Supporting Information for “Sample size considerations for stepped wedge designs with subclusters” by Davis-Plourde, Taljaard, and Li

Contents

Web Appendix A

LEMMA 1: Eigenvalues for $\boldsymbol{R}_i$

As shown in Web Appendix A of Li et al. (2018b): Theorem 8.3.4 and 8.4.4 (Graybill, 1983) states that any $u \times u$ exchangeable matrix $\boldsymbol{A} = x\boldsymbol{I}_u + y\boldsymbol{J}_u$ is invertible if and only if $x \neq 0$ and $x + uy \neq 0$. The inverse is,

$$\boldsymbol{A}^{-1} = \frac{1}{x}\boldsymbol{I}_u - \frac{y}{x(x+uy)}\boldsymbol{J}_u, \tag{1}$$

and the determinant is,

$$\det(\boldsymbol{A}) = x^{u-1}(x+uy). \tag{2}$$

A Corollary is that $\boldsymbol{A}$ has two eigenvalues, x with multiplicity $u-1$ and $x+uy$ with multiplicity 1, provided $uy \neq 0$.

Under a closed-cohort design at both the subcluster and subject levels (design variant A), we let $\boldsymbol{B} = \boldsymbol{I}_K \otimes \{(1 - \alpha_0 - \lambda)\boldsymbol{I}_N + (\alpha_0 - \rho_0)\boldsymbol{J}_N\} + \boldsymbol{J}_K \otimes \rho_0\boldsymbol{J}_N$ and $\boldsymbol{C} = \boldsymbol{I}_K \otimes \{(\alpha_2 - \alpha_1)\boldsymbol{I}_N + (\alpha_1 - \rho_1)\boldsymbol{J}_N\} + \boldsymbol{J}_K \otimes \rho_1\boldsymbol{J}_N$. The eigenvalues of $\boldsymbol{R}_i$ are given by the roots of the characteristic equation,

$$\begin{aligned}
0 &= \det(\boldsymbol{R}_i - \lambda\boldsymbol{I}_{TKN}) \\
&= \det(\boldsymbol{I}_T \otimes (\boldsymbol{B} - \boldsymbol{C}) + \boldsymbol{J}_T \otimes \boldsymbol{C}) \\
&= \det(\boldsymbol{B} - \boldsymbol{C})^{T-1}\det(\boldsymbol{B} + (T-1)\boldsymbol{C})
\end{aligned}$$

where the last quantity is given by Theorem 8.9.1 (Graybill, 1983). Since $\boldsymbol{B}-\boldsymbol{C}$ and $\boldsymbol{B}+(T-1)\boldsymbol{C}$ are exchangeable we can apply equation (2) to get the following six eigenvalues,

$$\begin{aligned}
\lambda_1 &= 1 - \alpha_0 - \alpha_2 + \alpha_1 \\
\lambda_2 &= 1 - \alpha_0 - \alpha_2 + \alpha_1 + N(\alpha_0 - \alpha_1 - \rho_0 + \rho_1) \\
\lambda_3 &= 1 - \alpha_0 - \alpha_2 + \alpha_1 + N\{\alpha_0 - \alpha_1 + (K-1)(\rho_0 - \rho_1)\} \\
\lambda_4 &= 1 - \alpha_0 + (T-1)(\alpha_2 - \alpha_1) \\
\lambda_5 &= 1 - \alpha_0 + (T-1)(\alpha_2 - \alpha_1) + N\{\alpha_0 - \rho_0 + (T-1)(\alpha_1 - \rho_1)\} \\
\lambda_6 &= 1 - \alpha_0 + (T-1)(\alpha_2 - \alpha_1) + N[\alpha_0 + (T-1)\alpha_1 + (K-1)\{\rho_0 + (T-1)\rho_1\}],
\end{aligned} \tag{3}$$

with algebraic multiplicities $(T-1)K(N-1)$, $(T-1)(K-1)$, $T-1$, $K(N-1)$, $K-1$, and 1 respectively. Eigenvalue expressions and their respective multiplicities under each design variant (A, B, and C) can be found in Web Table 1.

LEMMA 2: Derivation of $\boldsymbol{R}_i^{-1}$

By section 2.1 of Leiva (2007) we know 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

$$\boldsymbol{A}^{-1} = \boldsymbol{I}_u \otimes (\boldsymbol{B} - \boldsymbol{C})^{-1} + \boldsymbol{J}_u \otimes \frac{1}{u}\Big[\{\boldsymbol{B} + (u-1)\boldsymbol{C}\}^{-1} - (\boldsymbol{B} - \boldsymbol{C})^{-1}\Big]. \tag{4}$$

Closed-cohort design at both the subcluster and subject levels (design A)

Recall that we have the following correlation matrix, $\boldsymbol{R}_i = \boldsymbol{I}_T \otimes (\boldsymbol{B} - \boldsymbol{C}) + \boldsymbol{J}_T \otimes \boldsymbol{C}$ with $\boldsymbol{B} = (1-\alpha_0)\boldsymbol{I}_{KN} + (\alpha_0 - \rho_0)\boldsymbol{I}_K \otimes \boldsymbol{J}_N + \rho_0 \boldsymbol{J}_{KN}$ and $\boldsymbol{C} = (\alpha_2 - \alpha_1)\boldsymbol{I}_{KN} + (\alpha_1 - \rho_1)\boldsymbol{I}_K \otimes \boldsymbol{J}_N + \rho_1 \boldsymbol{J}_{KN}$. By equation (4) we know that,

$$\boldsymbol{R}_i^{-1} = \boldsymbol{I}_T \otimes (\boldsymbol{B} - \boldsymbol{C})^{-1} + \boldsymbol{J}_T \otimes \frac{1}{T}[\{\boldsymbol{B} + (T-1)\boldsymbol{C}\}^{-1} - (\boldsymbol{B} - \boldsymbol{C})^{-1}],$$

if $\boldsymbol{B} - \boldsymbol{C}$ and $\boldsymbol{B} + (u-1)\boldsymbol{C}$ are non-singular matrices. Looking at each required term, $\boldsymbol{B} - \boldsymbol{C}$ and $\boldsymbol{B} + (T-1)\boldsymbol{C}$, in terms of the eigenvalues we have:

$$\boldsymbol{B} - \boldsymbol{C} = \boldsymbol{I}_K \otimes \left(\lambda_1 \boldsymbol{I}_N + \frac{\lambda_2 - \lambda_1}{N}\boldsymbol{J}_N\right) + \boldsymbol{J}_K \otimes \frac{\lambda_3 - \lambda_2}{KN}\boldsymbol{J}_N$$

$$\boldsymbol{B} + (T-1)\boldsymbol{C} = \boldsymbol{I}_K \otimes \left(\lambda_4 \boldsymbol{I}_N + \frac{\lambda_5 - \lambda_4}{N}\boldsymbol{J}_N\right) + \boldsymbol{J}_K \otimes \frac{\lambda_6 - \lambda_5}{KN}\boldsymbol{J}_N.$$

Both terms are non-singular matrices and exchangeable. Therefore, each term can be inverted using equations (4) and (1) giving us a closed-form expression of $\boldsymbol{R}_i^{-1}$.

$$\begin{aligned}(\boldsymbol{B} - \boldsymbol{C})^{-1} &= \boldsymbol{I}_K \otimes \left(\frac{1}{\lambda_1}\boldsymbol{I}_N - \frac{\lambda_2 - \lambda_1}{N\lambda_1\lambda_2}\boldsymbol{J}_N\right) + \boldsymbol{J}_K \otimes \frac{\lambda_2 - \lambda_3}{KN\lambda_2\lambda_3}\boldsymbol{J}_N \\ \{\boldsymbol{B} + (T-1)\boldsymbol{C}\}^{-1} &= \boldsymbol{I}_K \otimes \left(\frac{1}{\lambda_4}\boldsymbol{I}_N - \frac{\lambda_5 - \lambda_4}{N\lambda_4\lambda_5}\boldsymbol{J}_N\right) + \boldsymbol{J}_K \otimes \frac{\lambda_5 - \lambda_6}{KN\lambda_5\lambda_6}\boldsymbol{J}_N\end{aligned}$$

$$\begin{aligned}\boldsymbol{R}_i^{-1} = \boldsymbol{I}_T &\otimes \left\{\boldsymbol{I}_K \otimes \left(\frac{1}{\lambda_1}\boldsymbol{I}_N - \frac{\lambda_2 - \lambda_1}{N\lambda_1\lambda_2}\boldsymbol{J}_N\right) + \boldsymbol{J}_K \otimes \frac{\lambda_2 - \lambda_3}{KN\lambda_2\lambda_3}\boldsymbol{J}_N\right\} \\ &+ \boldsymbol{J}_T \otimes \frac{1}{T}\left[\boldsymbol{I}_K \otimes \left\{\left(\frac{1}{\lambda_4} - \frac{1}{\lambda_1}\right)\boldsymbol{I}_N + \left(\frac{\lambda_2 - \lambda_1}{N\lambda_1\lambda_2} - \frac{\lambda_5 - \lambda_4}{N\lambda_4\lambda_5}\right)\boldsymbol{J}_N\right\} + \boldsymbol{J}_K \otimes \frac{1}{K}\left(\frac{\lambda_5 - \lambda_6}{N\lambda_5\lambda_6} - \frac{\lambda_2 - \lambda_3}{N\lambda_2\lambda_3}\right)\boldsymbol{J}_N\right].\end{aligned}$$

Note that this expression for $\boldsymbol{R}_i^{-1}$ is the same across all design variants using the eigenvalues shown in Web Table 1.

THEOREM 1: Derivation of var($\widehat{\delta}$) & relationship with ICCs

THEOREM 1: Assuming known variance components, the closed-form variance of the intervention effect estimator based on a linear mixed model with a Gaussian outcome is

$$\mathrm{var}(\widehat{\delta}) = \frac{\sigma^2}{IKNtr(\boldsymbol{\Omega})} \times \frac{T\lambda_6\lambda_3}{T\lambda_6 - \{1 + (T-1)\tau_X\}(\lambda_6 - \lambda_3)}, \quad (5)$$

where $\boldsymbol{\Omega} = I^{-1}\sum_{i=1}^{I}\boldsymbol{X}_i\boldsymbol{X}_i^\top - (I^{-1}\sum_{i=1}^{I}\boldsymbol{X}_i)(I^{-1}\sum_{i=1}^{I}\boldsymbol{X}_i^\top)$ is the covariance matrix of the intervention vector under a specific design and $\tau_X = \{(T-1)tr(\boldsymbol{\Omega})\}^{-1}\{\boldsymbol{1}^\top\boldsymbol{\Omega}\boldsymbol{1} - tr(\boldsymbol{\Omega})\} \in [-1, 1]$ 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). With all other design parameters fixed, larger values of the within-period ICCs, $\{\alpha_0, \rho_0\}$, are always associated with larger required sample size, whereas larger values of the between-period ICCs, $\{\alpha_1, \rho_1, \alpha_2\}$, are associated with smaller required sample size when $\tau_X < (\lambda_6^2 - \lambda_3^2)/\{\lambda_6^2 + (T-1)\lambda_3^2\}$.

Proof: Let $\boldsymbol{1}_u$ be a $u \times 1$ vector of ones. We define $\boldsymbol{X}_i$ as the treatment randomization schedule for cluster i and $\boldsymbol{Z}_i = (\boldsymbol{I}_T, \boldsymbol{X}_i) \otimes \boldsymbol{1}_K \otimes \boldsymbol{1}_N$. We know that var($\widehat{\delta}$) is the lower-right corner element of $\sigma^2(\sum_{i=1}^{I}\boldsymbol{Z}_i^\top\boldsymbol{R}_i^{-1}\boldsymbol{Z}_i)^{-1}$. In addition,

$$\sum_{i=1}^{I}\boldsymbol{Z}_i^\top\boldsymbol{R}_i^{-1}\boldsymbol{Z}_i = \sum_{i=1}^{I}\left[\begin{pmatrix}\boldsymbol{I}_T \\ \boldsymbol{X}_i^\top\end{pmatrix} \otimes \boldsymbol{1}_K^\top \otimes \boldsymbol{1}_N^\top\right]\boldsymbol{R}_i^{-1}\left[\begin{pmatrix}\boldsymbol{I}_T & \boldsymbol{X}_i\end{pmatrix} \otimes \boldsymbol{1}_K \otimes \boldsymbol{1}_N\right] = \begin{bmatrix}\Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22}\end{bmatrix}$$

where Ω_{11} is of dimension $T \times T$, $\Omega_{12} = \Omega_{21}^\top$ is of dimension $T \times 1$, and Ω_{22} is a scalar. Block matrix inversion gives us var($\widehat{\delta}$) $= (\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1}$. Recall we have the following inverse correlation matrix,

$$\begin{aligned}\boldsymbol{R}_i^{-1} &= \boldsymbol{I}_T \otimes \left\{\boldsymbol{I}_K \otimes \left(\frac{1}{\lambda_1}\boldsymbol{I}_N - \frac{\lambda_2 - \lambda_1}{N\lambda_1\lambda_2}\boldsymbol{J}_N\right) + \boldsymbol{J}_K \otimes \frac{\lambda_2 - \lambda_3}{KN\lambda_2\lambda_3}\boldsymbol{J}_N\right\} \\ &+ \boldsymbol{J}_T \otimes \frac{1}{T}\left[\boldsymbol{I}_K \otimes \left\{\left(\frac{1}{\lambda_4} - \frac{1}{\lambda_1}\right)\boldsymbol{I}_N + \left(\frac{\lambda_2 - \lambda_1}{N\lambda_1\lambda_2} - \frac{\lambda_5 - \lambda_4}{N\lambda_4\lambda_5}\right)\boldsymbol{J}_N\right\} + \boldsymbol{J}_K \otimes \frac{1}{K}\left(\frac{\lambda_5 - \lambda_6}{N\lambda_5\lambda_6} - \frac{\lambda_2 - \lambda_3}{N\lambda_2\lambda_3}\right)\boldsymbol{J}_N\right].\end{aligned}$$

Which is of the form,

$$\boldsymbol{R}_i^{-1} = \tilde{a}\boldsymbol{I}_T \otimes \boldsymbol{I}_K \otimes \boldsymbol{I}_N + \tilde{b}\boldsymbol{I}_T \otimes \boldsymbol{I}_K \otimes \boldsymbol{J}_N + \tilde{c}\boldsymbol{I}_T \otimes \boldsymbol{J}_K \otimes \boldsymbol{J}_N + \tilde{d}\boldsymbol{J}_T \otimes \boldsymbol{I}_K \otimes \boldsymbol{I}_N + \tilde{e}\boldsymbol{J}_T \otimes \boldsymbol{I}_K \otimes \boldsymbol{J}_N + \tilde{f}\boldsymbol{J}_T \otimes \boldsymbol{J}_K \otimes \boldsymbol{J}_N$$

where $\tilde{a} = \frac{1}{\lambda_1}$, $\tilde{b} = -\frac{\lambda_2 - \lambda_1}{N\lambda_2\lambda_1}$, $\tilde{c} = \frac{\lambda_2 - \lambda_3}{KN\lambda_2\lambda_3}$, $\tilde{d} = \frac{1}{T\lambda_4} - \frac{1}{T\lambda_1}$, $\tilde{e} = \frac{\lambda_2 - \lambda_1}{TN\lambda_2\lambda_1} - \frac{\lambda_5 - \lambda_4}{TN\lambda_4\lambda_5}$, $\tilde{f} = \frac{\lambda_5 - \lambda_6}{TKN\lambda_5\lambda_6} - \frac{\lambda_2 - \lambda_3}{TKN\lambda_2\lambda_3}$.

First, let's generate Ω_{11},

$$\begin{aligned}\Omega_{11} &= \sum_{i=1}^{I}\left\{(\boldsymbol{I}_T \otimes \mathbf{1}_K^\top \otimes \mathbf{1}_N^\top)R_i^{-1}(\boldsymbol{I}_T \otimes \mathbf{1}_K \otimes \mathbf{1}_N)\right\}\\ &= \sum_{i=1}^{I} KN\{(\tilde{a} + N\tilde{b} + KN\tilde{c})\boldsymbol{I}_T + (\widetilde{d} + N\tilde{e} + KN\widetilde{f})\boldsymbol{J}_T\}\\ &= IKN\left(\frac{1}{\lambda_3}\boldsymbol{I}_T + \frac{\lambda_3 - \lambda_6}{T\lambda_6\lambda_3}\boldsymbol{J}_T\right).\end{aligned}$$

We need the inverse of Ω_{11} for our block matrix inversion formula. Using equation (1) we know that,

$$\Omega_{11}^{-1} = \frac{1}{IKN}\left(\lambda_3\boldsymbol{I}_T + \frac{\lambda_6 - \lambda_3}{T}\boldsymbol{J}_T\right).$$

Similarly, we can generate $\Omega_{12} = \Omega_{21}^\top$,

$$\begin{aligned}\Omega_{12} &= \sum_{i=1}^{I}\left\{(\boldsymbol{I}_T \otimes \mathbf{1}_K^\top \otimes \mathbf{1}_N^\top)R_i^{-1}(\boldsymbol{X}_i \otimes \mathbf{1}_K \otimes \mathbf{1}_N)\right\}\\ &= \sum_{i=1}^{I} KN\left\{(\tilde{a} + N\tilde{b} + KN\tilde{c})\boldsymbol{X}_i + (\widetilde{d} + N\tilde{e} + KN\widetilde{f})\sum_{j=1}^{T} X_{ij}\mathbf{1}_T\right\}\\ &= KN\left(\frac{1}{\lambda_3}\sum_{i=1}^{I}\boldsymbol{X}_i + \frac{\lambda_3 - \lambda_6}{T\lambda_6\lambda_3}\sum_{i=1}^{I}\sum_{j=1}^{T} X_{ij}\mathbf{1}_T\right).\end{aligned}$$

Finally, we can generate Ω_{22},

$$\begin{aligned}\Omega_{22} &= \sum_{i=1}^{I}\left\{(\boldsymbol{X}_i^\top \otimes \mathbf{1}_K^\top \otimes \mathbf{1}_N^\top)R_i^{-1}(\boldsymbol{X}_i \otimes \mathbf{1}_K \otimes \mathbf{1}_N)\right\}\\ &= \sum_{i=1}^{I} KN\left\{(\tilde{a} + N\tilde{b} + KN\tilde{c})\sum_{j=1}^{T} X_{ij}^2 + (\widetilde{d} + N\tilde{e} + KN\widetilde{f})\left(\sum_{j=1}^{T} X_{ij}\right)^2\right\}\\ &= KN\left\{\frac{1}{\lambda_3}\sum_{i=1}^{I}\sum_{j=1}^{T} X_{ij}^2 + \frac{\lambda_3 - \lambda_6}{T\lambda_6\lambda_3}\sum_{i=1}^{I}\left(\sum_{j=1}^{T} X_{ij}\right)^2\right\}.\end{aligned}$$

Let $U = \sum_{i=1}^{I}\sum_{j=1}^{T} X_{ij} = \sum_{i=1}^{I}\sum_{j=1}^{T} X_{ij}^2$, $V = \sum_{i=1}^{I}(\sum_{j=1}^{T} X_{ij})^2$, and $W = \sum_{j=1}^{T}(\sum_{i=1}^{I} X_{ij})^2$. This gives us the following Ωs for our block inversion formula,

$$\begin{aligned}\Omega_{11}^{-1} &= \frac{1}{IKN}\left(\lambda_3\boldsymbol{I}_T + \frac{\lambda_6 - \lambda_3}{T}\boldsymbol{J}_T\right)\\ \Omega_{12} = \Omega_{21}^\top &= KN\left(\frac{1}{\lambda_3}\sum_{i=1}^{I}\boldsymbol{X}_i - \frac{\lambda_6 - \lambda_3}{T\lambda_6\lambda_3}U\mathbf{1}_T\right)\\ \Omega_{22} &= KN\left(\frac{1}{\lambda_3}U - \frac{\lambda_6 - \lambda_3}{T\lambda_6\lambda_3}V\right).\end{aligned}$$

Carrying out the block matrix inversion gives us,

$$\text{var}(\widehat{\delta}) = \frac{(\sigma^2/KN)IT\lambda_6\lambda_3}{(U^2 + ITU - TW - IV)\lambda_6 - (U^2 - IV)\lambda_3}. \tag{6}$$

Given that $\mathbf{1}^\top \mathbf{\Omega} \mathbf{1} = I^{-2}(IV - U^2)$ and $tr(\mathbf{\Omega}) = I^{-2}(IU - W)$, the variance can be rewritten as,

$$\text{var}(\widehat{\delta}) = \frac{\sigma^2}{IKNtr(\mathbf{\Omega})} \times \frac{T\lambda_6\lambda_3}{T\lambda_6 - \{1 + (T-1)\tau_X\}(\lambda_6 - \lambda_3)}.$$

Using the eigenvalue expressions under each design variant shown in Web Table 1, this expression for $\text{var}(\widehat{\delta})$ can be used across all design variants. Further, this variance expression can be used for longitudinal parallel or any type of crossover design through the specification of $tr(\mathbf{\Omega})$ and τ_X.

Relationship with ICCs

The partial derivative of $\text{var}(\widehat{\delta})$ with respect to ICC, θ, is

$$\frac{\partial}{\partial\theta}\text{var}(\widehat{\delta}) = \frac{\sigma^2 T}{IKNtr(\mathbf{\Omega})}\left[\frac{\lambda_3\frac{\partial}{\partial\theta}\lambda_6 + \lambda_6\frac{\partial}{\partial\theta}\lambda_3}{T\lambda_6 - \{1 + (T-1)\tau_X\}(\lambda_6 - \lambda_3)} - \frac{\lambda_6\lambda_3\left\{T\frac{\partial}{\partial\theta}\lambda_6 - \{1 + (T-1)\tau_X\}\left(\frac{\partial}{\partial\theta}\lambda_6 - \frac{\partial}{\partial\theta}\lambda_3\right)\right\}}{\{T\lambda_6 - \{1 + (T-1)\tau_X\}(\lambda_6 - \lambda_3)\}^2}\right]. \tag{7}$$

In order for $\boldsymbol{R}_i$ to be positive definite we have the constraint, $\min(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6) > 0$. We also assume that all ICCs are positive (as expected in SW-CRTs). The following exploration assumes a closed-cohort design at both the subcluster and subject levels (design variant A), but conclusions remain the same across design variants.

Relationship with within-period, within-subcluster, α_0

The partial derivative of λ_3 and λ_6 with respect to α_0 are,

$$\frac{\partial}{\partial\alpha_0}\lambda_3 = N - 1$$
$$\frac{\partial}{\partial\alpha_0}\lambda_6 = N - 1.$$

Looking at equation (7) we see that

$$\lambda_3\frac{\partial}{\partial\alpha_0}\lambda_6 + \lambda_6\frac{\partial}{\partial\alpha_0}\lambda_3 = (N-1)(\lambda_3 + \lambda_6)$$
$$T\frac{\partial}{\partial\alpha_0}\lambda_6 - \{1 + (T-1)\tau_X\}\left(\frac{\partial}{\partial\alpha_0}\lambda_6 - \frac{\partial}{\partial\alpha_0}\lambda_3\right) = T(N-1).$$

After some algebra this gives us,

$$\frac{\partial}{\partial \alpha_0}\text{var}(\widehat{\delta}) = \frac{\sigma^2 T(T-1)(N-1)}{IKNtr(\boldsymbol{\Omega})} \times \frac{\lambda_6^2 + \lambda_3^2 - (\lambda_6^2 - \lambda_3^2)\tau_X}{[T\lambda_6 - \{1 + (T-1)\tau_X\}(\lambda_6 - \lambda_3)]^2} > 0$$

Therefore, holding all other ICCs constant, an increase in α_0 leads to an increase in $\text{var}(\widehat{\delta})$.

