If you're a person with a disability or using assistive technology and are having difficulty accessing any of the content in this file, please contact Astrid Masfar at astrid.masfar@nih.gov.

Methods: Mind the Gap

Webinar Series

Overview of Statistical Models for the Design and Analysis of Stepped Wedge Cluster Randomized Trials

Presented by:

Fan Li, Ph.D.

Yale University School of Public Health

NIH National Institutes of Health

Office of Disease Prevention

Overview of Statistical Models for the Design and Analysis of Stepped Wedge Cluster Randomized Trials

Fan Li

Assistant Professor
Department of Biostatistics, and
Center for Methods in Implementation and Prevention Science
Yale University School of Public Health

Office of Disease Prevention: Mind the Gap Seminar Series

July 14, 2020

Yale SCHOOL OF PUBLIC HEALTH

Acknowledgement

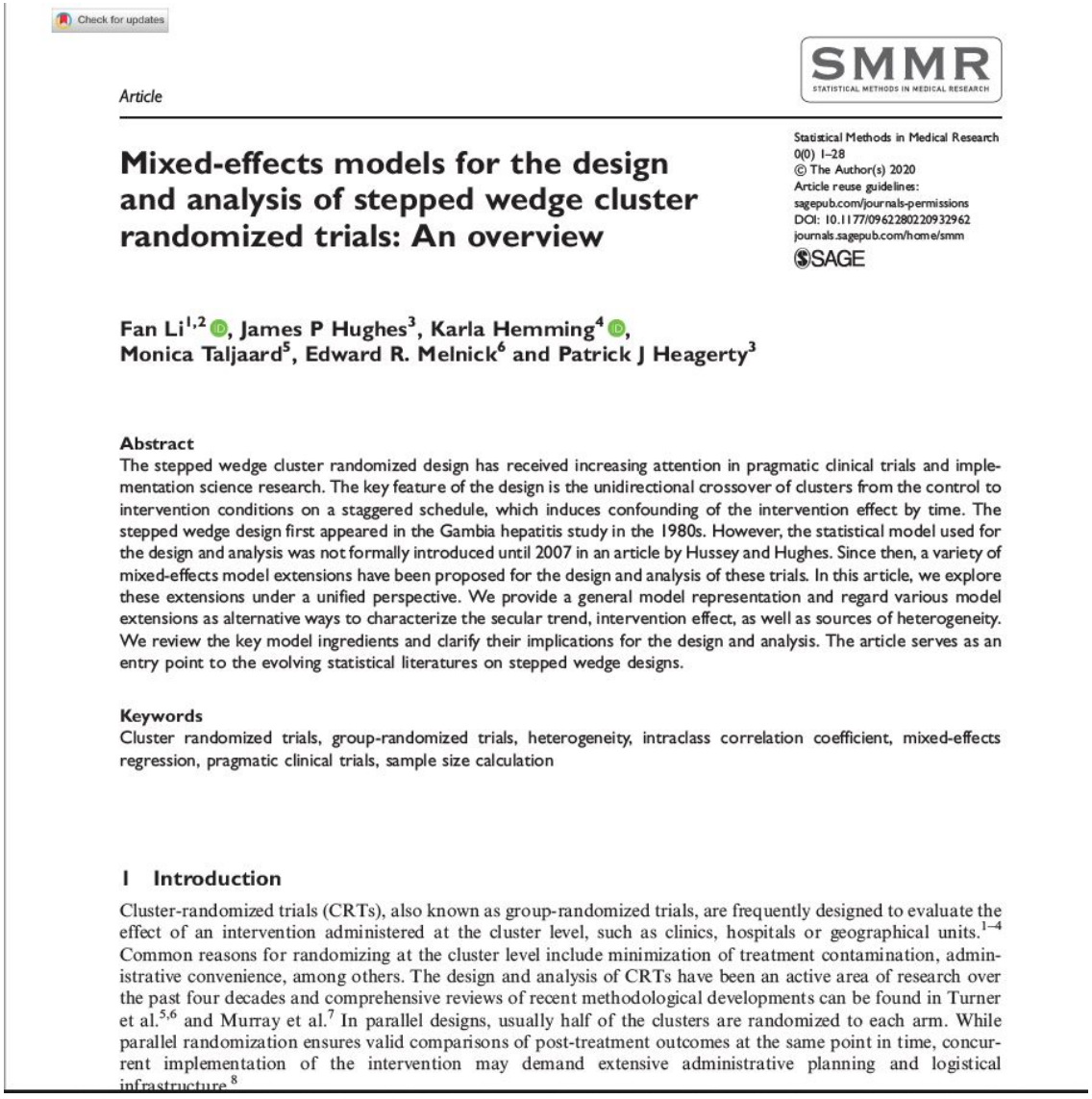
- NIH Collaboratory Biostatistics and Study Design Core Working Group
  - Drs. David Murray & Patrick Heagerty for discussions leading up to this article as part of the work from the working group
- Collaborators on the SMMR article[1]
  - James Hughes (UW), Patrick Heagerty (UW), Edward Melnick (Yale), Karla Hemming (UBir), Monica Taljaard (OHRI)
- Other collaborators
  - John Preisser (UNC), Elizabeth Turner (Duke), Paul Rathouz (UT Austin), Donna Spiegelman (Yale), Xin Zhou (Yale), Hengshi Yu (Umich), Zibo Tian (Yale), Jiachen Chen (Yale) . . .

