldsymbol{\Lambda}_y^{1/2}\right)^{-1}\boldsymbol{\omega} \\
&- (U^2 - IV)\boldsymbol{\omega}^\top\left(T\boldsymbol{\Lambda}_y^{1/2}\boldsymbol{\Gamma}_1\boldsymbol{\Lambda}_y^{1/2} + \boldsymbol{\Lambda}_y^{1/2}(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1)\boldsymbol{\Lambda}_y^{1/2} + \frac{1}{N}\boldsymbol{\Lambda}_y^{1/2}(\boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_0)\boldsymbol{\Lambda}_y^{1/2}\right)^{-1}\boldsymbol{\omega}\Bigg]^{-1} \\
= \frac{IT}{N}\Bigg[&(ITU - TW + U^2 - IV)\boldsymbol{\omega}^\top\boldsymbol{\Lambda}_y^{-1/2}\left(N(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1) + \boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_0\right)^{-1}\boldsymbol{\Lambda}_y^{-1/2}\boldsymbol{\omega} \\
&- (U^2 - IV)\boldsymbol{\omega}^\top\boldsymbol{\Lambda}_y^{-1/2}\left(TN\boldsymbol{\Gamma}_1 + N(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1) + \boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_0\right)^{-1}\boldsymbol{\Lambda}_y^{-1/2}\boldsymbol{\omega}\Bigg]^{-1} \\
= \frac{IT}{N}\Bigg[&\left(ITU - TW + U^2 - IV\right)\boldsymbol{\omega}^\top\boldsymbol{\Lambda}_y^{-1/2}\left(\boldsymbol{\Gamma}_2 - N\boldsymbol{\Gamma}_1 + (N-1)\boldsymbol{\Gamma}_0\right)^{-1}\boldsymbol{\Lambda}_y^{-1/2}\boldsymbol{\omega} \\
&- \left(U^2 - IV\right)\boldsymbol{\omega}^\top\boldsymbol{\Lambda}_y^{-1/2}\left(\boldsymbol{\Gamma}_2 + (T-1)N\boldsymbol{\Gamma}_1 + (N-1)\boldsymbol{\Gamma}_0\right)^{-1}\boldsymbol{\Lambda}_y^{-1/2}\boldsymbol{\omega}\Bigg]^{-1},
\end{aligned}$$

where $\boldsymbol{\omega} = \left(\sigma_{y1}(1-\rho_0^1)^{1/2}, \ldots, \sigma_{yL}(1-\rho_0^L)^{1/2}\right)^\top$.

Web Appendix D

Common ICCs across endpoints: derivation of covariance expression, $\boldsymbol{\Omega}_\delta$

The common ICC assumption leads to simplification of the variance expression $\boldsymbol{\Omega}_\delta$ by defining the three key ICC matrices with their explicit simple exchangeable forms:

$$\begin{aligned}
\boldsymbol{\Gamma}_0 &= (\rho_0 - \rho_{00})\boldsymbol{I}_L + \rho_{00}\boldsymbol{J}_L \\
\boldsymbol{\Gamma}_1 &= (\rho_1 - \rho_{11})\boldsymbol{I}_L + \rho_{11}\boldsymbol{J}_L \\
\boldsymbol{\Gamma}_2 &= (1 - \rho_2)\boldsymbol{I}_L + \rho_2\boldsymbol{J}_L.
\end{aligned}$$

Plugging in these explicit forms into our current variance expression under the ICC parameterization (shown below for ease)

$$\begin{aligned}
\boldsymbol{\Omega}_\delta =& \frac{IT}{N}\boldsymbol{\Lambda}_y^{1/2}\Bigg[\left(ITU - TW + U^2 - IV\right)(\boldsymbol{\Gamma}_2 - N\boldsymbol{\Gamma}_1 + (N-1)\boldsymbol{\Gamma}_0)^{-1} \\
&- \left(U^2 - IV\right)(\boldsymbol{\Gamma}_2 + (T-1)N\boldsymbol{\Gamma}_1 + (N-1)\boldsymbol{\Gamma}_0)^{-1}\Bigg]^{-1}\boldsymbol{\Lambda}_y^{1/2},
\end{aligned}$$

we get the following simplified expression

$$\begin{aligned}
\boldsymbol{\Omega}_\delta =& \frac{IT}{N}\boldsymbol{\Lambda}_y^{1/2}\Bigg[\left(ITU - TW + U^2 - IV\right)((1-\rho_2 - N(\rho_1 - \rho_{11}) + (N-1)(\rho_0 - \rho_{00}))\boldsymbol{I}_L + (\rho_2 - N\rho_{11} + (N-1)\rho_{00})\boldsymbol{J}_L)^{-1} \\
&- \left(U^2 - IV\right)((1-\rho_2 + (T-1)N(\rho_1 - \rho_{11}) + (N-1)(\rho_0 - \rho_{00}))\boldsymbol{I}_L + (\rho_2 + (T-1)N\rho_{11} + (N-1)\rho_{00})\boldsymbol{J}_L)^{-1}\Bigg]^{-1}\boldsymbol{\Lambda}_y^{1/2} \\
=& \frac{IT}{N}\boldsymbol{\Lambda}_y^{1/2}\Bigg[\left(ITU - TW + U^2 - IV\right)((\lambda_2 - \tau_2)\boldsymbol{I}_L + \tau_2\boldsymbol{J}_L)^{-1} - \left(U^2 - IV\right)((\lambda_3 - \tau_3)\boldsymbol{I}_L + \tau_3\boldsymbol{J}_L)^{-1}\Bigg]^{-1}\boldsymbol{\Lambda}_y^{1/2} \\
=& \frac{IT}{N}\boldsymbol{\Lambda}_y^{1/2}\Bigg[\left(ITU - TW + U^2 - IV\right)\left(\frac{1}{\lambda_2 - \tau_2}\boldsymbol{I}_L - \frac{\tau_2}{(\lambda_2 - \tau_2)(\lambda_2 + (L-1)\tau_2)}\boldsymbol{J}_L\right) \\
&- \left(U^2 - IV\right)\left(\frac{1}{\lambda_3 - \tau_3}\boldsymbol{I}_L - \frac{\tau_3}{(\lambda_3 - \tau_3)(\lambda_3 + (L-1)\tau_3)}\boldsymbol{J}_L\right)\Bigg]^{-1}\boldsymbol{\Lambda}_y^{1/2},
\end{aligned}$$

where $\lambda_2 = 1 + (N-1)\rho_0 - N\rho_1$, $\lambda_3 = 1 + (N-1)\rho_0 + (T-1)N\rho_1$, $\tau_2 = (N-1)\rho_{00} - N\rho_{11} + \rho_2$, and $\tau_3 = \tau_2 + TN\rho_{11}$. To continue simplifying our expression, let $x = ITU - TW + U^2 - IV$, $y = U^2 - IV$, $a_2 = \lambda_2 - \tau_2$, $a_3 = \lambda_3 - \tau_3$,

$b_2 = \lambda_2 + (L-1)\tau_2$, and $b_3 = \lambda_3 + (L-1)\tau_3$. This gives us

$$\begin{aligned}
\boldsymbol{\Omega}_\delta &= \frac{IT}{N}\boldsymbol{\Lambda}_y^{1/2}\left[x\left(\frac{1}{a_2}\boldsymbol{I}_L - \frac{\tau_2}{a_2 b_2}\boldsymbol{J}_L\right) - y\left(\frac{1}{a_3}\boldsymbol{I}_L - \frac{\tau_3}{a_3 b_3}\boldsymbol{J}_L\right)\right]^{-1}\boldsymbol{\Lambda}_y^{1/2} \\
&= \frac{IT}{N}\boldsymbol{\Lambda}_y^{1/2}\left[\frac{xa_3 - ya_2}{a_2a_3}\boldsymbol{I}_L + \frac{y\tau_3a_2b_2 - x\tau_2a_3b_3}{a_2a_3b_2b_3}\boldsymbol{J}_L\right]^{-1}\boldsymbol{\Lambda}_y^{1/2} \\
&= \frac{(IT/N)a_2a_3}{xa_3 - ya_2}\boldsymbol{\Lambda}_y^{1/2}\left[\boldsymbol{I}_L - \frac{y\tau_3a_2b_2 - x\tau_2a_3b_3}{(xa_3 - ya_2)\,b_2b_3 + L\,(y\tau_3a_2b_2 - x\tau_2a_3b_3)}\boldsymbol{J}_L\right]\boldsymbol{\Lambda}_y^{1/2}
\end{aligned}$$

Common ICCs across endpoints: proof of Theorem 1

For ease, we restate Theorem 1 below:

THEOREM **1.** *Under the parsimonious parameterization with common ICC values across endpoints, the l-th diagonal element of $\boldsymbol{\Omega}_\delta$ can be further written in the following analytical form*

$$\begin{aligned}
var(\hat{\delta}_l) =& \frac{(IT/N)\sigma_{yl}^2}{\left(ITU - TW + U^2 - IV\right)(\lambda_3 - \tau_3) - \left(U^2 - IV\right)(\lambda_2 - \tau_2)} \times \\
& \frac{\left(ITU - TW + U^2 - IV\right)\lambda_2(\lambda_3 - \tau_3)\{\lambda_3 + (L-1)\tau_3\} - \left(U^2 - IV\right)\lambda_3(\lambda_2 - \tau_2)\{\lambda_2 + (L-1)\tau_2\}}{\left(ITU - TW + U^2 - IV\right)\{\lambda_3 + (L-1)\tau_3\} - \left(U^2 - IV\right)\{\lambda_2 + (L-1)\tau_2\}}.
\end{aligned}$$

Furthermore, denote the variance of the l-th intervention effect estimator based on a univariate Hooper and Girling model (Hooper et al., 2016; Girling and Hemming, 2016) is

$$var^{HG}(\hat{\delta}_l) = \frac{(IT/N)\sigma_{yl}^2\lambda_2\lambda_3}{\left(ITU - TW + U^2 - IV\right)\lambda_3 - \left(U^2 - IV\right)\lambda_2}$$

and $var(\hat{\delta}_l) \le var^{HG}(\hat{\delta}_l)$ for any set of valid design parameters, with equality holds when $\tau_2\lambda_3 = \tau_3\lambda_2$ or $\rho_{00} = \rho_{11} = \rho_2 = 0$ (a special case when $\tau_2\lambda_3 = \tau_3\lambda_2$).