Relationship with within-period, between-subcluster, ρ_0

The partial derivative of λ_3 and λ_6 with respect to ρ_0 are,

$$\frac{\partial}{\partial \rho_0}\lambda_3 = N(K-1)$$
$$\frac{\partial}{\partial \rho_0}\lambda_6 = N(K-1).$$

Looking at equation (7) we see that

$$\lambda_3 \frac{\partial}{\partial \rho_0}\lambda_6 + \lambda_6 \frac{\partial}{\partial \rho_0}\lambda_3 = N(K-1)(\lambda_3 + \lambda_6)$$
$$T\frac{\partial}{\partial \rho_0}\lambda_6 - \{1 + (T-1)\tau_X\}\left(\frac{\partial}{\partial \rho_0}\lambda_6 - \frac{\partial}{\partial \rho_0}\lambda_3\right) = TN(K-1).$$

After some algebra this gives us,

$$\frac{\partial}{\partial \rho_0}\text{var}(\widehat{\delta}) = \frac{\sigma^2 T(T-1)(K-1)}{IKtr(\boldsymbol{\Omega})} \times \frac{\lambda_6^2 + \lambda_3^2 - (\lambda_6^2 - \lambda_3^2)\tau_X}{[T\lambda_6 - \{1 + (T-1)\tau_X\}(\lambda_6 - \lambda_3)]^2} > 0$$

Therefore, holding all other ICCs constant, an increase in ρ_0 leads to an increase in $\text{var}(\widehat{\delta})$.

Relationship with between-period, within-subcluster, α_1

The partial derivative of λ_3 and λ_6 with respect to α_1 are,

$$\frac{\partial}{\partial \alpha_1}\lambda_3 = -(N-1)$$
$$\frac{\partial}{\partial \alpha_1}\lambda_6 = (T-1)(N-1).$$

Looking at equation (7) we see that

$$\lambda_3 \frac{\partial}{\partial \alpha_1}\lambda_6 + \lambda_6 \frac{\partial}{\partial \alpha_1}\lambda_3 = (N-1)\{(T-1)\lambda_3 - \lambda_6\}$$
$$T\frac{\partial}{\partial \alpha_1}\lambda_6 - \{1 + (T-1)\tau_X\}\left(\frac{\partial}{\partial \alpha_1}\lambda_6 - \frac{\partial}{\partial \alpha_1}\lambda_3\right) = T(N-1)[(T-1) - \{1 + (T-1)\tau_X\}].$$

After some algebra this gives us,

$$\frac{\partial}{\partial \alpha_1}\text{var}(\widehat{\delta}) = \frac{\sigma^2 T(T-1)(N-1)}{IKNtr(\mathbf{\Omega})} \times \frac{-[\lambda_6^2 - \lambda_3^2 - \{\lambda_6^2 + (T-1)\lambda_3^2\}\tau_X]}{[T\lambda_6 - \{1 + (T-1)\tau_X\}(\lambda_6 - \lambda_3)]^2}$$

which is negative only under certain constraints (discussed below).

Relationship with between-period, between-subcluster, ρ_1

The partial derivative of λ_3 and λ_6 with respect to ρ_1 are,

$$\frac{\partial}{\partial \rho_1}\lambda_3 = -N(K-1)$$
$$\frac{\partial}{\partial \rho_1}\lambda_6 = N(K-1)(T-1).$$

Looking at equation (7) we see that

$$\lambda_3 \frac{\partial}{\partial \rho_1}\lambda_6 + \lambda_6 \frac{\partial}{\partial \rho_1}\lambda_3 = N(K-1)\{(T-1)\lambda_3 - \lambda_6\}$$
$$T\frac{\partial}{\partial \rho_1}\lambda_6 - \{1 + (T-1)\tau_X\}\Big(\frac{\partial}{\partial \rho_1}\lambda_6 - \frac{\partial}{\partial \rho_1}\lambda_3\Big) = TN(K-1)[(T-1) - \{1 + (T-1)\tau_X\}].$$

After some algebra this gives us,

$$\frac{\partial}{\partial \rho_1}\text{var}(\widehat{\delta}) = \frac{\sigma^2 T(T-1)(K-1)}{IKtr(\mathbf{\Omega})} \times \frac{-[\lambda_6^2 - \lambda_3^2 - \{\lambda_6^2 + (T-1)\lambda_3^2\}\tau_X]}{[T\lambda_6 - \{1 + (T-1)\tau_X\}(\lambda_6 - \lambda_3)]^2}$$

which is negative only under certain constraints (discussed below).

Relationship with within-subject auto-correlation, α_2

The partial derivative of λ_3 and λ_6 with respect to α_2 are,

$$\frac{\partial}{\partial \alpha_2}\lambda_3 = -1$$
$$\frac{\partial}{\partial \alpha_2}\lambda_6 = T-1.$$

Looking at equation (7) we see that

$$\lambda_3 \frac{\partial}{\partial \alpha_2}\lambda_6 + \lambda_6 \frac{\partial}{\partial \alpha_2}\lambda_3 = (T-1)\lambda_3 - \lambda_6$$
$$T\frac{\partial}{\partial \alpha_2}\lambda_6 - \{1 + (T-1)\tau_X\}\Big(\frac{\partial}{\partial \alpha_2}\lambda_6 - \frac{\partial}{\partial \alpha_2}\lambda_3\Big) = T[(T-1) - \{1 + (T-1)\tau_X\}].$$

After some algebra this gives us,

$$\frac{\partial}{\partial \alpha_2}\text{var}(\widehat{\delta}) = \frac{\sigma^2 T}{IKNtr(\mathbf{\Omega})} \times \frac{-[\lambda_6^2 - \lambda_3^2 - \{\lambda_6^2 + (T-1)\lambda_3^2\}\tau_X]}{[T\lambda_6 - \{1 + (T-1)\tau_X\}(\lambda_6 - \lambda_3)]^2}$$

which is negative only under certain constraints.

Constraint in which the partial derivatives for α_1, ρ_1, and α_2 are negative

The partial derivatives for α_1, ρ_1, and α_2 are negative if

$$\{\lambda_6^2 + (T-1)\lambda_3^2\}\tau_X < \lambda_6^2 - \lambda_3^2$$

$$\Rightarrow \tau_X < \frac{\lambda_6^2 - \lambda_3^2}{\lambda_6^2 + (T-1)\lambda_3^2}$$

where $0 < (\lambda_6^2 - \lambda_3^2)/\{\lambda_6^2 + (T-1)\lambda_3^2\} < 1$. Therefore, the partial derivatives for α_1, ρ_1, and α_2 are negative for all $\tau_X < 0$. For $\tau_X > 0$, the condition above must be verified.

Web Appendix B

Optimal cluster allocation for SW-CRTs with subclusters

The optimal treatment allocation minimizes the variance (5) through maximization of the denominator. In the absence of any subclusters, Lawrie et al. (2015) and Li et al. (2018a) proved that the most efficient SW-CRT allocated more clusters to the intervention during the second and last period, based on a simple exchangeable and block exchangeable correlation model. Specifically, Theorem 1 of Li et al. (2018a) focused on maximizing the denominator of the variance expression defined as, $Q(\boldsymbol{\pi}) = \{(IU - W)\gamma + (U^2 - IV)\xi\}/I^2$, where $\gamma = 1 + (N-1)\alpha_0 + (T-1)(N-1)\alpha_1 + (T-1)\alpha_2$ and $\xi = (N-1)\alpha_1 + \alpha_2$ with α_0 defined as the within-period within-cluster ICC; α_1 defined as the between-period within-cluster ICC; and α_2 defined as the within-subject auto-correlation (i.e. a closed-cohort SW-CRT in the absence of subclusters). U, V, and W are the same design constants described previously, but have been re-parameterized in terms of the proportion of clusters receiving the intervention at a specific period j, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_{T-1})$ with $\pi_1 = 0$ and $\sum_{j=2}^{T} \pi_j = 1$ by definition, such that, $U = I\sum_{j=1}^{T-1} j\pi_j$, $V = I\sum_{j=1}^{T-1} j^2\pi_j$, and $W = I^2\sum_{j=1}^{T-1}\left(\sum_{s=j}^{T-1}\pi_s\right)^2$. Li et al. (2018b) derived the eigenvalues for this block exchangeable correlation matrix and found that the variance expression only depends on two eigenvalues, $\lambda_3 = 1 + (N-1)(\alpha_0 - \alpha_1) - \alpha_2$ and $\lambda_4 = 1 + (N-1)\alpha_0 + (T-1)(N-1)\alpha_1 + (T-1)\alpha_2$. Right away we can see that $\gamma = \lambda_4$ and rewriting the Q function in terms of the eigenvalues gives us, $Q(\boldsymbol{\pi}) = \{(U^2 + ITU - TW - IV)\lambda_4 - (U^2 - IV)\lambda_3\}/I^2$. If we rewrite our variance expression (5) in the following form,

$$\mathrm{var}(\widehat{\delta}) = \frac{(\sigma^2/KN)IT\lambda_6\lambda_3}{(U^2 + ITU - TW - IV)\lambda_6 - (U^2 - IV)\lambda_3},$$

where $U = \sum_{i=1}^{I}\sum_{j=1}^{T} X_{ij}$, $V = \sum_{i=1}^{I}(\sum_{j=1}^{T} X_{ij})^2$, and $W = \sum_{j=1}^{T}(\sum_{i=1}^{I} X_{ij})^2$, our Q function (using the same re-parameterized functions for U, V, and W) is $Q(\boldsymbol{\pi}) = \{(U^2 + ITU - TW - IV)\lambda_6 - (U^2 - IV)\lambda_3\}/I^2$ and only differs from the Q function above through the eigenvalue expressions. Therefore, by a similar argument to Theorem 1 of Li et al. (2018a) we can show that the same conclusion holds in the presence of subclusters with the proposed extended block exchangeable correlation model. Specifically, among all possible randomization schemes that allocates $\pi_j I$ clusters during the jth period, the optimal allocation ratio that leads to the smallest $\mathrm{var}(\widehat{\delta})$ is given by

$$\pi_2 = \pi_T = \frac{3\lambda_6 + (T-3)\lambda_3}{2T\lambda_6}, \quad \pi_j = \frac{\lambda_6 - \lambda_3}{T\lambda_6}, \quad \forall\ j = 3, \ldots, T-1. \tag{8}$$

In other words, the allocation ratio in each period in the most efficient SW-CRT with subclusters depends on the five ICCs only through the two eigenvalues of the extended block exchangeable correlation matrix as well as the total number of periods.

Web Appendix C

Approximation of $\boldsymbol{V}_i$ and $\widetilde{\boldsymbol{V}}_i$ under a GLMM framework

Using individual-level outcomes

Recall our generalized linear mixed model (GLMM) is the following,

$$\mu_{ijkl} = g^{-1}(\eta_{ijkl}) = g^{-1}(\beta_j + \delta X_{ij} + b_i + c_{ik} + s_{ij} + \pi_{ijk} + \gamma_{ikl}),$$

where g is a link function, β_j represents the categorical secular trend, X_{ij} is the intervention status for cluster i at period j, δ is the intervention effect on the link function scale, $b_i \sim \mathrm{N}(0, \sigma_b^2)$ is the random cluster effect, $c_{ik} \sim \mathrm{N}(0, \sigma_c^2)$ is the random subcluster effect, $s_{ij} \sim \mathrm{N}(0, \sigma_s^2)$ is the random cluster-by-period interaction, $\pi_{ijk} \sim \mathrm{N}(0, \sigma_\pi^2)$ is the random subcluster-by-period interaction, and $\gamma_{ikl} \sim \mathrm{N}(0, \sigma_\gamma^2)$ is the random subject-level effect. We define the conditional variance of the outcome as $\phi\zeta(\mu_{ijkl})$, where ϕ is a common dispersion. Without loss of generality, we assume $\phi = 1$ but the following procedure applies to arbitrary $\phi > 0$.

We can re-write our GLMM using matrix notation. Let $\boldsymbol{Y}_{ij} = (Y_{ij11}, \ldots, Y_{ijKN})^\top$, $\boldsymbol{\mu}_{ij} = (\mu_{ij11}, \ldots, \mu_{ijKN})^\top$, $\boldsymbol{\eta}_{ij} = (\eta_{ij11}, \ldots, \eta_{ijKN})^\top$, $\boldsymbol{Z}_{ij} = (\boldsymbol{e}_j, X_{ij}) \otimes \boldsymbol{1}_{KN}$ ($\boldsymbol{e}_j$ is the jth row of $\boldsymbol{I}_T$), $\boldsymbol{\theta} = (\beta_1, \ldots, \beta_T, \delta)^\top$, $\boldsymbol{c}_{ij} = (c_{i1}, \ldots, c_{iK})^\top$,

$\boldsymbol{\pi}_{ij} = (\pi_{ij1}, \dots, \pi_{ijK})^\top$, and $\boldsymbol{\gamma}_{ij} = (\gamma_{i11}, \dots, \gamma_{iKN})^\top$. Using these vectors and matrices our GLMM becomes,

$$\boldsymbol{\mu}_{ij} = g^{-1}(\boldsymbol{\eta}_{ij}) = g^{-1}(\boldsymbol{Z}_{ij}\boldsymbol{\theta} + \mathbf{1}_{KN}b_i + (\boldsymbol{I}_K \otimes \mathbf{1}_N)\boldsymbol{c}_i + \mathbf{1}_{KN}\boldsymbol{s}_{ij} + (\boldsymbol{I}_K \otimes \mathbf{1}_N)\boldsymbol{\pi}_{ij} + \boldsymbol{\gamma}_i).$$

Next, we linearize the GLMM using a first-order Taylor expansion about the estimated fixed- and random-effects (Breslow and Clayton, 1993; Amatya and Bhaumik, 2018) such that,

$$\begin{aligned}\boldsymbol{Y}_{ij} =& \widehat{\boldsymbol{\mu}}_{ij} + \widehat{\boldsymbol{\Delta}}_{ij}\boldsymbol{Z}_{ij}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) + \widehat{\boldsymbol{\Delta}}_{ij}\mathbf{1}_{KN}(b_i - \widehat{b}_i) + \widehat{\boldsymbol{\Delta}}_{ij}(\boldsymbol{I}_K \otimes \mathbf{1}_N)(\boldsymbol{c}_{ij} - \widehat{\boldsymbol{c}}_{ij}) \\ &+ \widehat{\boldsymbol{\Delta}}_{ij}\mathbf{1}_{KN}(s_{ij} - \widehat{s}_{ij}) + \widehat{\boldsymbol{\Delta}}_{ij}(\boldsymbol{I}_K \otimes \mathbf{1}_N)(\boldsymbol{\pi}_{ij} - \widehat{\boldsymbol{\pi}}_{ij}) + \widehat{\boldsymbol{\Delta}}_{ij}(\boldsymbol{\gamma}_{ij} - \widehat{\boldsymbol{\gamma}}_{ij}) + \boldsymbol{\epsilon}_{ij}, \end{aligned} \quad (9)$$

where $\boldsymbol{\Delta}_{ij} = \text{diag}(\Delta_{ij11}, \dots, \Delta_{ijKN}) = \{\partial g^{-1}(\boldsymbol{\eta}_{ij})/\partial \boldsymbol{\eta}_{ij}^\top\}^{-1}$ is a diagonal matrix of derivatives, $\boldsymbol{\epsilon}_{ij} = (\epsilon_{ij11}, \dots, \epsilon_{ijKN})^\top$ and $\text{var}(\epsilon_{ijkl}) = \zeta(\mu_{ijkl})$. Therefore, if we define the vector of pseudo-outcomes as $\boldsymbol{Y}_{ij}^* = \widehat{\boldsymbol{\Delta}}_{ij}^{-1}(\boldsymbol{Y}_{ij} - \widehat{\boldsymbol{\mu}}_{ij}) + \widehat{\boldsymbol{\eta}}_{ij}$ and rearrange the terms in (9), we obtain an approximate linear mixed model with $\boldsymbol{Y}_{ij}^* = \boldsymbol{\eta}_{ij} + \boldsymbol{\epsilon}_{ij}^*$, with a modified random residual error $\boldsymbol{\epsilon}_{ij}^* = \widehat{\boldsymbol{\Delta}}_{ij}^{-1}\boldsymbol{\epsilon}_{ij}$. Define the collection of all pseudo-outcomes in cluster i as $\boldsymbol{Y}_i^* = (\boldsymbol{Y}_{i1}^{*\top}, \dots, \boldsymbol{Y}_{iT}^{*\top})^\top$, the covariance expression for the pseudo-observations, $\boldsymbol{V}_i$, is comprised of two parts, within-period and between-period, and has a block exchangeable matrix structure such that

$$\boldsymbol{V}_i = \boldsymbol{I}_T \otimes (\boldsymbol{B} - \boldsymbol{C}) + \boldsymbol{J}_T \otimes \boldsymbol{C},$$

where $\boldsymbol{B} = \text{var}(\boldsymbol{Y}_{ij}^*)$ and $\boldsymbol{C} = \text{cov}(\boldsymbol{Y}_{ij}^*, \boldsymbol{Y}_{ij'}^*)$ for $j \neq j'$. We can easily generate these covariance expressions to get

$$\begin{aligned}\boldsymbol{B} &= \text{var}(\boldsymbol{Y}_{ij}^*) = \boldsymbol{\Delta}_{ij}^{-1}\zeta(\boldsymbol{\mu}_{ij})\boldsymbol{\Delta}_{ij}^{-1} + (\sigma_b^2 + \sigma_s^2)\boldsymbol{J}_{KN} + (\sigma_c^2 + \sigma_\pi^2)(\boldsymbol{I}_K \otimes \boldsymbol{J}_N) + \sigma_\gamma^2(\boldsymbol{I}_K \otimes \boldsymbol{I}_N) \\ \boldsymbol{C} &= \text{cov}(\boldsymbol{Y}_{ij}^*, \boldsymbol{Y}_{ij'}^*) = \sigma_b^2\boldsymbol{J}_{KN} + \sigma_c^2(\boldsymbol{I}_K \otimes \boldsymbol{J}_N) + \sigma_\gamma^2(\boldsymbol{I}_K \otimes \boldsymbol{I}_N).\end{aligned}$$

Combining these expressions gives us

$$\begin{aligned}\boldsymbol{V}_i &= \boldsymbol{I}_T \otimes (\boldsymbol{B} - \boldsymbol{C}) + \boldsymbol{J}_T \otimes \boldsymbol{C} \\ &\approx \mathbb{E}\{\boldsymbol{\Delta}_i^{-1}\zeta(\boldsymbol{\mu}_i)\boldsymbol{\Delta}_i^{-1}\} + \sigma_s^2(\boldsymbol{I}_T \otimes \boldsymbol{J}_{KN}) + \sigma_\pi^2(\boldsymbol{I}_T \otimes \boldsymbol{I}_K \otimes \boldsymbol{J}_N) + \sigma_b^2\boldsymbol{J}_{TKN} + \sigma_c^2(\boldsymbol{J}_T \otimes \boldsymbol{I}_K \otimes \boldsymbol{J}_N) + \sigma_\gamma^2(\boldsymbol{J}_T \otimes \boldsymbol{I}_{KN}) \\ &= (\mathbf{E}_i \otimes \boldsymbol{I}_{KN}) + \sigma_s^2(\boldsymbol{I}_T \otimes \boldsymbol{J}_{KN}) + \sigma_\pi^2(\boldsymbol{I}_T \otimes \boldsymbol{I}_K \otimes \boldsymbol{J}_N) + \sigma_b^2\boldsymbol{J}_{TKN} + \sigma_c^2(\boldsymbol{J}_T \otimes \boldsymbol{I}_K \otimes \boldsymbol{J}_N) + \sigma_\gamma^2(\boldsymbol{J}_T \otimes \boldsymbol{I}_{KN}) \end{aligned} \quad (10)$$

with $\boldsymbol{\Delta}_i = \oplus_{j=1}^T \boldsymbol{\Delta}_{ij}$ where "$\oplus$" is a block diagonal operator with nonzero matrices along the diagonal and zero values elsewhere, and the expectation is taken over the distribution of all the random effects. We simplify this notation by defining $\boldsymbol{E}_i = \text{diag}\big[\mathbb{E}\{\Delta_{i111}^{-1}\zeta(\mu_{i111})\Delta_{i111}^{-1}\}, \dots, \mathbb{E}\{\Delta_{iT11}^{-1}\zeta(\mu_{iT11})\Delta_{iT11}^{-1}\}\big]$.

