Web-based Supplementary Materials for "Sample Size Determination for GEE Analyses of Stepped Wedge Cluster Randomized Trials" by Li, Turner and Preisser

Web Appendix A

Eigenvalues of $\boldsymbol{R}_i$

By Theorem 8.3.4 and 8.4.4 in Graybill (1983), any $u \times u$ exchangeable matrix $\boldsymbol{A} = x\boldsymbol{I} + y\boldsymbol{J}$ is invertible if and only if $x \neq 0$ and $x + uy \neq 0$. The inverse is

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

and the determinant is given by

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

A corollary of this result is that $\boldsymbol{A}$ has two eigenvalues, x and $x + uy$, with respective algebraic multiplicities $u-1$ and 1, provided $uy \neq 0$. Denote the $N_i \times N_i$ matrices $\boldsymbol{B} = (1-\alpha_0-\lambda)\boldsymbol{I}_{N_i} + \alpha_0\boldsymbol{J}_{N_i}$ and $\boldsymbol{C} = (\alpha_2 - \alpha_1)\boldsymbol{I}_{N_i} + \alpha_1\boldsymbol{J}_{N_i}$. The eigenvalues of $\boldsymbol{R}_i$ are given by the roots to the characteristic equation

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

where the last equality is given by Theorem 8.9.1 in Graybill (1983). Since both $\boldsymbol{B} - \boldsymbol{C}$ and $\boldsymbol{B} + (T-1)\boldsymbol{C}$ are of the exchangeable form, it follows from (A.2) that $\boldsymbol{R}_i$ has four eigenvalues, $\lambda_1 = 1 - \alpha_0 + \alpha_1 - \alpha_2$, $\lambda_2 = 1 - \alpha_0 - (T-1)(\alpha_1 - \alpha_2)$, $\lambda_{i3} = 1 + (N_i - 1)(\alpha_0 - \alpha_1) - \alpha_2$, $\lambda_{i4} = 1 + (N_i - 1)\alpha_0 + (T-1)(N_i - 1)\alpha_1 + (T-1)\alpha_2$, with respective multiplicities $(T-1)(N_i-1)$, $N_i - 1$, $T - 1$ and 1. For unbalanced N_i, λ_{i3} and λ_{i4} vary by cluster so depend on i; the subscript i will be suppressed under a balanced design with $N_i = N$.

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

For notational convenience, we write

$$\boldsymbol{R}_i = a\boldsymbol{I}_T \otimes \boldsymbol{I}_{N_i} + b\boldsymbol{J}_T \otimes \boldsymbol{I}_{N_i} + c\boldsymbol{I}_T \otimes \boldsymbol{J}_{N_i} + d\boldsymbol{J}_T \otimes \boldsymbol{J}_{N_i},$$

with a, b, c, d as functions of $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)$ given by equation (2) in the main text. We assume $\boldsymbol{R}_i$ is invertible, hence the coefficients of basis matrices will only take values such that $\min\{\lambda_1, \lambda_2, \lambda_{i3}, \lambda_{i4}\} > 0$. To derive an expression for the inverse, we first show that $\boldsymbol{R}_i^{-1}$ can be expanded by the same set of basis matrices used in constructing $\boldsymbol{R}_i$. To see this, we will use the matrix inversion result by Henderson and Searle (1981),

$$(\boldsymbol{A} + \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{I} + \boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{A}^{-1}. \tag{A.3}$$

Observe that $\boldsymbol{R}_i = \boldsymbol{L}_T \otimes \boldsymbol{I}_{N_i} + \boldsymbol{S}_T \otimes \boldsymbol{J}_{N_i}$, where both $\boldsymbol{L}_T = a\boldsymbol{I}_T + b\boldsymbol{J}_T$ and $\boldsymbol{S}_T = c\boldsymbol{I}_T + d\boldsymbol{J}_T$ are of the exchangeable form. It follows from (A.1) that $\boldsymbol{L}_T^{-1} = e\boldsymbol{I}_T + f\boldsymbol{J}_T$ is again of the exchangeable form for some e, f (and the same holds for $\boldsymbol{S}_T^{-1}$). By (A.3), the inverse of $\boldsymbol{R}_i$ is

$$\boldsymbol{R}_i^{-1} = (\boldsymbol{L}_T^{-1} \otimes \boldsymbol{I}_{N_i}) - \underbrace{(\boldsymbol{L}_T^{-1} \otimes \boldsymbol{I}_{N_i})(\boldsymbol{S}_T \otimes \boldsymbol{J}_{N_i})[\boldsymbol{I}_{TN_i} + (\boldsymbol{L}_T^{-1} \otimes \boldsymbol{I}_{N_i})(\boldsymbol{S}_T \otimes \boldsymbol{J}_{N_i})]^{-1}(\boldsymbol{L}_T^{-1} \otimes \boldsymbol{I}_{N_i})}_{\boldsymbol{G}_{TN_i}}.$$

Further $\boldsymbol{L}_T^{-1} \otimes \boldsymbol{I}_{N_i} = e\boldsymbol{I}_T \otimes \boldsymbol{I}_{N_i} + f\boldsymbol{J}_T \otimes \boldsymbol{I}_{N_i}$, and $\boldsymbol{S}_T \otimes \boldsymbol{J}_{N_i} = c\boldsymbol{I}_T \otimes \boldsymbol{J}_{N_i} + d\boldsymbol{J}_T \otimes \boldsymbol{J}_{N_i}$. Since $(\boldsymbol{J}_T)^2 = T\boldsymbol{J}_T$, we have

$$\begin{aligned}\boldsymbol{I}_{TN_i} + (\boldsymbol{L}_T^{-1} \otimes \boldsymbol{I}_{N_i})(\boldsymbol{S}_T \otimes \boldsymbol{J}_{N_i}) &= \boldsymbol{I}_{TN_i} + ec(\boldsymbol{I}_T \otimes \boldsymbol{J}_{N_i}) + (ed + fc + Tfd)(\boldsymbol{J}_T \otimes \boldsymbol{J}_{N_i}), \\ &= \boldsymbol{I}_T \otimes (\boldsymbol{I}_{N_i} + ec\boldsymbol{J}_{N_i}) + (ed + fc + Tfd)\boldsymbol{J}_{TN_i}\end{aligned}$$