[1] Li F. et al (2020) Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Stat Methods Med Res*

Funding Acknowledgement

Supported by

- Administrative supplement 3-UH3-DA047003-02S2 from the NIH Office of Disease Prevention & National Institute on Drug Abuse
- PCORI award ME-1507-31750 (PI: Heagerty)



Check for updates

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Article

Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview

Statistical Methods in Medical Research
0(0) 1–28

DOI: 10.1177/0962280220932962
journals.sagepub.com/home/smm
SAGE

Fan Li[1,2], James P Hughes[3], Karla Hemming[4], Monica Taljaard[5], Edward R. Melnick[6] and Patrick J Heagerty[3]

Abstract

The stepped wedge cluster randomized design has received increasing attention in pragmatic clinical trials and implementation science research. The key feature of the design is the unidirectional crossover of clusters from the control to intervention conditions on a staggered schedule, which induces confounding of the intervention effect by time. The stepped wedge design first appeared in the Gambia hepatitis study in the 1980s. However, the statistical model used for the design and analysis was not formally introduced until 2007 in an article by Hussey and Hughes. Since then, a variety of mixed-effects model extensions have been proposed for the design and analysis of these trials. In this article, we explore these extensions under a unified perspective. We provide a general model representation and regard various model extensions as alternative ways to characterize the secular trend, intervention effect, as well as sources of heterogeneity. We review the key model ingredients and clarify their implications for the design and analysis. The article serves as an entry point to the evolving statistical literatures on stepped wedge designs.

Keywords

Cluster randomized trials, group-randomized trials, heterogeneity, intraclass correlation coefficient, mixed-effects regression, pragmatic clinical trials, sample size calculation

1 Introduction

Cluster-randomized trials (CRTs), also known as group-randomized trials, are frequently designed to evaluate the effect of an intervention administered at the cluster level, such as clinics, hospitals or geographical units.[1–4] Common reasons for randomizing at the cluster level include minimization of treatment contamination, administrative convenience, among others. The design and analysis of CRTs have been an active area of research over the past four decades and comprehensive reviews of recent methodological developments can be found in Turner et al.[5,6] and Murray et al.[7] In parallel designs, usually half of the clusters are randomized to each arm. While parallel randomization ensures valid comparisons of post-treatment outcomes at the same point in time, concurrent implementation of the intervention may demand extensive administrative planning and logistical infrastructure.[8]

1. Introduction

Introduction

- Cluster randomized trials (CRTs) allocate clusters of individuals to intervention
  - minimize contamination
  - administrative convenience
  - usually in parallel design
- Stepped wedge (SW) design rolls out intervention in a staggered fashion
  - logistical constraints
  - perceived ethical benefit
- Other pros and cons discussed extensively by Hemming and Taljaard[2]

[2] Hemming, K., Taljaard, M. (2020). Reflection on modern methods: when is a stepped-wedge cluster randomized trial a good study design choice?. *Int. J. Epidemiol.*.