For example, if we have a binary outcome and our link function is the canonical logit, then g^{-1} is the inverse logit and $\boldsymbol{\Delta}$ is the derivative of the inverse logit which gives us

$$
\begin{aligned}
g^{-1}(\eta_{ijkl}) &= \frac{\exp(\eta_{ijkl})}{1+\exp(\eta_{ijkl})} \\
\boldsymbol{\Delta}_{ijkl}^{-1} &= \left\{\frac{\partial g^{-1}(\eta_{ijkl})}{\partial \eta_{ijkl}^{\top}}\right\}^{-1} = \frac{\{1+\exp(\eta_{ijkl})\}^2}{\exp(\eta_{ijkl})} \\
\zeta(\mu_{ijkl}) &= \mu_{ijkl}(1-\mu_{ijkl}) = \frac{\exp(\eta_{ijkl})}{\{1+\exp(\eta_{ijkl})\}^2}.
\end{aligned}
$$

Let $\zeta(\boldsymbol{\mu}_i)$ be a $TKN \times TKN$ diagonal matrix with elements $\exp(\eta_{ijkl})/\{1+\exp(\eta_{ijkl})\}^2$ and $\boldsymbol{\Delta}_i^{-1}$ be a $TKN \times TKN$ diagonal matrix with elements $\{1+\exp(\eta_{ijkl})\}^2/\exp(\eta_{ijkl})$, then $\boldsymbol{\Delta}_i^{-1}\zeta(\boldsymbol{\mu}_i)$ is $\boldsymbol{I}_{TKN}$. This means that, $\boldsymbol{\Delta}_i^{-1}\zeta(\boldsymbol{\mu}_i)\boldsymbol{\Delta}_i^{-1} = \boldsymbol{\Delta}_i^{-1}$. Looking at the diagonal elements of $\boldsymbol{\Delta}_i^{-1}$ further we can rewrite them as

$$
\boldsymbol{\Delta}_{ijkl}^{-1} = 2 + \exp(-\eta_{ijkl}) + \exp(\eta_{ijkl}).
$$

Next, we take the expectation of $\boldsymbol{\Delta}_{ijkl}^{-1}$ with respect to the random effects to get

$$
\begin{aligned}
\mathbb{E}(\boldsymbol{\Delta}_{ijkl}^{-1}) &= \mathbb{E}\{2+\exp(-\eta_{ijkl})+\exp(\eta_{ijkl})\} \\
&= 2 + \exp\left\{0.5\left(\sigma_b^2+\sigma_c^2+\sigma_s^2+\sigma_\pi^2+\sigma_\gamma^2\right)\right\}\{\exp(-\beta_j-\delta X_{ij})+\exp(\beta_j+\delta X_{ij})\} \\
&= 2 + 2\exp\left\{0.5\left(\sigma_b^2+\sigma_c^2+\sigma_s^2+\sigma_\pi^2+\sigma_\gamma^2\right)\right\}\cosh\left(\beta_j + X_{ij}\delta\right),
\end{aligned}
$$

where $\cosh(t) = (e^t + e^{-t})/2$ is the hyperbolic cosine function. Note that the expectation only depends on i and j. Therefore, we can express the unique values in matrix form as

$$
\mathbf{E}_i = 2\boldsymbol{I}_T + 2\exp\left\{0.5\left(\sigma_b^2+\sigma_c^2+\sigma_s^2+\sigma_\pi^2+\sigma_\gamma^2\right)\right\}\mathrm{diag}\{\cosh\left(\beta_1+X_{i1}\delta\right),\ldots,\cosh\left(\beta_T+X_{iT}\delta\right)\}.
$$

Using equation (10) we now have the following variance expression

$$
\begin{aligned}
\boldsymbol{V}_i \approx &\left[2\boldsymbol{I}_T + 2\exp\left\{0.5\left(\sigma_b^2+\sigma_c^2+\sigma_s^2+\sigma_\pi^2+\sigma_\gamma^2\right)\right\}\mathrm{diag}\left\{\cosh\left(\beta_1+X_{i1}\delta\right),\ldots,\cosh\left(\beta_T+X_{iT}\delta\right)\right\}\right] \otimes \boldsymbol{I}_{KN} \\
&+ \sigma_s^2(\boldsymbol{I}_T \otimes \boldsymbol{J}_{KN}) + \sigma_\pi^2(\boldsymbol{I}_T \otimes \boldsymbol{I}_K \otimes \boldsymbol{J}_N) + \sigma_b^2\boldsymbol{J}_{TKN} + \sigma_c^2(\boldsymbol{J}_T \otimes \boldsymbol{I}_K \otimes \boldsymbol{J}_N) + \sigma_\gamma^2(\boldsymbol{J}_T \otimes \boldsymbol{I}_{KN}), \quad (11)
\end{aligned}
$$

which depends on the conditional mean of the outcomes through the secular trend and intervention status, thus $\boldsymbol{V}_i$ will be cluster-specific. The variance of our treatment effect, $\mathrm{var}(\widehat{\delta})$, is the lower-right corner element of $\phi\left(\sum_{i=1}^{I}\boldsymbol{Z}_i^{\top}\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i\right)^{-1}$ which is equal to

$$
\mathrm{var}(\widehat{\delta}) = \phi\left\{\sum_{i=1}^{I}\mathbf{H}_i^{\top}\boldsymbol{V}_i^{-1}\mathbf{H}_i - \left(\sum_{i=1}^{I}\mathbf{H}_i^{\top}\boldsymbol{V}_i^{-1}\right)\boldsymbol{F}\left(\sum_{i=1}^{I}\boldsymbol{F}^{\top}\boldsymbol{V}_i^{-1}\boldsymbol{F}\right)^{-1}\boldsymbol{F}^{\top}\left(\sum_{i=1}^{I}\boldsymbol{V}_i^{-1}\mathbf{H}_i\right)\right\}^{-1}, \quad (12)
$$

where $\mathbf{H}_i = \boldsymbol{X}_i \otimes \mathbf{1}_{KN}$ and $\boldsymbol{F} = \boldsymbol{I}_T \otimes \mathbf{1}_{KN}$ and requires an algorithm to compute due to $\boldsymbol{V}_i$ being cluster-specific.

Using cluster-period means

One limitation of our variance expression (10) is that the dimension of our covariance matrix, $\boldsymbol{V}_i$, increases with increasing number of subclusters per cluster (K) and participants per subcluster (N) which in turn increases the computational burden. To reduce the computational burden we could use a cluster-period means approach. This approach reduces our covariance matrix from a $TKN \times TKN$ matrix to a $T \times T$ matrix. Using a cluster-period means approach the variance of our treatment effect, $\text{var}(\widehat{\delta})$, is the lower-right corner element of $\phi\left(\sum_{i=1}^{I} \boldsymbol{Z}_{2i}^{\top} \widetilde{\boldsymbol{V}}_i^{-1} \boldsymbol{Z}_{2i}\right)^{-1}$ with $\boldsymbol{Z}_{2i} = (\boldsymbol{I}_T, \boldsymbol{X}_i)$ which is equal to

$$\text{var}(\widehat{\delta}) = \phi\left\{\sum_{i=1}^{I} \boldsymbol{X}_i^{\top} \widetilde{\boldsymbol{V}}_i^{-1} \boldsymbol{X}_i - \left(\sum_{i=1}^{I} \boldsymbol{X}_i^{\top} \widetilde{\boldsymbol{V}}_i^{-1}\right)\left(\sum_{i=1}^{I} \widetilde{\boldsymbol{V}}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{I} \widetilde{\boldsymbol{V}}_i^{-1} \boldsymbol{X}_i\right)\right\}^{-1}, \tag{13}$$

where $\widetilde{\boldsymbol{V}}_i$ is the covariance matrix for $\overline{Y}_i^* = (\overline{Y}_{i1}^*, \ldots, \overline{Y}_{iT}^*)^{\top}$.

First, we focus on deriving $\widetilde{\boldsymbol{V}}_i$. This matrix is made up of within-period and between-period components. The within-period can be expressed as

$$\begin{aligned}
\boldsymbol{B} = \text{var}(\overline{Y}_{ij}^*) &= \text{var}\left(\frac{1}{KN}\sum_{k=1}^{K}\sum_{l=1}^{N} Y_{ijkl}^*\right) \\
&= \frac{\text{var}(Y_{ijkl}^*) + (N-1)\text{cov}(Y_{ijkl}^*, Y_{ijkl'}^*) + N(K-1)\text{cov}(Y_{ijkl}^*, Y_{ijk'l}^*)}{KN} \\
&= \frac{\mathbf{E}_i + \sigma_\gamma^2}{KN} + \frac{\sigma_c^2 + \sigma_\pi^2}{K} + \sigma_b^2 + \sigma_s^2.
\end{aligned}$$

The between-period can be expressed as

$$\begin{aligned}
\boldsymbol{C} = \text{cov}(\overline{Y}_{ij}^*, \overline{Y}_{ij'}^*) &= \text{cov}\left(\frac{1}{KN}\sum_{k=1}^{K}\sum_{l=1}^{N} Y_{ijkl}^*, \frac{1}{KN}\sum_{k=1}^{K}\sum_{l=1}^{N} Y_{ij'kl}^*\right) \\
&= \frac{\text{cov}(Y_{ijkl}^*, Y_{ij'kl}^*) + (N-1)\text{cov}(Y_{ijkl}^*, Y_{ij'kl'}^*) + N(K-1)\text{cov}(Y_{ijkl}^*, Y_{ij'k'l}^*)}{KN} \\
&= \sigma_b^2 + \frac{\sigma_c^2}{K} + \frac{\sigma_\gamma^2}{KN}.
\end{aligned}$$

Putting both of these components together we can generate our final covariance matrix for $\overline{Y}_i^*$,

$$\begin{aligned}
\widetilde{\boldsymbol{V}}_i &= (\boldsymbol{B} - \boldsymbol{C})\boldsymbol{I}_T + \boldsymbol{C}\boldsymbol{J}_T \\
&\approx \frac{\mathbf{E}_i}{KN} + \left(\frac{\sigma_\pi^2}{K} + \sigma_s^2\right)\boldsymbol{I}_T + \left(\sigma_b^2 + \frac{\sigma_c^2}{K} + \frac{\sigma_\gamma^2}{KN}\right)\boldsymbol{J}_T.
\end{aligned} \tag{14}$$

For example, if our outcome is binary with the canonical logit link then $\mathbf{E}_i$ is the same as previously derived giving us

$$\widetilde{\boldsymbol{V}}_i \approx (KN)^{-1}\left[2\boldsymbol{I}_T + 2\exp\left\{0.5\left(\sigma_b^2+\sigma_c^2+\sigma_s^2+\sigma_\pi^2+\sigma_\gamma^2\right)\right\}\mathrm{diag}\{\cosh(\beta_1+X_{i1}\delta),\ldots,\cosh(\beta_T+X_{iT}\delta)\}\right] + \left(\frac{\sigma_\pi^2}{K}+\sigma_s^2\right)\boldsymbol{I}_T + \left(\sigma_b^2+\frac{\sigma_c^2}{K}+\frac{\sigma_\gamma^2}{KN}\right)\boldsymbol{J}_T. \quad (15)$$

LEMMA 3: Equivalence of var($\widehat{\delta}$) approaches

LEMMA 3: The variance of the intervention effect estimator in the generalized linear mixed model is equivalently written as

$$\mathrm{var}(\widehat{\delta}) = \left\{\sum_{i=1}^{I}\boldsymbol{X}_i^\top\widetilde{\boldsymbol{V}}_i^{-1}\boldsymbol{X}_i - \left(\sum_{i=1}^{I}\boldsymbol{X}_i^\top\widetilde{\boldsymbol{V}}_i^{-1}\right)\left(\sum_{i=1}^{I}\widetilde{\boldsymbol{V}}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{I}\widetilde{\boldsymbol{V}}_i^{-1}\boldsymbol{X}_i\right)\right\}^{-1},$$

where $\widetilde{\boldsymbol{V}}_i = (KN)^{-1}\boldsymbol{E}_i + (K^{-1}\sigma_\pi^2+\sigma_s^2)\boldsymbol{I}_T + \{\sigma_b^2 + K^{-1}\sigma_c^2 + (KN)^{-1}\sigma_\gamma^2\}\boldsymbol{J}_T$ is the $T\times T$ matrix characterizing the covariance of the cluster-period means of the pseudo-outcomes $\overline{\boldsymbol{Y}}_i^* = (KN)^{-1}(\boldsymbol{I}_T\otimes\boldsymbol{1}_{KN}^\top)\boldsymbol{Y}_i^*$, and $\boldsymbol{E}_i = \mathrm{diag}\left[\mathbb{E}\{\Delta_{i111}^{-1}\zeta(\mu_{i111})\Delta_{i111}^{-1}\},\ldots,\mathbb{E}\{\Delta_{iT11}^{-1}\zeta(\mu_{iT11})\Delta_{iT11}^{-1}\}\right]$.

Proof: To prove Lemma 3 holds we need to show that the variance expression under individual-level outcomes is equivalent to the variance expression under a cluster-period means approach, that is that

$$\begin{aligned}\mathrm{var}(\widehat{\delta}) &= \left\{\sum_{i=1}^{I}\mathbf{H}_i^\top\boldsymbol{V}_i^{-1}\mathbf{H}_i - \left(\sum_{i=1}^{I}\mathbf{H}_i^\top\boldsymbol{V}_i^{-1}\right)\boldsymbol{F}\left(\sum_{i=1}^{I}\boldsymbol{F}^\top\boldsymbol{V}_i^{-1}\boldsymbol{F}\right)^{-1}\boldsymbol{F}^\top\left(\sum_{i=1}^{I}\boldsymbol{V}_i^{-1}\mathbf{H}_i\right)\right\}^{-1}\\ &= \left\{\sum_{i=1}^{I}\boldsymbol{X}_i^\top\widetilde{\boldsymbol{V}}_i^{-1}\boldsymbol{X}_i - \left(\sum_{i=1}^{I}\boldsymbol{X}_i^\top\widetilde{\boldsymbol{V}}_i^{-1}\right)\left(\sum_{i=1}^{I}\widetilde{\boldsymbol{V}}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{I}\widetilde{\boldsymbol{V}}_i^{-1}\boldsymbol{X}_i\right)\right\}^{-1}.\end{aligned}$$

We claim that the expressions for var($\widehat{\delta}$) are equivalent if for each cluster i,

$$(KN)^{-2}\boldsymbol{F}^\top\boldsymbol{V}_i^{-1}\boldsymbol{F} = (\boldsymbol{F}^\top\boldsymbol{V}_i\boldsymbol{F})^{-1}. \quad (16)$$

First, note that the covariance of Y_i^* can be written in the form $\boldsymbol{V}_i = (\boldsymbol{Q}_{i1}-\boldsymbol{Q}_{i2})\otimes\boldsymbol{I}_N + \boldsymbol{Q}_{i2}\otimes\boldsymbol{J}_N$ with $\boldsymbol{Q}_{i1}-\boldsymbol{Q}_{i2} = (\boldsymbol{E}_i+\sigma_\gamma^2\boldsymbol{J}_T)\otimes\boldsymbol{I}_K$ and $\boldsymbol{Q}_{2i} = (\sigma_\pi^2\boldsymbol{I}_T+\sigma_c^2\boldsymbol{J}_T)\otimes\boldsymbol{I}_K + (\sigma_s^2\boldsymbol{I}_T+\sigma_b^2\boldsymbol{J}_T)\otimes\boldsymbol{J}_K$. Therefore, the inverse can be obtained using formula (4) giving us

$$\boldsymbol{V}_i^{-1} = (\boldsymbol{Q}_{i1}-\boldsymbol{Q}_{i2})^{-1}\otimes\boldsymbol{I}_N + \frac{1}{N}\left[\{\boldsymbol{Q}_{i1}+(N-1)\boldsymbol{Q}_{i2}\}^{-1} - (\boldsymbol{Q}_{i1}-\boldsymbol{Q}_{i2})^{-1}\right]\otimes\boldsymbol{J}_N.$$

Looking at each sub-inverse we have

$$
\begin{aligned}
(\boldsymbol{Q}_{i1} - \boldsymbol{Q}_{i2})^{-1} &= (\boldsymbol{E}_i + \sigma_\gamma^2 \boldsymbol{J}_T)^{-1} \otimes \boldsymbol{I}_K \\
&= \boldsymbol{A}_i^{-1} \otimes \boldsymbol{I}_K
\end{aligned}
$$

$$
\{\boldsymbol{Q}_{i1} + (N-1)\boldsymbol{Q}_{i2}\}^{-1} = \left[\left\{\boldsymbol{E}_i + N\sigma_\pi^2 \boldsymbol{I}_T + (\sigma_\gamma^2 + N\sigma_c^2)\boldsymbol{J}_T\right\} \otimes \boldsymbol{I}_K + N(\sigma_s^2 \boldsymbol{I}_T + \sigma_b^2 \boldsymbol{J}_T) \otimes \boldsymbol{J}_K\right]^{-1}.
$$

We can easily rewrite $\boldsymbol{Q}_{i1} + (N-1)\boldsymbol{Q}_{i2}$ in the form of $(\boldsymbol{W}_{i1} - \boldsymbol{W}_{i2}) \otimes \boldsymbol{I}_K + \boldsymbol{W}_{i2} \otimes \boldsymbol{J}_K$ with $\boldsymbol{W}_{i1} - \boldsymbol{W}_{i2} = \boldsymbol{E}_i + N\sigma_\pi^2 \boldsymbol{I}_T + (\sigma_\gamma^2 + N\sigma_c^2)\boldsymbol{J}_T$ and $\boldsymbol{W}_{i2} = N(\sigma_s^2 \boldsymbol{I}_T + \sigma_b^2 \boldsymbol{J}_T)$, therefore, to generate the inverse we can apply formula (4) again to get

$$
\begin{aligned}
\{\boldsymbol{Q}_{i1} + (N-1)\boldsymbol{Q}_{i2}\}^{-1} &= (\boldsymbol{W}_{i1} - \boldsymbol{W}_{i2})^{-1} \otimes \boldsymbol{I}_K + \frac{1}{K}\left[\{\boldsymbol{W}_{i1} + (K-1)\boldsymbol{W}_{i2}\}^{-1} - (\boldsymbol{W}_{i1} - \boldsymbol{W}_{i2})^{-1}\right] \otimes \boldsymbol{J}_K \\
&= \boldsymbol{B}_i^{-1} \otimes \boldsymbol{I}_K + \frac{1}{K}\left(\boldsymbol{C}_i^{-1} - \boldsymbol{B}_i^{-1}\right) \otimes \boldsymbol{J}_K.
\end{aligned}
$$

Putting these components together gives us

$$
\boldsymbol{V}_i^{-1} = \boldsymbol{A}_i^{-1} \otimes \boldsymbol{I}_K \otimes \boldsymbol{I}_N + N^{-1}(\boldsymbol{B}_i^{-1} - \boldsymbol{A}_i^{-1}) \otimes \boldsymbol{I}_K \otimes \boldsymbol{J}_N + (KN)^{-1}(\boldsymbol{C}_i^{-1} - \boldsymbol{B}_i^{-1}) \otimes \boldsymbol{J}_K \otimes \boldsymbol{J}_N.
$$

Now that we have an expression for the inverse we can calculate the following directly,

$$
\boldsymbol{F}^\top \boldsymbol{V}_i^{-1} \boldsymbol{F} = KN\left\{\boldsymbol{E}_i + (N\sigma_\pi^2 + KN\sigma_s^2)\boldsymbol{I}_T + (\sigma_\gamma^2 + N\sigma_c^2 + KN\sigma_b^2)\boldsymbol{J}_T\right\}^{-1}.
$$

Comparing this result to the following,

$$
(\boldsymbol{F}^\top \boldsymbol{V}_i \boldsymbol{F})^{-1} = (KN)^{-1}\left\{\boldsymbol{E}_i + (N\sigma_\pi^2 + KN\sigma_s^2)\boldsymbol{I}_T + (\sigma_\gamma^2 + N\sigma_c^2 + KN\sigma_b^2)\boldsymbol{J}_T\right\}^{-1},
$$

we see that equation (16) holds. By definition, $\widetilde{\boldsymbol{V}}_i = (KN)^{-2}\boldsymbol{F}^\top \boldsymbol{V}_i \boldsymbol{F}$, therefore, by equation (16) we know that $\widetilde{\boldsymbol{V}}_i^{-1} = \boldsymbol{F}^\top \boldsymbol{V}_i^{-1} \boldsymbol{F}$. We can further show that

$$
\begin{aligned}
\mathbf{H}_i^\top \boldsymbol{V}_i^{-1} \mathbf{H}_i &= KN\boldsymbol{X}_i^\top \left\{\boldsymbol{E}_i + (N\sigma_\pi^2 + KN\sigma_s^2)\boldsymbol{I}_T + (\sigma_\gamma^2 + N\sigma_c^2 + KN\sigma_b^2)\boldsymbol{J}_T\right\}^{-1} \boldsymbol{X}_i \\
&= \boldsymbol{X}_i^\top \widetilde{\boldsymbol{V}}_i^{-1} \boldsymbol{X}_i
\end{aligned}
$$

and

$$
\mathbf{H}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{F} = KN\boldsymbol{X}_i^\top \left\{\boldsymbol{E}_i + (N\sigma_\pi^2 + KN\sigma_s^2)\boldsymbol{I}_T + (\sigma_\gamma^2 + N\sigma_c^2 + KN\sigma_b^2)\boldsymbol{J}_T\right\}^{-1}
$$

$$= \boldsymbol{X}_i^\top \widetilde{\boldsymbol{V}}_i^{-1}.$$

Therefore, we have shown that

$$\begin{aligned}\mathrm{var}(\widehat{\delta}) &= \left\{ \sum_{i=1}^{I} \mathbf{H}_i^\top \boldsymbol{V}_i^{-1} \mathbf{H}_i - \left(\sum_{i=1}^{I} \mathbf{H}_i^\top \boldsymbol{V}_i^{-1} \right) \boldsymbol{F} \left(\sum_{i=1}^{I} \boldsymbol{F}^\top \boldsymbol{V}_i^{-1} \boldsymbol{F} \right)^{-1} \boldsymbol{F}^\top \left(\sum_{i=1}^{I} \boldsymbol{V}_i^{-1} \mathbf{H}_i \right) \right\}^{-1} \\ &= \left\{ \sum_{i=1}^{I} \boldsymbol{X}_i^\top \widetilde{\boldsymbol{V}}_i^{-1} \boldsymbol{X}_i - \left(\sum_{i=1}^{I} \boldsymbol{X}_i^\top \widetilde{\boldsymbol{V}}_i^{-1} \right) \left(\sum_{i=1}^{I} \widetilde{\boldsymbol{V}}_i^{-1} \right)^{-1} \left(\sum_{i=1}^{I} \widetilde{\boldsymbol{V}}_i^{-1} \boldsymbol{X}_i \right) \right\}^{-1}.\end{aligned}$$

Web Appendix D

Count and gamma outcomes under a GLMM framework

Count outcome with canonical log link

Under a count outcome with canonical log link, we have $\zeta(\mu_{ijkl}) = \mu_{ijkl} = \exp(\eta_{ijkl})$ and $\boldsymbol{\Delta}_{ijkl}^{-1} = \exp(-\eta_{ijkl})$. Therefore, $\boldsymbol{\Delta}_{ijkl}^{-1} \zeta(\mu_{ijkl}) \boldsymbol{\Delta}_{ijkl}^{-1} = \boldsymbol{\Delta}_{ijkl}^{-1} = \exp(-\eta_{ijkl})$. Taking the expectation with respect to the random effects and using the property of the Gaussian moment generating function allows us to obtain

$$\mathbb{E}\{\exp(-\eta_{ijkl})\} = \exp\left(\frac{\sigma_b^2 + \sigma_c^2 + \sigma_s^2 + \sigma_\pi^2 + \sigma_\gamma^2}{2} \right) \exp\left(-\beta_j - X_{ij}\delta\right).$$

Therefore,

$$\boldsymbol{E}_i = \exp\left(\frac{\sigma_b^2 + \sigma_c^2 + \sigma_s^2 + \sigma_\pi^2 + \sigma_\gamma^2}{2} \right) \mathrm{diag}\{\exp(-\beta_1 - X_{i1}\delta), \ldots, \exp(-\beta_T - X_{iT}\delta)\}.$$

Gamma outcome with canonical inverse link

Under a gamma distribution with canonical inverse link function, we have $\zeta(\mu_{ijkl}) = \mu_{ijkl}^2 = 1/\eta_{ijkl}^2$ and $\boldsymbol{\Delta}_{ijkl}^{-1} = -\eta_{ijkl}^2$. Therefore, $\boldsymbol{\Delta}_{ijkl}^{-1} \zeta(\mu_{ijkl}) \boldsymbol{\Delta}_{ijkl}^{-1} = -\boldsymbol{\Delta}_{ijkl}^{-1} = \eta_{ijkl}^2$. Taking the expectation with respect to the random effects we have

$$\mathbb{E}(\eta_{ijkl}^2) = (\beta_j + X_{ij}\delta)^2 + \sigma_b^2 + \sigma_c^2 + \sigma_s^2 + \sigma_\pi^2 + \sigma_\gamma^2.$$

Therefore,

$$\boldsymbol{E}_i = (\sigma_b^2 + \sigma_c^2 + \sigma_s^2 + \sigma_\pi^2 + \sigma_\gamma^2)\boldsymbol{I}_T + \mathrm{diag}\{(\beta_1 + X_{i1}\delta)^2, \ldots, (\beta_T + X_{iT}\delta)^2\}.$$

Web Appendix E

Gaussian outcomes

Under a Gaussian outcome with unequal cluster sizes our induced correlation matrix given by

$$\begin{aligned}\boldsymbol{R}_i &= (1-\alpha_0-\alpha_2+\alpha_1)\boldsymbol{I}_{TKN} + (\alpha_0-\rho_0-\alpha_1+\rho_1)\boldsymbol{I}_{TK}\otimes\boldsymbol{J}_N + (\rho_0-\rho_1)\boldsymbol{I}_T\otimes\boldsymbol{J}_{KN}\\ &\quad + (\alpha_2-\alpha_1)\boldsymbol{J}_T\otimes\boldsymbol{I}_{KN} + (\alpha_1-\rho_1)\boldsymbol{J}_T\otimes\boldsymbol{I}_K\otimes\boldsymbol{J}_N + \rho_1\boldsymbol{J}_{TKN},\end{aligned}$$

becomes

$$\begin{aligned}\boldsymbol{R}_i &= (1-\alpha_0-\alpha_2+\alpha_1)\boldsymbol{I}_{TK_iN_i} + (\alpha_0-\rho_0-\alpha_1+\rho_1)\boldsymbol{I}_{TK_i}\otimes\boldsymbol{J}_{N_i} + (\rho_0-\rho_1)\boldsymbol{I}_T\otimes\boldsymbol{J}_{K_iN_i}\\ &\quad + (\alpha_2-\alpha_1)\boldsymbol{J}_T\otimes\boldsymbol{I}_{K_iN_i} + (\alpha_1-\rho_1)\boldsymbol{J}_T\otimes\boldsymbol{I}_{K_i}\otimes\boldsymbol{J}_{N_i} + \rho_1\boldsymbol{J}_{TK_iN_i},\end{aligned}$$

which is a $TK_iN_i \times TK_iN_i$ matrix. Thus, $\sigma^2(\sum_{i=1}^I \boldsymbol{Z}_i^\top \boldsymbol{R}_i^{-1}\boldsymbol{Z}_i)^{-1}$ must be derived numerically due to $\boldsymbol{R}_i$ being cluster-specific, a process simplified by using a cluster-period means approach.