Proof:

Given our expression for $\boldsymbol{\Omega}_\delta$ under the common ICCs assumption (shown above), the l-th diagonal element is

$$\begin{aligned}
\mathrm{var}(\hat{\delta}_l) &= \left(\frac{IT\sigma_{yl}^2}{N}\right)\left(\frac{a_2a_3}{xa_3 - ya_2}\right)\left(1 - \frac{y\tau_3a_2b_2 - x\tau_2a_3b_3}{(xa_3 - ya_2)\,b_2b_3 + L\,(y\tau_3a_2b_2 - x\tau_2a_3b_3)}\right) \\
&= \left(\frac{(IT/N)\sigma_{yl}^2}{xa_3 - ya_2}\right)\left(\frac{x(b_2 - (L-1)\tau_2)a_3b_3 - y(b_3 - (L-1)\tau_3)a_2b_2}{x(b_2 - L\tau_2)a_3b_3 - y(b_3 - L\tau_3)a_2b_2}\right)a_2a_3 \\
&= \left(\frac{(IT/N)\sigma_{yl}^2}{xa_3 - ya_2}\right)\left(\frac{x\lambda_2a_3b_3 - y\lambda_3a_2b_2}{xb_3 - yb_2}\right) \\
&= \frac{(IT/N)\sigma_{yl}^2}{(ITU - TW + U^2 - IV)(\lambda_3 - \tau_3) - (U^2 - IV)(\lambda_2 - \tau_2)} \times
\end{aligned}$$

$$\frac{(ITU-TW+U^2-IV)\lambda_2(\lambda_3-\tau_3)\{\lambda_3+(L-1)\tau_3\}-(U^2-IV)\lambda_3(\lambda_2-\tau_2)\{\lambda_2+(L-1)\tau_2\}}{(ITU-TW+U^2-IV)\{\lambda_3+(L-1)\tau_3\}-(U^2-IV)\{\lambda_2+(L-1)\tau_2\}},$$

which matches our variance expression in Theorem 1. Under the Hooper and Girling LMM, the variance of the l-th outcome is

$$\mathrm{var}^{\mathrm{HG}}(\widehat{\delta}_l)=\frac{(IT/N)\sigma_{yl}^2\lambda_2\lambda_3}{\left(ITU-TW+U^2-IV\right)\lambda_3-\left(U^2-IV\right)\lambda_2}.$$

Therefore, the variance ratio comparing the MLMM to the LMM is

$$\begin{aligned}\frac{\mathrm{var}(\widehat{\delta}_l)}{\mathrm{var}^{\mathrm{HG}}(\widehat{\delta}_l)}&=\left(\frac{\left(ITU-TW+U^2-IV\right)\lambda_3-\left(U^2-IV\right)\lambda_2}{(IT/N)\sigma_{yl}^2\lambda_2\lambda_3}\right)\left(\frac{(IT/N)\sigma_{yl}^2}{(ITU-TW+U^2-IV)(\lambda_3-\tau_3)-(U^2-IV)(\lambda_2-\tau_2)}\right)\times\\&\frac{(ITU-TW+U^2-IV)\lambda_2(\lambda_3-\tau_3)\{\lambda_3+(L-1)\tau_3\}-(U^2-IV)\lambda_3(\lambda_2-\tau_2)\{\lambda_2+(L-1)\tau_2\}}{(ITU-TW+U^2-IV)\{\lambda_3+(L-1)\tau_3\}-(U^2-IV)\{\lambda_2+(L-1)\tau_2\}}\\&=\frac{\left(ITU-TW+U^2-IV\right)\lambda_3-\left(U^2-IV\right)\lambda_2}{(ITU-TW+U^2-IV)\{\lambda_3+(L-1)\tau_3\}-(U^2-IV)\{\lambda_2+(L-1)\tau_2\}}\times\\&\frac{(ITU-TW+U^2-IV)\lambda_3^{-1}(\lambda_3-\tau_3)\{\lambda_3+(L-1)\tau_3\}-(U^2-IV)\lambda_2^{-1}(\lambda_2-\tau_2)\{\lambda_2+(L-1)\tau_2\}}{(ITU-TW+U^2-IV)(\lambda_3-\tau_3)-(U^2-IV)(\lambda_2-\tau_2)}.\end{aligned}$$

We can again use $x=ITU-TW+U^2-IV$, $y=U^2-IV$, $a_2=\lambda_2-\tau_2$, $a_3=\lambda_3-\tau_3$, $b_2=\lambda_2+(L-1)\tau_2$, and $b_3=\lambda_3+(L-1)\tau_3$. This gives us

$$\begin{aligned}&=\frac{x\lambda_3-y\lambda_2}{xb_3-yb_2}\left(\frac{x\lambda_3^{-1}a_3b_3-y\lambda_2^{-1}a_2b_2}{xa_3-ya_2}\right)\\&=\frac{x^2a_3b_3-xy\lambda_3\lambda_2^{-1}a_2b_2-xy\lambda_3^{-1}\lambda_2a_3b_3+y^2a_2b_2}{x^2a_3b_3-xya_2b_3-xya_3b_2+y^2a_2b_2}.\end{aligned}$$

To evaluate whether this ratio is less than or greater than one we can take the difference of the numerator and denominator.

$$\begin{aligned}\text{num. - den.}&=x^2a_3b_3-xy\lambda_3\lambda_2^{-1}a_2b_2-xy\lambda_3^{-1}\lambda_2a_3b_3+y^2a_2b_2-(x^2a_3b_3-xya_2b_3-xya_3b_2+y^2a_2b_2)\\&=-xy(\{\lambda_2+(L-1)\tau_2\}(-\lambda_3\lambda_2^{-1}\tau_2+\tau_3)+\{\lambda_3+(L-1)\tau_3\}(-\lambda_3^{-1}\lambda_2\tau_3+\tau_2))\\&=-xy(2\tau_2\tau_3\lambda_2\lambda_3-\lambda_3^2\tau_2^2-\lambda_2^2\tau_3^2)((L-1)/\lambda_2\lambda_3)\\&=xy(\lambda_3\tau_2-\lambda_2\tau_3)^2((L-1)/\lambda_2\lambda_3).\end{aligned}$$

We know each term is positive except for $x=ITU-TW+U^2-IV$ and $y=U^2-IV$. Starting with $x=ITU-TW+U^2-IV$ and as shown in Theorem 1 of Davis-Plourde et al. (2021), we can rewrite this term using $\mathbf{1}_T^\top\mathbf{\Omega}\mathbf{1}_T=I^{-2}(IV-U^2)$ and $\mathrm{tr}(\mathbf{\Omega})=I^{-2}(IU-W)$ where $\mathbf{\Omega}=I^{-1}\sum_{i=1}^I\boldsymbol{X}_i\boldsymbol{X}_i^\top-(I^{-1}\sum_{i=1}^I\boldsymbol{X}_i)(I^{-1}\sum_{i=1}^I\boldsymbol{X}_i^\top)$ is the covariance matrix of the intervention vector under a specific design and $\tau_X=\{(T-1)\mathrm{tr}(\mathbf{\Omega})\}^{-1}\{\mathbf{1}_T^\top\mathbf{\Omega}\mathbf{1}_T-\mathrm{tr}(\mathbf{\Omega})\}\in$

$[-1, 1]$ is the generalized ICC of the intervention, which is the ratio of average covariance over the average variance and measures the similarity between the intervention status for each cluster in different periods (Kistner and Muller, 2004). This gives us

$$\begin{aligned} x = ITU - TW + U^2 - IV &= I^2 T \text{tr}(\boldsymbol{\Omega}) - I^2 \mathbf{1}_T^\top \boldsymbol{\Omega} \mathbf{1}_T \\ &= I^2 \left((T-1)\text{tr}(\boldsymbol{\Omega}) - \{ \mathbf{1}_T^\top \boldsymbol{\Omega} \mathbf{1}_T - \text{tr}(\boldsymbol{\Omega}) \} \right) \\ &= I^2 (T-1) \text{tr}(\boldsymbol{\Omega}) \left(1 - \tau_X\right) \geq 0. \end{aligned}$$

Also shown in Davis-Plourde et al. (2021), typically under a standard SW-CRT design $\tau_X \in (0, 1)$ thus $x > 0$. Next, let's evaluate the sign of y given by $y = U^2 - IV = \left(\sum_{i=1}^{I} \sum_{j=1}^{T} X_{ij}\right)^2 - I \sum_{i=1}^{I} \left(\sum_{j=1}^{T} X_{ij}\right)^2$. By the Cauchy-Schwarz inequality we know that

$$\begin{aligned} &\left(\sum_{i=1}^{I} \left\{ 1 \times \sum_{j=1}^{T} X_{ij} \right\}\right)^2 - \left(\sum_{i=1}^{I} 1^2\right) \left(\sum_{i=1}^{I} \left\{ \sum_{j=1}^{T} X_{ij} \right\}^2 \right) \leq 0 \\ &\Rightarrow \left(\sum_{i=1}^{I} \sum_{j=1}^{T} X_{ij}\right)^2 - I \sum_{i=1}^{I} \left(\sum_{j=1}^{T} X_{ij}\right)^2 \leq 0 \\ &\Rightarrow y = U^2 - IV \leq 0. \end{aligned}$$

Additionally, y can be rewritten as $y = -I^2 \mathbf{1}_T^\top \boldsymbol{\Omega} \mathbf{1}_T < 0$. Therefore, the difference between the numerator and denominator is less than or equal to zero which leads to a variance ratio of less than or equal to one. Thus, $\text{var}(\hat{\delta}_l) \leq \text{var}^{\text{HG}}(\hat{\delta}_l)$ for any set of valid design parameters, with equality holds when $\tau_2 \lambda_3 = \tau_3 \lambda_2$ or $\rho_{00} = \rho_{11} = \rho_2 = 0$ (a special case when $\tau_2 \lambda_3 = \tau_3 \lambda_2$).