whose inverse is $\boldsymbol{I}_{TN_i} + l\boldsymbol{I}_T \otimes \boldsymbol{J}_{N_i} + m\boldsymbol{J}_T \otimes \boldsymbol{J}_{N_i}$ for some l, m by using (A.1) and (A.3) (i.e. treating $\boldsymbol{I}_T \otimes (\boldsymbol{I}_{N_i} + ec\boldsymbol{J}_{N_i})$ as $\boldsymbol{A}$ and $(ed+fc+Tfd)\boldsymbol{J}_{TN_i}$ as $\boldsymbol{B}$; then $\boldsymbol{A}^{-1}$ is derived from (A.1) and plugged into the right hand side of (A.3)). Therefore, $\boldsymbol{G}_{TN_i}$ must be of the form $r\boldsymbol{I}_T \otimes \boldsymbol{J}_{N_i} + s\boldsymbol{J}_T \otimes \boldsymbol{J}_{N_i}$, from which we conclude

$$\boldsymbol{R}_i^{-1} = \tilde{a}\boldsymbol{I}_T \otimes \boldsymbol{I}_{N_i} + \tilde{b}\boldsymbol{J}_T \otimes \boldsymbol{I}_{N_i} + \tilde{c}\boldsymbol{I}_T \otimes \boldsymbol{J}_{N_i} + \tilde{d}\boldsymbol{J}_T \otimes \boldsymbol{J}_{N_i},$$

for some coefficients $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}$. We then use the method of undetermined coefficients to derive an explicit expression for the inverse. Solving the full-rank system of equations impled by $\boldsymbol{R}_i^{-1}\boldsymbol{R}_i = \boldsymbol{I}_{TN_i}$, we obtain $\tilde{a} = 1/a$, $\tilde{b} = -b/[a(a+Tb)]$, $\tilde{c} = -c/[a(a+N_ic)]$ and $\tilde{d} = bc/[a(a+Tb)(a+N_ic)] + (bc-ad)/[(a+Tb)(a+N_ic)\{(a+N_ic)+T(b+N_id)\}]$. Expressing a, b, c, d as functions of the correlation parameters, we obtain a closed-form matrix inverse

$$\begin{aligned}\boldsymbol{R}_i^{-1} =& \frac{1}{\lambda_1}\boldsymbol{I}_{TN_i} - \frac{\alpha_2-\alpha_1}{\lambda_1\lambda_2}\boldsymbol{J}_T \otimes \boldsymbol{I}_{N_i} - \frac{\alpha_0-\alpha_1}{\lambda_1\lambda_{i3}}\boldsymbol{I}_T \otimes \boldsymbol{J}_{N_i} + \\ & \left\{\frac{(\alpha_2-\alpha_1)(\alpha_0-\alpha_1)}{\lambda_1\lambda_2\lambda_{i3}} + \frac{\alpha_2\alpha_0-\alpha_1}{\lambda_2\lambda_{i3}\lambda_{i4}}\right\}\boldsymbol{J}_{TN_i}.\end{aligned} \tag{A.4}$$

The significance of the closed-form expression (A.4) is in the computational savings arising from not having to invert $\boldsymbol{R}_i$, which would otherwise be costly for large TN_i.

Web Appendix B

Matrix-adjusted Estimating Equations

The matrix-adjusted estimating equations (MAEE) for estimating $\boldsymbol{\alpha}$ have been described by Lu et al. (2007) and Preisser et al. (2008) for binary responses. We briefly review the relevant technical details and further extend MAEE to accommodate continuous responses.

For each pair of responses l and l' ($1 \le l, l' \le TN_i$) within cluster i, write

$$\eta_{ill'}(\mu_{il}, \mu_{il'}, \phi) = \frac{(y_{il}-\mu_{il})(y_{il'}-\mu_{il'})}{\phi v^{1/2}(\mu_{il})v^{1/2}(\mu_{il'})}. \tag{B.1}$$

For $l \neq l'$, $\eta_{ill'}$ is interpreted as the sample correlation; for $l = l'$, η_{ill} is interpreted as the squared standardized residual. Without loss of clarity, the subscript jk in the main text is now replaced by l to simplify the notation. If response y_{il} is binary and $l \neq l'$, Prentice (1988) assumed a binomial model ($\phi = 1$) and gave $E(\eta_{ill'}) = \rho_{ill'}$, and

$$\text{var}(\eta_{ill'}) = w_{ill'} = 1 + \frac{(1-2\mu_{il})(1-2\mu_{il'})\rho_{ill'}}{\{\mu_{il}\mu_{il'}(1-\mu_{il})(1-\mu_{il'})\}^{1/2}} - \rho_{ill'}^2,$$

where $\rho_{ill'}$ is the correlation parameter at (l, l')th position in matrix $\boldsymbol{R}_i$. In a cohort stepped wedge cluster randomized trial, $\rho_{ill'} = \rho_{ill'}(\boldsymbol{\alpha})$ is a function of the correlation parameters and will be equal to either α_0, α_1 or α_2 depending on the coordinate (l, l'). Define $\boldsymbol{\rho}_i = (\rho_{i12}, \rho_{i13}, \ldots, \rho_{i,(TN_i-1),TN_i})'$, $\boldsymbol{E}_i = \partial\boldsymbol{\rho}_i/\partial\boldsymbol{\alpha}'$, and $\boldsymbol{W}_i = \text{diag}(w_{i12}, w_{i13}, \ldots, w_{i,(TN_i-1),TN_i})$. By using an independent working correlation, the matrix-adjusted $\boldsymbol{\alpha}$-estimating equations are constructed as

$$\sum_{i=1}^{I} \boldsymbol{E}_i' \boldsymbol{W}_i^{-1}(\tilde{\boldsymbol{\eta}}_i - \boldsymbol{\rho}_i(\boldsymbol{\alpha})) = \mathbf{0}, \tag{B.2}$$

where $\tilde{\boldsymbol{\eta}}_i = (\tilde{\eta}_{i12}, \tilde{\eta}_{i13}, \ldots, \tilde{\eta}_{i,(TN_i-1),TN_i})'$ is a vector of bias-adjusted sample correlations with element

$$\tilde{\eta}_{ill'} = [\boldsymbol{A}_i^{-1/2}(\boldsymbol{I} - \boldsymbol{H}_i)^{-1}\boldsymbol{A}_i^{1/2}]_{l\cdot}[\hat{\boldsymbol{R}}_i]_{\cdot l'}. \tag{B.3}$$