Under a cluster-period means approach our correlation matrix is

$$\widetilde{\boldsymbol{R}}_i = \frac{1-\alpha_2+(N_i-1)(\alpha_0-\alpha_1)+N_i(K_i-1)(\rho_0-\rho_1)}{K_iN_i}\boldsymbol{I}_T + \frac{\alpha_2+(N_i-1)\alpha_1+N_i(K_i-1)\rho_1}{K_iN_i}\boldsymbol{J}_T,$$

which is a $T\times T$ exchangeable matrix with diagonal elements $\{1+(N_i-1)\alpha_0+N_i(K_i-1)\rho_0\}/(K_iN_i)$ and off-diagonal elements $\{\alpha_2+(N_i-1)\alpha_1+N_i(K_i-1)\rho_1\}/(K_iN_i)$. Re-writing $\widetilde{\boldsymbol{R}}_i$ in terms of the eigenvalues (Web Table 1) and using equation (1) we have

$$\widetilde{\boldsymbol{R}}_i^{-1} = \frac{K_iN_i}{\lambda_{3i}}\boldsymbol{I}_T - \frac{K_iN_i(\lambda_{6i}-\lambda_{3i})}{T\lambda_{3i}\lambda_{6i}}\boldsymbol{J}_T, \tag{17}$$

where $\lambda_{3i} = 1-\alpha_0-\alpha_2+\alpha_1+N_i\{\alpha_0-\alpha_1+(K_i-1)(\rho_0-\rho_1)\}$ and $\lambda_{6i} = 1-\alpha_0+(T-1)(\alpha_2-\alpha_1)+N_i\{\alpha_0+(T-1)\alpha_1+(K_i-1)(\rho_0+(T-1)\rho_1)\}$ under design variant A. Therefore, $\mathrm{var}(\widehat{\delta})$ will still depend on the ICCs through the same two eigenvalues of the extended block exchangeable correlation matrix as seen in Web Appendix A. The variance of the intervention effect estimate under a cluster-period means approach is the $(T+1,T+1)$th element of

$$\left(\sum_{i=1}^I \boldsymbol{Z}_{2i}^\top\widetilde{\boldsymbol{R}}_i^{-1}\boldsymbol{Z}_{2i}\right)^{-1} = \left(\sum_{i=1}^I \begin{pmatrix}\boldsymbol{I}_T\\ \boldsymbol{X}_i^\top\end{pmatrix}\widetilde{\boldsymbol{R}}_i^{-1}\begin{pmatrix}\boldsymbol{I}_T & \boldsymbol{X}_i\end{pmatrix}\right)^{-1} = \begin{bmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{bmatrix}^{-1},$$

where $\boldsymbol{Z}_{2i} = (\boldsymbol{I}_T, \boldsymbol{X}_i)$, Ω_{11} is of dimension $T \times T$, $\Omega_{12} = \Omega_{21}^\top$ is of dimension $T \times 1$, and Ω_{22} is a scalar. For a given cluster i we have

$$\Omega_{11i} = \widetilde{\boldsymbol{R}}_i^{-1}$$

$$\Omega_{12i} = \frac{K_i N_i}{\lambda_{3i}} \boldsymbol{X}_i - \frac{K_i N_i(\lambda_{6i} - \lambda_{3i})}{T\lambda_{3i}\lambda_{6i}} \sum_{j=1}^{T} X_{ij} \mathbf{1}_T$$

$$\Omega_{22i} = \frac{K_i N_i}{\lambda_{3i}} \sum_{j=1}^{T} X_{ij}^2 - \frac{K_i N_i(\lambda_{6i} - \lambda_{3i})}{T\lambda_{3i}\lambda_{6i}} \left(\sum_{j=1}^{T} X_{ij} \right)^2$$

Equivalence of var($\widehat{\delta}$) approaches under equal cluster sizes

We already know according to Lemma 3 that the individual-level outcomes approach is equivalent to the cluster-period means approach under equal cluster sizes. Therefore, we should be able to derive the same variance expression (6) from Web Appendix A using cluster-period means.

If we have equal cluster sizes, $K_i = K$ and $N_i = N$, then we would have

$$\Omega_{11} = \sum_{i=1}^{I} \Omega_{11i} = I \left\{ \frac{KN}{\lambda_3} \boldsymbol{I}_T - \frac{KN(\lambda_6 - \lambda_3)}{T\lambda_3\lambda_6} \boldsymbol{J}_T \right\}.$$

We can apply equation (1) to get the inverse

$$\Omega_{11}^{-1} = (IKN)^{-1} \left(\lambda_3 \boldsymbol{I}_T + \frac{\lambda_6 - \lambda_3}{T} \boldsymbol{J}_T \right).$$

We would also have

$$\Omega_{12} = \sum_{i=1}^{I} \Omega_{12i} = \frac{KN}{\lambda_3} \sum_{i=1}^{I} \boldsymbol{X}_i - \frac{KN(\lambda_6 - \lambda_3)}{T\lambda_3\lambda_6} U \mathbf{1}_T$$

$$\Omega_{22} = \sum_{i=1}^{I} \Omega_{22i} = \frac{KN}{\lambda_3} U - \frac{KN(\lambda_6 - \lambda_3)}{T\lambda_3\lambda_6} V.$$

Using block matrix inversion we can generate $\text{var}(\widehat{\delta}) = \left(\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}\right)^{-1}$ which gives us

$$\text{var}(\widehat{\delta}) = \frac{(\sigma^2/KN)IT\lambda_6\lambda_3}{(U^2 + ITU - TW - IV)\lambda_6 - (U^2 - IV)\lambda_3},$$

which is the same variance expression (6) derived in Web Appendix A using individual-level outcomes. Therefore, the cluster-period means approach is equivalent to using individual-level outcomes when cluster sizes are equal as expected. Since $\widetilde{\boldsymbol{R}}_i$ is cluster-specific when cluster sizes are unequal we cannot derive a closed-form expression. In this case, the proof for Lemma 3A below is sufficient.

Non-Gaussian outcomes

Under non-Gaussian outcomes we already utilize numerical methods using cluster-period means which can be easily modified to accommodate cluster-specific sizes. The covariance of the cluster-period means of the pseudo-outcomes becomes

$$\widetilde{\boldsymbol{V}}_i = (K_iN_i)^{-1}\boldsymbol{E}_i + (K_i^{-1}\sigma_\pi^2 + \sigma_s^2)\boldsymbol{I}_T + \{\sigma_b^2 + K_i^{-1}\sigma_c^2 + (K_iN_i)^{-1}\sigma_\gamma^2\}\boldsymbol{J}_T, \tag{18}$$

with the same $\boldsymbol{E}_i$ and variance components as defined in Web Appendix C. Under the individual-level outcomes approach with unequal cluster sizes $\boldsymbol{V}_i$ becomes

$$\boldsymbol{V}_i = (\mathbf{E}_i \otimes \boldsymbol{I}_{K_iN_i}) + \sigma_s^2(\boldsymbol{I}_T \otimes \boldsymbol{J}_{K_iN_i}) + \sigma_\pi^2(\boldsymbol{I}_T \otimes \boldsymbol{I}_{K_i} \otimes \boldsymbol{J}_{N_i}) + \sigma_b^2\boldsymbol{J}_{TK_iN_i} + \sigma_c^2(\boldsymbol{J}_T \otimes \boldsymbol{I}_{K_i} \otimes \boldsymbol{J}_{N_i}) + \sigma_\gamma^2(\boldsymbol{J}_T \otimes \boldsymbol{I}_{K_iN_i}),$$

which is a $TK_iN_i \times TK_iN_i$ matrix. These variance expressions will be used in Lemma 3A below.

LEMMA 3A: Equivalence of $\text{var}(\widehat{\delta})$ approaches under unequal cluster sizes

LEMMA 3A: The variance of the intervention effect estimator in the generalized linear mixed model under unequal cluster sizes is equivalently written as

$$\text{var}(\widehat{\delta}) = \left\{\sum_{i=1}^{I}\boldsymbol{X}_i^\top\widetilde{\boldsymbol{V}}_i^{-1}\boldsymbol{X}_i - \left(\sum_{i=1}^{I}\boldsymbol{X}_i^\top\widetilde{\boldsymbol{V}}_i^{-1}\right)\left(\sum_{i=1}^{I}\widetilde{\boldsymbol{V}}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{I}\widetilde{\boldsymbol{V}}_i^{-1}\boldsymbol{X}_i\right)\right\}^{-1},$$

where $\widetilde{\boldsymbol{V}}_i = (K_iN_i)^{-1}\boldsymbol{E}_i + (K_i^{-1}\sigma_\pi^2 + \sigma_s^2)\boldsymbol{I}_T + \{\sigma_b^2 + K_i^{-1}\sigma_c^2 + (K_iN_i)^{-1}\sigma_\gamma^2\}\boldsymbol{J}_T$ is the $T \times T$ matrix characterizing the covariance of the cluster-period means of the pseudo-outcomes $\overline{\boldsymbol{Y}}_i^* = (K_iN_i)^{-1}(\boldsymbol{I}_T \otimes \boldsymbol{1}_{K_iN_i}^\top)\boldsymbol{Y}_i^*$, and $\boldsymbol{E}_i = \text{diag}\left[\mathbb{E}\{\Delta_{i111}^{-1}\zeta(\mu_{i111})\Delta_{i111}^{-1}\}, \ldots, \mathbb{E}\{\Delta_{iT11}^{-1}\zeta(\mu_{iT11})\Delta_{iT11}^{-1}\}\right]$.

Proof: To prove Lemma 3A holds we need to show that the variance expression under individual-level outcomes is equivalent to the variance expression under a cluster-period means approach, that is that,

$$\begin{aligned}\text{var}(\widehat{\delta}) &= \left\{\sum_{i=1}^{I}\mathbf{H}_i^\top \boldsymbol{V}_i^{-1}\mathbf{H}_i - \left(\sum_{i=1}^{I}\mathbf{H}_i^\top \boldsymbol{V}_i^{-1}\right)\boldsymbol{F}_i\left(\sum_{i=1}^{I}\boldsymbol{F}_i^\top \boldsymbol{V}_i^{-1}\boldsymbol{F}_i\right)^{-1}\boldsymbol{F}_i^\top\left(\sum_{i=1}^{I}\boldsymbol{V}_i^{-1}\mathbf{H}_i\right)\right\}^{-1} \\ &= \left\{\sum_{i=1}^{I}\boldsymbol{X}_i^\top \widetilde{\boldsymbol{V}}_i^{-1}\boldsymbol{X}_i - \left(\sum_{i=1}^{I}\boldsymbol{X}_i^\top \widetilde{\boldsymbol{V}}_i^{-1}\right)\left(\sum_{i=1}^{I}\widetilde{\boldsymbol{V}}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{I}\widetilde{\boldsymbol{V}}_i^{-1}\boldsymbol{X}_i\right)\right\}^{-1}.\end{aligned}$$

where $\mathbf{H}_i = \boldsymbol{X}_i \otimes \mathbf{1}_{K_i N_i}$ and $\boldsymbol{F}_i = \boldsymbol{I}_T \otimes \mathbf{1}_{K_i N_i}$. We claim that the expressions for $\text{var}(\widehat{\delta})$ are equivalent if for each cluster i,

$$(K_i N_i)^{-2}\boldsymbol{F}_i^\top \boldsymbol{V}_i^{-1}\boldsymbol{F}_i = (\boldsymbol{F}_i^\top \boldsymbol{V}_i \boldsymbol{F}_i)^{-1}. \tag{19}$$

First, note that the covariance of Y_i^* can be written in the form $\boldsymbol{V}_i = (\boldsymbol{Q}_{i1} - \boldsymbol{Q}_{i2}) \otimes \boldsymbol{I}_{N_i} + \boldsymbol{Q}_{i2} \otimes \boldsymbol{J}_{N_i}$ with $\boldsymbol{Q}_{i1} - \boldsymbol{Q}_{i2} = (\boldsymbol{E}_i + \sigma_\gamma^2 \boldsymbol{J}_T) \otimes \boldsymbol{I}_{K_i}$ and $\boldsymbol{Q}_{2i} = (\sigma_\pi^2 \boldsymbol{I}_T + \sigma_c^2 \boldsymbol{J}_T) \otimes \boldsymbol{I}_{K_i} + (\sigma_s^2 \boldsymbol{I}_T + \sigma_b^2 \boldsymbol{J}_T) \otimes \boldsymbol{J}_{K_i}$. Therefore, the inverse can be obtained using formula (4) giving us

$$\boldsymbol{V}_i^{-1} = (\boldsymbol{Q}_{i1} - \boldsymbol{Q}_{i2})^{-1} \otimes \boldsymbol{I}_{N_i} + \frac{1}{N_i}\left[\{\boldsymbol{Q}_{i1} + (N_i - 1)\boldsymbol{Q}_{i2}\}^{-1} - (\boldsymbol{Q}_{i1} - \boldsymbol{Q}_{i2})^{-1}\right] \otimes \boldsymbol{J}_{N_i}.$$

Looking at each sub-inverse we have

$$\begin{aligned}(\boldsymbol{Q}_{i1} - \boldsymbol{Q}_{i2})^{-1} &= (\boldsymbol{E}_i + \sigma_\gamma^2 \boldsymbol{J}_T)^{-1} \otimes \boldsymbol{I}_{K_i} \\ &= \boldsymbol{A}_i^{-1} \otimes \boldsymbol{I}_{K_i}\end{aligned}$$

$$\{\boldsymbol{Q}_{i1} + (N_i - 1)\boldsymbol{Q}_{i2}\}^{-1} = [\{\boldsymbol{E}_i + N_i \sigma_\pi^2 \boldsymbol{I}_T + (\sigma_\gamma^2 + N_i \sigma_c^2)\boldsymbol{J}_T\} \otimes \boldsymbol{I}_{K_i} + N_i(\sigma_s^2 \boldsymbol{I}_T + \sigma_b^2 \boldsymbol{J}_T) \otimes \boldsymbol{J}_{K_i}]^{-1}.$$

We can easily rewrite $\boldsymbol{Q}_{i1} + (N_i - 1)\boldsymbol{Q}_{i2}$ in the form of $(\boldsymbol{W}_{i1} - \boldsymbol{W}_{i2}) \otimes \boldsymbol{I}_{K_i} + \boldsymbol{W}_{i2} \otimes \boldsymbol{J}_{K_i}$ with $\boldsymbol{W}_{i1} - \boldsymbol{W}_{i2} = \boldsymbol{E}_i + N_i \sigma_\pi^2 \boldsymbol{I}_T + (\sigma_\gamma^2 + N_i \sigma_c^2)\boldsymbol{J}_T$ and $\boldsymbol{W}_{i2} = N_i(\sigma_s^2 \boldsymbol{I}_T + \sigma_b^2 \boldsymbol{J}_T)$, therefore, to generate the inverse we can apply formula (4) again to get

$$\begin{aligned}\{\boldsymbol{Q}_{i1} + (N_i - 1)\boldsymbol{Q}_{i2}\}^{-1} &= (\boldsymbol{W}_{i1} - \boldsymbol{W}_{i2})^{-1} \otimes \boldsymbol{I}_{K_i} + \frac{1}{K_i}\left[\{\boldsymbol{W}_{i1} + (K_i - 1)\boldsymbol{W}_{i2}\}^{-1} - (\boldsymbol{W}_{i1} - \boldsymbol{W}_{i2})^{-1}\right] \otimes \boldsymbol{J}_{K_i} \\ &= \boldsymbol{B}_i^{-1} \otimes \boldsymbol{I}_{K_i} + \frac{1}{K_i}\left(\boldsymbol{C}_i^{-1} - \boldsymbol{B}_i^{-1}\right) \otimes \boldsymbol{J}_{K_i}.\end{aligned}$$

Putting these components together gives us

$$\boldsymbol{V}_i^{-1} = \boldsymbol{A}_i^{-1} \otimes \boldsymbol{I}_{K_i} \otimes \boldsymbol{I}_{N_i} + N_i^{-1}(\boldsymbol{B}_i^{-1} - \boldsymbol{A}_i^{-1}) \otimes \boldsymbol{I}_{K_i} \otimes \boldsymbol{J}_{N_i} + (K_i N_i)^{-1}(\boldsymbol{C}_i^{-1} - \boldsymbol{B}_i^{-1}) \otimes \boldsymbol{J}_{K_i} \otimes \boldsymbol{J}_{N_i}.$$

Now that we have an expression for the inverse we can calculate the following directly,

$$\boldsymbol{F}_i^\top \boldsymbol{V}_i^{-1}\boldsymbol{F}_i = K_iN_i\{\boldsymbol{E}_i + (N_i\sigma_\pi^2 + K_iN_i\sigma_s^2)\boldsymbol{I}_T + (\sigma_\gamma^2 + N_i\sigma_c^2 + K_iN_i\sigma_b^2)\boldsymbol{J}_T\}^{-1}.$$

Comparing this result to the following,

$$(\boldsymbol{F}_i^\top \boldsymbol{V}_i\boldsymbol{F}_i)^{-1} = (K_iN_i)^{-1}\{\boldsymbol{E}_i + (N_i\sigma_\pi^2 + K_iN_i\sigma_s^2)\boldsymbol{I}_T + (\sigma_\gamma^2 + N_i\sigma_c^2 + K_iN_i\sigma_b^2)\boldsymbol{J}_T\}^{-1},$$

we see that equation (19) holds. By definition, $\widetilde{\boldsymbol{V}}_i = (K_iN_i)^{-2}\boldsymbol{F}_i^\top\boldsymbol{V}_i\boldsymbol{F}_i$, therefore, by equation (19) we know that $\widetilde{\boldsymbol{V}}_i^{-1} = \boldsymbol{F}_i^\top\boldsymbol{V}_i^{-1}\boldsymbol{F}_i$. We can further show that

$$\begin{aligned}\mathbf{H}_i^\top \boldsymbol{V}_i^{-1}\mathbf{H}_i &= K_iN_i\boldsymbol{X}_i^\top\{\boldsymbol{E}_i + (N_i\sigma_\pi^2 + K_iN_i\sigma_s^2)\boldsymbol{I}_T + (\sigma_\gamma^2 + N_i\sigma_c^2 + K_iN_i\sigma_b^2)\boldsymbol{J}_T\}^{-1}\boldsymbol{X}_i \\ &= \boldsymbol{X}_i^\top\widetilde{\boldsymbol{V}}_i^{-1}\boldsymbol{X}_i\end{aligned}$$

and

$$\begin{aligned}\mathbf{H}_i^\top \boldsymbol{V}_i^{-1}\boldsymbol{F}_i &= K_iN_i\boldsymbol{X}_i^\top\{\boldsymbol{E}_i + (N_i\sigma_\pi^2 + K_iN_i\sigma_s^2)\boldsymbol{I}_T + (\sigma_\gamma^2 + N_i\sigma_c^2 + K_iN_i\sigma_b^2)\boldsymbol{J}_T\}^{-1} \\ &= \boldsymbol{X}_i^\top\widetilde{\boldsymbol{V}}_i^{-1}.\end{aligned}$$

Therefore, we have shown that,

$$\begin{aligned}\text{var}(\widehat{\delta}) &= \left\{\sum_{i=1}^I \mathbf{H}_i^\top\boldsymbol{V}_i^{-1}\mathbf{H}_i - \left(\sum_{i=1}^I \mathbf{H}_i^\top\boldsymbol{V}_i^{-1}\right)\boldsymbol{F}_i\left(\sum_{i=1}^I \boldsymbol{F}_i^\top\boldsymbol{V}_i^{-1}\boldsymbol{F}_i\right)^{-1}\boldsymbol{F}_i^\top\left(\sum_{i=1}^I \boldsymbol{V}_i^{-1}\mathbf{H}_i\right)\right\}^{-1} \\ &= \left\{\sum_{i=1}^I \boldsymbol{X}_i^\top\widetilde{\boldsymbol{V}}_i^{-1}\boldsymbol{X}_i - \left(\sum_{i=1}^I \boldsymbol{X}_i^\top\widetilde{\boldsymbol{V}}_i^{-1}\right)\left(\sum_{i=1}^I \widetilde{\boldsymbol{V}}_i^{-1}\right)^{-1}\left(\sum_{i=1}^I \widetilde{\boldsymbol{V}}_i^{-1}\boldsymbol{X}_i\right)\right\}^{-1}.\end{aligned}$$

Additional details regarding the procedure for deriving var$(\widehat{\delta})$ under unequal cluster sizes

We provide the specific steps used in deriving var$(\widehat{\delta})$ under unequal cluster sizes.

1. *Specify inputs*: Specify the number of clusters (I), number of periods (T), average number of subclusters per cluster ($\overline{K}$), average number of subjects per subcluster ($\overline{N}$), coefficient of variation for the number of subclusters (CV_K) and number of subjects (CV_N), ICCs (α_0,α_1,α_2,ρ_0,ρ_1) or equivalently the variance parameters ($\sigma_b^2, \sigma_c^2, \sigma_s^2, \sigma_\pi^2, \sigma_\gamma^2$), effect size on the link function scale (δ), and period effects on the link function scale (β_j).

