Sample size considerations for stepped wedge designs with subclusters

Kendra Plourde, PhD
Yale University School of Public Health
kendra.plourde@yale.edu

February 14, 2022

What is a Cluster Randomized Trial?

In a cluster randomized trial (CRT) the unit of randomization is a *cluster* instead of an *individual* in a randomized controlled trial (RCT).

One might choose a CRT over a RCT if:

1. There is a high risk of contamination.
2. The goal is evaluate how a practice-wide change affects patient outcomes.

What is a Stepped Wedge Cluster Randomized Trial?

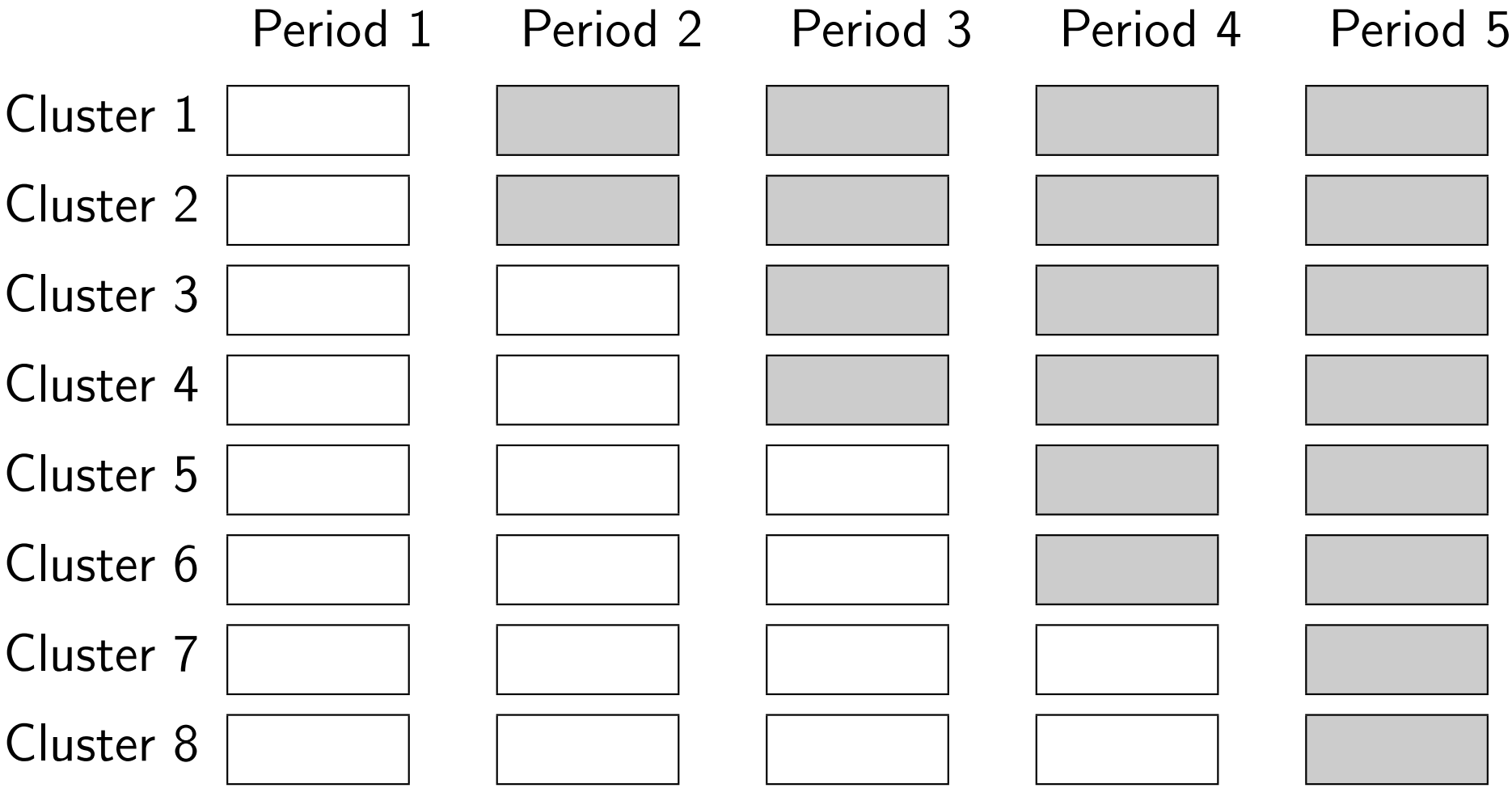
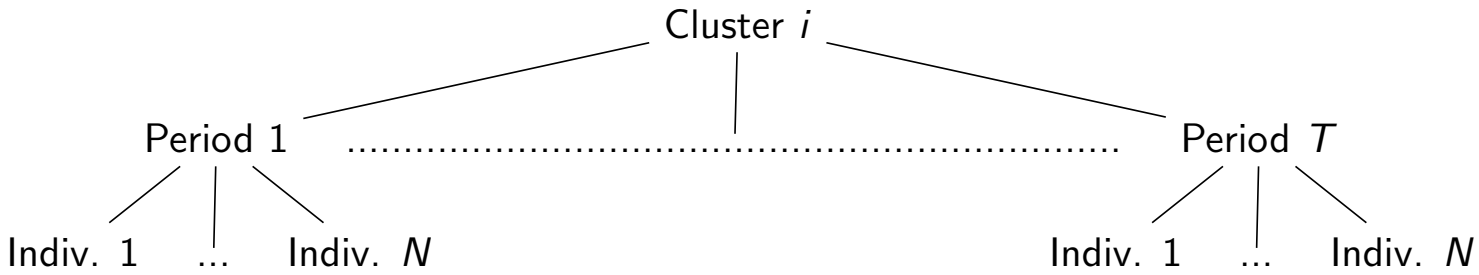