In (B.3), $[\boldsymbol{M}]_{l\cdot}$ and $[\boldsymbol{M}]_{\cdot l'}$ represent the lth row and l'th column of the matrix $\boldsymbol{M}$, respectively, $\boldsymbol{H}_i$ is the ith cluster leverage (Preisser and Qaqish, 1996), and $\hat{\boldsymbol{R}}_i$ is the estimated working correlation matrix with its (l, l')th element given by $\eta_{ill'}(\hat{\mu}_{il}, \hat{\mu}_{il'}, \phi = 1)$. In a cross-sectional stepped wedge design, where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$ and both values tend to be small, we could follow Preisser et al. (2008) to further approximate (B.3) by $\tilde{\eta}_{ill'} \approx [(\boldsymbol{I} - \boldsymbol{H}_i)^{-1}]_{l\cdot}[\hat{\boldsymbol{R}}_i]_{\cdot l'}$. However, since α_2 is not guaranteed to be small in a cohort stepped wedge design, correction (B.3) will still be recommended. Joint estimation for model parameters based on the $\boldsymbol{\theta}$-estimating equations and $\boldsymbol{\alpha}$-estimating equations follows the iterative steps outlined in Prentice (1988).

If response y_{il} is continuous, we can assume a Gaussian model such that $v(\mu_{ijk}) = 1$ and ϕ will be interpreted as the common variance. For $l \neq l'$, we could show $E(\eta_{ill'}) = \rho_{ill'}$ and $\text{var}(\eta_{ill'}) = w_{ill'} = 1 + \rho_{ill'}^2$. The latter equality is derived by computing the higher-order moments from the bivariate normal moment generating function of y_{il} and $y_{il'}$ (the algebraic details are omitted here for brevity). Estimating equations (B.2) will be used to estimate the correlation parameters provided the appropriate values of $w_{ill'}$ be substituted in $\boldsymbol{W}_i$. Notably in this case, the bias-adjusted sample correlation becomes

$$\tilde{\eta}_{ill'} = [(\boldsymbol{I} - \boldsymbol{H}_i)^{-1}]_{l\cdot}[\hat{\boldsymbol{R}}_i]_{\cdot l'},$$

because $\boldsymbol{A}_i = \phi\boldsymbol{I}$ no longer depends on the marginal means and $\boldsymbol{A}_i^{-1/2}$, $\boldsymbol{A}_i^{1/2}$ cancel out. Joint estimation for model parameters proceeds in the same way as in the binary case. An additional step will be required in the iterative procedure to update the nuisance ϕ, and we use a moment-based approach similar to Liang and Zeger (1986). From iteration s to $s+1$, the update is given by

$$\hat{\phi}^{(s+1)} = \hat{\phi}^{(s)} \frac{\sum_{i=1}^{I} \sum_{l=1}^{TN_i} \tilde{\eta}_{ill}}{T\sum_{i=1}^{I} N_i - (T+1)}.$$

Web Appendix C

Derivation of var$(\hat{\delta})$

Denote $\mathbf{1}_u$ as a $u \times 1$ vector of ones and $\boldsymbol{I}_u$ as a $u \times u$ identity matrix, the design matrix of cluster i is $\boldsymbol{Z}_i = (\boldsymbol{I}_T, \boldsymbol{X}_i) \otimes \mathbf{1}_{N_i}$. Assuming $N_i = N$ (hence λ_{i3} and λ_{i4} no longer depend on i and will be simply written as λ_3 and λ_4) and $\text{var}(\boldsymbol{y}_i) = \boldsymbol{V}_i$ as in Shih (1997), we observe that $\text{var}(\hat{\delta})$ equals to the lower-right corner element of $(\sum_{i=1}^{I} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} \boldsymbol{D}_i)^{-1} = \phi(\sum_{i=1}^{I} \boldsymbol{Z}_i' \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i)^{-1}$. Further

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

where $\boldsymbol{\Omega}_{11}$ is of dimension $T \times T$, $\boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}_{21}'$ is of dimension $T \times 1$, and $\boldsymbol{\Omega}_{22}$ is a scalar. Recall

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

where the coefficients of the basis matrices correspond to those in (A.4). We have $\boldsymbol{\Omega}_{11} = \sum_{i=1}^{I} [N(\tilde{a} + N\tilde{c}) \boldsymbol{I}_T + N(\tilde{b} + N\tilde{d}) \boldsymbol{J}_T]$, and

$$\begin{aligned} \tilde{a} + N\tilde{c} &= \frac{1}{\lambda_1} - \frac{N(\alpha_0 - \alpha_1)}{\lambda_1 \lambda_3} = \frac{1}{\lambda_3}, \\ \tilde{b} + N\tilde{d} &= -\frac{\alpha_2 - \alpha_1}{\lambda_1 \lambda_2} + N \left\{ \frac{(\alpha_2 - \alpha_1)(\alpha_0 - \alpha_1)}{\lambda_1 \lambda_2 \lambda_3} + \frac{\alpha_2 \alpha_0 - \alpha_1}{\lambda_2 \lambda_3 \lambda_4} \right\} = -\frac{\lambda_4 - \lambda_3}{T \lambda_3 \lambda_4}. \end{aligned}$$