Analytical Models

- Unique features requires of SW designs require more complex considerations on analytical models
- Mixed-effects models
  - seminal methods paper of Hussey and Hughes (2007)[3]
  - fixed-effects for time & intervention
  - random-effects for clustering
- Among other modeling alternatives, mixed-effects models are more accessible from standard software, and are most widely used in SW-CRTs (Barker et al. 2016, BMC Med. Res. Methodol.)

[3] Hussey MA, Hughes JP (2007) Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials*

Analytical Models - Cont'd

- Many extensions of basic model over past decade
- Systematic reviews (Davey et al. 2015; Martin et al. 2016; Barker et al. 2016; Grayling et al. 2017)
  - statistical methods for the sample size determination varied across studies
  - insufficient details on modeling assumptions were provided
- reproducibility and sensitivity
- Integrate the toolkit of analytical models for SW-CRTs
  - essential ingredients?
  - common variants?
  - identify areas that need further development or assessment

CONSORT extension to SW-CRTs

- CONSORT item 7a: Sample size[4]
  - *Extension for SW-CRTs – . . . Method of calculation and relevant parameters with sufficient detail so the calculation can be replicated. Assumptions made about correlations between outcomes of participants from the same cluster*.
- CONSORT item 12a & b: Statistical methods
  - *Extension for SW-CRTs – . . . Statistical methods used to compare treatment conditions for primary and secondary outcomes including how time effects, clustering, and repeated measures were taken into account.*.
- High-level, general model representation
  - introduce model variants
  - clarify assumptions and implications

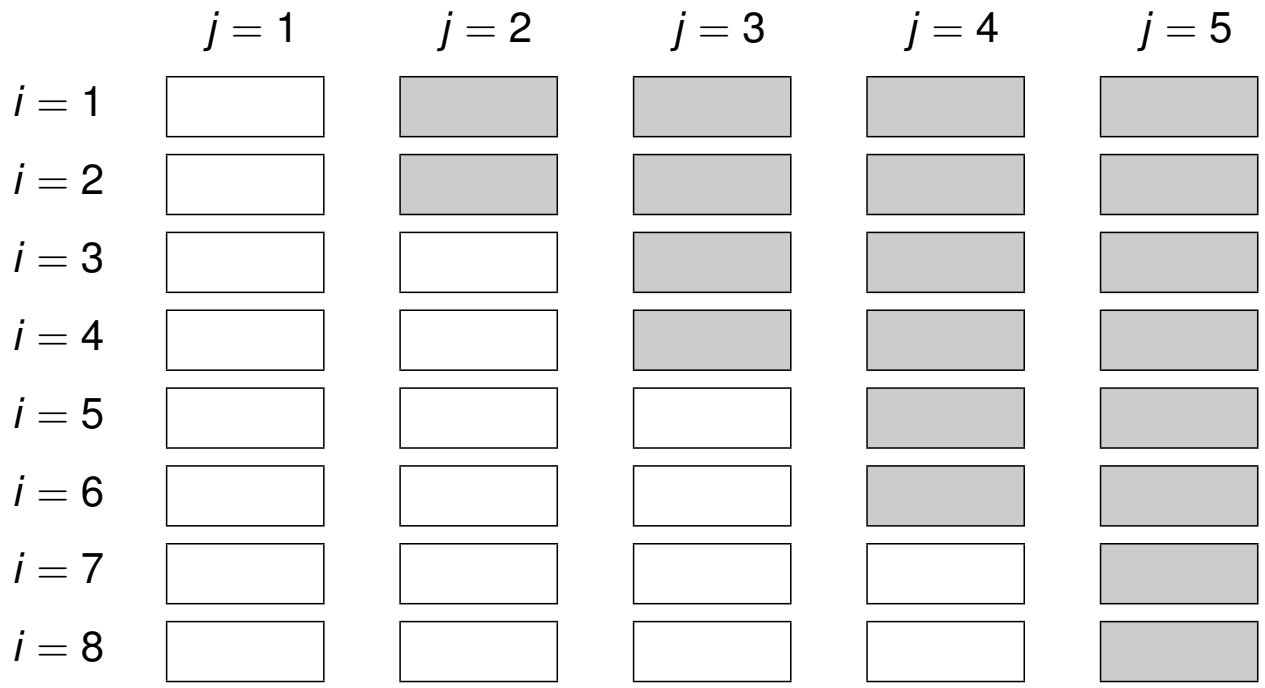
[4] Hemming K (2018) Reporting of stepped wedge cluster randomised trials: Extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ*