Web Appendix E

Common ICCs and intervention effects: derivation of variance expression, $\text{var}(\widehat{\delta'})$

Recall that if we assume common ICCs we have

$$\begin{aligned} \boldsymbol{\Gamma}_0 &= (\rho_0 - \rho_{00}) \boldsymbol{I}_L + \rho_{00} \boldsymbol{J}_L \\ \boldsymbol{\Gamma}_1 &= (\rho_1 - \rho_{11}) \boldsymbol{I}_L + \rho_{11} \boldsymbol{J}_L \\ \boldsymbol{\Gamma}_2 &= (1 - \rho_2) \boldsymbol{I}_L + \rho_2 \boldsymbol{J}_L. \end{aligned}$$

And if we assume common intervention effects, then the variance of the intervention effect estimator becomes

$$\text{var}(\widehat{\delta'}) = \frac{IT}{N} \Bigg[\left(ITU - TW + U^2 - IV \right) \boldsymbol{\omega}^\top \boldsymbol{\Lambda}_y^{-1/2} \left(\boldsymbol{\Gamma}_2 - N \boldsymbol{\Gamma}_1 + (N-1) \boldsymbol{\Gamma}_0 \right)^{-1} \boldsymbol{\Lambda}_y^{-1/2} \boldsymbol{\omega}$$

$$-\left(U^2 - IV\right)\boldsymbol{\omega}^\top \boldsymbol{\Lambda}_y^{-1/2} \left(\boldsymbol{\Gamma}_2 + (T-1)N\boldsymbol{\Gamma}_1 + (N-1)\boldsymbol{\Gamma}_0\right)^{-1} \boldsymbol{\Lambda}_y^{-1/2}\boldsymbol{\omega}\Bigg]^{-1}$$

$$= \frac{IT}{N}\Bigg[\left(ITU - TW + U^2 - IV\right)\boldsymbol{\omega}^\top \boldsymbol{\Lambda}_y^{-1/2} \left((\lambda_2 - \tau_2)\boldsymbol{I}_L + \tau_2 \boldsymbol{J}_L\right)^{-1} \boldsymbol{\Lambda}_y^{-1/2}\boldsymbol{\omega}$$

$$-\left(U^2 - IV\right)\boldsymbol{\omega}^\top \boldsymbol{\Lambda}_y^{-1/2} \left((\lambda_3 - \tau_3)\boldsymbol{I}_L + \tau_3 \boldsymbol{J}_L\right)^{-1} \boldsymbol{\Lambda}_y^{-1/2}\boldsymbol{\omega}\Bigg]^{-1}$$

$$= \frac{IT}{N}\Bigg[\left(ITU - TW + U^2 - IV\right)\left((\lambda_2 - \tau_2)\left(\boldsymbol{\omega}^\top\right)^{-1}\boldsymbol{\Lambda}_y\boldsymbol{\omega}^{-1} + \tau_2\left(\boldsymbol{\omega}^\top\right)^{-1}\boldsymbol{\Lambda}_y^{1/2}\boldsymbol{J}_L\boldsymbol{\Lambda}_y^{1/2}\boldsymbol{\omega}^{-1}\right)^{-1}$$

$$-\left(U^2 - IV\right)\left((\lambda_3 - \tau_3)\left(\boldsymbol{\omega}^\top\right)^{-1}\boldsymbol{\Lambda}_y\boldsymbol{\omega}^{-1} + \tau_3\left(\boldsymbol{\omega}^\top\right)^{-1}\boldsymbol{\Lambda}_y^{1/2}\boldsymbol{J}_L\boldsymbol{\Lambda}_y^{1/2}\boldsymbol{\omega}^{-1}\right)^{-1}\Bigg]^{-1}$$

$$= \frac{IT}{LN\lambda_1}\Bigg[\left(ITU - TW + U^2 - IV\right)(\lambda_2 + (L-1)\tau_2)^{-1} - \left(U^2 - IV\right)(\lambda_3 + (L-1)\tau_3)^{-1}\Bigg]^{-1}$$

$$= \frac{(IT/(LN\lambda_1))(\lambda_2 + (L-1)\tau_2)(\lambda_3 + (L-1)\tau_3)}{(ITU - TW + U^2 - IV)(\lambda_3 + (L-1)\tau_3) - (U^2 - IV)(\lambda_2 + (L-1)\tau_2)},$$

where $\lambda_2 = 1 + (N-1)\rho_0 - N\rho_1$, $\lambda_3 = 1 + (N-1)\rho_0 + (T-1)N\rho_1$, $\tau_2 = (N-1)\rho_{00} - N\rho_{11} + \rho_2$, and $\tau_3 = \tau_2 + TN\rho_{11}$ as discussed previously. Under the common intervention effects model $\boldsymbol{\omega} = \left(\sigma_{y1}(1-\rho_0^1)^{1/2}, \ldots, \sigma_{yL}(1-\rho_0^L)^{1/2}\right)^\top$, but under the common ICCs assumption $\rho_0^l = \rho_0$, therefore $\boldsymbol{\omega} = \left(\sigma_{y1}\lambda_1^{1/2}, \ldots, \sigma_{yL}\lambda_1^{1/2}\right)^\top$ where same as λ_2 and λ_3, $\lambda_1 = 1 - \rho_0$ is a distinct eigenvalue of the (endpoint-specific) nested exchangeable correlation structure (Li et al., 2018) defined for cross-sectional SW-CRTs with a univariate outcome.

Common ICCs and intervention effects: proof of Theorem 2

For ease, we restate Theorem 2 below:

THEOREM **2.** *Under the parsimonious parameterization with common ICC values and a common intervention effect across endpoints, the variance of the l-th intervention effect estimator (unscaled) under model (5), i.e. $\delta_l = \sigma_{yl}\lambda_1^{1/2}\delta'$, is*

$$var^{both}(\hat{\delta}_l) = \frac{(IT/(LN))\sigma_{yl}^2(\lambda_2 + (L-1)\tau_2)(\lambda_3 + (L-1)\tau_3)}{(ITU - TW + U^2 - IV)(\lambda_3 + (L-1)\tau_3) - (U^2 - IV)(\lambda_2 + (L-1)\tau_2)}.$$

As shown in Theorem 1, under the parsimonious parameterization with common ICC values across endpoints, the l-th diagonal element of $\boldsymbol{\Omega}_\delta$ is denoted by

$$var^{ICC}(\hat{\delta}_l) = \frac{(IT/N)\sigma_{yl}^2}{\left(ITU - TW + U^2 - IV\right)(\lambda_3 - \tau_3) - \left(U^2 - IV\right)(\lambda_2 - \tau_2)} \times$$
$$\frac{\left(ITU - TW + U^2 - IV\right)\lambda_2(\lambda_3 - \tau_3)\{\lambda_3 + (L-1)\tau_3\} - \left(U^2 - IV\right)\lambda_3(\lambda_2 - \tau_2)\{\lambda_2 + (L-1)\tau_2\}}{\left(ITU - TW + U^2 - IV\right)\{\lambda_3 + (L-1)\tau_3\} - \left(U^2 - IV\right)\{\lambda_2 + (L-1)\tau_2\}},$$

and $var^{both}(\hat{\delta}_l) < var^{ICC}(\hat{\delta}_l)$ for any set of valid design parameters.

Proof:

$$\begin{aligned}
\frac{\text{var}^{\text{both}}(\widehat{\delta}_l)}{\text{var}^{\text{ICC}}(\widehat{\delta}_l)} &= \frac{(IT/(LN))\sigma_{yl}^2(\lambda_2+(L-1)\tau_2)(\lambda_3+(L-1)\tau_3)}{(ITU-TW+U^2-IV)(\lambda_3+(L-1)\tau_3)-(U^2-IV)(\lambda_2+(L-1)\tau_2)} \\
&\times \frac{\left(ITU-TW+U^2-IV\right)(\lambda_3-\tau_3)-\left(U^2-IV\right)(\lambda_2-\tau_2)}{(IT/N)\sigma_{yl}^2} \\
&\times \frac{\left(ITU-TW+U^2-IV\right)\{\lambda_3+(L-1)\tau_3\}-\left(U^2-IV\right)\{\lambda_2+(L-1)\tau_2\}}{(ITU-TW+U^2-IV)\,\lambda_2(\lambda_3-\tau_3)\{\lambda_3+(L-1)\tau_3\}-(U^2-IV)\,\lambda_3(\lambda_2-\tau_2)\{\lambda_2+(L-1)\tau_2\}} \\
&= \frac{(1/L)(\lambda_2+(L-1)\tau_2)(\lambda_3+(L-1)\tau_3)\left(\left(ITU-TW+U^2-IV\right)(\lambda_3-\tau_3)-\left(U^2-IV\right)(\lambda_2-\tau_2)\right)}{(ITU-TW+U^2-IV)\,\lambda_2(\lambda_3-\tau_3)\{\lambda_3+(L-1)\tau_3\}-(U^2-IV)\,\lambda_3(\lambda_2-\tau_2)\{\lambda_2+(L-1)\tau_2\}}
\end{aligned}$$