It follows directly that

$$\boldsymbol{\Omega}_{11} = IN \left\{ \frac{1}{\lambda_3} \boldsymbol{I}_T - \frac{\lambda_4 - \lambda_3}{T \lambda_3 \lambda_4} \boldsymbol{J}_T \right\},$$

whose inverse, from an application of (A.1), is

$$\boldsymbol{\Omega}_{11}^{-1} = \frac{1}{IN} \left\{ \lambda_3 \boldsymbol{I}_T + \frac{\lambda_4 - \lambda_3}{T} \boldsymbol{J}_T \right\}.$$

Similarly, $\mathbf{\Omega}_{12} = \mathbf{\Omega}_{21}' = \sum_{i=1}^{I}\{N(\tilde{a}+N\tilde{c})\boldsymbol{X}_i + N(\tilde{b}+N\tilde{d})(\sum_{j=1}^{T}X_{ij}\mathbf{1}_T)\} = N\sum_{i=1}^{I}\boldsymbol{X}_i/\lambda_3 - N(\lambda_4-\lambda_3)U\mathbf{1}_T/(T\lambda_3\lambda_4)$, where $U=\sum_{i=1}^{I}\sum_{j=1}^{T}X_{ij}=\sum_{i=1}^{I}\sum_{j=1}^{T}X_{ij}^2$. Finally, we could obtain $\mathbf{\Omega}_{22}=\sum_{i=1}^{I}\{N(\tilde{a}+N\tilde{c})(\sum_{j=1}^{T}X_{ij}^2)+N(\tilde{b}+N\tilde{d})(\sum_{j=1}^{T}X_{ij})^2\}=NU/\lambda_3-N(\lambda_4-\lambda_3)V/(T\lambda_3\lambda_4)$, where $V=\sum_{i=1}^{I}(\sum_{j=1}^{T}X_{ij})^2$.

Block matrix inversion gives $\text{var}(\hat{\delta})=(\mathbf{\Omega}_{22}-\mathbf{\Omega}_{21}\mathbf{\Omega}_{11}^{-1}\mathbf{\Omega}_{12})^{-1}$. We next define $W=\sum_{j=1}^{T}(\sum_{i=1}^{I}X_{ij})^2$ and further observe that $\sum_{i=1}^{I}\sum_{j\neq k}X_{ij}X_{ik}=V-U$, $\sum_{j\neq k}(\sum_{i=1}^{I}X_{ij})(\sum_{i=1}^{I}X_{ik})=U^2-W$. After some algebra simplifying the expression for $\text{var}(\hat{\delta})$, we obtain

$$\begin{aligned}\text{var}(\hat{\delta}) &= \frac{(\phi/N)I\lambda_3\lambda_4}{(U^2+ITU-TW-IV)\{\alpha_2+(N-1)\alpha_1\}+(IU-W)\lambda_3}\\ &= \frac{(\phi/N)IT\lambda_3\lambda_4}{(U^2+ITU-TW-IV)\lambda_4-(U^2-IV)\lambda_3},\end{aligned}$$

where the last equality holds since $\lambda_4=\lambda_3+T\{\alpha_2+(N-1)\alpha_1\}$.

Limit of var($\hat{\delta}$) as Cohort Size $N\to\infty$

We note here that the following discussion is inspired by a comment from an anonymous referee. In cluster randomized trials, the limiting factor for power is usually the number of clusters I instead of the number of individuals per cluster N (Murray, 1998). However, the above variance expression involves the (ϕ/N) multiplier, and may leave an impression that we could always increase the number of enrolled individuals (cohort size N) and make the variance arbitrarily small (hence the limiting power goes to 1 regardless of number of clusters). To better understand the variance expression, we derive its limit as $N\to\infty$. Notice that $\lambda_3/N\to\alpha_0-\alpha_1$, $\lambda_4/N\to\alpha_0+(T-1)\alpha_1$. Some simplification algebra gives

$$\lim_{N\to\infty}\text{var}(\hat{\delta})=\frac{\phi I(\alpha_0-\alpha_1)\{\alpha_0+(T-1)\alpha_1\}}{(IU-W)\{\alpha_0+(T-1)\alpha_1\}+(U^2-IV)\alpha_1},$$

where the design constants U, V, W are free of N by definition. First, the limit of $\text{var}(\hat{\delta})$ does not depend on the within-individual correlation α_2, which implies that the same limit applies to a cross-sectional design (where $\alpha_1=\alpha_2$, and N refers to the cluster-period size). Second, when $\alpha_0\neq\alpha_1$ and $\alpha_0+\alpha_1/(T-1)\neq 0$, the limit of the variance is a positive constant determined by available design resources I, T and two correlation values (α_0,α_1), and so cannot be made arbitrarily small. Finally, the limit variance is 0 when the within-period and the inter-period correlations are equal $(\alpha_0=\alpha_1)$ or $\alpha_0+\alpha_1/(T-1)=0$ (this latter condition is impossible for positive correlation values). This last statement confirms (based on our analytical results from a marginal model) the important point made by Taljaard et al. (2016) that "the underlying assumption of equal intracluster (within-period) and inter-period correlations may not be plausible ... under this assumption, the required number of clusters will always tend to 1 as the cluster size increases indefinitely."

Web Appendix D: Web Tables

Web Table 1: Convergence rates (out of 1000) for GEE analyses of simulated continuous responses.

Effect Size[a]	$\boldsymbol{\alpha}$[b]	I	N	T	Convergence (Size)[c]	Convergence (Power)[d]
0.65	A1	9	11	4	988	988
0.65	A1	8	24	3	949	945
0.65	A1	15	4	4	1000	1000
0.65	A1	10	16	3	992	994
0.65	A1	12	6	4	999	997
0.4	A1	12	14	5	999	999
0.4	A1	20	6	5	1000	999
0.4	A1	15	12	4	1000	1000
0.4	A1	21	8	4	1000	1000
0.4	A1	15	8	6	1000	1000
0.4	A2	16	12	5	1000	1000
0.4	A2	18	15	4	1000	1000
0.4	A2	15	10	6	1000	1000
0.4	A2	18	6	7	1000	1000
0.4	A2	15	25	4	1000	1000
0.35	A3	18	10	4	1000	1000
0.35	A3	16	9	5	975	975
0.35	A3	20	6	5	1000	998
0.25	A3	25	8	6	1000	1000
0.25	A3	24	7	7	943	935

[a] Effect size, $\delta/\phi^{1/2}$, of simulation scenarios for studying power.