2. *Simulate cluster sizes*: For each cluster $i \in \{1, \ldots, I\}$, simulate the number of subclusters (K_i) and subjects per subcluster (N_i) such that the average number of subclusters per cluster is $\overline{K}$ and average number of

subjects per subcluster is $\overline{N}$. In particular, we can assume $K_i \sim f_K(K_i; \overline{K}, \text{CV}_K)$ and $N_i \sim f_N(N_i; \overline{N}, \text{CV}_N)$, where $f(\bullet; a, b)$ represents a valid density or mass function with mean a and coefficient of variation b. Any distribution could be designated for f_K and f_N. For example, we could assume $K_i \sim \text{Gamma}(\text{shape} = \text{CV}_K^{-2}, \text{rate} = \overline{K}^{-1}\text{CV}_K^{-2})$ and $N_i \sim \text{Gamma}(\text{shape} = \text{CV}_N^{-2}, \text{rate} = \overline{N}^{-1}\text{CV}_N^{-2})$ with K_i and N_i rounded to the nearest integers in each draw. Each simulation replicate, $\mathcal{O} = \{(K_i, N_i), i = 1, \ldots, I\}$, represents a specific design with unequal cluster sizes. Repeat this step R (i.e. $R = 1000$) times and record each replicate using $\mathcal{O}^{(r)}$ for $r \in \{1, \ldots, R\}$.

3. *Generate var*$(\widehat{\delta}|\mathcal{O}^{(r)})$ *for each replicate* $\mathcal{O}^{(r)}$: Using a cluster-period means approach (Lemma 3A) we generate $\text{var}(\widehat{\delta}|\mathcal{O}^{(r)})$ for each replicate $\mathcal{O}^{(r)}$. Under a Gaussian outcome, we generate $\widetilde{\boldsymbol{R}}_i^{-1}$ (17) and use this with $\sigma^2(\sum_{i=1}^{I} \boldsymbol{Z}_{2i}^\top \widetilde{\boldsymbol{R}}_i^{-1} \boldsymbol{Z}_{2i})^{-1}$ and $\boldsymbol{Z}_{2i} = (\boldsymbol{I}_T, \boldsymbol{X}_i)$ to generate $\text{var}(\widehat{\delta}|\mathcal{O}^{(r)})$ which is the $(T+1, T+1)$th element. Under a non-Gaussian outcome, we generate $\widetilde{\boldsymbol{V}}_i$ (18) and use $\phi(\sum_{i=1}^{I} \boldsymbol{Z}_{2i}^\top \widetilde{\boldsymbol{V}}_i^{-1} \boldsymbol{Z}_{2i})^{-1}$ with $\boldsymbol{Z}_{2i} = (\boldsymbol{I}_T, \boldsymbol{X}_i)$ to generate $\text{var}(\widehat{\delta}|\mathcal{O}^{(r)})$ which is the $(T+1, T+1)$th element.

4. *Generate var*$(\widehat{\delta})$: Calculate $\text{var}(\widehat{\delta})$ by taking the average over all $\text{var}(\widehat{\delta}|\mathcal{O}^{(r)})$, $\text{var}(\widehat{\delta}) \approx R^{-1} \sum_{r=1}^{R} \text{var}(\widehat{\delta}|\mathcal{O}^{(r)})$.

$\text{Var}(\widehat{\delta})$ can then be used with the power formula to aid in sample size calculations of SW-CRTs with subclusters and unequal cluster sizes.

Web Appendix F

Derivation of variance components using ICCs under each design variant

Investigators generally have a better idea of what ICCs to expect in a trial rather than individual variance components. Here, we derive the variance components given the ICCs under each design variant.

Closed-cohort on subcluster and individual levels (design A)

Under this design we have the following ICCs,

$$\begin{aligned}
\alpha_0 &= \frac{\sigma_b^2 + \sigma_c^2 + \sigma_s^2 + \sigma_\pi^2}{\sigma_b^2 + \sigma_c^2 + \sigma_s^2 + \sigma_\pi^2 + \sigma_\gamma^2 + \sigma_\epsilon^2} \\
\rho_0 &= \frac{\sigma_b^2 + \sigma_s^2}{\sigma_b^2 + \sigma_c^2 + \sigma_s^2 + \sigma_\pi^2 + \sigma_\gamma^2 + \sigma_\epsilon^2} \\
\alpha_1 &= \frac{\sigma_b^2 + \sigma_c^2}{\sigma_b^2 + \sigma_c^2 + \sigma_s^2 + \sigma_\pi^2 + \sigma_\gamma^2 + \sigma_\epsilon^2} \\
\rho_1 &= \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2 + \sigma_s^2 + \sigma_\pi^2 + \sigma_\gamma^2 + \sigma_\epsilon^2} \\
\alpha_2 &= \frac{\sigma_b^2 + \sigma_c^2 + \sigma_\gamma^2}{\sigma_b^2 + \sigma_c^2 + \sigma_s^2 + \sigma_\pi^2 + \sigma_\gamma^2 + \sigma_\epsilon^2}.
\end{aligned}$$

Note that $\alpha_0 + \alpha_2 - \alpha_1 = (\sigma_b^2 + \sigma_c^2 + \sigma_s^2 + \sigma_\pi^2 + \sigma_\gamma^2)/(\sigma_b^2 + \sigma_c^2 + \sigma_s^2 + \sigma_\pi^2 + \sigma_\gamma^2 + \sigma_\epsilon^2)$. Assuming σ_ϵ^2 is known, we can solve for the total variance of the random effects, $x = \sigma_b^2 + \sigma_c^2 + \sigma_s^2 + \sigma_\pi^2 + \sigma_\gamma^2$,

$$\alpha_0 + \alpha_2 - \alpha_1 = \frac{x}{x + \sigma_\epsilon^2}$$
$$\Rightarrow x = \frac{\sigma_\epsilon^2}{1 - \alpha_0 - \alpha_2 + \alpha_1}.$$

We can use this to solve for σ_b^2,

$$\rho_1 = \frac{\sigma_b^2(1 - \alpha_0 - \alpha_2 + \alpha_1)}{\sigma_\epsilon^2}$$
$$\Rightarrow \sigma_b^2 = \frac{\rho_1 \sigma_\epsilon^2}{1 - \alpha_0 - \alpha_2 + \alpha_1}.$$

Now we can solve for σ_s^2 and σ_c^2,

$$\rho_0 = \frac{(1 - \alpha_0 - \alpha_2 + \alpha_1)\left(\frac{\rho_1 \sigma_\epsilon^2}{1 - \alpha_0 - \alpha_2 + \alpha_1} + \sigma_s^2\right)}{\sigma_\epsilon^2}$$
$$\Rightarrow \sigma_s^2 = \frac{\sigma_\epsilon^2(\rho_0 - \rho_1)}{1 - \alpha_0 - \alpha_2 + \alpha_1},$$

$$\alpha_1 = \frac{(1 - \alpha_0 - \alpha_2 + \alpha_1)\left(\frac{\rho_1 \sigma_\epsilon^2}{1 - \alpha_0 - \alpha_2 + \alpha_1} + \sigma_c^2\right)}{\sigma_\epsilon^2}$$
$$\Rightarrow \sigma_c^2 = \frac{\sigma_\epsilon^2(\alpha_1 - \rho_1)}{1 - \alpha_0 - \alpha_2 + \alpha_1}.$$

Now we can generate σ_π^2 and σ_γ^2,

$$\alpha_0 = \frac{(1 - \alpha_0 - \alpha_2 + \alpha_1)\left\{\frac{\rho_1 \sigma_\epsilon^2}{1 - \alpha_0 - \alpha_2 + \alpha_1} + \frac{\sigma_\epsilon^2(\alpha_1 - \rho_1)}{1 - \alpha_0 - \alpha_2 + \alpha_1} + \frac{\sigma_\epsilon^2(\rho_0 - \rho_1)}{1 - \alpha_0 - \alpha_2 + \alpha_1} + \sigma_\pi^2\right\}}{\sigma_\epsilon^2}$$
$$\Rightarrow \sigma_\pi^2 = \frac{\sigma_\epsilon^2(\alpha_0 - \alpha_1 - \rho_0 + \rho_1)}{1 - \alpha_0 - \alpha_2 + \alpha_1},$$

$$\alpha_2 = \frac{(1 - \alpha_0 - \alpha_2 + \alpha_1)\left\{\frac{\rho_1 \sigma_\epsilon^2}{1 - \alpha_0 - \alpha_2 + \alpha_1} + \frac{\sigma_\epsilon^2(\alpha_1 - \rho_1)}{1 - \alpha_0 - \alpha_2 + \alpha_1} + \sigma_\gamma^2\right\}}{\sigma_\epsilon^2}$$
$$\Rightarrow \sigma_\gamma^2 = \frac{\sigma_\epsilon^2(\alpha_2 - \alpha_1)}{1 - \alpha_0 - \alpha_2 + \alpha_1}.$$

Closed-cohort on subcluster level and cross-sectional at individual level (design B)

If we change the individual level from closed-cohort to cross-sectional ($\alpha_2 = \alpha_1$) then we would have the following,

$$\sigma_b^2 = \frac{\rho_1\sigma_\epsilon^2}{1-\alpha_0}$$
$$\sigma_s^2 = \frac{\sigma_\epsilon^2(\rho_0-\rho_1)}{1-\alpha_0}$$
$$\sigma_c^2 = \frac{\sigma_\epsilon^2(\alpha_1-\rho_1)}{1-\alpha_0}$$
$$\sigma_\pi^2 = \frac{\sigma_\epsilon^2(\alpha_0-\alpha_1-\rho_0+\rho_1)}{1-\alpha_0}.$$

Cross-sectional on subcluster and individual levels (design C)

If in addition we allow the subcluster level to be cross-sectional ($\alpha_2 = \alpha_1 = \rho_1$) then we would have the following,

$$\sigma_b^2 = \frac{\rho_1\sigma_\epsilon^2}{1-\alpha_0}$$
$$\sigma_s^2 = \frac{\sigma_\epsilon^2(\rho_0-\rho_1)}{1-\alpha_0}$$
$$\sigma_\pi^2 = \frac{\sigma_\epsilon^2(\alpha_0-\rho_0)}{1-\alpha_0}.$$

Web Appendix G

Design Effect

Derivation of design effect

An investigator could use the variance expression (5) combined with the power formula to determine the required sample size for a longitudinal CRT with subclusters. Alternatively, one could focus on the design effect of a longitudinal CRT with subclusters compared to an individually randomized trial to ease sample size calculations. The sample mean difference has a variance of $4\sigma^2/(IKN)$, therefore, the variance ratio (design effect) under a longitudinal CRT with subclusters to individual randomization is

$$\begin{aligned}\text{design effect} &= \frac{1}{4tr(\mathbf{\Omega})} \times \frac{T\lambda_6\lambda_3}{T\lambda_6 - \{1+(T-1)\tau_X\}(\lambda_6-\lambda_3)} \\ &= \frac{I^2T\lambda_6\lambda_3}{4(U^2+ITU-TW-IV)\lambda_6 - 4(U^2-IV)\lambda_3}. \qquad (20)\end{aligned}$$

This expression can be generalized to longitudinal parallel or any type of crossover design and we provide the stepped wedge design as an example. Similar to Woertman et al. (2013) and Li et al. (2018b), we assume an equal number of clusters cross over to intervention at each step ($m_s = m$) and an equal number of measurements are taken after each step ($c_s = c$) for each step $s = 1, ..., S$. We will then have $I = Sm$ clusters, $T = b + Sc$ periods (where b is the number of measures taken at baseline), and constants $U = \frac{1}{2}S(S+1)mc$, $W = \left(\frac{1}{3}S^3 + \frac{1}{2}S^2 + \frac{1}{6}S\right)m^2c$, and $V = \left(\frac{1}{3}S^3 + \frac{1}{2}S^2 + \frac{1}{6}S\right)mc^2$. Therefore, the design effect under a SW-CRT with subclusters design to individual randomization is

$$\text{design effect} = \frac{3}{2c(S-1/S)}\left\{\frac{(b+Sc)\lambda_3\lambda_6}{(Sc/2)\lambda_3 + (b+Sc/2)\lambda_6}\right\}. \tag{21}$$

Our design effect generalizes to the design effect derived in Li et al. (2018b), Woertman et al. (2013) based on the Hussey and Hughes (2007) model, and Hooper et al. (2016). To determine the total number of participants required under a SW-CRT with subclusters, an investigator could first generate the required sample size under individual randomization, multiply by the design effect (21), and then round up to the nearest integer or multiple of the number of clusters (I) for a balanced design. Given the total number of participants, the number of clusters (I), subclusters per cluster (K), and participants per subcluster (N) can be determined. Finally, investigators can explore the effect of each ICC on the multilevel SW-CRT design effect across design variants (A,B,C) using our RShiny app (Davis-Plourde, 2021).

Application to Lumbar Imaging with Reporting of Epidemiology (LIRE) trial

In our LIRE application we used the variance expression (5) coupled with the power formula to generate the required sample size per subcluster (N). Here, we use the design effect to calculate the total required number of subjects and clusters (I). The LIRE trial is a closed-cohort design on the subcluster level but cross-sectional design at the subject level (design B) SW-CRT containing $I = 100$ practices consisting of a total of 1700 primary care providers (PCPs) over $T = 6$ periods. Each practice is a cluster and each PCP represents a subcluster. While the number of PCPs per practice varied, we assume $K = 17$ PCPs per practice for illustration. The primary outcome was log-transformed spine-related relative value units (RVUs), a continuous composite measure of back pain. Assuming the median and total variance of RVU is approximately 3.56 and 2.5, respectively (Jarvik et al., 2020); a 5% reduction due to treatment corresponds to a standardized effect size of around -0.1. Based on preliminary data, an overall ICC was estimated to be 0.013 with a 95% confidence interval of (0.00, 0.046). We therefore assume the within-period within-PCP ICC to be the upper bound of the preliminary estimates, $\alpha_0 = 0.046$, and a slightly smaller within-period between-PCP ICC of $\rho_0 = 0.04$. Assuming a CAC of 0.5 further gives us $\alpha_1 = 0.023$ and $\rho_1 = 0.02$. We calculate the required number of subjects and clusters to achieve at least 80% power for a two-sided 5% test. Under individual randomization and assuming 98 degrees of freedom, a total of 9,860 participants are required to achieve 87.5% power. Assuming 77 participants are to be recruited at each of the 17 PCPs per practice gives us a design effect of 13.3, meaning 130,789 participants are required and 100 practices are needed.

Web Appendix H

LIRE trial: sample size determination under alternative cluster level designs

Closed-cohort design at both the subcluster and subject levels (design A)

In this example we treat the LIRE trial as a closed-cohort design on both the subcluster level and the subject level (design A). The study planned to randomize $I = 100$ practices consisting of a total of 1700 primary care providers (PCPs) over $T = 6$ periods; each practice is a cluster and each PCP represents a subcluster. While the number of PCPs per practice varied, we assume $K = 17$ PCPs per practice for illustration. The primary outcome was log-transformed spine-related relative value units (RVUs), a continuous composite measure of back pain. Assuming the median and total variance of RVU is approximately 3.56 and 2.5, respectively (Jarvik et al., 2020); a 5% reduction in median due to treatment corresponds to a standardized effect size of around -0.1. Based on preliminary data, an overall ICC was estimated to be 0.013 with a 95% confidence interval of (0.00, 0.046). We therefore assume the within-period within-PCP ICC to be the upper bound of the preliminary estimates, $\alpha_0 = 0.046$, and a slightly smaller within-period between-PCP ICC of $\rho_0 = 0.04$. Assuming a CAC of 0.5 further gives us $\alpha_1 = 0.023$ and $\rho_1 = 0.02$. We also set $\alpha_2 = 0.1$, and calculate the required number of subjects per PCP (N) to achieve at least 80% power for a two-sided 5% test. Based on the power formula and our closed-form variance expression (5), we found having $N = 72$ participants per PCP produced 87.5% power. To assess the sensitivity of our power calculation to ICC specifications, we looked at power trends for varying $\alpha_0 \in (0, 0.1)$ with various ratios of ρ_0/α_0 across varying levels of CAC (0.2, 0.5, 0.8) and $\alpha_2 = (0.1, 0.5)$. In concordance with our findings in Theorem 1, we found that higher values of within-period ICCs (α_0 and ρ_0) and lower values of between-period ICCs (α_1, ρ_1, and α_2) correspond to more conservative power predictions (Web Figure 3), thus we are confident that our ICC specifications have likely produced a conservative power estimate.

Cross-sectional design at both the subcluster and subject levels (design C)

In this example we treat the LIRE trial as a cross-sectional design on both the subcluster level and the subject level (design C). The study planned to randomize $I = 100$ practices consisting of a total of 1700 primary care providers (PCPs) over $T = 6$ periods; each practice is a cluster and each PCP represents a subcluster. While the number of PCPs per practice varied, we assume $K = 17$ PCPs per practice for illustration. The primary outcome was log-transformed spine-related relative value units (RVUs), a continuous composite measure of back pain. Assuming the median and total variance of RVU is approximately 3.56 and 2.5, respectively (Jarvik et al., 2020); a 5% reduction due to treatment corresponds to a standardized effect size of around -0.1. Based on preliminary data, an overall ICC was estimated to be 0.013 with a 95% confidence interval of (0.00, 0.046). We therefore assume the within-period within-PCP ICC to be the upper bound of the preliminary estimates, $\alpha_0 = 0.046$, and a slightly smaller within-period between-PCP ICC of $\rho_0 = 0.04$. Assuming a CAC of 0.5 further gives us $\rho_1 = 0.02$. We calculate the required number of subjects per PCP (N) to achieve at least 80% power for a two-sided 5% test. Based on the power formula and our closed-form variance expression (5), we found having $N = 99$ participants per PCP produced 87.5% power. To assess the sensitivity of our power calculation to ICC specifications, we looked at power trends for varying $\alpha_0 \in (0, 0.1)$ with various ratios of ρ_0/α_0 across varying levels of CAC (0.2, 0.5, 0.8). In concordance with our findings in Theorem 1, we found that higher values of within-period ICCs (α_0 and ρ_0) and lower values of between-period ICCs (ρ_1) correspond to more conservative power predictions (Web Figure 4), thus we are confident that our ICC specifications have likely produced a conservative power estimate.

Web Appendix I

LIRE trial: sample size determination adjusting for between-cluster size imbalances

In this example we consider a closed-cohort design at the subcluster level and cross-sectional design at the subject level (design B). The LIRE trial planned to randomize $I = 100$ practices consisting of a total of 1700 primary care providers (PCPs) over $T = 6$ periods; each practice is a cluster and each PCP represents a subcluster. The primary outcome was log-transformed spine-related relative value units (RVUs), a continuous composite measure of back pain. Assuming the median and total variance of RVU is approximately 3.56 and 2.5, respectively (Jarvik et al., 2020); a 5% reduction in median due to treatment corresponds to a standardized effect size of around -0.1. Based on preliminary data, an overall ICC was estimated to be 0.013 with a 95% confidence interval of (0.00, 0.046). We therefore assume the within-period within-PCP ICC to be the upper bound of the preliminary estimates, $\alpha_0 = 0.046$, and a slightly smaller within-period between-PCP ICC of $\rho_0 = 0.04$. Assuming a CAC of 0.5 further gives us $\alpha_1 = 0.023$ and $\rho_1 = 0.02$. To estimate the CVs, we first need to estimate the average number of PCPs and subjects per PCP over time within each practice, then we calculate the variance of the averages denoted by $\sigma^2_{cs.K}$ and $\sigma^2_{cs.N}$, and finally we generate our CV estimates using $\text{CV}_K = \sigma_{cs.K}/\overline{K}$ and $\text{CV}_N = \sigma_{cs.N}/\overline{N}$ (assuming cluster sizes follow a gamma distribution as outlined in Web Appendix E). Overall, the number of PCPs per practice varied from 2 to 106 with a specific breakdown of the number of PCPs per practice provided in stratified ranges in Table 2 of Jarvik et al. (2015). Using the provided Table, we assume the midpoint of each given range is the average number of PCPs per cluster which gives us $\overline{K} = 18$ and $\text{CV}_K = 1.0$. Unfortunately, a breakdown of the number of subjects per PCP was not provided. Therefore, we used the subject totals by site (Jarvik et al., 2020) with Table 2 (Jarvik et al., 2015) to estimate $\overline{N} = 126$ and $\text{CV}_N = 1.1$. We calculate the required number of practices (I) to achieve at least 80% power for a two-sided 5% test. Based on the power formula and our variance algorithm for unequal cluster sizes, we found having $I = 110$ practices produced 87.0% power. To assess the sensitivity of our predicted power to CV specification, we looked at power trends for varying $\text{CV}_K \in \{0.8, 0.9, 1.0, 1.1, 1.2\}$ and

$\mathrm{CV}_N \in \{0.9, 1.0, 1.1, 1.2, 1.3\}$. The results of our sensitivity analysis are shown in Web Table 11, we found that the predicted power varied between 83.1% and 89.1% which is still within our goal of at least 80% power. Further, there appears to be a monotone relationship between power and CVs. Specifically, an increase in CV (either CV_K or CV_N) is associated with a decrease in power.

Web Appendix J

EPT study: sample size determination under alternative cluster level designs

Closed-cohort design at both the subcluster and subject levels (design A)

In this example we treat the Washington State EPT study as a closed-cohort design on both the subcluster level and the subject level (design A). The study included $I = 24$ local health jurisdictions (LHJ) that were randomly assigned to intervention at one of four steps ($T = 5$). Each LHJ includes clinics that provide subject-level outcomes over time. A total of 219 clinics participated in chlamydia testing, but to be conservative we assume the number of clinics per LHJ is $K = 5$. In the design of this study, investigators aimed to detect a prevalence ratio of 0.7 and assumed a baseline prevalence of 0.05. Because the outcome is rare, we assume the effect size expressed as an odds ratio can be approximated by the prevalence ratio, and obtain the required number of participants per clinic (N) to achieve at least 80% power for a two-sided 5% test. Without distinguishing between clusters and subclusters, Li et al. (2021) estimated the within-period ICC to be 0.007 and the between-period ICC to be 0.004 based on marginal models. We consider these values to be the within-period between-clinic and between-period within-clinic ICCs such that $\rho_0 = 0.007$ and $\alpha_1 = 0.004$, and set the remaining ICCs to be $\alpha_0 = 0.008$ and $\rho_1 = 0.0035$ (corresponding to a CAC of 0.5). We also set $\alpha_2 = 0.2$ and assume a slightly decreasing time effect as in our simulations, and find based on the power formula and variance expressions that including $N = 66$ participants per clinic gives us 89.5% power. As a sensitivity analysis, we considered a larger decreasing period effect such that $\beta_j - \beta_{j+1} = 1 \times (0.5)^{j-1}$ for $j \geq 1$, which increased $N = 218$ to attain 89.5% power. On the other hand, using a smaller decreasing period effect, $\beta_j - \beta_{j+1} = 0.01 \times (0.5)^{j-1}$ for $j \geq 1$, reduced our required number of participants per clinic to $N = 59$ to achieve 89.6% power. Similarly, we assessed the sensitivity of our power calculation to ICC specifications by examining the power trends for varying $\alpha_0 \in (0, 0.05)$ with various ratios of ρ_0/α_0 across varying levels of CAC (0.2, 0.5, 0.8), α_2 (0.1, 0.5), and time trends ($\beta_j - \beta_{j+1} = \{0.01, 0.1, 1\} \times (0.5)^{j-1}$ for $j \geq 1$), however, the highest time trend is omitted since it produced contour plots that were less than 70% power regardless of α_0, ratio of ρ_0/α_0, CAC, and α_2 (Web Figure 5). We found that higher values of within-period ICCs (α_0 and ρ_0), lower values of between-period ICCs (α_1, ρ_1), and higher values of within-subject auto-correlation (α_2) correspond to more conservative power predictions, thus we are confident that our ICC specifications have likely produced a conservative power estimate.