Figure: A schematic illustration of a stepped wedge cluster randomized trial with 8 clusters and 5 periods. Each white cell indicates a cluster-period under the control condition and each gray cell indicates a cluster-period under the intervention condition.

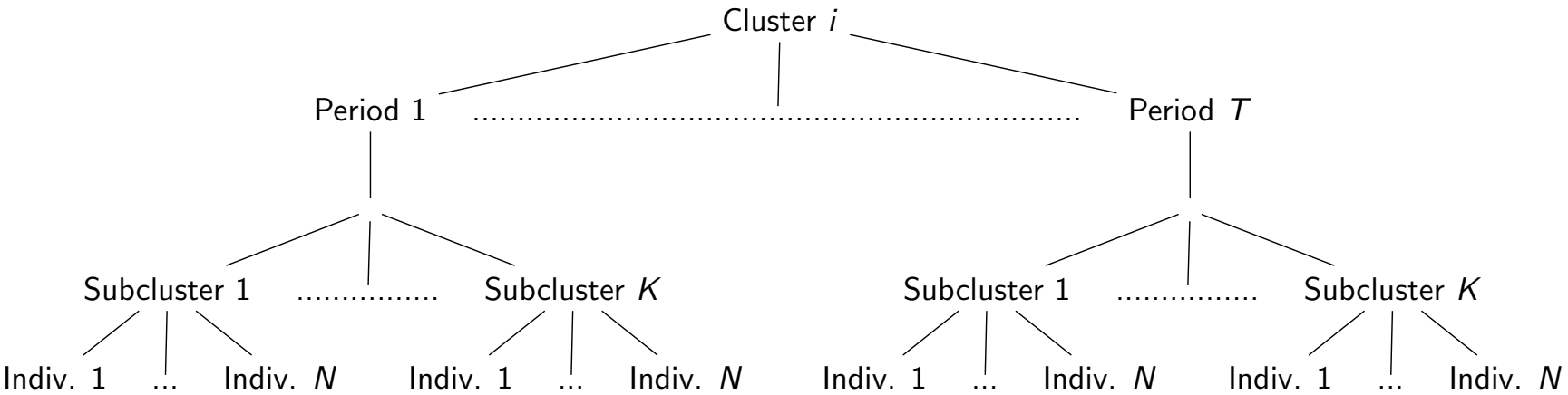
Example of Multilevel SW-CRT

SW-CRT without multiple levels of clustering:



Example: Individuals are nested in hospitals (cluster).