[b] A1: $\boldsymbol{\alpha} = (0.03, 0.015, 0.2)$; A2: $\boldsymbol{\alpha} = (0.1, 0.05, 0.2)$; A3: $\boldsymbol{\alpha} = (0.01, 0.005, 0.4)$.

[c] Convergence rates (out of 1000) in simulation scenarios for studying test size.

[d] Convergence rates (out of 1000) in simulation scenarios for studying power.

Web Table 2: Convergence rates (out of 1000) for GEE analyses of simulated binary responses.

Baseline[a]	Effect Size [b]	$\boldsymbol{\alpha}$[c]	I	N	T	Convergence (Size)[d]	Convergence (Power)[e]
0.75	0.25	A1	12	6	4	980	970
0.75	0.25	A1	14	8	3	978	977
0.7	0.3	A1	9	15	4	980	978
0.7	0.3	A1	12	7	5	968	976
0.7	0.3	A1	10	20	3	934	933
0.65	0.45	A1	15	12	4	1000	1000
0.65	0.45	A1	15	8	6	1000	1000
0.65	0.45	A1	16	8	5	998	998
0.65	0.45	A1	12	14	5	997	998
0.65	0.45	A1	18	10	4	999	999
0.7	0.45	A2	15	12	6	1000	1000
0.7	0.45	A2	16	14	5	1000	996
0.7	0.45	A2	18	18	4	999	997
0.7	0.45	A2	21	12	4	1000	999
0.7	0.45	A2	20	7	6	1000	1000
0.65	0.6	A3	21	15	4	999	1000
0.65	0.6	A3	20	12	5	1000	1000
0.65	0.6	A3	15	25	4	999	998
0.65	0.6	A3	24	9	5	986	982
0.65	0.6	A3	16	16	5	999	991

[a] Baseline prevalence, $e^{\beta_1}/(1+e^{\beta_1})$.

[b] Effect size in odds ratio, e^{δ}, of simulation scenarios for studying power.

[c] A1: $\boldsymbol{\alpha} = (0.03, 0.015, 0.2)$; A2: $\boldsymbol{\alpha} = (0.1, 0.05, 0.2)$; A3: $\boldsymbol{\alpha} = (0.01, 0.005, 0.4)$.

[d] Convergence rates (out of 1000) in simulation scenarios for studying test size.

[e] Convergence rates (out of 1000) in simulation scenarios for studying power.

Web Table 3: Simulation scenarios, predicted power based on z-test and t-test, along with the corresponding empirical power of GEE analyses using different variance estimators for binary responses. Bias-corrected estimation of correlation parameters uses MAEE.

						z-test			t-test			
Baseline[a]	Effect Size[b]	$\boldsymbol{\alpha}$[c]	I	N	T	Pred[d]	MB[e]	BC2[g]	Pred[d]	MB[e]	BC1[f]	BC3[h]
0.75	0.25	A1	12	6	4	0.936	0.957	0.915	0.850	0.901	0.881	0.877
0.75	0.25	A1	14	8	3	0.905	0.942	0.898	0.840	0.897	0.871	0.847
0.70	0.30	A1	9	15	4	0.973	0.983	0.946	0.835	0.899	0.868	0.870
0.70	0.30	A1	12	7	5	0.970	0.976	0.946	0.893	0.928	0.919	0.921
0.70	0.30	A1	10	20	3	0.939	0.947	0.895	0.835	0.870	0.847	0.827
0.65	0.45	A1	15	12	4	0.880	0.884	0.843	0.807	0.831	0.820	0.821
0.65	0.45	A1	15	8	6	0.929	0.933	0.904	0.852	0.877	0.869	0.878
0.65	0.45	A1	16	8	5	0.890	0.914	0.879	0.820	0.873	0.847	0.852
0.65	0.45	A1	12	14	5	0.931	0.939	0.897	0.822	0.855	0.852	0.857
0.65	0.45	A1	18	10	4	0.892	0.886	0.869	0.840	0.854	0.857	0.856
0.70	0.45	A2	15	12	6	0.910	0.914	0.877	0.825	0.840	0.839	0.847
0.70	0.45	A2	16	14	5	0.890	0.890	0.843	0.820	0.828	0.810	0.817
0.70	0.45	A2	18	18	4	0.870	0.880	0.849	0.814	0.845	0.828	0.830
0.70	0.45	A2	21	12	4	0.858	0.878	0.835	0.812	0.831	0.827	0.828
0.70	0.45	A2	20	7	6	0.903	0.911	0.895	0.854	0.876	0.874	0.878
0.65	0.60	A3	21	15	4	0.872	0.873	0.842	0.828	0.834	0.830	0.831
0.65	0.60	A3	20	12	5	0.886	0.895	0.860	0.838	0.858	0.842	0.845
0.65	0.60	A3	15	25	4	0.902	0.894	0.855	0.835	0.836	0.827	0.828
0.65	0.60	A3	24	9	5	0.860	0.860	0.822	0.820	0.817	0.812	0.812
0.65	0.60	A3	16	16	5	0.896	0.908	0.861	0.827	0.846	0.827	0.835

[a] Baseline prevalence, $e^{\beta_1}/(1+e^{\beta_1})$.

[b] Effect size in odds ratio, e^{δ}.

[c] A1: $\boldsymbol{\alpha} = (0.03, 0.015, 0.2)$; A2: $\boldsymbol{\alpha} = (0.1, 0.05, 0.2)$; A3: $\boldsymbol{\alpha} = (0.01, 0.005, 0.4)$.

[d] Pred: Predicted power.