2. General Model Representation

Terminology

- Consider a *complete* stepped wedge CRT with I participating clusters followed over J ($J \geq 3$) time periods
- *Cross-sectional* (CS)
  - different individuals observed in each cluster over time
  - assume N_{ij} individuals are included during period j in cluster i
- *Closed-cohort* (CC)
  - individuals identified at the start of the trial and scheduled for repeated outcome assessment
  - N_i as the cohort size in cluster i as repeated measurements are taken from the same individuals
- *Open-cohort* (OC): a mix of the two
- Cluster starts out in the control; sets of clusters randomized to intervention until all clusters exposed

Cluster-Period Diagram



- schematic illustration with $I = 8$ clusters and $J = 5$ periods. Each white cell indicates a cluster-period under the control condition and each gray cell indicates a cluster-period under the intervention condition. There are in total $S = 4$ distinct intervention sequences
- each one of the 4 distinct intervention sequences is fully determined by the time period during which the intervention is first implemented

Outcome Model

- $Y_{ijk}(s)$: the outcome of individual k during period j in cluster i, had cluster i received an intervention sequence s
- Mean model

$$g[\mu_{ijk}(s)] = \boldsymbol{F}_i(j,s)'\boldsymbol{\theta} + \boldsymbol{R}_{ik}(j,s)'\boldsymbol{\alpha}_i,$$

  - $\mu_{ijk}(s)$ conditional mean of $Y_{ijk}(s)$, g link function
  - $\boldsymbol{F}_i(j,s)'\boldsymbol{\theta}$: the group-average outcome trajectory and $\boldsymbol{\theta}$ includes the parameter of interest (i.e., the intervention effect)
  - $\boldsymbol{R}_{ik}(j,s)'\boldsymbol{\alpha}_i$: the cluster-specific, time-specific and/or individual-specific departure from the group average
- A GLM, but borrow the *"potential outcome"* language to clearly indicate the dependence of elements on intervention sequence s[5]

[5] Sitlani CM, Heagerty PJ et al. (2012) Longitudinal structural mixed models for the analysis of surgical trials with noncompliance. *Stat Med*

Outcome Model - Cont'd

- Separate $\boldsymbol{F}_i(j,s)'\boldsymbol{\theta} = \boldsymbol{F}^0(j)'\boldsymbol{\beta} + F_i^1(j,s)\Delta(j,s)$
  - baseline component $\boldsymbol{F}^0(j)$ characterizing the background secular trend in the absence of intervention
  - a time-dependent intervention component $F_i^1(j,s) = \mathbb{I}_{[j \geq s]}$
  - $\boldsymbol{\beta}$ is the parameter encoding the secular trend
  - $\Delta(j,s)$ is the change in the mean outcome at period j due to sequence s
- General representation of outcome model

$$g[\mu_{ijk}(s)] = \underbrace{\boldsymbol{F}^0(j)'\boldsymbol{\beta}}_{\text{secular trend}} + \underbrace{F_i^1(j,s)\Delta(j,s)}_{\text{intervention effect}} + \underbrace{\boldsymbol{R}_{ik}(j,s)'\boldsymbol{\alpha}_i}_{\text{heterogeneity}}.$$

- Useful for conceptualizing model elements

Outcome Model - Cont'd

- $Y_{ijk}(s)$ is then assumed to follow a parametric distribution with mean $\mu_{ijk}(s)$ and variance as a function of $\mu_{ijk}(s)$
- Continuous, normally distributed outcome, obtain the *linear mixed model* (LMM)

$$Y_{ijk}(s) = \boldsymbol{F}^0(j)'\boldsymbol{\beta} + F_i^1(j,s)\Delta(j,s) + \boldsymbol{R}_{ik}(j,s)'\boldsymbol{\alpha}_i + \epsilon_{ijk}$$
$$\boldsymbol{\alpha}_i \sim f(\boldsymbol{\alpha}_i;\boldsymbol{\Theta}), \quad \epsilon_{ijk} \sim N(0,\sigma_\epsilon^2)$$