SW-CRT with multiple levels of clustering:



Example: Individuals are nested in primary care providers (subcluster) which are nested in hospitals (cluster).

Motivating Studies: LIRE & EPT Trials

The Lumbar Imaging with Reporting of Epidemiology (LIRE) trial (Jarvik et al., 2015).

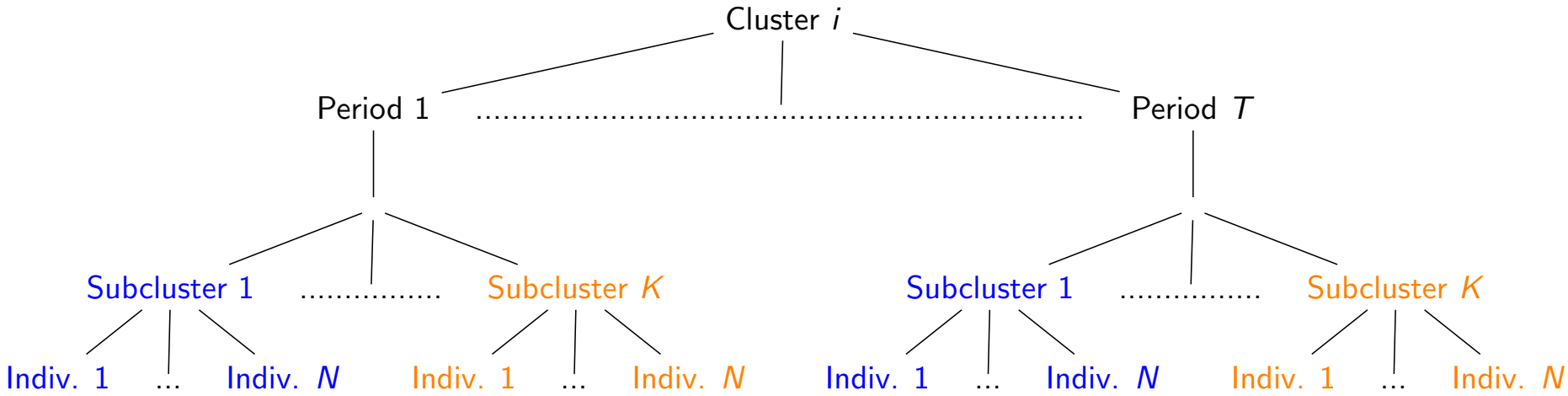
- Randomized clinics consisting of primary care providers.
- Outcome of interest was a composite measure of back pain (Gaussian).

The Washington Expedited Partner Therapy (EPT) trial (Golden et al., 2015).

- Randomized local health jurisdictions containing clinics.
- Outcome of interest was chlamydia reinfection (Binary).

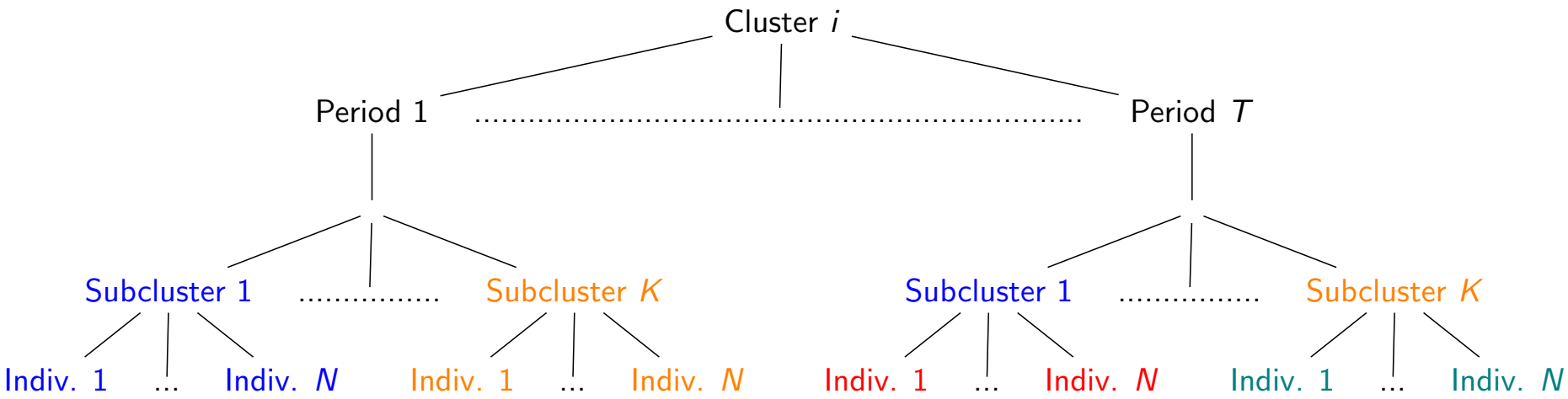
Design Variants under Multilevel Clustering