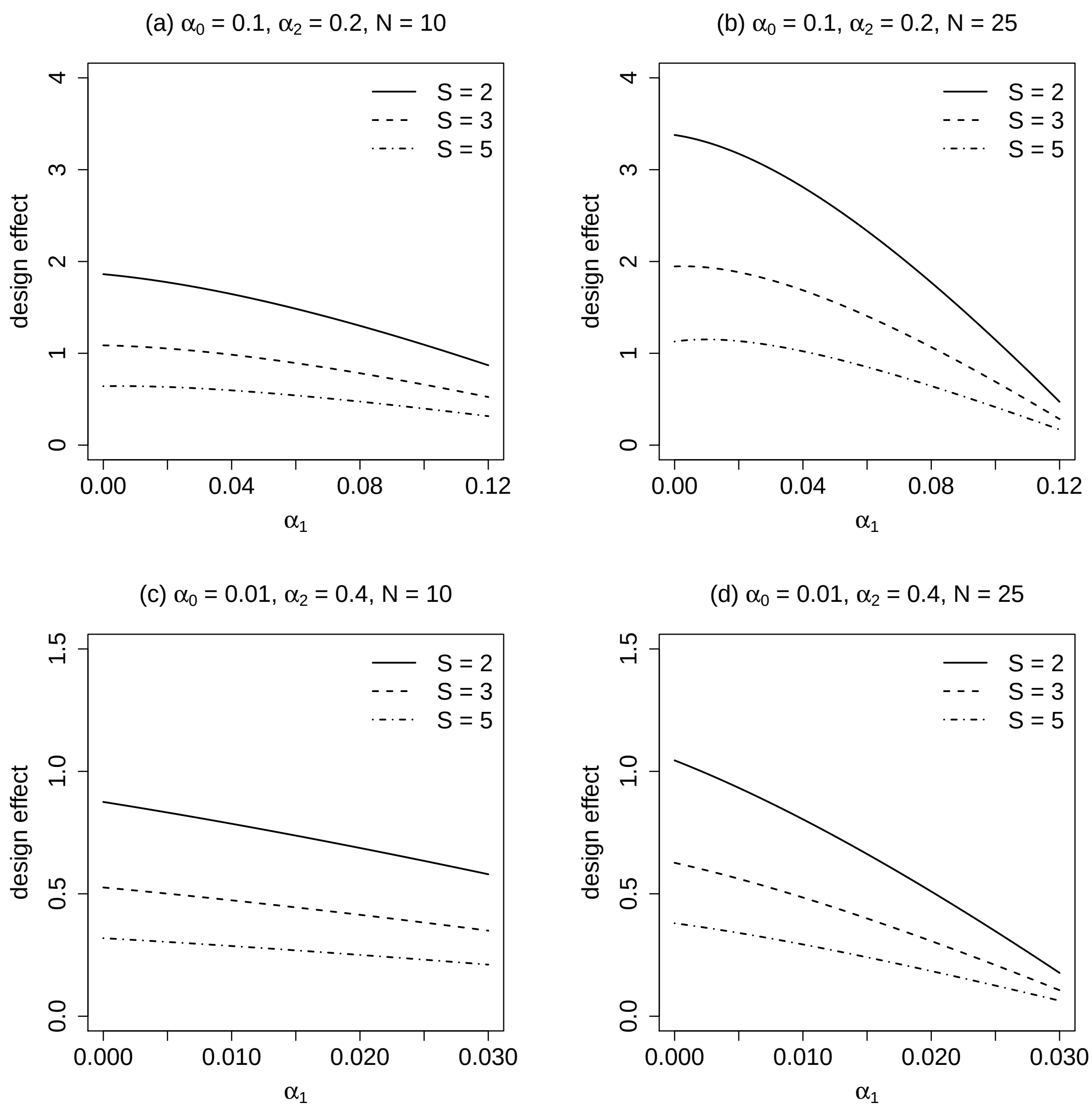
[e] MB: Model-based variance.

[f] BC1: Bias-corrected sandwich variance of Kauermann and Carroll (2001).

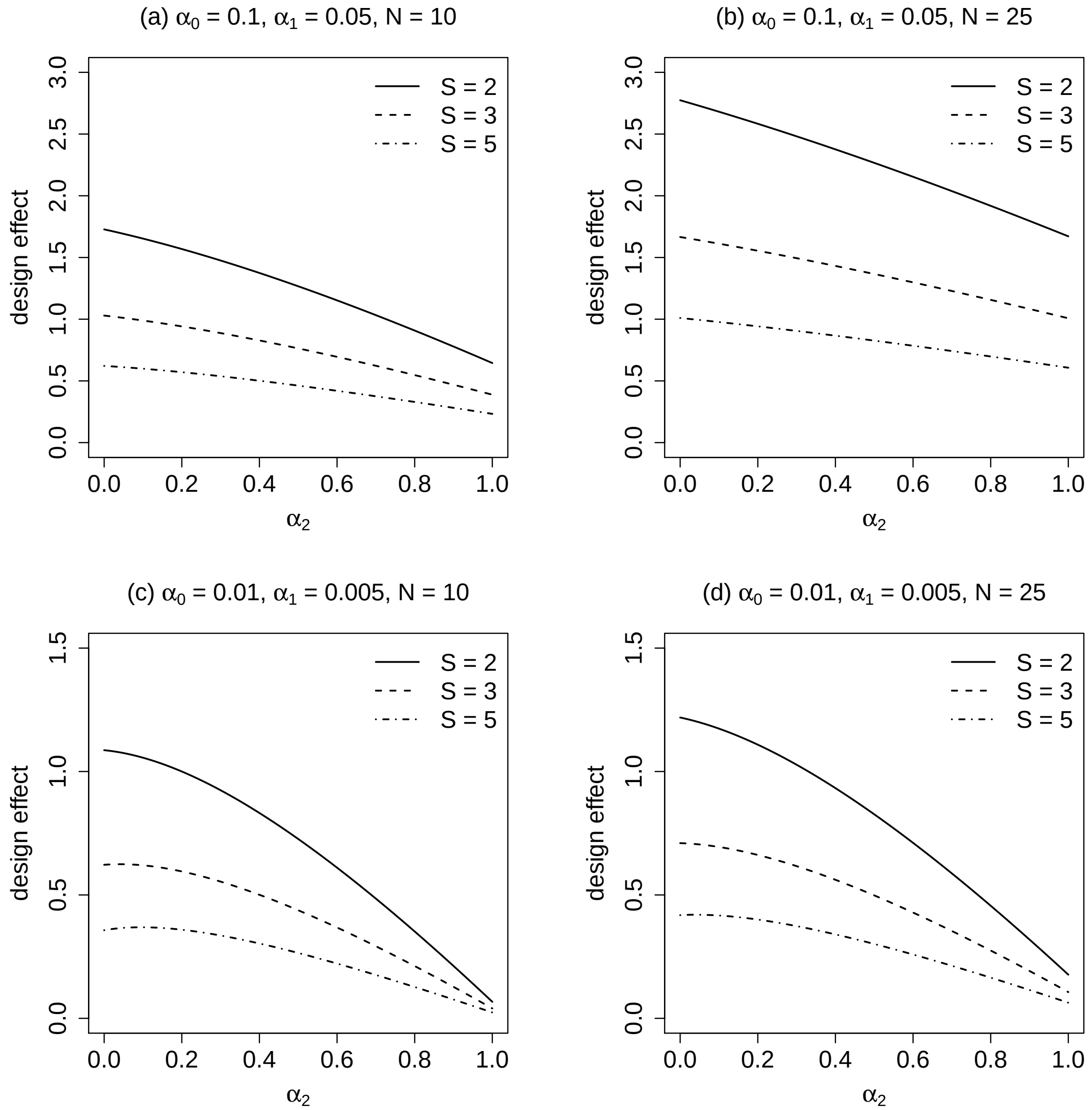
[g] BC2: Bias-corrected sandwich variance of Mancl and DeRouen (2001).