Again using $x = ITU - TW + U^2 - IV$, $y = U^2 - IV$, $a_2 = \lambda_2 - \tau_2$, $a_3 = \lambda_3 - \tau_3$, $b_2 = \lambda_2 + (L-1)\tau_2$, and $b_3 = \lambda_3 + (L-1)\tau_3$. This gives us

$$\begin{aligned}
&= \frac{(1/L)b_2b_3\,(xa_3 - ya_2)}{x\lambda_2a_3b_3 - y\lambda_3a_2b_2} \\
&= (1/L)\frac{(xa_3b_3)b_2 - (ya_2b_2)b_3}{(xa_3b_3)\lambda_2 - (ya_2b_2)\lambda_3} \\
&= (1/L)\frac{(xa_3b_3)\{\lambda_2 + (L-1)\tau_2\} - (ya_2b_2)\{\lambda_3 + (L-1)\tau_3\}}{(xa_3b_3)\lambda_2 - (ya_2b_2)\lambda_3} \\
&= (1/L)\frac{(xa_3b_3)\lambda_2 - (ya_2b_2)\lambda_3 + (L-1)(xa_3b_3)\tau_2 - (L-1)(ya_2b_2)\tau_3}{(xa_3b_3)\lambda_2 - (ya_2b_2)\lambda_3} \\
&= (1/L)\left(1 + \frac{(L-1)\{(xa_3b_3)\tau_2 - (ya_2b_2)\tau_3\}}{(xa_3b_3)\lambda_2 - (ya_2b_2)\lambda_3}\right).
\end{aligned}$$

We are interested in determining when this ratio is less than 1 which is when

$$\begin{aligned}
&(1/L)\left(1 + \frac{(L-1)\{(xa_3b_3)\tau_2 - (ya_2b_2)\tau_3\}}{(xa_3b_3)\lambda_2 - (ya_2b_2)\lambda_3}\right) < 1 \\
&\Rightarrow \frac{(xa_3b_3)\tau_2 - (ya_2b_2)\tau_3}{(xa_3b_3)\lambda_2 - (ya_2b_2)\lambda_3} < 1.
\end{aligned}$$

We know that $\tau_2 < \lambda_2$ and $\tau_3 < \lambda_3$ by definition and we already showed that $x > 0$ and $y < 0$ under a SW-CRT, therefore the numerator is always less than the denominator which means the variance ratio is always less than 1.

Web Appendix F

Closed-cohort design: derivation of $\widetilde{\boldsymbol{V}}_i$

Model under a Cluster-Period Means Approach

Again, we can simplify our approach using cluster-period means (Li et al., 2021; Davis-Plourde et al., 2021). Let $\overline{\boldsymbol{Y}}_{ij} = (\overline{Y}_{ij1}, \ldots, \overline{Y}_{ijL})^\top$ where $\overline{Y}_{ijl} = \frac{1}{N}\sum_{k=1}^N Y_{ijkl}$ and let $\overline{\gamma}_i = (\overline{\gamma}_{i1}, \ldots, \overline{\gamma}_{iL})^\top$ where $\overline{\gamma}_{il} = \frac{1}{N}\sum_{k=1}^N \gamma_{ikl}$. Also let

$\overline{\boldsymbol{\epsilon}}_{ij} = (\overline{\epsilon}_{ij1}, \dots, \overline{\epsilon}_{ijL})^\top$ with $\overline{\epsilon}_{ijl} = \frac{1}{N}\sum_{k=1}^{N} \epsilon_{ijkl}$. Our MLMM is equivalent to

$$\overline{\boldsymbol{Y}}_{ij} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_j + X_{ij}\boldsymbol{\delta} + \boldsymbol{b}_i + \boldsymbol{s}_{ij} + \overline{\boldsymbol{\gamma}}_i + \overline{\boldsymbol{\epsilon}}_{ij}.$$

Within-Period Block of $\widetilde{\boldsymbol{V}}_i$

The variance of $\overline{Y}_{ijl}$ is

$$\begin{aligned}
\mathrm{var}(\overline{Y}_{ijl}) &= \frac{1}{N}\left(\mathrm{var}(Y_{ijkl}) + (N-1)\mathrm{cov}(Y_{ijkl}, Y_{ijk'l})\right) \\
&= \frac{1}{N}\left(\sigma_{bl}^2 + \sigma_{sl}^2 + \sigma_{\gamma l}^2 + \sigma_{\epsilon l}^2 + (N-1)(\sigma_{bl}^2 + \sigma_{sl}^2)\right) \\
&= \sigma_{bl}^2 + \sigma_{sl}^2 + \frac{\sigma_{\gamma l}^2 + \sigma_{\epsilon l}^2}{N}.
\end{aligned}$$

The covariance between outcomes in the same period is

$$\begin{aligned}
\mathrm{cov}(\overline{Y}_{ijl}, \overline{Y}_{ijl'}) &= \frac{1}{N}\left(\mathrm{cov}(Y_{ijkl}, Y_{ijkl'}) + (N-1)\mathrm{cov}(Y_{ijkl}, Y_{ijk'l'})\right) \\
&= \frac{1}{N}\left(\sigma_{bll'} + \sigma_{sll'} + \sigma_{\gamma ll'} + \sigma_{\epsilon ll'} + (N-1)(\sigma_{bll'} + \sigma_{sll'})\right) \\
&= \sigma_{bll'} + \sigma_{sll'} + \frac{\sigma_{\gamma ll'} + \sigma_{\epsilon ll'}}{N}.
\end{aligned}$$

Therefore, the within-period block is $\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\left(\boldsymbol{\Sigma}_\gamma + \boldsymbol{\Sigma}_\epsilon\right)$.

Between-Period Block of $\widetilde{\boldsymbol{V}}_i$

The within outcome covariance between periods is

$$\begin{aligned}
\mathrm{cov}(\overline{Y}_{ijl}, \overline{Y}_{ij'l}) &= \frac{1}{N}\left(\mathrm{cov}(Y_{ijkl}, Y_{ij'kl}) + (N-1)\mathrm{cov}(Y_{ijkl}, Y_{ij'k'l})\right) \\
&= \frac{1}{N}\left(\sigma_{bl}^2 + \sigma_{\gamma l}^2 + (N-1)\sigma_{bl}^2\right) \\
&= \sigma_{bl}^2 + \frac{\sigma_{\gamma l}^2}{N}.
\end{aligned}$$

The between outcome covariance between periods is

$$\begin{aligned}
\mathrm{cov}(\overline{Y}_{ijl}, \overline{Y}_{ij'l'}) &= \frac{1}{N}\left(\mathrm{cov}(Y_{ijkl}, Y_{ij'kl'}) + (N-1)\mathrm{cov}(Y_{ijkl}, Y_{ij'k'l'})\right) \\
&= \frac{1}{N}\left(\sigma_{bll'} + \sigma_{\gamma ll'} + (N-1)\sigma_{bll'}\right) \\
&= \sigma_{bll'} + \frac{\sigma_{\gamma ll'}}{N}.
\end{aligned}$$

Therefore, the between-period block is $\boldsymbol{\Sigma}_b + \frac{1}{N}\boldsymbol{\Sigma}_\gamma$.

Combining Within-Period and Between-Period Blocks to obtain $\widetilde{\boldsymbol{V}}_i$

Combining the within-period and between-period block gives us

$$\begin{aligned}\widetilde{\boldsymbol{V}}_i &= \boldsymbol{I}_T \otimes \left(\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\left(\boldsymbol{\Sigma}_\gamma + \boldsymbol{\Sigma}_\epsilon\right) - \left(\boldsymbol{\Sigma}_b + \frac{1}{N}\boldsymbol{\Sigma}_\gamma\right)\right) + \boldsymbol{J}_T \otimes \left(\boldsymbol{\Sigma}_b + \frac{1}{N}\boldsymbol{\Sigma}_\gamma\right) \\ &= \boldsymbol{I}_T \otimes \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right) + \boldsymbol{J}_T \otimes \left(\boldsymbol{\Sigma}_b + \frac{1}{N}\boldsymbol{\Sigma}_\gamma\right).\end{aligned}$$

Closed-cohort design: derivation of $\widetilde{\boldsymbol{V}}_i^{-1}$

The inverse of $\widetilde{\boldsymbol{V}}_i$ can be computed using Leiva (2007) since $\widetilde{\boldsymbol{V}}_i$ is exchangeable giving us

$$\begin{aligned}\widetilde{\boldsymbol{V}}_i^{-1} &= \boldsymbol{I}_T \otimes \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\left(\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\left(\boldsymbol{\Sigma}_\gamma + \boldsymbol{\Sigma}_\epsilon\right) + (T-1)\left(\boldsymbol{\Sigma}_b + \frac{1}{N}\boldsymbol{\Sigma}_\gamma\right)\right)^{-1} - \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\right] \\ &= \boldsymbol{I}_T \otimes \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{T}{N}\boldsymbol{\Sigma}_\gamma + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1} - \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\right].\end{aligned}$$

Closed-cohort design: derivation of $\boldsymbol{\Omega}_\delta$

Again let $\boldsymbol{Z}_i = (\boldsymbol{I}_T, \boldsymbol{X}_i) \otimes \boldsymbol{I}_L$ where $\boldsymbol{X}_i$ is the randomization schedule for cluster i. The covariance matrix for the model estimators are $(\sum_{i=1}^{I} \boldsymbol{Z}_i^\top \widetilde{\boldsymbol{V}}_i^{-1}\boldsymbol{Z}_i)^{-1}$ where

$$\sum_{i=1}^{I}\boldsymbol{Z}_i^\top \widetilde{\boldsymbol{V}}_i^{-1}\boldsymbol{Z}_i = \sum_{i=1}^{I}\left[\begin{pmatrix}\boldsymbol{I}_T \\ \boldsymbol{X}_i^\top\end{pmatrix} \otimes \boldsymbol{I}_L\right]\widetilde{\boldsymbol{V}}_i^{-1}\left[\begin{pmatrix}\boldsymbol{I}_T & \boldsymbol{X}_i\end{pmatrix} \otimes \boldsymbol{I}_L\right] = \begin{bmatrix}\boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22}\end{bmatrix},$$

where $\boldsymbol{\Omega}_{11}$ is a $TL \times TL$ matrix, $\boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}_{21}^\top$ is a $TL \times L$ matrix, and $\boldsymbol{\Omega}_{22}$ is a $L \times L$ matrix. Block matrix inversion gives us $\boldsymbol{\Omega}_\delta = \boldsymbol{\Omega}_{22}^{-1} = (\boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12})^{-1}$.