- Heterogeneity parameter $\boldsymbol{\alpha}_i$ induces within-cluster correlations, or intraclass correlation coefficients (ICCs)
- Equate $Y_{ijk} = Y_{ijk}(s)$, if cluster i receives sequence s
- Current literature has focused on a continuous outcome, review existing linear mixed model variants as special cases of the general representation

3. Modeling Considerations & Implications

The Hussey and Hughes (HH) Model

- $Y_{ijk} = \mu + \beta_j + \delta X_{ij} + \alpha_i + \epsilon_{ijk}$[6]
  - μ grand mean, β_j is the jth period effect ($\beta_1 = 0$)
  - X_{ij} intervention indicator, δ the intervention effect
  - $\alpha_i \sim N(0, \tau^2_\alpha)$, $\epsilon_{ijk} \sim N(0, \sigma^2_\epsilon)$ independent of α_i
- Secular trend $\boldsymbol{F}^0(j)'\boldsymbol{\beta} = \mu + \beta_2 \mathbb{I}_{[j=2]} + \ldots + \beta_J \mathbb{I}_{[j=J]}$
  - Intervention effect, $\Delta(j, s) = \delta$, does not depend on the time interval during which the intervention was initiated
- Heterogeneity, $\boldsymbol{R}_{ik}(j, s)'\boldsymbol{\alpha}_i = \alpha_i$, captures the cluster-specific departure from the average but is assumed to be homogeneous across time periods and intervention sequences

[6] Hussey MA, Hughes JP (2007) Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials*

The Hussey and Hughes Model - Cont'd

- Analogous to parallel CRT, the single cluster random effect postulates a *simple exchangeable* correlation structure
- Common ICC: $\rho = \tau_\alpha^2/(\tau_\alpha^2 + \sigma_\epsilon^2)$
- Sample size calculation

$$\text{var}(\hat{\delta}) = \frac{(\sigma_{\text{tot}}^2/N)IJ\lambda_1\lambda_2}{(U^2 + IJU - JW - IV)\lambda_2 - (U^2 - IV)\lambda_1},$$

  - $\sigma_{\text{tot}}^2 = \tau_\alpha^2 + \sigma_\epsilon^2$, $U = \sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}$, $W = \sum_{j=1}^{J}(\sum_{i=1}^{I} X_{ij})^2$ and $V = \sum_{i=1}^{I}(\sum_{j=1}^{J} X_{ij})^2$ are design constants
  - $\lambda_1 = 1 - \rho$, $\lambda_2 = 1 + (JN - 1)\rho$ eigenvalues of corr structure
- Many subsequent development based on the HH model
- $\lim_{N\to\infty} \text{var}(\hat{\delta}) = 0$, but most widely used[7]

[7] Taljaard M et al. (2016) Substantial risks associated with few clusters in cluster randomized and stepped wedge designs. *Clinical Trials*

Modeling Secular Trend

- Generally, $\boldsymbol{F}^0(j)'\boldsymbol{\beta} = \beta_1 B_1(j) + \ldots + \beta_p B_p(j)$
  - $\boldsymbol{F}^0(j) = (B_1(j), \ldots, B_p(j))'$ p-dimensional basis function
- Alternative approaches include linear specification $\boldsymbol{F}^0(j) = (1, j)'$ or polynomial specification etc
- From an efficiency perspective, favor the dimension of $\boldsymbol{\beta}$ controlled, except that
  - var$(\hat{\delta})$ invariant to time parameterization as long $\left(\sum_{i=1}^{I} X_{i1}, \ldots, \sum_{i=1}^{I} X_{iJ}\right)'$ lay in the column space of $\boldsymbol{F}^0 = (\boldsymbol{F}^0(1), \ldots, \boldsymbol{F}^0(J))'$[8]— *balanced* allocation
- From a bias perspective, natural to consider a nonparametric representation of $\boldsymbol{F}^0(j)'\boldsymbol{\beta}$ (as in HH model)

[8] Grantham KL et al. (2019) Time parameterizations in cluster randomized trial planning. *Am Stat*.