[h] BC3: Bias-corrected sandwich variance of Fay and Graubard (2001).

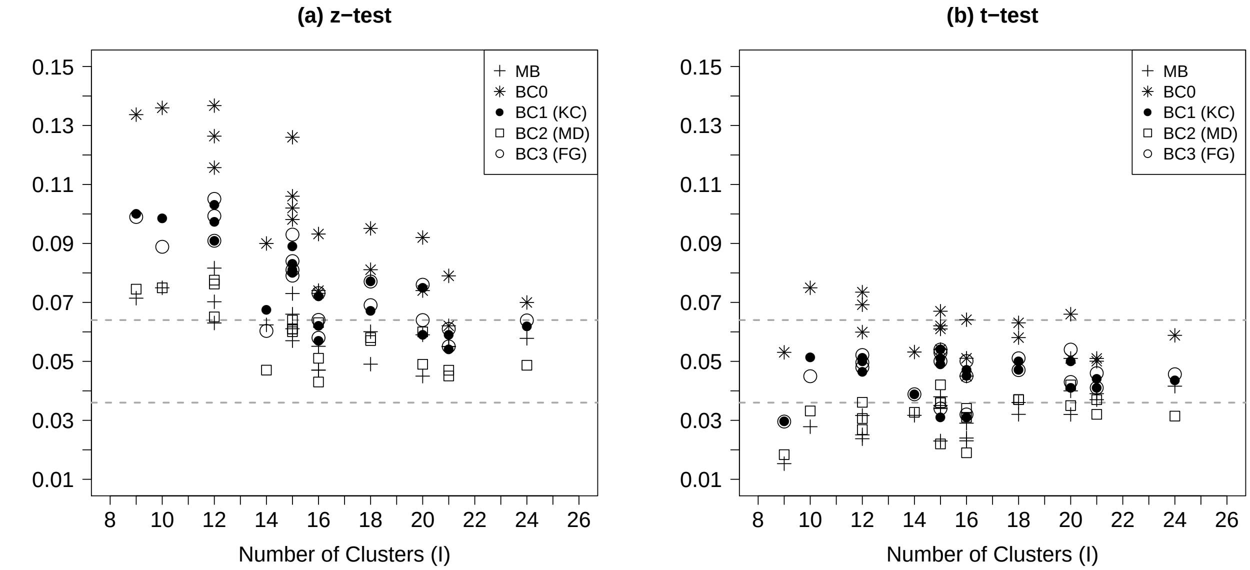
Web Appendix E: Web Figures



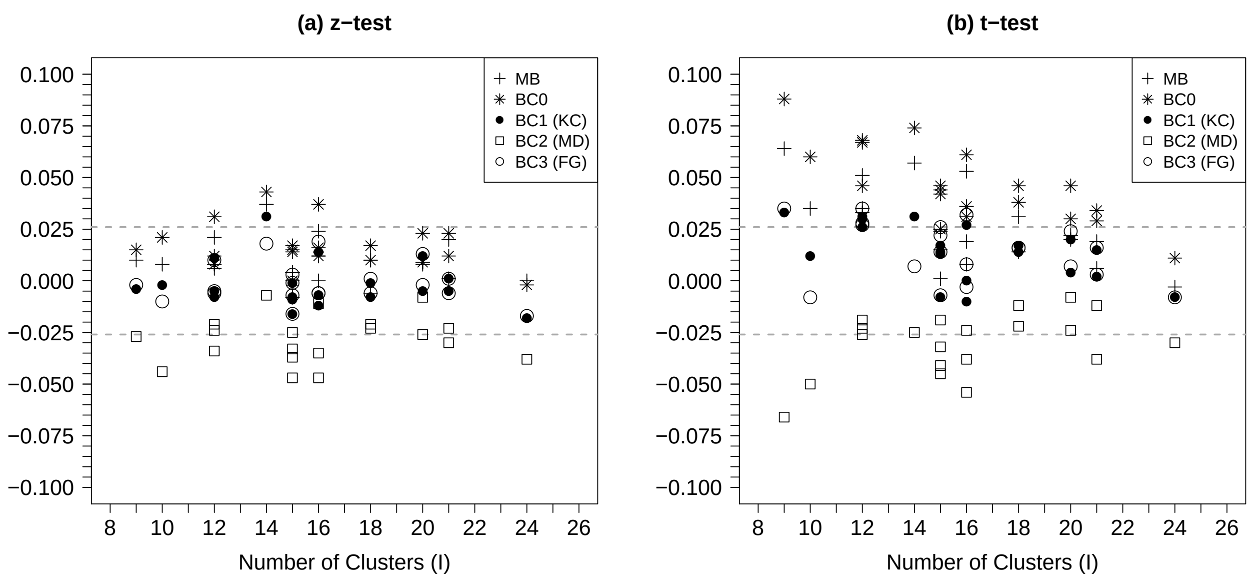
Web Figure 1: The design effect as a function of the inter-period correlation, α_1, in cohort stepped wedge scenarios with $(\alpha_0, \alpha_2) = \{(0.1, 0.2), (0.01, 0.4)\}$, $N = \{10, 25\}$, $S = \{2, 3, 5\}$ steps, $b = 1$ baseline measurement and $c_s = 1$ measurement after each step. Any displayed combination of $(\alpha_0, \alpha_1, \alpha_2)$ ensures a positive definite correlation structure.