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



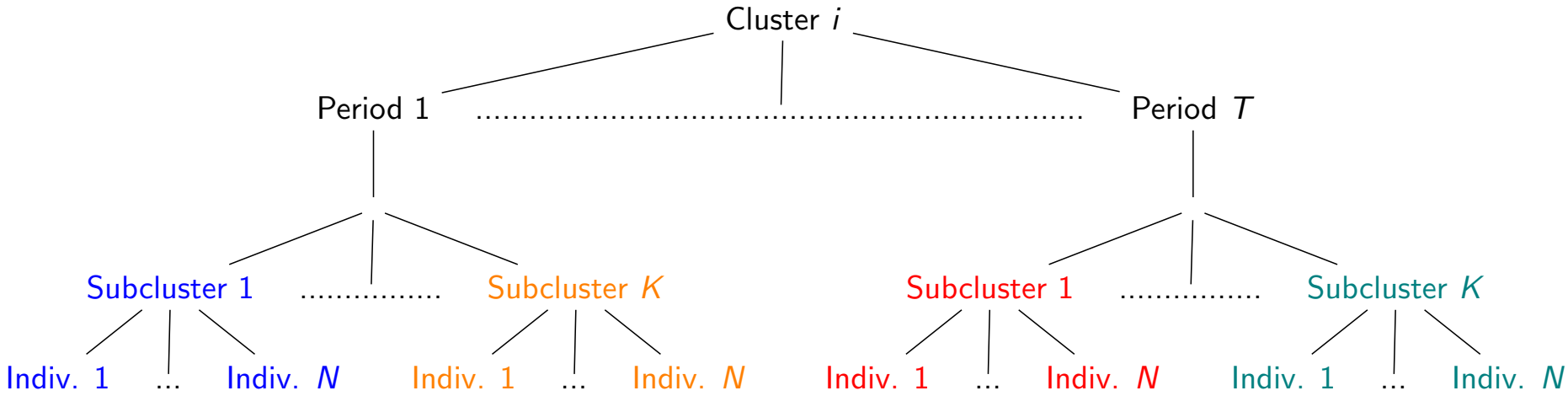
Design Variants under Multilevel Clustering

(B) Closed-cohort design on the subcluster level but a cross-sectional design at the subject level



Design Variants under Multilevel Clustering

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



Current Methods

We want to analyze a trial where participant l is nested in subcluster k sampled at time period j within cluster i.

- Hussey and Hughes (2007) used a linear mixed model (LMM) to take into account correlations within a cluster using a single random effect.

$$Y_{ijkl} = \beta_j + X_{ij}\delta + b_i + \epsilon_{ijkl}$$

- To differentiate the within- and between-period correlations Hooper et al. (2016) extended the LMM (Hussey and Hughes, 2007).

$$Y_{ijkl} = \beta_j + X_{ij}\delta + b_i + s_{ij} + \epsilon_{ijkl}$$

- To take into account multiple levels of clustering Teerenstra et al. (2019) extended the LMM (Hussey and Hughes, 2007).

$$Y_{ijkl} = \beta_j + X_{ij}\delta + b_i + c_{ik} + \epsilon_{ijkl}$$

Gap in Knowledge

Key problem:

- Current methods available are too simplistic. Differentiating the within and between-period correlations are necessary to avoid underpowered trials in SW-CRTs (Taljaard et al., 2016).
- Extensive simulation studies are required in order to explore power estimates across various possible correlation parameters.