Modeling the Intervention Effect (a)

- $\Delta(j, s) = \delta$ is a constant/averaged intervention effect

(a) *constant intervention effect* $\Delta(j, s) = \delta$

$i = 1$	0	δ	δ	δ	δ
$i = 2$	0	0	δ	δ	δ
$i = 3$	0	0	0	δ	δ
$i = 4$	0	0	0	0	δ

- Easy to work with, especially in the design stage
- Does not allow the strengthening or weakening of effect over time

Modeling the Intervention Effect (b)

- $\Delta(j, s)$ can depend on period j and sequence s[9]
- *Linear time-on-treatment effect*

$$\Delta(j, s) = \delta_0 + \delta_1(j - s), \text{or}, \quad \Delta(j, s) = \delta(j - s + 1)$$

(b) linear time-on-treatment $\Delta(j, s) = \delta_0 + \delta_1(j - s)$

$i = 1$	0	δ_0	$\delta_0 + \delta_1$	$\delta_0 + 2\delta_1$	$\delta_0 + 3\delta_1$
$i = 2$	0	0	δ_0	$\delta_0 + \delta_1$	$\delta_0 + 2\delta_1$
$i = 3$	0	0	0	δ_0	$\delta_0 + \delta_1$
$i = 4$	0	0	0	0	δ_0

- Considers strengthening or weakening of effect over time
- Model in analyzing longitudinal parallel CRTs

[9] Hughes JP et al. (2015) Current issues in the design and analysis of stepped wedge trials. *Contemp Clin Trials*

Modeling the Intervention Effect (c)

- *Delayed treatment effect* [9]

$$\Delta(j,s) = \delta\pi_0 \mathbb{I}_{[j=s]} + \delta \mathbb{I}_{[j>s]},$$

(c) *delayed treatment effect* $\Delta(j,s) = \delta\pi_0 \mathbb{I}_{[j=s]} + \delta \mathbb{I}_{[j>s]}$

$i=1$	0	$\pi_0\delta$	δ	δ	δ
$i=2$	0	0	$\pi_0\delta$	δ	δ
$i=3$	0	0	0	$\pi_0\delta$	δ
$i=4$	0	0	0	0	$\pi_0\delta$

- *General delayed treatment effect*

$$\Delta(j,s) = \delta\pi_0 \mathbb{I}_{[j=s]} + \delta\pi_1 \mathbb{I}_{[j=s+1]} + \ldots + \delta\pi_{J-s} \mathbb{I}_{[j=J]}.$$

- Prior knowledge or assumptions on π_{j-s} ($\pi_{j-s} = 0$ if $j < s$)

Modeling the Intervention Effect (d)

- *General time-on-treatment effect*[9]

$$\Delta(j,s) = \delta_{j-s} = \delta_0 \mathbb{I}_{[j=s]} + \delta_1 \mathbb{I}_{[j=s+1]} + \ldots + \delta_{J-s} \mathbb{I}_{[j=J]}.$$

(d) *general time-on-treatment* $\Delta(j,s) = \delta_{j-s}$

$i=1$	0	δ_0	δ_1	δ_2	δ_3
$i=2$	0	0	δ_0	δ_1	δ_2
$i=3$	0	0	0	δ_0	δ_1
$i=4$	0	0	0	0	δ_0

- Interpretable global tests
  - H_0: $\delta_0 = \delta_1 = \ldots = \delta_{J-2} = 0$ (no intervention effect)
  - H_0: $\delta_0 = \delta_1 = \ldots = \delta_{J-2}$ (constant intervention effect)
  - H_0: $\delta_1 - \delta_0 = \delta_2 - \delta_1 = \ldots$ (linear time-on-treatment)

Preclude for Modeling Heterogeneity

- There has been extensive discussions of alternative strategies for modeling the random-effects structure in stepped wedge trials
- Centered on extensions to the Hussey and Hughes model
  - constant intervention effect
  - categorical time parameterization
- review variants of random-effects structures by assuming a linear link, categorical secular trend (with one exception) as well as a time-invariant intervention effect