Cross-sectional design at both the subcluster and subject levels (design C)

In this example we treat the Washington State EPT study as a cross-sectional design on both the subcluster level and the subject level (design C). The study included $I = 24$ local health jurisdictions (LHJ) that were randomly assigned to intervention at one of four steps ($T = 5$). Each LHJ includes clinics that provide subject-level outcomes over time. A total of 219 clinics participated in chlamydia testing, but to be conservative we assume the number of clinics per LHJ is $K = 5$. In the design of this study, investigators aimed to detect a prevalence ratio of 0.7 and assumed a baseline prevalence of 0.05. Because the outcome is rare, we assume the effect size expressed as an odds ratio can be approximated by the prevalence ratio, and obtain the required number of participants per clinic (N) to achieve at least 80% power for a two-sided 5% test. Without distinguishing between clusters and subclusters, Li et al. (2021) estimated the within-period ICC to be 0.007 based on marginal models. We consider this value to be the within-period between-clinic ICC such that $\rho_0 = 0.007$ and set the remaining ICCs to be $\alpha_0 = 0.008$ and $\rho_1 = 0.0035$ (corresponding to a CAC of 0.5). We assume a slightly decreasing time effect as in our simulations, and find based on the power formula and variance expressions that including $N = 42$ participants per clinic gives us 89.5% power. As a sensitivity analysis, we considered a larger decreasing period effect such that $\beta_j - \beta_{j+1} = 1 \times (0.5)^{j-1}$ for $j \geq 1$, which increased $N = 139$ to attain 89.5% power. On the other hand, using a smaller decreasing period effect, $\beta_j - \beta_{j+1} = 0.01 \times (0.5)^{j-1}$ for $j \geq 1$, reduced our required number of participants per clinic to $N = 37$ to achieve 89.3% power. Similarly, we assessed the sensitivity of our power calculation to ICC specifications by examining the power trends for varying $\alpha_0 \in (0, 0.05)$ with various ratios of ρ_0/α_0 across varying levels of CAC (0.2, 0.5, 0.8) and time trends ($\beta_j - \beta_{j+1} = \{0.01, 0.1, 1\} \times (0.5)^{j-1}$ for $j \geq 1$) (Web Figure 6). We found that higher values of within-period ICCs (α_0 and ρ_0) and lower values of between-period ICCs (ρ_1) correspond to more conservative power predictions, thus we are confident that our ICC specifications have likely produced a conservative power estimate.

Web Appendix K

EPT study: sample size determination adjusting for between-cluster size imbalances

In this example we consider a closed-cohort design at the subcluster level and cross-sectional design at the subject level (design B). The Washington State EPT study included $I = 24$ local health jurisdictions (LHJ) that were randomly assigned to intervention at one of four steps ($T = 5$). Each LHJ includes clinics that provide subject-level outcomes over time. A total of 219 clinics participated in chlamydia testing. The average number of clinics per LHJ is $\overline{K} = 4$ and the average number of subjects per clinic is $\overline{N} = 79$. To estimate the CVs, we first compute the average number of clinics and subjects per clinic over time within each LHJ, then we calculate the variance of the averages denoted by $\sigma^2_{cs.K}$ and $\sigma^2_{cs.N}$, and finally we generate our CV estimates using $\text{CV}_K = \sigma_{cs.K}/\overline{K}$ and $\text{CV}_N = \sigma_{cs.N}/\overline{N}$ (assuming cluster sizes follow a gamma distribution as outlined in Web Appendix E). Using this methodology we estimate $\text{CV}_K = 0.69$ and $\text{CV}_N = 0.79$. In the design of this study, investigators aimed to detect a prevalence ratio of 0.7 and assumed a baseline prevalence of 0.05. Because the outcome is rare, we assume the effect size expressed as an odds ratio can be approximated by the prevalence ratio, and obtain the required number of LHJs (I) to achieve at least 80% power for a two-sided 5% test. Without distinguishing between clusters and subclusters, Li et al. (2021) estimated the within-period ICC to be 0.007 and the between-period ICC to be 0.004 based on marginal models. We consider these values to be the within-period between-clinic and between-period within-clinic ICCs such that $\rho_0 = 0.007$ and $\alpha_1 = 0.004$, and set the remaining ICCs to be $\alpha_0 = 0.008$ and $\rho_1 = 0.0035$ (corresponding to a CAC of 0.5). We also assume a slightly decreasing time effect as in our simulations. Using the variance algorithm for unequal cluster sizes and power formula we found that including $I = 20$ clusters gives us 88.5% power. To assess the sensitivity of our predicted power to CV specification, we looked at power trends for varying $\text{CV}_K \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$ and $\text{CV}_N \in \{0.6, 0.7, 0.8, 0.9, 1.0\}$. The results of our sensitivity analysis are shown in Web Table 12, we found that the predicted power varied between 83.3% and 91.1% which is still within our goal of at least 80% power. Further, there appears to be a non-monotone relationship between the CVs and power. In some cases higher CVs are associated with higher power and in some cases higher CVs are associated with lower power. Understanding the complex relationship between the CVs and power is beyond the scope of this study and we leave this open for future exploration.

Web Appendix L

We extend the methodology used in Li et al. (2018b) for Gaussian and non-Gaussian outcomes under a GEE framework in SW-CRTs without subclusters. Let Y_{ijkl} be the outcome of interest for individual $l = 1, \ldots, N_{ijk}$ nested in subcluster $k = 1, \ldots, K_{ij}$ nested in cluster $i = 1, \ldots, I$ during period $j = 1, \ldots, T$. For simplicity, we assume $N_{ijk} = N$ and $K_{ij} = K$ for all i, j, and k. The mean model is given by

$$\tilde{\mu}_{ijkl} = g^{-1}(\tilde{\eta}_{ijkl}) = g^{-1}(\beta_j + \delta X_{ij}), \tag{22}$$

where g is a link function, β_j represents the categorical secular trend, X_{ij} is the intervention status for cluster i at period j (equal to 1 if exposed under intervention and 0 otherwise), and δ is the intervention effect of interest on the link function scale. We define the variance of the outcome as $\phi\zeta(\tilde{\mu}_{ijkl})$, where ϕ is a common dispersion and $\zeta(\tilde{\mu}_{ijkl})$ is the variance function. For example, the variance function of a binary outcome is parameterized as $\zeta(\tilde{\mu}_{ijkl}) = \tilde{\mu}_{ijkl}(1 - \tilde{\mu}_{ijkl})$. Without loss of generality, we assume $\phi = 1$ but the following procedure applies to arbitrary $\phi > 0$. We define the covariance matrix under the GEE framework as $\boldsymbol{M}_i = \boldsymbol{A}_i^{1/2}\boldsymbol{R}_i\boldsymbol{A}_i^{1/2}$ where $\boldsymbol{A}_i$ is a $TKN \times TKN$ diagonal matrix with elements $\zeta(\tilde{\mu}_{ijkl})$ and $\boldsymbol{R}_i$ is the $TKN \times TKN$ extended block exchangeable correlation matrix as shown under Gaussian outcomes. Using the model-based variance matrix, $\text{var}(\hat{\delta})$ is the lower-right corner element of $\phi\left(\sum_{i=1}^{I} \boldsymbol{D}_i^\top \boldsymbol{M}_i^{-1}\boldsymbol{D}_i\right)^{-1}$ where $\boldsymbol{D}_i = \partial g^{-1}(\tilde{\boldsymbol{\eta}}_i)/\partial\boldsymbol{\theta}^\top$ with $\tilde{\boldsymbol{\eta}}_i = \{\tilde{\eta}_{i111}, \ldots, \tilde{\eta}_{iTKN}\}$.

For example, if we have a binary outcome and our link function is the canonical logit, then $\boldsymbol{A}_i$ is a diagonal matrix with elements $\exp(\tilde{\eta}_{ijkl})/\{1+\exp(\tilde{\eta}_{ijkl})\}^2$ and $\boldsymbol{D}_i = \boldsymbol{A}_i\boldsymbol{Z}_i$ where $\boldsymbol{Z}_i = (\boldsymbol{I}_T, \boldsymbol{X}_i) \otimes \mathbf{1}_{KN}$ and $\boldsymbol{X}_i$ is the randomization schedule for cluster i. The model-based variance of our treatment effect, $\text{var}(\widehat{\delta})$, is then the lower-right corner element of

$$\begin{aligned}
&\phi\left(\sum_{i=1}^{I} \boldsymbol{D}_i^\top \boldsymbol{M}_i^{-1} \boldsymbol{D}_i\right)^{-1} \\
&= \phi\left(\sum_{i=1}^{I} \boldsymbol{Z}_i^\top \boldsymbol{A}_i \boldsymbol{A}_i^{-1/2} \boldsymbol{R}_i^{-1} \boldsymbol{A}_i^{-1/2} \boldsymbol{A}_i \boldsymbol{Z}_i\right)^{-1} \\
&= \phi\left(\sum_{i=1}^{I} \boldsymbol{Z}_i^\top \boldsymbol{W}_i \boldsymbol{Z}_i\right)^{-1},
\end{aligned}$$

where $\boldsymbol{W}_i = \boldsymbol{A}_i^{1/2}\boldsymbol{R}_i^{-1}\boldsymbol{A}_i^{1/2}$ which is equivalent to

$$\text{var}(\widehat{\delta}) = \phi\left\{\sum_{i=1}^{I} \mathbf{H}_i^\top \boldsymbol{W}_i \mathbf{H}_i - \left(\sum_{i=1}^{I} \mathbf{H}_i^\top \boldsymbol{W}_i\right) \boldsymbol{F} \left(\sum_{i=1}^{I} \boldsymbol{F}^\top \boldsymbol{W}_i \boldsymbol{F}\right)^{-1} \boldsymbol{F}^\top \left(\sum_{i=1}^{I} \boldsymbol{W}_i \mathbf{H}_i\right)\right\}^{-1},$$

where $\mathbf{H}_i = \boldsymbol{X}_i \otimes \mathbf{1}_{KN}$ and $\boldsymbol{F} = \boldsymbol{I}_T \otimes \mathbf{1}_{KN}$ and requires an algorithm to compute due to $\boldsymbol{W}_i$ being cluster-specific.

Using cluster-period means

We can simplify our GEE approach and decrease computation time by using cluster-period means (Tian et al., 2021). Under this approach, the model-based variance of our treatment effect, $\text{var}(\widehat{\delta})$, is the lower-right corner element of $\phi\left(\sum_{i=1}^{I} \overline{\boldsymbol{D}}_i^\top \widetilde{\boldsymbol{M}}_i^{-1} \overline{\boldsymbol{D}}_i\right)^{-1}$ where $\overline{\boldsymbol{D}}_i = \partial g^{-1}(\overline{\overline{\boldsymbol{\eta}}}_i)/\partial \boldsymbol{\theta}^\top$ with $\overline{\overline{\boldsymbol{\eta}}}_i = \{\overline{\overline{\eta}}_{i1}, \ldots, \overline{\overline{\eta}}_{iT}\}$ and $\widetilde{\boldsymbol{M}}_i$ is the covariance matrix for $\overline{\boldsymbol{Y}}_i = (\overline{Y}_{i1}, \ldots, \overline{Y}_{iT})^\top$. Here, $\widetilde{\boldsymbol{M}}_i = \widetilde{\boldsymbol{A}}_i^{1/2}\widetilde{\boldsymbol{R}}_i\widetilde{\boldsymbol{A}}_i^{1/2}$ where $\widetilde{\boldsymbol{A}}_i$ is a $T \times T$ diagonal matrix with elements $\zeta(\tilde{\mu}_{ij11})/(KN)$ and $\widetilde{\boldsymbol{R}}_i$ is a $T \times T$ working correlation matrix for the cluster-period means given by

$$\widetilde{\boldsymbol{R}}_i = \frac{1 - \alpha_2 + (N-1)(\alpha_0 - \alpha_1) + N(K-1)(\rho_0 - \rho_1)}{KN}\boldsymbol{I}_T + \frac{\alpha_2 + (N-1)\alpha_1 + N(K-1)\rho_1}{KN}\boldsymbol{J}_T,$$

which is an exchangeable matrix with diagonal elements $\{1 + (N-1)\alpha_0 + N(K-1)\rho_0\}/(KN)$ and off-diagonal elements $\{\alpha_2 + (N-1)\alpha_1 + N(K-1)\rho_1\}/(KN)$. Re-writing $\widetilde{\boldsymbol{R}}_i$ in terms of the eigenvalues (Web Table 1) and using equation (1) we have

$$\widetilde{\boldsymbol{R}}_i^{-1} = \frac{KN}{\lambda_3}\boldsymbol{I}_T - \frac{KN(\lambda_6 - \lambda_3)}{T\lambda_3\lambda_6}\boldsymbol{J}_T.$$

Therefore, $\text{var}(\widehat{\delta})$ will depend on the ICCs through two eigenvalues of the extended block exchangeable correlation matrix; the same two eigenvalues found under Gaussian outcomes.

For example, if we have a binary outcome and our link function is the canonical logit, then $\widetilde{\boldsymbol{A}}_i$ is a diagonal matrix with elements $\exp(\tilde{\eta}_{ij11})/[KN\{1+\exp(\tilde{\eta}_{ij11})\}^2]$ and $\overline{\boldsymbol{D}}_i = \widetilde{\boldsymbol{A}}_i \boldsymbol{Z}_{2i}$ with $\boldsymbol{Z}_{2i} = (\boldsymbol{I}_T, \boldsymbol{X}_i)$. The model-based variance of our treatment effect, $\text{var}(\widehat{\delta})$, is then the lower-right corner element of $\phi\left(\sum_{i=1}^{I} \boldsymbol{Z}_{2i}^\top \widetilde{\boldsymbol{W}}_i \boldsymbol{Z}_{2i}\right)^{-1}$ with $\widetilde{\boldsymbol{W}}_i = \widetilde{\boldsymbol{A}}_i^{1/2} \widetilde{\boldsymbol{R}}_i^{-1} \widetilde{\boldsymbol{A}}_i^{1/2}$ which is equal to

$$\text{var}(\widehat{\delta}) = \phi \left\{ \sum_{i=1}^{I} \boldsymbol{X}_i^\top \widetilde{\boldsymbol{W}}_i \boldsymbol{X}_i - \left(\sum_{i=1}^{I} \boldsymbol{X}_i^\top \widetilde{\boldsymbol{W}}_i \right) \left(\sum_{i=1}^{I} \widetilde{\boldsymbol{W}}_i \right)^{-1} \left(\sum_{i=1}^{I} \widetilde{\boldsymbol{W}}_i \boldsymbol{X}_i \right) \right\}^{-1} .$$

Web Appendix M: Web Figures

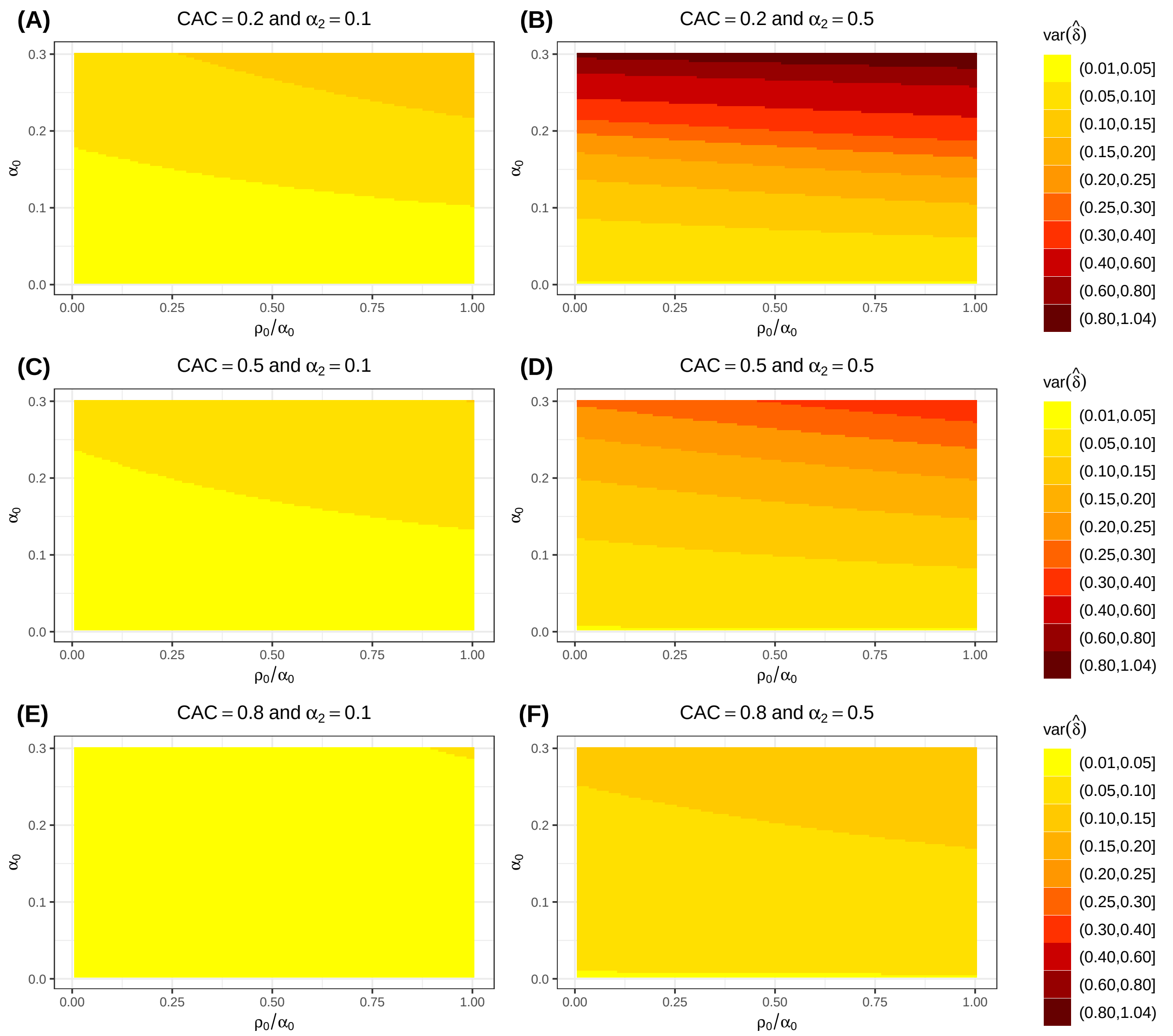

Figure 1: Contour plots illustrating the relationship between intracluster correlation coefficients (ICCs) and var($\hat{\delta}$) for binary outcomes with canonical logit link. ICCs include: within-period within-subcluster α_0; between-period within-subcluster α_1; within-period between-subcluster ρ_0; between-period between-subcluster ρ_1; and within-subject auto-correlation α_2. Various α_0 specifications are shown on the y-axis and various ρ_0 specifications are shown on the x-axis as a ratio of α_0. Between-period specifications are denoted by the cluster auto-correlation coefficient (CAC). Darker colors correspond to higher values of variance.

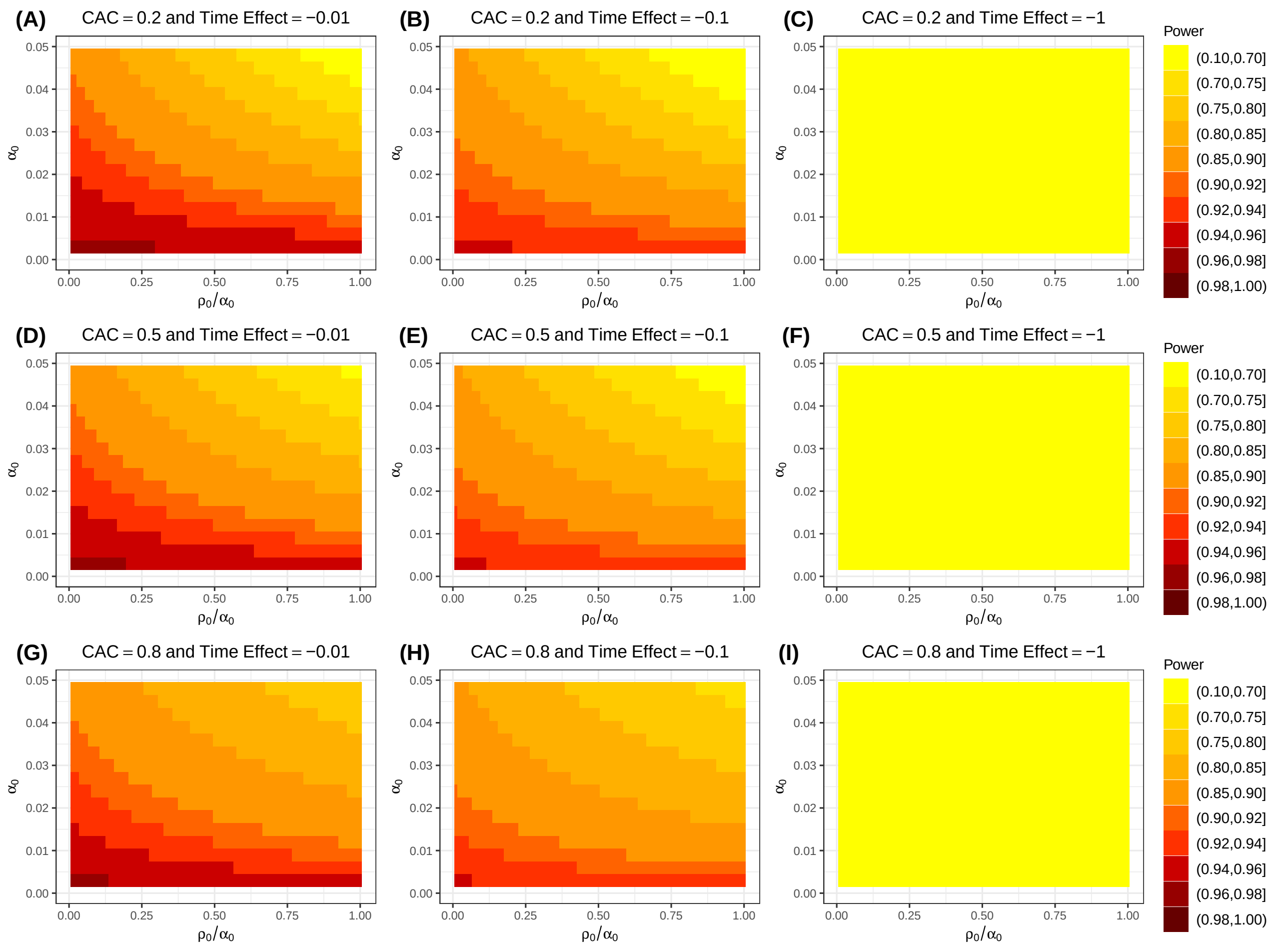

Figure 2: Contour plots illustrating the relationship between intracluster correlation coefficients (ICCs) and power in our application study of the Washington State Expedited Partner Therapy trial. ICCs include: within-period within-subcluster α_0; between-period within-subcluster α_1; within-period between-subcluster ρ_0; and between-period between-subcluster ρ_1. Various α_0 specifications are shown on the y-axis and various ρ_0 specifications are shown on the x-axis as a ratio of α_0. Between-period specifications are denoted by the cluster auto-correlation coefficient (CAC). Time effects $\{-0.01, -0.1, -1\}$ correspond to $\beta_j - \beta_{j+1} = \{0.01, 0.1, 1\} \times (0.5)^{j-1}$ for $j \geq 1$. Darker colors correspond to higher values of power.