We can rewrite $\widetilde{\boldsymbol{V}}_i$ as,

$$\widetilde{\boldsymbol{V}}_i^{-1} = \boldsymbol{I}_T \otimes \boldsymbol{A} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right].$$

where $\boldsymbol{A} = \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}$ and $\boldsymbol{B} = \left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{T}{N}\boldsymbol{\Sigma}_\gamma + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}$. Using this expression we can generate our Ωs.

Recall in Web Appendix B that we already derived expressions for each Ω in terms of $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$\boldsymbol{\Omega}_{22} = U\boldsymbol{A} + \frac{V}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right]$$

$$\begin{aligned}\boldsymbol{\Omega}_{12} &= \boldsymbol{\Omega}_{21}^{\top} = \sum_{i=1}^{I} \boldsymbol{X}_i \otimes \boldsymbol{A} + U\mathbf{1}_T \otimes \frac{1}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right] \\ \boldsymbol{\Omega}_{11}^{-1} &= \left(I\widetilde{\boldsymbol{V}}_i^{-1}\right)^{-1} = \frac{1}{I}\widetilde{\boldsymbol{V}}_i = \frac{1}{I}\left(\boldsymbol{I}_T \otimes \boldsymbol{A}^{-1} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\boldsymbol{B}^{-1} - \boldsymbol{A}^{-1}\right]\right).\end{aligned}$$

This gives us

$$\begin{aligned}\boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12} &= \frac{1}{I}\left(W\boldsymbol{A} + \frac{U^2}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right]\right) \\ \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12} &= \frac{1}{I}\left(IU\boldsymbol{A} + \frac{IV}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right] - \left(W\boldsymbol{A} + \frac{U^2}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right]\right)\right) \\ &= \frac{1}{IT}\left((ITU - TW - IV + U^2)\boldsymbol{A} - (U^2 - IV)\boldsymbol{B}\right) \\ \boldsymbol{\Omega}_{\delta} &= IT\left[(ITU - TW - IV + U^2)\boldsymbol{A} - (U^2 - IV)\boldsymbol{B}\right]^{-1} \\ &= IT\left[(ITU - TW + U^2 - IV)\left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_{\epsilon}\right)^{-1} - (U^2 - IV)\left(T\boldsymbol{\Sigma}_b + \frac{T}{N}\boldsymbol{\Sigma}_{\gamma} + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_{\epsilon}\right)^{-1}\right]^{-1},\end{aligned}$$

where $U = \sum_{j=1}^{T} X_{ij}$, $V = \sum_{i=1}^{I}\left(\sum_{j=1}^{T} X_{ij}\right)^2$, and $W = \sum_{j=1}^{T}\left(\sum_{i=1}^{I} X_{ij}\right)^2$. To rewrite this expression in terms of the ICCs we can use Web Table 1. Specifically, we note that $\sigma_{bl}^2 = \sigma_{yl}^2\rho_1^l$, $\sigma_{bll'} = \sigma_{yl}\sigma_{yl'}\rho_1^{ll'}$, $\sigma_{sl}^2 = \sigma_{yl}^2(\rho_0^l - \rho_1^l)$, $\sigma_{sll'} = \sigma_{yl}\sigma_{yl'}(\rho_0^{ll'} - \rho_1^{ll'})$, $\sigma_{\gamma l}^2 = \sigma_{yl}^2(\rho_2^l - \rho_1^l)$, $\sigma_{\gamma ll'} = \sigma_{yl}\sigma_{yl'}(\rho_{2,1}^{ll'} - \rho_1^{ll'})$, $\sigma_{\epsilon l}^2 = \sigma_{yl}^2(1 - \rho_2^l - \rho_0^l + \rho_1^l)$ and $\sigma_{\epsilon ll'} = \sigma_{yl}\sigma_{yl'}(\rho_{2,0}^{ll'} - \rho_{2,1}^{ll'} - \rho_0^{ll'} + \rho_1^{ll'})$. We can further let $\boldsymbol{\Gamma}_0$, $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, and $\boldsymbol{\Gamma}_{2'}$ be defined as

$$\boldsymbol{\Gamma}_0 = \begin{pmatrix} \rho_0^1 & \rho_0^{12} & \dots & \rho_0^{1L} \\ \rho_0^{12} & \rho_0^2 & \dots & \rho_0^{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_0^{1L} & \rho_0^{2L} & \dots & \rho_0^L \end{pmatrix}, \ \boldsymbol{\Gamma}_1 = \begin{pmatrix} \rho_1^1 & \rho_1^{12} & \dots & \rho_1^{1L} \\ \rho_1^{12} & \rho_1^2 & \dots & \rho_1^{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_1^{1L} & \rho_1^{2L} & \dots & \rho_1^L \end{pmatrix}, \ \boldsymbol{\Gamma}_2 = \begin{pmatrix} 1 & \rho_{2,0}^{12} & \dots & \rho_{2,0}^{1L} \\ \rho_{2,0}^{12} & 1 & \dots & \rho_{2,0}^{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{2,0}^{1L} & \rho_{2,0}^{2L} & \dots & 1 \end{pmatrix}, \ \boldsymbol{\Gamma}_{2'} = \begin{pmatrix} \rho_2^1 & \rho_{2,1}^{12} & \dots & \rho_{2,1}^{1L} \\ \rho_{2,1}^{12} & \rho_2^2 & \dots & \rho_{2,1}^{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{2,1}^{1L} & \rho_{2,1}^{2L} & \dots & \rho_2^L \end{pmatrix}.$$

Defining the diagonal matrix of outcome variances as $\boldsymbol{\Lambda}_y = \text{diag}(\sigma_{y1}^2, \dots, \sigma_{yL}^2)$, we can further rewrite the covariance matrix of the intervention effect estimators in terms of the ICCs through the realization that $\boldsymbol{\Sigma}_b = \boldsymbol{\Lambda}_y^{1/2}\boldsymbol{\Gamma}_1\boldsymbol{\Lambda}_y^{1/2}$, $\boldsymbol{\Sigma}_s = \boldsymbol{\Lambda}_y^{1/2}\left(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1\right)\boldsymbol{\Lambda}_y^{1/2}$, $\boldsymbol{\Sigma}_{\gamma} = \boldsymbol{\Lambda}_y^{1/2}\left(\boldsymbol{\Gamma}_{2'} - \boldsymbol{\Gamma}_1\right)\boldsymbol{\Lambda}_y^{1/2}$, and $\boldsymbol{\Sigma}_{\epsilon} = \boldsymbol{\Lambda}_y^{1/2}\left(\boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_{2'} - \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1\right)\boldsymbol{\Lambda}_y^{1/2}$ giving us

$$\begin{aligned}\boldsymbol{\Omega}_{\delta} = IT\Bigg[&(ITU - TW + U^2 - IV)\left(\boldsymbol{\Lambda}_y^{1/2}\left(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1\right)\boldsymbol{\Lambda}_y^{1/2} + \frac{1}{N}\boldsymbol{\Lambda}_y^{1/2}\left(\boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_{2'} - \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1\right)\boldsymbol{\Lambda}_y^{1/2}\right)^{-1} \\ &- (U^2 - IV)\left(T\boldsymbol{\Lambda}_y^{1/2}\boldsymbol{\Gamma}_1\boldsymbol{\Lambda}_y^{1/2} + \frac{T}{N}\boldsymbol{\Lambda}_y^{1/2}\left(\boldsymbol{\Gamma}_{2'} - \boldsymbol{\Gamma}_1\right)\boldsymbol{\Lambda}_y^{1/2} + \boldsymbol{\Lambda}_y^{1/2}\left(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1\right)\boldsymbol{\Lambda}_y^{1/2} + \frac{1}{N}\boldsymbol{\Lambda}_y^{1/2}\left(\boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_{2'} - \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1\right)\boldsymbol{\Lambda}_y^{1/2}\right)^{-1}\Bigg]^{-1} \\ = \frac{IT}{N}\boldsymbol{\Lambda}_y^{1/2}\Bigg[&(ITU - TW + U^2 - IV)\left(N\left(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1\right) + \boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_{2'} - \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1\right)^{-1} \\ &- (U^2 - IV)\left(TN\boldsymbol{\Gamma}_1 + T\left(\boldsymbol{\Gamma}_{2'} - \boldsymbol{\Gamma}_1\right) + N\left(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1\right) + \boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_{2'} - \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1\right)^{-1}\Bigg]^{-1}\boldsymbol{\Lambda}_y^{1/2}\end{aligned}$$

$$= \frac{IT}{N}\boldsymbol{\Lambda}_y^{1/2}\Bigg[(ITU - TW + U^2 - IV)\left((N-1)(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1) + \boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_{2'}\right)^{-1}$$

$$- (U^2 - IV)\left((T-1)(N-1)\boldsymbol{\Gamma}_1 + (T-1)\boldsymbol{\Gamma}_{2'} + (N-1)\boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_2\right)^{-1}\Bigg]^{-1}\boldsymbol{\Lambda}_y^{1/2}$$