Key point of this talk:

- We provide a closed-form variance expression for Gaussian outcomes thus eliminating the need for extensive simulations studies.

Statistical Model

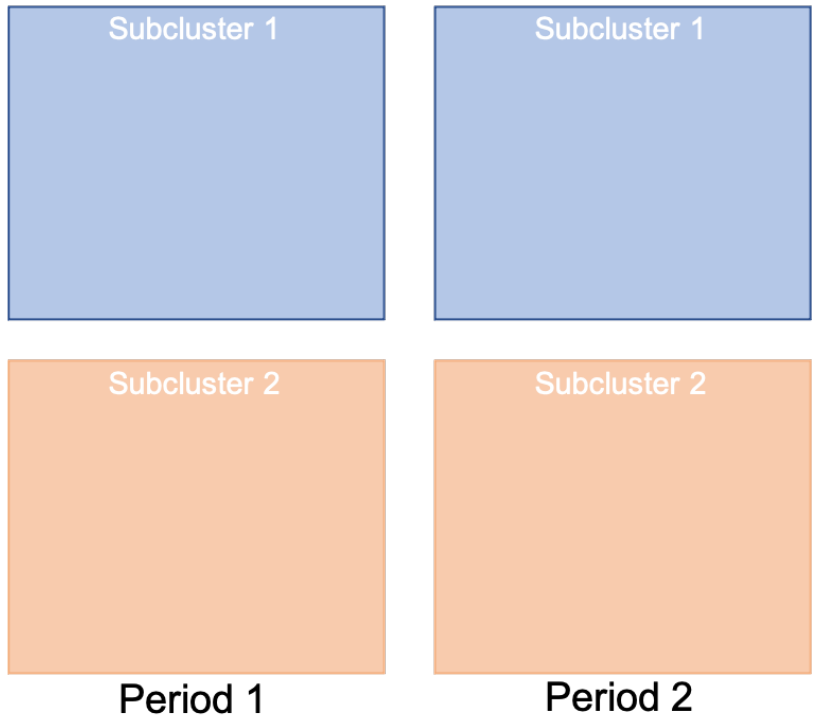
Given our outcome of interest, Y_{ijkl}, for individual $l = 1, ..., N$ nested in subcluster $k = 1, ..., K$ nested in period $j = 1, ..., T$ and cluster $i = 1, ..., I$. We are interested in the following model,

$$\text{LMM: } Y_{ijkl} = \beta_j + X_{ij}\delta + \underbrace{b_i + c_{ik}}_{\text{(sub)cluster}} + \underbrace{s_{ij} + \pi_{ijk}}_{\text{period interactions}} + \underbrace{\gamma_{ikl}}_{\text{within-person}} + \epsilon_{ijkl}$$

β_j is the effect of period j (time effect).
X_{ij} is the intervention indicator for cluster i at period j.
δ is the intervention effect.
$b_i \sim \text{Normal}(0, \sigma_b^2)$ is the random cluster effect.
$c_{ik} \sim \text{Normal}(0, \sigma_c^2)$ is the random subcluster effect.
$s_{ij} \sim \text{Normal}(0, \sigma_s^2)$ is the random cluster-by-period effect.
$\pi_{ijk} \sim \text{Normal}(0, \sigma_\pi^2)$ is the random subcluster-by-period effect.
$\gamma_{ikl} \sim \text{Normal}(0, \sigma_\gamma^2)$ is the random participant effect (if closed-cohort).
$\epsilon_{ijkl} \sim \text{Normal}(0, \sigma_\epsilon^2)$ is the error.

ICCs under Statistical Model

Under this model we have the following intracluster correlation coefficients (ICC),



where the total variance is $\sigma^2 = \sigma_b^2 + \sigma_c^2 + \sigma_s^2 + \sigma_\pi^2 + \sigma_\gamma^2 + \sigma_\epsilon^2$.