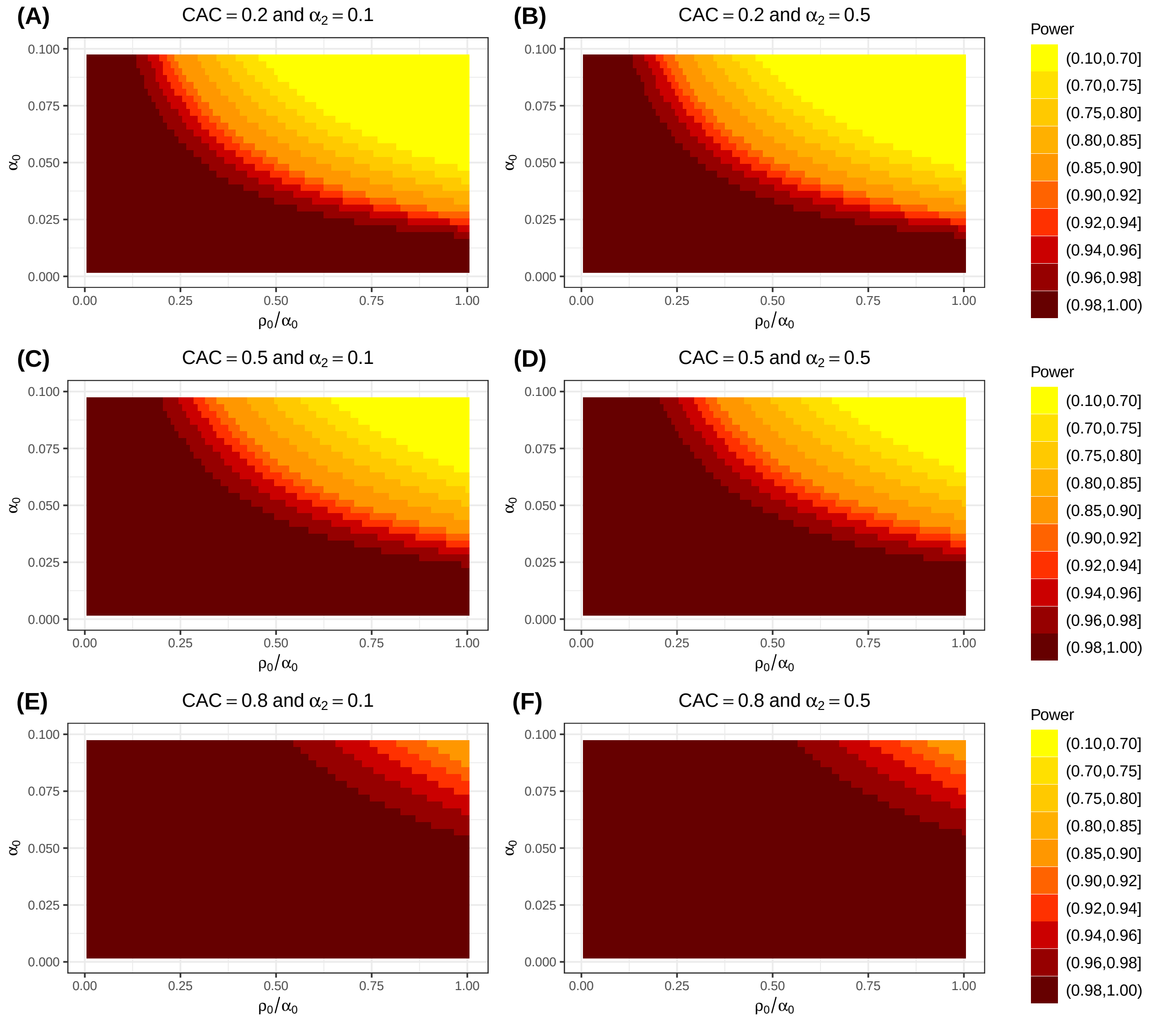

Figure 3: Contour plots illustrating the relationship between intracluster correlation coefficients (ICCs) and power in our application study of the Lumbar Imaging with Reporting of Epidemiology trial assuming a closed-cohort design at the subcluster and subject levels. ICCs include: within-period within-subcluster α_0; between-period within-subcluster α_1; within-period between-subcluster ρ_0; between-period between-subcluster ρ_1; and within-subject auto-correlation α_2. Various α_0 specifications are shown on the y-axis and various ρ_0 specifications are shown on the x-axis as a ratio of α_0. Between-period specifications are denoted by the cluster auto-correlation coefficient (CAC). Darker colors correspond to higher values of power.

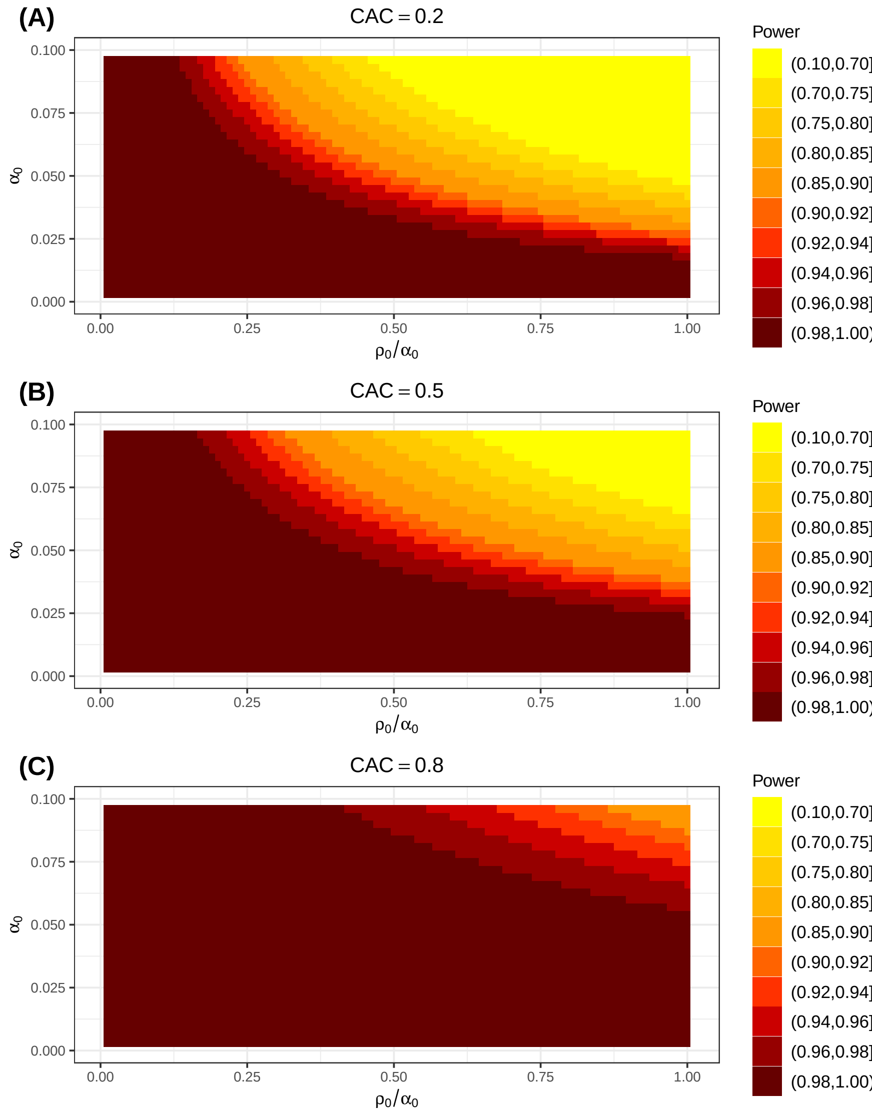

Figure 4: Contour plots illustrating the relationship between intracluster correlation coefficients (ICCs) and power in our application study of the Lumbar Imaging with Reporting of Epidemiology trial assuming a cross-sectional design at the subcluster and subject levels. ICCs include: within-period within-subcluster α_0; within-period between-subcluster ρ_0; and between-period between-subcluster ρ_1. Various α_0 specifications are shown on the y-axis and various ρ_0 specifications are shown on the x-axis as a ratio of α_0. Between-period specifications are denoted by the cluster auto-correlation coefficient (CAC). Darker colors correspond to higher values of power.

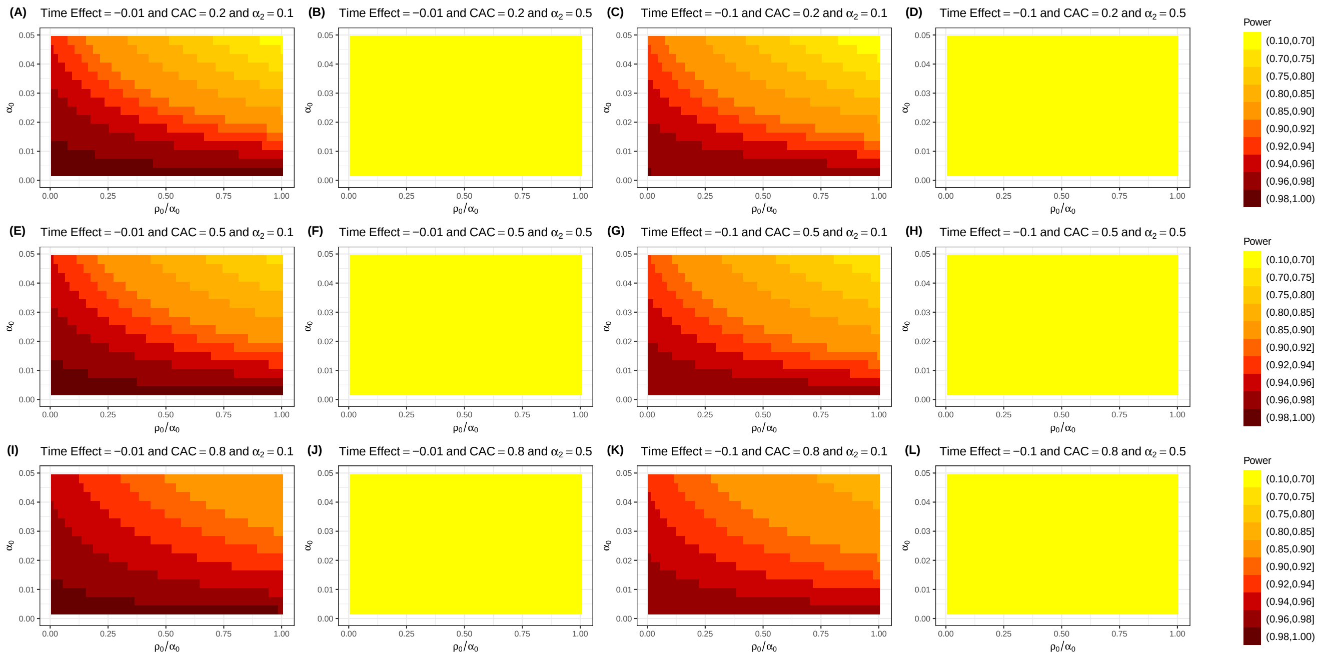

Figure 5: Contour plots illustrating the relationship between intracluster correlation coefficients (ICCs) and power in our application study of the Washington State Expedited Partner Therapy trial assuming closed-cohort design at the subcluster and subject levels. ICCs include: within-period within-subcluster α_0; between-period within-subcluster α_1; within-period between-subcluster ρ_0; between-period between-subcluster ρ_1; and within-subject auto-correlation α_2. Various α_0 specifications are shown on the y-axis and various ρ_0 specifications are shown on the x-axis as a ratio of α_0. Between-period specifications are denoted by the cluster auto-correlation coefficient (CAC). Time effects $\{-0.01, -0.1\}$ correspond to $\beta_j - \beta_{j+1} = \{0.01, 0.1\} \times (0.5)^{j-1}$ for $j \geq 1$. Darker colors correspond to higher values of power.

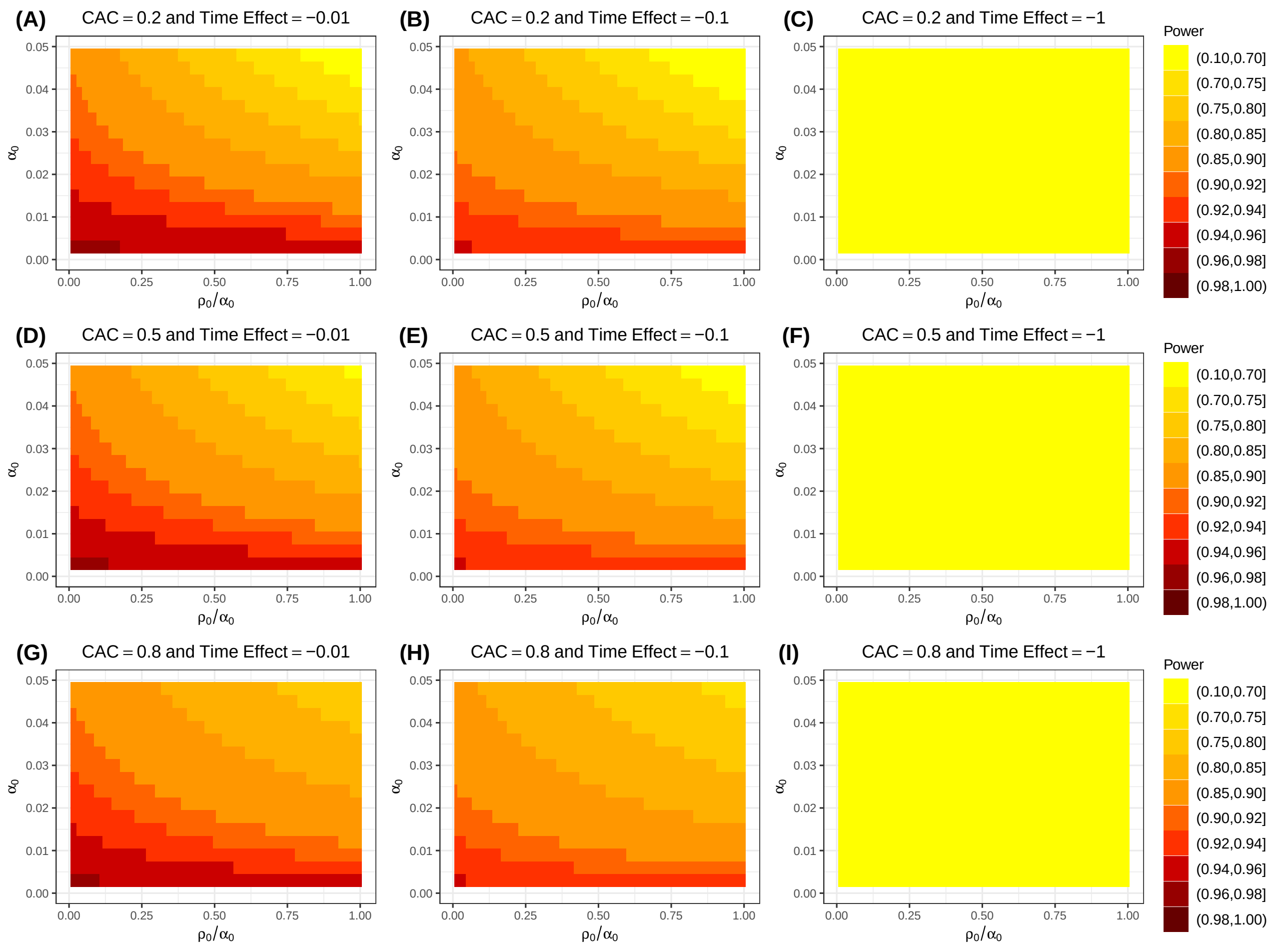

Figure 6: Contour plots illustrating the relationship between intracluster correlation coefficients (ICCs) and power in our application study of the Washington State Expedited Partner Therapy trial assuming a cross-sectional design at subcluster and subject levels. ICCs include: within-period within-subcluster α_0; within-period between-subcluster ρ_0; and between-period between-subcluster ρ_1. Various α_0 specifications are shown on the y-axis and various ρ_0 specifications are shown on the x-axis as a ratio of α_0. Between-period specifications are denoted by the cluster auto-correlation coefficient (CAC). Time effects $\{-0.01, -0.1, -1\}$ correspond to $\beta_j - \beta_{j+1} = \{0.01, 0.1, 1\} \times (0.5)^{j-1}$ for $j \geq 1$. Darker colors correspond to higher values of power.

Web Appendix N: Web Tables

Table 1: Eigenvalues (λ) expressed as functions of intracluster correlation coefficients (ICCs): within-period within-subcluster (α_0), between-period within-subcluster (α_1), within-subject auto-correlation (α_2), within-period between-subcluster (ρ_0), and between-period between-subcluster (ρ_1) under each design variant: (A) Closed-cohort design at both the subcluster and subject levels, (B) Closed-cohort design on the subcluster level but a cross-sectional design at the subject level, and (C) Cross-sectional design at both the subcluster and subject level with T periods, K subclusters per cluster, and N subjects per subcluster.

	Multiplicity	Eigenvalue expression by design variant
λ_1	$(T-1)K(N-1)$	(A) $1-\alpha_0-\alpha_2+\alpha_1$
	$TK(N-1)$	(B) $1-\alpha_0$
		(C) $1-\alpha_0$
λ_2	$(T-1)(K-1)$	(A) $1-\alpha_0-\alpha_2+\alpha_1+N(\alpha_0-\alpha_1-\rho_0+\rho_1)$
		(B) $1-\alpha_0+N(\alpha_0-\alpha_1-\rho_0+\rho_1)$
	$T(K-1)$	(C) $1-\alpha_0+N(\alpha_0-\rho_0)$
λ_3	$T-1$	(A) $1-\alpha_0-\alpha_2+\alpha_1+N\{\alpha_0-\alpha_1+(K-1)(\rho_0-\rho_1)\}$
		(B) $1-\alpha_0+N\{\alpha_0-\alpha_1+(K-1)(\rho_0-\rho_1)\}$
		(C) $1-\alpha_0+N\{\alpha_0+(K-1)\rho_0-K\rho_1\}$
λ_4	$K(N-1)$	(A) $1-\alpha_0+(T-1)(\alpha_2-\alpha_1)$
		(B) $1-\alpha_0=\lambda_1$
		(C) $1-\alpha_0=\lambda_1$
λ_5	$K-1$	(A) $1-\alpha_0+(T-1)(\alpha_2-\alpha_1)+N\{\alpha_0-\rho_0+(T-1)(\alpha_1-\rho_1)\}$
		(B) $1-\alpha_0+N\{\alpha_0-\rho_0+(T-1)(\alpha_1-\rho_1)\}$
		(C) $1-\alpha_0+N(\alpha_0-\rho_0)=\lambda_2$
λ_6	1	(A) $1-\alpha_0+(T-1)(\alpha_2-\alpha_1)+N[\alpha_0+(T-1)\alpha_1+(K-1)\{\rho_0+(T-1)\rho_1\}]$
		(B) $1-\alpha_0+N[\alpha_0+(T-1)\alpha_1+(K-1)\{\rho_0+(T-1)\rho_1\}]$
		(C) $1-\alpha_0+N\{\alpha_0+(K-1)\rho_0+(T-1)K\rho_1\}$

Table 2: Design constants under stepped wedge (SW), parallel longitudinal (Parallel), and repeated crossover (CXO) cluster randomized trial designs for a given number of periods (T).

	SW		Parallel		CXO	
T	$tr(\Omega)$	τ_X	$tr(\Omega)$	τ_X	$tr(\Omega)$	τ_X
4	0.44	0.17	0.89	1.00	0.89	-0.33
5	0.63	0.25	1.25	1.00	1.25	-0.20
6	0.80	0.30	1.44	1.00	1.44	-0.20
7	0.97	0.33	1.75	1.00	1.75	-0.14

Table 3: Predicted power (Predicted) obtained from sample size formula (correctly specifying CAC=0.5) compared to predicted power (Naive) assuming equal within- and between-period ICCs (incorrectly assuming CAC=1) for given effect size δ/σ, number of clusters I, subclusters per cluster K, participants per subcluster N, periods T, within-period intracluster correlations for within- and between-subcluster (α_0, ρ_0), and between-period intracluster correlations for within- and between-subcluster (α_1, ρ_1) assuming a cluster autocorrelation of 0.5, when outcome is Gaussian.

δ/σ	(α_0, ρ_0)	(α_1, ρ_1)	I	K	N	T	Predicted	Naive
0.1	(0.03, 0.0075)	(0.015, 0.00375)	24	6	15	7	85.3	93.9
	(0.01, 0.0025)	(0.005, 0.00125)	30	6	15	4	82.2	84.5
			24	5	10	7	81.4	81.0
0.2	(0.1, 0.025)	(0.05, 0.0125)	24	6	10	4	83.3	98.5
			18	3	12	7	81.8	97.0
	(0.03, 0.0075)	(0.015, 0.00375)	18	3	15	4	80.0	86.8
			15	3	10	6	80.8	84.2
	(0.01, 0.0025)	(0.005, 0.00125)	12	6	10	4	82.6	83.5
			10	4	10	6	80.0	79.7
0.25	(0.1, 0.025)	(0.05, 0.0125)	21	4	10	4	84.6	97.6
			18	2	10	7	83.5	95.1
	(0.03, 0.0075)	(0.015, 0.00375)	15	4	8	4	81.4	84.9
			12	2	10	7	80.2	82.5
	(0.01, 0.0025)	(0.005, 0.00125)	24	2	8	4	84.3	84.2
			10	3	9	6	83.6	82.9
0.35	(0.1, 0.025)	(0.05, 0.0125)	12	4	9	4	83.2	96.4
			10	3	8	6	82.9	94.3
	(0.03, 0.0075)	(0.015, 0.00375)	9	3	12	4	83.5	88.1
			16	2	5	5	84.0	84.2
	(0.01, 0.0025)	(0.005, 0.00125)	9	3	9	4	82.9	83.0
			8	3	7	5	80.0	79.5
0.4	(0.1, 0.025)	(0.05, 0.0125)	18	2	7	4	86.2	93.9
			12	2	8	5	82.0	92.2
	(0.03, 0.0075)	(0.015, 0.00375)	9	3	8	4	82.5	85.1
			8	3	7	5	83.5	85.0
	(0.01, 0.0025)	(0.005, 0.00125)	15	2	5	4	83.3	83.0
			12	2	5	5	85.1	84.6
0.5	(0.1, 0.025)	(0.05, 0.0125)	12	2	7	4	84.7	92.9
			12	2	4	5	82.5	87.0
	(0.03, 0.0075)	(0.015, 0.00375)	9	2	8	4	85.4	87.1

Table 4: Predicted power (Predicted) obtained from sample size formula (correctly specifying CAC=0.5) compared to predicted power (Naive) assuming equal within- and between-period ICCs (incorrectly assuming CAC=1) for given effect size $\exp(\delta)$, number of clusters I, subclusters per cluster K, participants per subcluster N, periods T, within-period intracluster correlations for within- and between-subcluster (α_0, ρ_0), and between-period intracluster correlations for within- and between-subcluster (α_1, ρ_1) assuming a cluster autocorrelation of 0.5, when outcome is binary with canonical logit link.