Web Appendix G

Closed-cohort design & common intervention effects: derivation of variance expression, var($\widehat{\delta'}$)

In Web Appendix C we derived the variance expression for a cross-sectional SW-CRT using

$$\widetilde{\boldsymbol{V}}_i^{-1} = \boldsymbol{I}_T \otimes \boldsymbol{A} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right],$$

where $\boldsymbol{A} = \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}$ and $\boldsymbol{B} = \left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}$ giving us

$$\mathrm{var}(\widehat{\delta'}) = IT\left[(ITU - TW + U^2 - IV)\boldsymbol{\sigma}_\epsilon^\top \boldsymbol{A}\boldsymbol{\sigma}_\epsilon + (IV - U^2)\boldsymbol{\sigma}_\epsilon^\top \boldsymbol{B}\boldsymbol{\sigma}_\epsilon\right]^{-1}.$$

We can use the same approach under a closed-cohort design. As shown in Web Appendix E, we have

$$\widetilde{\boldsymbol{V}}_i^{-1} = \boldsymbol{I}_T \otimes \boldsymbol{A} + \boldsymbol{J}_T \otimes \frac{1}{T}\left[\boldsymbol{B} - \boldsymbol{A}\right],$$

where $\boldsymbol{A} = \left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}$ and $\boldsymbol{B} = \left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{T}{N}\boldsymbol{\Sigma}_\gamma + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}$ under a closed-cohort design. Plugging these expressions for $\boldsymbol{A}$ and $\boldsymbol{B}$ into the expression above gives us the variance of the common intervention effect under a closed-cohort design

$$\mathrm{var}(\widehat{\delta'}) = IT\left[(ITU - TW + U^2 - IV)\boldsymbol{\sigma}_\epsilon^\top\left(\boldsymbol{\Sigma}_s + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\boldsymbol{\sigma}_\epsilon + (IV - U^2)\boldsymbol{\sigma}_\epsilon^\top\left(T\boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_s + \frac{T}{N}\boldsymbol{\Sigma}_\gamma + \frac{1}{N}\boldsymbol{\Sigma}_\epsilon\right)^{-1}\boldsymbol{\sigma}_\epsilon\right]^{-1}.$$

Or equivalently

$$\mathrm{var}(\widehat{\delta'}) = \frac{IT}{N}\Bigg[\left(ITU - TW + U^2 - IV\right)\boldsymbol{\omega}^\top\boldsymbol{\Lambda}_y^{-1/2}\left\{(N-1)(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1) + \boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_{2'}\right\}^{-1}\boldsymbol{\Lambda}_y^{-1/2}\boldsymbol{\omega}$$

$$- \left(U^2 - IV\right)\boldsymbol{\omega}^\top\boldsymbol{\Lambda}_y^{-1/2}\left\{(T-1)(N-1)\boldsymbol{\Gamma}_1 + (T-1)\boldsymbol{\Gamma}_{2'} + (N-1)\boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_2\right\}^{-1}\boldsymbol{\Lambda}_y^{-1/2}\boldsymbol{\omega}\Bigg]^{-1},$$

where $\boldsymbol{\omega} = \left(\sigma_{y1}\left(1-\rho_0^1+\rho_1^1-\rho_2^1\right)^{1/2},\ldots,\sigma_{yL}\left(1-\rho_0^L+\rho_1^L-\rho_2^L\right)^{1/2}\right)^\top$ and remaining parameters are the same as previously described.

Web Appendix H

Closed-cohort design & common ICCs: derivation of covariance expression, $\boldsymbol{\Omega}_\delta$

The common ICC assumption leads to simplification of the variance expression $\boldsymbol{\Omega}_\delta$ by defining the three key ICC matrices with their explicit simple exchangeable forms:

$$
\begin{aligned}
\boldsymbol{\Gamma}_0 &= (\rho_0-\rho_{00})\boldsymbol{I}_L+\rho_{00}\boldsymbol{J}_L\\
\boldsymbol{\Gamma}_1 &= (\rho_1-\rho_{11})\boldsymbol{I}_L+\rho_{11}\boldsymbol{J}_L\\
\boldsymbol{\Gamma}_2 &= (1-\rho_{2,0})\boldsymbol{I}_L+\rho_{2,0}\boldsymbol{J}_L\\
\boldsymbol{\Gamma}_2' &= (\rho_2-\rho_{2,1})\boldsymbol{I}_L+\rho_{2,1}\boldsymbol{J}_L.
\end{aligned}
$$

Plugging in these explicit forms into our current variance expression under the ICC parameterization (shown below for ease)

$$
\begin{aligned}
\boldsymbol{\Omega}_\delta = \frac{IT}{N}\boldsymbol{\Lambda}_y^{1/2}\Big[&(ITU-TW+U^2-IV)\left\{(N-1)(\boldsymbol{\Gamma}_0-\boldsymbol{\Gamma}_1)+\boldsymbol{\Gamma}_2-\boldsymbol{\Gamma}_{2'}\right\}^{-1}\\
&-(U^2-IV)\left\{(T-1)(N-1)\boldsymbol{\Gamma}_1+(T-1)\boldsymbol{\Gamma}_{2'}+(N-1)\boldsymbol{\Gamma}_0+\boldsymbol{\Gamma}_2\right\}^{-1}\Big]^{-1}\boldsymbol{\Lambda}_y^{1/2},
\end{aligned}
$$

gives us

$$
\begin{aligned}
\boldsymbol{\Omega}_\delta = \frac{IT}{N}\boldsymbol{\Lambda}_y^{1/2}\Big[&(ITU-TW+U^2-IV)\big\{\left[1-\rho_2-\rho_{2,0}+\rho_{2,1}+(N-1)(\rho_0-\rho_{00}-\rho_1+\rho_{11})\right]\boldsymbol{I}_L\\
&+\left[\rho_{2,0}-\rho_{2,1}+(N-1)(\rho_{00}-\rho_{11})\right]\boldsymbol{J}_L\big\}^{-1}\\
&-(U^2-IV)\big\{\left[1-\rho_{2,0}+(N-1)(\rho_0-\rho_{00})+(T-1)(\rho_2-\rho_{2,1})+(T-1)(N-1)(\rho_1-\rho_{11})\right]\boldsymbol{I}_L\\
&+\left[\rho_{2,0}+(N-1)\rho_{00}+(T-1)\rho_{2,1}+(T-1)(N-1)\rho_{11}\right]\boldsymbol{J}_L\big\}^{-1}\Big]^{-1}\boldsymbol{\Lambda}_y^{1/2}\\
= \frac{IT}{N}\boldsymbol{\Lambda}_y^{1/2}\Big[&(ITU-TW+U^2-IV)\left\{(\lambda_3-\tau_3)\boldsymbol{I}_L+\tau_3\boldsymbol{J}_L\right\}^{-1}-(U^2-IV)\left\{(\lambda_4-\tau_4)\boldsymbol{I}_L+\tau_4\boldsymbol{J}_L\right\}^{-1}\Big]^{-1}\boldsymbol{\Lambda}_y^{1/2},
\end{aligned}
$$

where $\lambda_3 = 1+(N-1)(\rho_0-\rho_1)-\rho_2$ and $\lambda_4 = 1+(N-1)\rho_0+(T-1)(N-1)\rho_1+(T-1)\rho_2$ are two distinct eigenvalues of the (endpoint-specific) block exchangeable correlation structure (Hooper et al., 2016; Girling and Hemming, 2016; Li et al., 2021). Further, $\tau_3 = \rho_{2,0}-\rho_{2,1}+(N-1)(\rho_{00}-\rho_{11})$ and $\tau_4 = \tau_3+T\left(\rho_{2,1}+(N-1)\rho_{11}\right)$

characterize the impact of the between-endpoint ICCs on the variance of intervention effect estimators through the MLMM. In the special case where all endpoints are completely independent such that $\rho_{00} = \rho_{11} = \rho_{2,0} = \rho_{2,1} = 0$, $\boldsymbol{\Omega}_\delta$ becomes a diagonal matrix and each element becomes identical to the variance expression developed in Hooper, Girling, and Li (Hooper et al., 2016; Girling and Hemming, 2016; Li et al., 2021) for closed-cohort SW-CRTs with a univariate outcome. Our variance expression above is similar to the one derived in Web Appendix D for cross-sectional designs. Using the same approach, letting $x = ITU - TW + U^2 - IV$, $y = U^2 - IV$, $a_3 = \lambda_3 - \tau_3$, $a_4 = \lambda_4 - \tau_4$, $b_3 = \lambda_3 + (L-1)\tau_3$, and $b_4 = \lambda_4 + (L-1)\tau_4$. This gives us

$$\boldsymbol{\Omega}_\delta = \frac{(IT/N)a_3a_4}{xa_4 - ya_3}\boldsymbol{\Lambda}_y^{1/2}\left[\boldsymbol{I}_L - \frac{y\tau_4a_3b_3 - x\tau_3a_4b_4}{(xa_4 - ya_3)\,b_3b_4 + L\,(y\tau_4a_3b_3 - x\tau_3a_4b_4)}\boldsymbol{J}_L\right]\boldsymbol{\Lambda}_y^{1/2}$$

Closed-cohort design & common ICCs across endpoints: proof of Theorem 3

An extension of Theorem 1 to closed-cohort designs, denoted Theorem 3, is provided below:

Theorem **3.** *Under the parsimonious parameterization with common ICC values across endpoints and a closed-cohort design, the l-th diagonal element of $\boldsymbol{\Omega}_\delta$ can be further written in the following analytical form*

$$var(\hat{\delta}_l) = \frac{(IT/N)\sigma_{yl}^2}{\left(ITU - TW + U^2 - IV\right)(\lambda_4 - \tau_4) - \left(U^2 - IV\right)(\lambda_3 - \tau_3)} \times \\ \frac{\left(ITU - TW + U^2 - IV\right)\lambda_3(\lambda_4 - \tau_4)\left\{\lambda_4 + (L-1)\tau_4\right\} - \left(U^2 - IV\right)\lambda_4(\lambda_3 - \tau_3)\left\{\lambda_3 + (L-1)\tau_3\right\}}{\left(ITU - TW + U^2 - IV\right)\left\{\lambda_4 + (L-1)\tau_4\right\} - \left(U^2 - IV\right)\left\{\lambda_3 + (L-1)\tau_3\right\}}.$$

Furthermore, denote the variance of the l-th intervention effect estimator based on a univariate Hooper and Girling model (Hooper et al., 2016; Girling and Hemming, 2016; Li et al., 2021) is

$$var^{HG}(\hat{\delta}_l) = \frac{(IT/N)\sigma_{yl}^2\lambda_3\lambda_4}{\left(ITU - TW + U^2 - IV\right)\lambda_4 - \left(U^2 - IV\right)\lambda_3}$$

and $var(\hat{\delta}_l) \le var^{HG}(\hat{\delta}_l)$ for any set of valid design parameters, with equality holds when $\tau_3\lambda_4 = \tau_4\lambda_3$ or $\rho_{00} = \rho_{11} = \rho_{2,0} = \rho_{2,1} = 0$ (a special case when $\tau_3\lambda_4 = \tau_4\lambda_3$).

Proof:

Given our expression for $\boldsymbol{\Omega}_\delta$ under the common ICCs assumption (shown above) and using the same approach as Theorem 1, the l-th diagonal element is

$$\begin{aligned}\mathrm{var}(\widehat{\delta}_l) &= \left(\frac{IT\sigma_{yl}^2}{N}\right)\left(\frac{a_3a_4}{xa_4 - ya_3}\right)\left(1 - \frac{y\tau_4a_3b_3 - x\tau_3a_4b_4}{(xa_4 - ya_3)\,b_3b_4 + L\,(y\tau_4a_3b_3 - x\tau_3a_4b_4)}\right) \\ &= \frac{(IT/N)\sigma_{yl}^2}{(ITU - TW + U^2 - IV)(\lambda_4 - \tau_4) - (U^2 - IV)(\lambda_3 - \tau_3)} \times\end{aligned}$$

$$\frac{(ITU - TW + U^2 - IV)\lambda_3(\lambda_4 - \tau_4)\{\lambda_4 + (L-1)\tau_4\} - (U^2 - IV)\lambda_4(\lambda_3 - \tau_3)\{\lambda_3 + (L-1)\tau_3\}}{(ITU - TW + U^2 - IV)\{\lambda_4 + (L-1)\tau_4\} - (U^2 - IV)\{\lambda_3 + (L-1)\tau_3\}},$$

which matches our variance expression in Theorem 3. Under the Hooper and Girling LMM, the variance of the l-th outcome is

$$\text{var}^{\text{HG}}(\widehat{\delta}_l) = \frac{(IT/N)\sigma_{yl}^2\lambda_3\lambda_4}{\left(ITU - TW + U^2 - IV\right)\lambda_4 - \left(U^2 - IV\right)\lambda_3}.$$

Therefore, the variance ratio comparing the MLMM to the LMM is

$$\begin{aligned}
\frac{\text{var}(\widehat{\delta}_l)}{\text{var}^{\text{HG}}(\widehat{\delta}_l)} &= \left(\frac{\left(ITU - TW + U^2 - IV\right)\lambda_4 - \left(U^2 - IV\right)\lambda_3}{(IT/N)\sigma_{yl}^2\lambda_3\lambda_4}\right)\left(\frac{(IT/N)\sigma_{yl}^2}{(ITU - TW + U^2 - IV)(\lambda_4 - \tau_4) - (U^2 - IV)(\lambda_3 - \tau_3)}\right)\times \\
&\quad \frac{(ITU - TW + U^2 - IV)\lambda_3(\lambda_4 - \tau_4)\{\lambda_4 + (L-1)\tau_4\} - (U^2 - IV)\lambda_4(\lambda_3 - \tau_3)\{\lambda_3 + (L-1)\tau_3\}}{(ITU - TW + U^2 - IV)\{\lambda_4 + (L-1)\tau_4\} - (U^2 - IV)\{\lambda_3 + (L-1)\tau_3\}} \\
&= \frac{\left(ITU - TW + U^2 - IV\right)\lambda_4 - \left(U^2 - IV\right)\lambda_3}{(ITU - TW + U^2 - IV)\{\lambda_4 + (L-1)\tau_4\} - (U^2 - IV)\{\lambda_3 + (L-1)\tau_3\}}\times \\
&\quad \frac{(ITU - TW + U^2 - IV)\lambda_4^{-1}(\lambda_4 - \tau_4)\{\lambda_4 + (L-1)\tau_4\} - (U^2 - IV)\lambda_3^{-1}(\lambda_3 - \tau_3)\{\lambda_3 + (L-1)\tau_3\}}{(ITU - TW + U^2 - IV)(\lambda_4 - \tau_4) - (U^2 - IV)(\lambda_3 - \tau_3)}.
\end{aligned}$$

We can again use $x = ITU - TW + U^2 - IV$, $y = U^2 - IV$, $a_3 = \lambda_3 - \tau_3$, $a_4 = \lambda_4 - \tau_4$, $b_3 = \lambda_3 + (L-1)\tau_3$, and $b_4 = \lambda_4 + (L-1)\tau_4$. This gives us

$$\begin{aligned}
&= \frac{x\lambda_4 - y\lambda_3}{xb_4 - yb_3}\left(\frac{x\lambda_4^{-1}a_4b_4 - y\lambda_3^{-1}a_3b_3}{xa_4 - ya_3}\right) \\
&= \frac{x^2a_4b_4 - xy\lambda_4\lambda_3^{-1}a_3b_3 - xy\lambda_4^{-1}\lambda_3a_4b_4 + y^2a_3b_3}{x^2a_4b_4 - xya_3b_4 - xya_4b_3 + y^2a_3b_3}.
\end{aligned}$$

To evaluate whether this ratio is less than or greater than one we can take the difference of the numerator and denominator.

$$\begin{aligned}
\text{num. - den.} &= x^2a_4b_4 - xy\lambda_4\lambda_3^{-1}a_3b_3 - xy\lambda_4^{-1}\lambda_3a_4b_4 + y^2a_3b_3 - (x^2a_4b_4 - xya_3b_4 - xya_4b_3 + y^2a_3b_3) \\
&= -xy\left(2\tau_3\tau_4\lambda_3\lambda_4 - \lambda_4^2\tau_3^2 - \lambda_3^2\tau_4^2)((L-1)/\lambda_3\lambda_4\right) \\
&= xy(\lambda_4\tau_3 - \lambda_3\tau_4)^2((L-1)/\lambda_3\lambda_4).
\end{aligned}$$

We already showed in Web Appendix D that $x > 0$ and $y < 0$ in a SW-CRT design. Therefore, the difference between the numerator and denominator is less than or equal to zero which leads to a variance ratio of less than or equal to one. Thus, $\text{var}(\hat{\delta}_l) \leq \text{var}^{\text{HG}}(\hat{\delta}_l)$ for any set of valid design parameters, with equality holds when $\tau_3\lambda_4 = \tau_4\lambda_3$ or $\rho_{00} = \rho_{11} = \rho_{2,0} = \rho_{2,1} = 0$ (a special case when $\tau_3\lambda_4 = \tau_4\lambda_3$).

Web Appendix I

Closed-cohort design & common ICCs and intervention effects: derivation of variance expression, var($\widehat{\delta}'$)

Recall that if we assume common ICCs we have

$$
\begin{aligned}
\boldsymbol{\Gamma}_0 &= (\rho_0 - \rho_{00})\boldsymbol{I}_L + \rho_{00}\boldsymbol{J}_L \\
\boldsymbol{\Gamma}_1 &= (\rho_1 - \rho_{11})\boldsymbol{I}_L + \rho_{11}\boldsymbol{J}_L \\
\boldsymbol{\Gamma}_2 &= (1 - \rho_{2,0})\boldsymbol{I}_L + \rho_{2,0}\boldsymbol{J}_L \\
\boldsymbol{\Gamma}_2' &= (\rho_2 - \rho_{2,1})\boldsymbol{I}_L + \rho_{2,1}\boldsymbol{J}_L.
\end{aligned}
$$