Calculating Power

The power to detect a treatment effect $\delta \neq 0$ with nominal type I error rate α is

$$\text{power} \approx \Phi_t \left(t_{\alpha/2,\text{DoF}}; \text{DoF}, |\delta| / \sqrt{\text{var}(\hat{\delta})} \right),$$

where $\Phi_t(t; \text{DoF}, \Lambda)$ is the cumulative t-distribution function with DoF degrees of freedom and noncentrality parameter Λ and $t_{\alpha/2,\text{DoF}}$ is the upper $\alpha/2$th quantile of the central t-distribution.

We used $DoF = I - 2$ which has been found to control type I error rate well (Ford and Westgate, 2020).

We assume an equal number of subclusters and participants across all clusters at all time periods such that $K_{ij} = K$ and $N_{ijk} = N$.

Variance of Intervention Effect

We can generate the variance of our intervention effect, var($\hat{\delta}$), using the feasible generalized least squares (FGLS) estimator

$$\sigma^2(\Sigma_{i=1}^{I} \boldsymbol{Z}_i' \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i)^{-1},$$

where $\boldsymbol{Z}_i$ is the design matrix and $\boldsymbol{R}_i$ is the induced correlation matrix for cluster i.

Given our correlation parameters we can express our extended block exchangeable correlation matrix, $\boldsymbol{R}_i$, as $\boldsymbol{R}_i = \boldsymbol{I}_T \otimes (\boldsymbol{B} - \boldsymbol{C}) + \boldsymbol{J}_T \otimes \boldsymbol{C}$ where

$$\boldsymbol{B} = (1 - \alpha_0)\boldsymbol{I}_{KN} + (\alpha_0 - \rho_0)\boldsymbol{I}_K \otimes \boldsymbol{J}_N + \rho_0 \boldsymbol{J}_{KN}$$
$$\boldsymbol{C} = (\alpha_2 - \alpha_1)\boldsymbol{I}_{KN} + (\alpha_1 - \rho_1)\boldsymbol{I}_K \otimes \boldsymbol{J}_N + \rho_1 \boldsymbol{J}_{KN}$$

Variance of Intervention Effect

$\boldsymbol{R}_i$ has six eigenvalues Graybill (1983),

$$\lambda_1 = 1 - \alpha_0 - \alpha_2 + \alpha_1$$
$$\lambda_2 = 1 - \alpha_0 - \alpha_2 + \alpha_1 + N(\alpha_0 - \alpha_1 - \rho_0 + \rho_1)$$
$$\lambda_3 = 1 - \alpha_0 - \alpha_2 + \alpha_1 + N(\alpha_0 - \alpha_1 + (K-1)(\rho_0 - \rho_1))$$
$$\lambda_4 = 1 - \alpha_0 + (T-1)(\alpha_2 - \alpha_1)$$
$$\lambda_5 = 1 - \alpha_0 + (T-1)(\alpha_2 - \alpha_1) + N(\alpha_0 - \rho_0 + (T-1)(\alpha_1 - \rho_1))$$
$$\lambda_6 = 1 - \alpha_0 + (T-1)(\alpha_2 - \alpha_1) + N(\alpha_0 + (T-1)\alpha_1 + (K-1)(\rho_0 + (T-1)\rho_1))$$

Using Leiva (2007) we can generate a closed-form expression of $\boldsymbol{R}_i^{-1}$

$$\boldsymbol{R}_i^{-1} = \frac{1}{\lambda_1}\boldsymbol{I}_{TKN} - \frac{\lambda_2 - \lambda_1}{N\lambda_1\lambda_2}\boldsymbol{I}_{TK} \otimes \boldsymbol{J}_N + \frac{\lambda_2 - \lambda_3}{KN\lambda_2\lambda_3}\boldsymbol{I}_T \otimes \boldsymbol{J}_{KN} + \frac{1}{T}\left(\frac{1}{\lambda_4} - \frac{1}{\lambda_1}\right)\boldsymbol{J}_T \otimes \boldsymbol{I}_{KN}$$
$$+ \frac{1}{T}\left(\frac{\lambda_2 - \lambda_1}{N\lambda_1\lambda_2} - \frac{\lambda_5 - \lambda_4}{N\lambda_4\lambda_5}\right)\boldsymbol{J}_T \otimes \boldsymbol{I}_K \otimes \boldsymbol{J}_N + \frac{1}{TK}\left(\frac{\lambda_5 - \lambda_6}{N\lambda_5\lambda_6} - \frac{\lambda_2 - \lambda_3}{N\lambda_2\lambda_3}\right)\boldsymbol{J}_{TKN}.$$