$\exp(\delta)$	(α_0, ρ_0)	(α_1, ρ_1)	I	K	N	T	Predicted	Naive
0.8	(0.03, 0.0075)	(0.015, 0.00375)	18	6	15	7	80.7	88.0
	(0.01, 0.0025)	(0.005, 0.00125)	27	6	15	4	84.2	85.4
			25	4	12	6	81.0	80.4
0.75	(0.1, 0.025)	(0.05, 0.0125)	25	6	15	6	82.8	98.6
			24	5	15	7	83.1	98.2
	(0.03, 0.0075)	(0.015, 0.00375)	27	5	12	4	80.6	85.4
			30	3	10	6	83.3	84.9
	(0.01, 0.0025)	(0.005, 0.00125)	21	6	10	4	80.5	80.9
			12	4	15	7	81.5	81.1
0.7	(0.1, 0.025)	(0.05, 0.0125)	30	5	14	4	82.3	97.4
			18	4	15	7	81.7	97.0
	(0.03, 0.0075)	(0.015, 0.00375)	18	6	10	4	80.6	85.1
			15	3	15	6	81.2	85.3
	(0.01, 0.0025)	(0.005, 0.00125)	18	4	12	4	82.3	82.6
			20	2	15	5	81.8	81.5
0.65	(0.1, 0.025)	(0.05, 0.0125)	21	6	12	4	83.6	97.6
			18	3	12	7	84.1	95.3
	(0.03, 0.0075)	(0.015, 0.00375)	24	3	10	4	85.0	87.1
			20	2	10	6	83.7	84.5
	(0.01, 0.0025)	(0.005, 0.00125)	15	4	10	4	82.7	82.8
			12	3	14	5	85.2	85.0
0.6	(0.1, 0.025)	(0.05, 0.0125)	18	5	10	4	82.3	95.1
			12	3	15	7	82.8	96.4
	(0.03, 0.0075)	(0.015, 0.00375)	16	2	12	5	83.9	85.8
			15	2	10	6	84.0	84.9
	(0.01, 0.0025)	(0.005, 0.00125)	21	2	10	4	85.5	85.3
			12	3	8	5	80.0	79.4
0.5	(0.1, 0.025)	(0.05, 0.0125)	15	3	10	4	83.2	93.3
			16	2	9	5	82.5	90.4
	(0.03, 0.0075)	(0.015, 0.00375)	15	2	9	4	84.1	85.3

Table 5: Estimated required number of clusters I, subclusters per cluster K, participants per subcluster N, periods T, empirical type I error (Test Size), empirical power (Empirical), and predicted power (Predicted) obtained from sample size formula for given effect size δ/σ, within-period intracluster correlations for within- and between-subcluster (α_0, ρ_0), and between-period intracluster correlations for within- and between-subcluster (α_1, ρ_1) assuming a cluster autocorrelation of 0.5, when outcome is Gaussian using (unrestricted) maximum likelihood (n=1000).

δ/σ	(α_0, ρ_0)	(α_1, ρ_1)	I	K	N	T	Test Size	Empirical	Predicted
0.1	(0.03, 0.0075)	(0.015, 0.00375)	24	6	15	7	3.8	89.2	85.3
	(0.01, 0.0025)	(0.005, 0.00125)	30	6	15	4	5.1	82.9	82.2
			24	5	10	7	4.6	83.5	81.4
0.2	(0.1, 0.025)	(0.05, 0.0125)	24	6	10	4	5.5	86.4	83.3
			18	3	12	7	4.1	85.2	81.8
	(0.03, 0.0075)	(0.015, 0.00375)	18	3	15	4	3.5	82.8	80.0
			15	3	10	6	4.6	82.7	80.8
	(0.01, 0.0025)	(0.005, 0.00125)	12	6	10	4	3.5	84.5	82.6
			10	4	10	6	2.9	81.9	80.0
0.25	(0.1, 0.025)	(0.05, 0.0125)	21	4	10	4	6.0	85.0	84.6
			18	2	10	7	5.2	84.6	83.5
	(0.03, 0.0075)	(0.015, 0.00375)	15	4	8	4	3.1	84.3	81.4
			12	2	10	7	4.1	82.8	80.2
	(0.01, 0.0025)	(0.005, 0.00125)	24	2	8	4	4.5	84.1	84.3
			10	3	9	6	2.0	85.9	83.6
0.35	(0.1, 0.025)	(0.05, 0.0125)	12	4	9	4	3.5	87.0	83.2
			10	3	8	6	2.6	86.5	82.9
	(0.03, 0.0075)	(0.015, 0.00375)	9	3	12	4	3.1	87.4	83.5
			16	2	5	5	3.3	88.2	84.0
	(0.01, 0.0025)	(0.005, 0.00125)	9	3	9	4	2.7	84.8	82.9
			8	3	7	5	2.2	81.5	80.0
0.4	(0.1, 0.025)	(0.05, 0.0125)	18	2	7	4	3.5	89.2	86.2
			12	2	8	5	4.1	83.2	82.0
	(0.03, 0.0075)	(0.015, 0.00375)	9	3	8	4	1.7	84.1	82.5
			8	3	7	5	1.7	87.1	83.5
	(0.01, 0.0025)	(0.005, 0.00125)	15	2	5	4	3.5	82.0	83.3
			12	2	5	5	1.9	87.6	85.1
0.5	(0.1, 0.025)	(0.05, 0.0125)	12	2	7	4	4.1	84.3	84.7
			12	2	4	5	3.7	86.8	82.5
	(0.03, 0.0075)	(0.015, 0.00375)	9	2	8	4	2.0	89.2	85.4

Table 6: Estimated required number of clusters I, average number of subclusters per cluster $\overline{K}$, average number of participants per subcluster $\overline{N}$, periods T, empirical type I error (Test Size), empirical power (Empirical), and predicted power (Predicted) obtained from sample size formula for given effect size δ/σ, within-period intracluster correlations for within- and between-subcluster (α_0, ρ_0), between-period intracluster correlations for within- and between-subcluster (α_1, ρ_1) assuming a cluster autocorrelation of 0.5, coefficient of variation on the subcluster level CV_K, and coefficient of variation on the subject level CV_N, when outcome is Gaussian (n=1000).

δ/σ	(α_0, ρ_0)	(α_1, ρ_1)	I	$\overline{K}$	$\overline{N}$	T	CV_K	CV_N	Test Size	Empirical	Predicted
0.1	(0.03, 0.0075)	(0.015, 0.00375)	24	6	15	7	0	0	3.6	88.2	85.3
							0	0.25	3.6	85.0	84.2
							0	0.5	4.0	83.2	82.0
							0	0.75	3.7	83.9	80.4
							0	1.0	5.3	78.0	78.8
							0.25	0	3.8	84.2	82.3
							0.25	0.25	5.0	82.4	80.9
							0.25	0.5	5.1	80.7	79.7
							0.25	0.75	3.4	79.3	78.4
							0.25	1.0	4.4	75.7	75.3
							0.5	0	4.6	80.3	81.1
							0.5	0.25	4.9	79.6	77.8
							0.5	0.5	4.1	77.1	78.8
							0.5	0.75	4.5	77.2	77.2
							0.5	1.0	3.7	76.9	74.5
0.2	(0.03, 0.0075)	(0.015, 0.00375)	15	3	10	6	0	0	4.0	81.1	80.8
							0	0.25	2.8	80.5	78.8
							0	0.5	2.9	76.2	77.2
							0	0.75	3.4	73.4	77.8
							0	1.0	3.1	74.6	75.7
							0.25	0	3.4	74.2	74.0
							0.25	0.25	4.1	72.5	72.5
							0.25	0.5	3.2	73.1	71.4
							0.25	0.75	4.1	69.8	72.7
							0.25	1.0	4.2	67.9	65.2
							0.5	0	3.9	76.8	76.8
							0.5	0.25	4.0	72.5	73.6
							0.5	0.5	3.6	73.7	76.4
							0.5	0.75	3.4	70.7	70.9
							0.5	1.0	2.6	72.2	70.3

Table 7: Estimated required number of clusters I, average number of subclusters per cluster $\overline{K}$, average number of participants per subcluster $\overline{N}$, periods T, empirical type I error (Test Size), empirical power (Empirical), and predicted power (Predicted) obtained from sample size formula for given effect size δ/σ, within-period intracluster correlations for within- and between-subcluster (α_0, ρ_0), between-period intracluster correlations for within- and between-subcluster (α_1, ρ_1) assuming a cluster autocorrelation of 0.5, coefficient of variation on the subcluster level CV_K, and coefficient of variation on the subject level CV_N, when outcome is Gaussian (n=1000). *(Continued)*

δ/σ	(α_0, ρ_0)	(α_1, ρ_1)	I	$\overline{K}$	$\overline{N}$	T	CV_K	CV_N	Test Size	Empirical	Predicted
							0	0	1.8	85.4	85.1
							0	0.25	2.8	81.8	82.3
							0	0.5	1.9	81.9	79.4
							0	0.75	2.3	83.3	85.5
							0	1.0	2.2	84.0	87.0
							0.25	0	2.2	86.1	85.1
							0.25	0.25	2.1	79.5	82.1
0.4	(0.01, 0.0025)	(0.005, 0.00125)	12	2	5	5	0.25	0.5	3.0	80.3	84.7
							0.25	0.75	2.6	82.9	82.3
							0.25	1.0	1.8	84.0	84.9
							0.5	0	1.7	88.1	87.6
							0.5	0.25	3.4	83.4	84.7
							0.5	0.5	2.5	83.2	84.5
							0.5	0.75	1.9	86.9	86.9
							0.5	1.0	2.4	86.8	88.7
							0	0	3.3	82.6	84.7
							0	0.25	2.8	82.6	83.0
							0	0.5	3.6	80.6	80.4
							0	0.75	3.2	81.4	80.8
							0	1.0	3.0	78.6	81.4
							0.25	0	2.5	85.3	86.0
							0.25	0.25	3.8	83.7	82.2
0.5	(0.1, 0.025)	(0.05, 0.0125)	12	2	7	4	0.25	0.5	2.5	83.8	80.7
							0.25	0.75	2.3	82.8	79.3
							0.25	1.0	3.4	80.4	82.3
							0.5	0	2.5	88.2	87.7
							0.5	0.25	2.6	86.8	86.5
							0.5	0.5	3.9	83.0	81.9
							0.5	0.75	3.7	83.8	83.3
							0.5	1.0	3.4	83.8	83.2

Table 8: Estimated required number of clusters I, subclusters per cluster K, participants per subcluster N, periods T, empirical type I error (Test Size), empirical power (Empirical), and predicted power (Predicted) obtained from sample size formula for given effect size $\exp(\delta)$, within-period intracluster correlations for within- and between-subcluster (α_0, ρ_0), and between-period intracluster correlations for within- and between-subcluster (α_1, ρ_1) assuming a cluster autocorrelation of 0.5, when outcome is binary with canonical logit link using penalized quasi-likelihood (n=1000).

$\exp(\delta)$	(α_0, ρ_0)	(α_1, ρ_1)	I	K	N	T	Test Size	Empirical	Predicted
0.8	(0.03, 0.0075)	(0.015, 0.00375)	18	6	15	7	4.4	82.1	80.7
	(0.01, 0.0025)	(0.005, 0.00125)	27	6	15	4	5.5	85.5	84.2
			25	4	12	6	2.8	81.9	81.0
0.75	(0.1, 0.025)	(0.05, 0.0125)	25	6	15	6	4.9	85.0	82.8
			24	5	15	7	4.5	87.0	83.1
	(0.03, 0.0075)	(0.015, 0.00375)	27	5	12	4	5.2	83.2	80.6
			30	3	10	6	5.0	83.8	83.3
	(0.01, 0.0025)	(0.005, 0.00125)	21	6	10	4	2.7	81.9	80.5
			12	4	15	7	2.4	85.1	81.5
0.7	(0.1, 0.025)	(0.05, 0.0125)	30	5	14	4	3.9	81.6	82.3
			18	4	15	7	4.0	86.1	81.7
	(0.03, 0.0075)	(0.015, 0.00375)	18	6	10	4	4.6	82.4	80.6
			15	3	15	6	4.5	83.3	81.2
	(0.01, 0.0025)	(0.005, 0.00125)	18	4	12	4	3.8	83.4	82.3
			20	2	15	5	4.1	81.8	81.8
0.65	(0.1, 0.025)	(0.05, 0.0125)	21	6	12	4	4.1	87.7	83.6
			18	3	12	7	3.8	87.0	84.1
	(0.03, 0.0075)	(0.015, 0.00375)	24	3	10	4	4.7	86.4	85.0
			20	2	10	6	3.2	86.1	83.7
	(0.01, 0.0025)	(0.005, 0.00125)	15	4	10	4	2.6	85.4	82.7
			12	3	14	5	3.2	85.9	85.2
0.6	(0.1, 0.025)	(0.05, 0.0125)	18	5	10	4	3.4	84.6	82.3
			12	3	15	7	4.6	87.6	82.8
	(0.03, 0.0075)	(0.015, 0.00375)	16	2	12	5	4.6	85.6	83.9
			15	2	10	6	3.8	86.9	84.0
	(0.01, 0.0025)	(0.005, 0.00125)	21	2	10	4	3.9	84.4	85.5
			12	3	8	5	2.5	81.9	80.0
0.5	(0.1, 0.025)	(0.05, 0.0125)	15	3	10	4	2.6	86.3	83.2
			16	2	9	5	3.3	83.5	82.5
	(0.03, 0.0075)	(0.015, 0.00375)	15	2	9	4	2.6	88.3	84.1

Table 9: Estimated required number of clusters I, average number of subclusters per cluster $\overline{K}$, average number of participants per subcluster $\overline{N}$, periods T, empirical type I error (Test Size), empirical power (Empirical), and predicted power (Predicted) obtained from sample size formula for given effect size $\exp(\delta)$, within-period intracluster correlations for within- and between-subcluster (α_0, ρ_0), between-period intracluster correlations for within- and between-subcluster (α_1, ρ_1) assuming a cluster autocorrelation of 0.5, coefficient of variation on the subcluster level CV_K, and coefficient of variation on the subject level CV_N, when outcome is binary with canonical logit link (n=1000).

$\exp(\delta)$	(α_0, ρ_0)	(α_1, ρ_1)	I	$\overline{K}$	$\overline{N}$	T	CV_K	CV_N	Test Size	Empirical	Predicted
							0	0	4.2	82.2	80.7
							0	0.25	4.2	81.9	79.2
							0	0.5	5.4	81.0	78.5
							0	0.75	4.8	80.7	75.7
							0	1.0	5.3	76.2	73.2
							0.25	0	4.0	77.6	77.1
							0.25	0.25	4.5	76.7	76.5
0.8	(0.03, 0.0075)	(0.015, 0.00375)	18	6	15	7	0.25	0.5	3.6	77.8	77.0
							0.25	0.75	4.9	76.9	73.9
							0.25	1.0	3.9	72.3	68.9
							0.5	0	5.3	80.8	75.5
							0.5	0.25	4.1	80.1	74.1
							0.5	0.5	5.4	74.6	75.0
							0.5	0.75	5.4	73.5	74.1
							0.5	1.0	4.4	74.4	67.2
							0	0	5.2	84.0	83.3
							0	0.25	3.8	81.8	80.9
							0	0.5	4.9	80.9	80.6
							0	0.75	4.9	80.9	78.3
							0	1.0	5.3	78.6	79.0
							0.25	0	3.2	79.8	77.2
							0.25	0.25	3.4	75.9	74.7
0.75	(0.03, 0.0075)	(0.015, 0.00375)	30	3	10	6	0.25	0.5	5.7	74.6	75.6
							0.25	0.75	5.9	72.3	73.1
							0.25	1.0	3.6	74.6	72.4
							0.5	0	4.4	82.4	80.4
							0.5	0.25	4.1	79.8	77.9
							0.5	0.5	4.9	79.7	74.1
							0.5	0.75	3.6	75.4	75.0
							0.5	1.0	4.1	75.7	75.1

Table 10: Estimated required number of clusters I, average number of subclusters per cluster $\overline{K}$, average number of participants per subcluster $\overline{N}$, periods T, empirical type I error (Test Size), empirical power (Empirical), and predicted power (Predicted) obtained from sample size formula for given effect size $\exp(\delta)$, within-period intracluster correlations for within- and between-subcluster (α_0, ρ_0), between-period intracluster correlations for within- and between-subcluster (α_1, ρ_1) assuming a cluster autocorrelation of 0.5, coefficient of variation on the subcluster level CV_K, and coefficient of variation on the subject level CV_N, when outcome is binary with canonical logit link (n=1000). *(Continued)*

$\exp(\delta)$	(α_0, ρ_0)	(α_1, ρ_1)	I	$\overline{K}$	$\overline{N}$	T	CV_K	CV_N	Test Size	Empirical	Predicted
							0	0	3.1	85.4	85.2
							0	0.25	3.8	85.5	83.8
							0	0.5	2.7	83.5	82.1
							0	0.75	2.4	81.8	80.7
							0	1.0	2.1	79.7	81.7
							0.25	0	1.8	79.5	80.2
							0.25	0.25	3.4	76.8	76.1
0.65	(0.01, 0.0025)	(0.005, 0.00125)	12	3	14	5	0.25	0.5	3.3	78.2	80.5
							0.25	0.75	3.8	74.7	74.1
							0.25	1.0	3.9	73.9	74.9
							0.5	0	4.1	82.6	82.1
							0.5	0.25	2.5	80.8	81.2
							0.5	0.5	2.5	79.1	80.1
							0.5	0.75	3.3	76.6	82.7
							0.5	1.0	2.1	75.5	79.4
							0	0	3.0	86.9	83.2
							0	0.25	5.1	83.8	81.2
							0	0.5	3.9	83.5	80.7
							0	0.75	2.8	83.0	78.5
							0	1.0	3.9	80.2	75.6
							0.25	0	4.4	80.9	78.1
							0.25	0.25	3.7	80.7	78.2
0.5	(0.1, 0.025)	(0.05, 0.0125)	15	3	10		0.25	0.5	3.6	75.7	75.0
							0.25	0.75	5.1	76.6	68.4
							0.25	1.0	4.3	73.8	74.7
						4	0.5	0	5.2	81.7	79.9
							0.5	0.25	3.7	81.6	77.9
							0.5	0.5	4.3	81.5	76.3
							0.5	0.75	3.4	75.4	77.7
							0.5	1.0	5.8	78.4	76.3

Table 11: Predicted power for the LIRE trial obtained from sample size formula for given between-cluster size imbalances measured by the coefficient of variation on the primary care provider level CV_K and subject level CV_N.

	Predicted Power				
	$CV_N = 0.9$	$CV_N = 1.0$	$CV_N = 1.1$	$CV_N = 1.2$	$CV_N = 1.3$
$CV_K = 0.8$	89.1	88.4	87.5	86.4	85.9
$CV_K = 0.9$	88.6	87.8	87.0	86.4	85.4
$CV_K = 1.0$	88.5	87.7	87.0	86.0	84.9
$CV_K = 1.1$	87.7	87.2	85.0	84.5	83.2
$CV_K = 1.2$	87.4	86.6	84.8	83.9	83.1

Table 12: Predicted power for the Washington EPT Study obtained from sample size formula for given between-cluster size imbalances measured by the coefficient of variation on the clinic level CV_K and subject level CV_N.

	Predicted Power				
	$CV_N = 0.6$	$CV_N = 0.7$	$CV_N = 0.8$	$CV_N = 0.9$	$CV_N = 1.0$
$CV_K = 0.5$	83.3	86.9	86.2	86.2	85.0
$CV_K = 0.6$	88.9	89.9	88.9	88.1	87.6
$CV_K = 0.7$	88.8	89.3	88.3	86.5	85.9
$CV_K = 0.8$	88.9	89.5	88.5	86.7	86.1
$CV_K = 0.9$	86.0	90.6	91.1	89.7	89.2

References

Amatya, A. and Bhaumik, D. K. (2018). Sample size determination for multilevel hierarchical designs using generalized linear mixed models. *Biometrics* **74,** 673–684.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88,** 9–25.

Davis-Plourde, K. (2021). Stepped wedge cluster randomized trial with three levels of clustering in each period: Design effect. `https://kendra-davis-plourde.shinyapps.io/SWCRT-3Level-DesignEffect/`.

Graybill, F. A. (1983). Matrices with applications in statistics. *Wadsworth, California, USA* .

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.

Hussey, M. A. and Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials* **28,** 182–191.

Jarvik, J. G., Comstock, B. A., James, K. T., Avins, A. L., Bresnahan, B. W., Deyo, R. A., et al. (2015). Lumbar imaging with reporting of epidemiology (LIRE)—protocol for a pragmatic cluster randomized trial. *Contemporary Clinical Trials* **45,** 157–163.

Jarvik, J. G., Meier, E. N., James, K. T., Gold, L. S., Tan, K. W., Kessler, L. G., et al. (2020). The effect of including benchmark prevalence data of common imaging findings in spine image reports on health care utilization among adults undergoing spine imaging: a stepped-wedge randomized clinical trial. *JAMA Network Open* **3,** e2015713–e2015713.

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.

Lawrie, J., Carlin, J. B., and Forbes, A. B. (2015). Optimal stepped wedge designs. *Statistics and Probability Letters* **99,** 210–214.

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

Li, F., Turner, E. L., and Preisser, J. S. (2018a). Optimal allocation of clusters in cohort stepped wedge designs. *Statistics and Probability Letters* **137,** 257–263.

Li, F., Turner, E. L., and Preisser, J. S. (2018b). 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.

Tian, Z., Preisser, J., Esserman, D., Turner, E., Rathouz, P., and Li, F. (2021). Impact of unequal cluster sizes for GEE analyses of stepped wedge cluster randomized trials with binary outcomes. *Biometrical Journal* page doi:10.1002/bimj.202100112.

Woertman, W., de Hoop, E., Moerbeek, M., Zuidema, S. U., Gerritsen, D. L., and Teerenstra, S. (2013). Stepped wedge designs could reduce the required sample size in cluster randomized trials. *Journal of Clinical Epidemiology* **66,** 752–758.