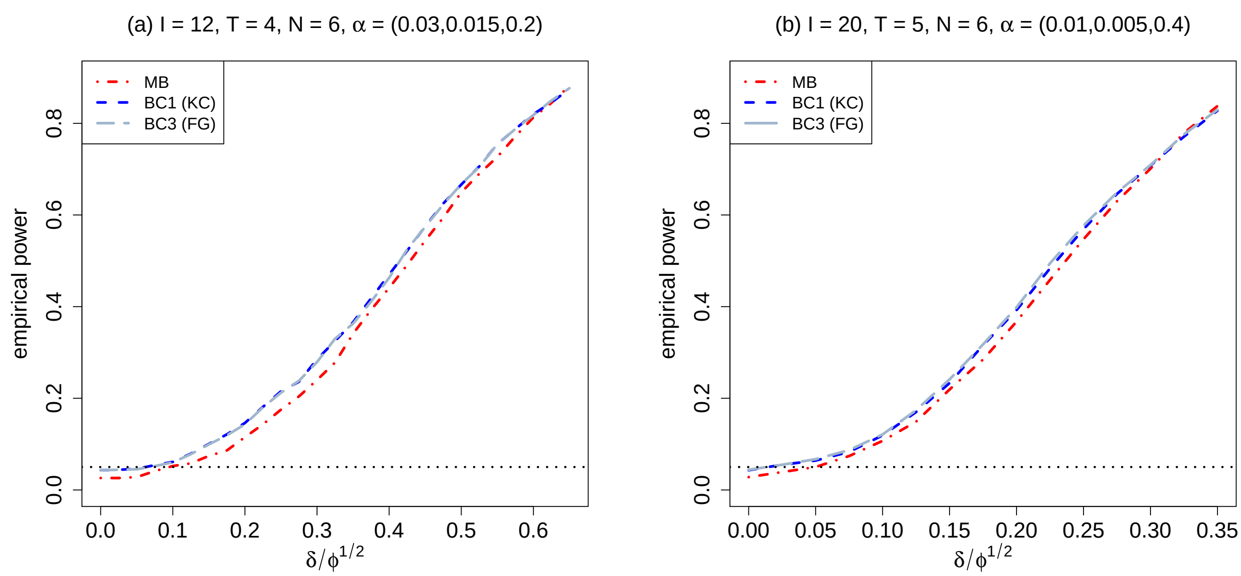
Web Figure 2: The design effect as a function of the within-individual correlation, α_2, in cohort stepped wedge scenarios with $(\alpha_0, \alpha_1) = \{(0.1, 0.05), (0.01, 0.005)\}$, $N = \{10, 25\}$, $S = \{2, 3, 5\}$ steps, $b = 1$ baseline measurement and $c_s = 1$ measurement after each step. Any displayed combination of $(\alpha_0, \alpha_1, \alpha_2)$ ensures a positive definite correlation structure.



Web Figure 3: Empirical type I error rates for GEE-based (a) z-tests and (b) t-tests for binary responses. MB: model-based variance; BC0: uncorrected sandwich variance; BC1: KC-corrected sandwich variance; BC2: MD-corrected sandwich variance; BC3: FG-corrected sandwich variance.



Web Figure 4: Differences between the empirical power and the predicted power of GEE-based (a) z-tests and (b) t-tests for binary responses. MB: model-based variance; BC0: uncorrected sandwich variance; BC1: KC-corrected sandwich variance; BC2: MD-corrected sandwich variance; BC3: FG-corrected sandwich variance.



Web Figure 5: Illustrative comparisons of empirical power curves for GEE t-tests with continuous responses. MB: model-based variance; BC1: KC-corrected sandwich variance; BC3: FG-corrected sandwich variance. The dotted horizontal line indicates the nominal size at 0.05.

References

Graybill, F. A. (1983). *Matrices with Applications in Statistics.* Brooks/Cole, Wadsworth, Belmont, CA.

Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *Siam Review* **23,** 53–60.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Lu, B., Preisser, J. S., Qaqish, B. F., Suchindran, C., Bangdiwala, S. I., and Wolfson, M. (2007). A comparison of two bias-corrected covariance estimators for generalized estimating equations. *Biometrics* **63,** 935–941.

Murray, D. M. (1998). *Design and Analysis of Group-Randomized Trials.* Oxford University Press, New York, NY.

Preisser, J. S., Lu, B., and Qaqish, B. F. (2008). Finite sample adjustments in estimating equations and covariance estimators for intracluster correlations. *Statistics in Medicine* **27,** 5764–5785.

Preisser, J. S. and Qaqish, B. F. (1996). Deletion diagnostics for generalised estimating equations. *Biometrika* **83,** 551–562.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44,** 1033–1048.

Shih, W. J. (1997). Sample size and power calculations for periodontal and other studies with clustered samples using the method of generalized estimating equations. *Biometrical Journal* **39,** 899–908.

Taljaard, M., Teerenstra, S., Ivers, N. M., and Fergusson, D. A. (2016). Substantial risks associated with few clusters in cluster randomized and stepped wedge designs. *Clinical Trials* **13,** 459–463.