And if we assume common intervention effects, then the variance of the intervention effect estimator becomes

$$
\begin{aligned}
\mathrm{var}(\widehat{\delta}') &= \frac{IT}{N}\Big[\left(ITU - TW + U^2 - IV\right)\boldsymbol{\omega}^\top \boldsymbol{\Lambda}_y^{-1/2}\left\{(N-1)(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1) + \boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_{2'}\right\}^{-1}\boldsymbol{\Lambda}_y^{-1/2}\boldsymbol{\omega} \\
&\quad - \left(U^2 - IV\right)\boldsymbol{\omega}^\top \boldsymbol{\Lambda}_y^{-1/2}\left\{(T-1)(N-1)\boldsymbol{\Gamma}_1 + (T-1)\boldsymbol{\Gamma}_{2'} + (N-1)\boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_2\right\}^{-1}\boldsymbol{\Lambda}_y^{-1/2}\boldsymbol{\omega}\Big]^{-1} \\
&= \frac{IT}{N}\Big[\left(ITU - TW + U^2 - IV\right)\boldsymbol{\omega}^\top \boldsymbol{\Lambda}_y^{-1/2}\left\{(\lambda_3 - \tau_3)\boldsymbol{I}_L + \tau_3\boldsymbol{J}_L\right\}^{-1}\boldsymbol{\Lambda}_y^{-1/2}\boldsymbol{\omega} \\
&\quad - \left(U^2 - IV\right)\boldsymbol{\omega}^\top \boldsymbol{\Lambda}_y^{-1/2}\left\{(\lambda_4 - \tau_4)\boldsymbol{I}_L + \tau_4\boldsymbol{J}_L\right\}^{-1}\boldsymbol{\Lambda}_y^{-1/2}\boldsymbol{\omega}\Big]^{-1} \\
&= \frac{IT}{LN\lambda_1}\Big[\left(ITU - TW + U^2 - IV\right)\left(\lambda_3 + (L-1)\tau_3\right)^{-1} - \left(U^2 - IV\right)\left(\lambda_4 + (L-1)\tau_4\right)^{-1}\Big]^{-1} \\
&= \frac{(IT/(LN\lambda_1))(\lambda_3 + (L-1)\tau_3)(\lambda_4 + (L-1)\tau_4)}{(ITU - TW + U^2 - IV)(\lambda_4 + (L-1)\tau_4) - (U^2 - IV)(\lambda_3 + (L-1)\tau_3)},
\end{aligned}
$$

where again we have $\lambda_3 = 1 + (N-1)(\rho_0 - \rho_1) - \rho_2$, $\lambda_4 = 1 + (N-1)\rho_0 + (T-1)(N-1)\rho_1 + (T-1)\rho_2$, $\tau_3 = \rho_{2,0} - \rho_{2,1} + (N-1)(\rho_{00} - \rho_{11})$, and $\tau_4 = \tau_3 + T\left(\rho_{2,1} + (N-1)\rho_{11}\right)$. Under the common ICCs assumption, $\boldsymbol{\omega} = \left(\sigma_{y1}\lambda_1^{1/2}, \ldots, \sigma_{yL}\lambda_1^{1/2}\right)^\top$ where same as λ_3 and λ_4, $\lambda_1 = 1 - \rho_0 + \rho_1 - \rho_2$ is a distinct eigenvalue of the (endpoint-specific) block exchangeable correlation structure (Li et al., 2018) defined for closed-cohort SW-CRTs with a univariate outcome.

Closed-cohort design & common ICCs and intervention effects: proof of Theorem 4

An extension of Theorem 2 to closed-cohort designs, denoted Theorem 4, is provided below:

Theorem **4.** *Under the parsimonious parameterization with common ICC values and a common intervention effect* *across endpoints under a closed-cohort design, the variance of the l-th intervention effect estimator (unscaled) under model (11), i.e.* $\delta_l = \sigma_{yl}\lambda_1^{1/2}\delta'$, *is*

$$var^{both}(\hat{\delta}_l) = \frac{(IT/(LN))\sigma_{yl}^2(\lambda_3 + (L-1)\tau_3)(\lambda_4 + (L-1)\tau_4)}{(ITU - TW + U^2 - IV)(\lambda_4 + (L-1)\tau_4) - (U^2 - IV)(\lambda_3 + (L-1)\tau_3)}.$$

As shown in Theorem 3, under the parsimonious parameterization with common ICC values across endpoints, the l-th diagonal element of $\mathbf{\Omega}_\delta$ *is denoted by*

$$var^{ICC}(\hat{\delta}_l) = \frac{(IT/N)\sigma_{yl}^2}{\left(ITU - TW + U^2 - IV\right)(\lambda_4 - \tau_4) - \left(U^2 - IV\right)(\lambda_3 - \tau_3)} \times \frac{\left(ITU - TW + U^2 - IV\right)\lambda_3(\lambda_4 - \tau_4)\{\lambda_4 + (L-1)\tau_4\} - \left(U^2 - IV\right)\lambda_4(\lambda_3 - \tau_3)\{\lambda_3 + (L-1)\tau_3\}}{\left(ITU - TW + U^2 - IV\right)\{\lambda_4 + (L-1)\tau_4\} - \left(U^2 - IV\right)\{\lambda_3 + (L-1)\tau_3\}},$$

and $var^{both}(\hat{\delta}_l) < var^{ICC}(\hat{\delta}_l)$ *for any set of valid design parameters.*

Proof:

$$\begin{aligned}
\frac{\mathrm{var}^{\mathrm{both}}(\hat{\delta}_l)}{\mathrm{var}^{\mathrm{ICC}}(\hat{\delta}_l)} &= \frac{(IT/(LN))\sigma_{yl}^2(\lambda_3 + (L-1)\tau_3)(\lambda_4 + (L-1)\tau_4)}{(ITU - TW + U^2 - IV)(\lambda_4 + (L-1)\tau_4) - (U^2 - IV)(\lambda_3 + (L-1)\tau_3)} \\
&\times \frac{\left(ITU - TW + U^2 - IV\right)(\lambda_4 - \tau_4) - \left(U^2 - IV\right)(\lambda_3 - \tau_3)}{(IT/N)\sigma_{yl}^2} \\
&\times \frac{\left(ITU - TW + U^2 - IV\right)\{\lambda_4 + (L-1)\tau_4\} - \left(U^2 - IV\right)\{\lambda_3 + (L-1)\tau_3\}}{(ITU - TW + U^2 - IV)\lambda_3(\lambda_4 - \tau_4)\{\lambda_4 + (L-1)\tau_4\} - (U^2 - IV)\lambda_4(\lambda_3 - \tau_3)\{\lambda_3 + (L-1)\tau_3\}} \\
&= \frac{(1/L)(\lambda_3 + (L-1)\tau_3)(\lambda_4 + (L-1)\tau_4)\left(\left(ITU - TW + U^2 - IV\right)(\lambda_4 - \tau_4) - \left(U^2 - IV\right)(\lambda_3 - \tau_3)\right)}{(ITU - TW + U^2 - IV)\lambda_3(\lambda_4 - \tau_4)\{\lambda_4 + (L-1)\tau_4\} - (U^2 - IV)\lambda_4(\lambda_3 - \tau_3)\{\lambda_3 + (L-1)\tau_3\}}
\end{aligned}$$

Again using $x = ITU - TW + U^2 - IV$, $y = U^2 - IV$, $a_3 = \lambda_3 - \tau_3$, $a_4 = \lambda_4 - \tau_4$, $b_3 = \lambda_3 + (L-1)\tau_3$, and $b_4 = \lambda_4 + (L-1)\tau_4$. This gives us

$$\begin{aligned}
&= \frac{(1/L)b_3b_4(xa_4 - ya_3)}{x\lambda_3a_4b_4 - y\lambda_4a_3b_3} \\
&= (1/L)\frac{(xa_4b_4)b_3 - (ya_3b_3)b_4}{(xa_4b_4)\lambda_3 - (ya_3b_3)\lambda_4} \\
&= (1/L)\frac{(xa_4b_4)\{\lambda_3 + (L-1)\tau_3\} - (ya_3b_3)\{\lambda_4 + (L-1)\tau_4\}}{(xa_4b_4)\lambda_3 - (ya_3b_3)\lambda_4} \\
&= (1/L)\frac{(xa_4b_4)\lambda_3 - (ya_3b_3)\lambda_4 + (L-1)(xa_4b_4)\tau_3 - (L-1)(ya_3b_3)\tau_4}{(xa_4b_4)\lambda_3 - (ya_3b_3)\lambda_4} \\
&= (1/L)\left(1 + \frac{(L-1)\{(xa_4b_4)\tau_3 - (ya_3b_3)\tau_4\}}{(xa_4b_4)\lambda_3 - (ya_3b_3)\lambda_4}\right).
\end{aligned}$$

We are interested in determining when this ratio is less than 1 which is when

$$(1/L)\left(1 + \frac{(L-1)\{(xa_4b_4)\tau_3 - (ya_3b_3)\tau_4\}}{(xa_4b_4)\lambda_3 - (ya_3b_3)\lambda_4}\right) < 1$$

$$\Rightarrow \frac{(xa_4b_4)\tau_3 - (ya_3b_3)\tau_4}{(xa_4b_4)\lambda_3 - (ya_3b_3)\lambda_4} < 1.$$

We know that $\tau_3 < \lambda_3$ and $\tau_4 < \lambda_4$ by definition and we already showed that $x > 0$ and $y < 0$ under a SW-CRT, therefore the numerator is always less than the denominator which means the variance ratio is always less than 1.

References

Davis-Plourde, K., Taljaard, M., and Li, F. (2021). Sample size considerations for stepped wedge designs with subclusters. *Biometrics* .

Girling, A. J. and Hemming, K. (2016). Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. *Statistics in Medicine* **35,** 2149–2166.

Hooper, R., Teerenstra, S., de Hoop, E., and Eldridge, S. (2016). Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. *Statistics in Medicine* **35,** 4718–4728.

Kistner, E. O. and Muller, K. E. (2004). Exact distributions of intraclass correlation and cronbach's alpha with gaussian data and general covariance. *Psychometrika* **69,** 459–474.

Leiva, R. (2007). Linear discrimination with equicorrelated training vectors. *Journal of Multivariate Analysis* **98,** 384–409.

Li, F., Hughes, J. P., Hemming, K., Taljaard, M., Melnick, E. R., and Heagerty, P. J. (2021). Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Statistical Methods in Medical Research* **30,** 612–639.

Li, F., Turner, E. L., and Preisser, J. S. (2018). Sample size determination for GEE analyses of stepped wedge cluster randomized trials. *Biometrics* **74,** 1450–1458.

Li, F., Yu, H., Rathouz, P. J., Turner, E. L., and Preisser, J. S. (2021). Marginal modeling of cluster-period means and intraclass correlations in stepped wedge designs with binary outcomes. *Biostatistics* page doi:10.1093/biostatistics/kxaa056.