Variance of Intervention Effect

Using $\boldsymbol{R}_i^{-1}$ we can derive a closed-form expression for the variance of the intervention effect

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

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 $W = \sum_{j=1}^{T}(\sum_{i=1}^{I} X_{ij})^2$ are the same design constants used by Hussey and Hughes (2007) and others.

$$\lambda_3 = 1 - \alpha_0 - \alpha_2 + \alpha_1 + N(\alpha_0 - \alpha_1 + (K-1)(\rho_0 - \rho_1))$$
$$\lambda_6 = 1 - \alpha_0 + (T-1)(\alpha_2 - \alpha_1) + N(\alpha_0 + (T-1)\alpha_1 + (K-1)(\rho_0 + (T-1)\rho_1))$$

Connection to other design variants:

1. Closed-cohort on subcluster level and cross-sectional on participant level ($\alpha_2 = \alpha_1$)
2. Cross-sectional on subcluster (and participant) level ($\alpha_2 = \alpha_1 = \rho_1$)

This expression can be used for each of the three design variants and any type of longitudinal CRT (parallel or crossover designs).

Design Effect

The variance ratio under a multilevel cluster randomized trial design to individual randomization is

$$\text{design effect} = \frac{I^2 T \lambda_6 \lambda_3}{4(U^2 + ITU - TW - IV)\lambda_6 - 4(U^2 - IV)\lambda_3}.$$

$$\lambda_3 = 1 - \alpha_0 - \alpha_2 + \alpha_1 + N(\alpha_0 - \alpha_1 + (K-1)(\rho_0 - \rho_1))$$
$$\lambda_6 = 1 - \alpha_0 + (T-1)(\alpha_2 - \alpha_1) + N(\alpha_0 + (T-1)\alpha_1 + (K-1)(\rho_0 + (T-1)\rho_1))$$

- Design effect increases with increasing within-period ICCs (α_0 and ρ_0).
- Design effect *typically* increases with decreasing between-period ICCs (α_1, ρ_1, and α_2).

Simulation Study: Overview

To validate our sample size methodology we conducted a simulation study. We assumed design variant (B), closed-cohort on the subcluster level and cross-sectional at the individual level ($\alpha_1 = \alpha_2$).

- Number of clusters, I, varied between 8 and 30.
- Number of subclusters, K, varied between 2 and 6.
- Subcluster sizes, N, up to 15.
- Number of periods, T, varied between 4 and 7.
- Standardized effect sizes, δ/σ, ranged between 0.1 and 0.5 (for Gaussian outcomes).
- Three sets of ICCs representing small and large correlations.

Simulation Study: Results (n=1000)

δ/σ	α_0	α_1	ρ_0	ρ_1	I	K	N	T	Test Size	Empirical	Predicted
0.1	0.03	0.015	0.0075	0.00375	24	6	15	7	3.6	88.2	85.3
0.1	0.01	0.005	0.0025	0.00125	30	6	15	4	4.3	82.5	82.2
0.2	0.1	0.05	0.025	0.0125	24	6	10	4	4.5	85.4	83.3
0.2	0.03	0.015	0.0075	0.00375	15	3	10	6	4.0	81.1	80.9
0.2	0.01	0.005	0.0025	0.00125	10	4	10	6	2.7	79.4	80.2
0.25	0.1	0.05	0.025	0.0125	21	4	10	4	5.2	83.7	84.6
0.25	0.03	0.015	0.0075	0.00375	12	2	10	7	3.5	81.0	80.3
0.25	0.01	0.005	0.0025	0.00125	24	2	8	4	4.1	83.7	84.3
0.35	0.1	0.05	0.025	0.0125	10	3	8	6	1.9	83.6	83.1
0.35	0.03	0.015	0.0075	0.00375	9	3	12	4	2.4	84.3	83.8
0.35	0.01	0.005	0.0025	0.00125	8	3	7	5	1.7	77.7	80.4
0.4	0.1	0.05	0.025	0.0125	18	2	7	4	3.2	87.9	86.2
0.4	0.03	0.015	0.0075	0.00375	8	3	7	5	1.1	84.3	83.9
0.4	0.01	0.005	0.0025	0.00125	15	2	5	4	3.3	81.2	83.3
0.5	0.1	0.05	0.025	0.0125	12	2	4	5	3.2	84.3	82.6
0.5	0.03	0.015	0.0075	0.00375	9	2	8	4	1.8	87.4	85.8

Time effect used in simulation: $\beta_1 = 0$ with $\beta_{j+1} - \beta_j = 0.1 \times (0.5)^{j-1}$ for $j \geq 1$.

Application to LIRE Trial

The LIRE trial (Jarvik et al., 2015) randomized 100 clinics consisting of a total of 1700 primary care providers (PCP) over 6 periods.

- Assuming an equal number of PCPs per clinic we have: $I = 100$, $K = 17$, and $T = 6$.

The outcome of interest was spine-related RVUs, a composite measure of back pain. Assuming an effect size of -0.1 and total variance of 2.5, we are interested in calculating the required number of patients per PCP, N, to achieve at least 80% power at the 5% nominal test size.

- Using the ICC estimates from the study design, we assume the following: $\alpha_0 = 0.046$, $\alpha_1 = 0.023$, $\rho_0 = 0.040$, $\rho_1 = 0.020$.
- Using our closed-form expression for $\text{var}(\hat{\delta})$ and the power formula we found that having 77 participants per PCP, $N = 77$, produced 87.5% power.

Additional & Future Work

Additional extensions include:

- Non-Gaussian outcomes under a GLMM.
- Unequal cluster sizes.

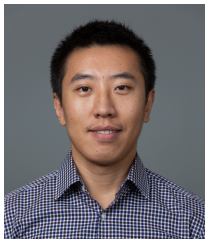
This work was recently published in *Biometrics Methodology* and is available online: `https://onlinelibrary.wiley.com/doi/10.1111/biom.13596`.

Future work in this area includes:

- Extending the current methodology to accommodate a decaying correlation structure.
- Open enrollment.

Collaborators & Funding Source

Thank you to my collaborators:



Dr. Fan Li, Yale University School of Public Health

Dr. Monica Taljaard, Ottawa Hospital Research Institute

Funding Source: Supported by the National Institute of Aging (NIA) of the National Institutes of Health under Award Number U54AG063546, which funds NIA Imbedded Pragmatic Alzheimer's Disease and AD-Related Dementias Clinical Trials Collaboratory (NIA IMPACT Collaboratory). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References I

Ford, W. P. and Westgate, P. M. (2020). Maintaining the validity of inference in small-sample stepped wedge cluster randomized trials with binary outcomes when using generalized estimating equations. *Statistics in Medicine*, 39(21):2779–2792.

Golden, M. R., Kerani, R. P., Stenger, M., Hughes, J. P., Aubin, M., Malinski, C., and Holmes, K. K. (2015). Uptake and population-level impact of expedited partner therapy (ept) on chlamydia trachomatis and neisseria gonorrhoeae: the washington state community-level randomized trial of ept. *PLoS Med*, 12(1):e1001777.

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(26):4718–4728.

Hussey, M. A. and Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary clinical trials*, 28(2):182–191.

Jarvik, J. G., Comstock, B. A., James, K. T., Avins, A. L., Bresnahan, B. W., Deyo, R. A., Luetmer, P. H., Friedly, J. L., Meier, E. N., Cherkin, D. C., et al. (2015). Lumbar imaging with reporting of epidemiology (lire)—protocol for a pragmatic cluster randomized trial. *Contemporary clinical trials*, 45:157–163.

Leiva, R. (2007). Linear discrimination with equicorrelated training vectors. *Journal of multivariate analysis*, 98(2):384–409.

References II

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(4):459–463.

Teerenstra, S., Taljaard, M., Haenen, A., Huis, A., Atsma, F., Rodwell, L., and Hulscher, M. (2019). Sample size calculation for stepped-wedge cluster-randomized trials with more than two levels of clustering. *Clinical Trials*, 16(3):225–236.

Thank you!