Modeling Heterogeneity in CS Designs

- Example extensions to the Hussey and Hughes model cross-sectional designs; all models assume a continuous outcome and an identity link function[10]

Extension	Feature	Example references
Nested Exchangeable*	Distinguish between within-period and between-period ICCs	Hooper et al. (2016)[56] Girling and Hemming (2016)[38]
Exponential Decay*	Allow the between-period ICC to decay at an exponential rate over time	Kasza et al. (2017)[57] Kasza and Forbes (2018)[61]
Random Intervention	Include random cluster-specific intervention effects, and ICC depends on intervention status	Hughes et al. (2015)[55] Hemming et al. (2017)[47]
Random Coefficient	Include random cluster-specific time slopes; ICC tends to be an increasing function of distance in time	Murray et al. (1998)[58]

[10] Li F. et al (2020) Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Stat Methods Med Res*

Nested Exchangeable Correlation Model

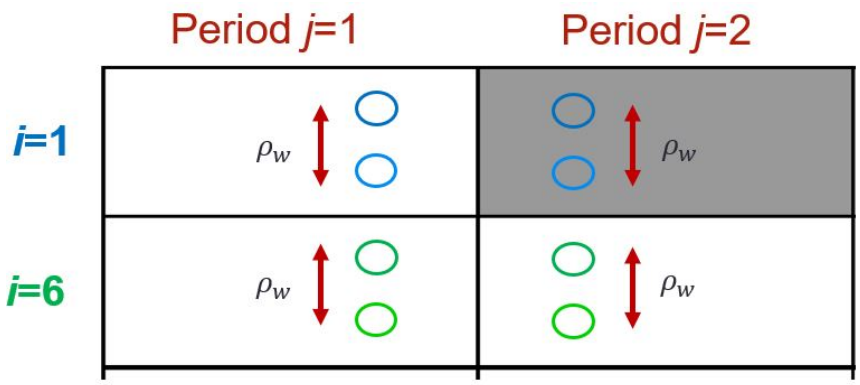
$$Y_{ijk} = \mu + \beta_j + \delta X_{ij} + \alpha_i + \gamma_{ij} + \epsilon_{ijk}$$

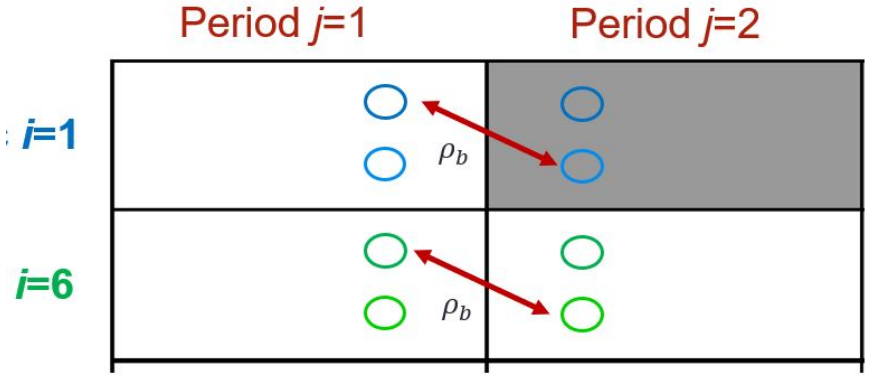
- $\gamma_{ij} \sim N(0, \tau_\gamma^2)$, the random cluster-by-time interaction
- $\boldsymbol{R}_{ik}(j, s)'\boldsymbol{\alpha}_i = \alpha_i + \gamma_{ij}$, therefore allows the deviation from the group average to be both cluster-specific and period-specific
- distinguishes between *within-period* ICC and *between-period* ICC[11]

$$\text{corr}[Y_{ijk}(s), Y_{ilm}(s)] = \begin{cases} \rho_w = (\tau_\alpha^2 + \tau_\gamma^2)/(\tau_\alpha^2 + \tau_\gamma^2 + \sigma_\epsilon^2), & j = l \\ \rho_b = \tau_\alpha^2/(\tau_\alpha^2 + \tau_\gamma^2 + \sigma_\epsilon^2), & j \neq l, \end{cases}$$

[11] Hooper R et al (2016) Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. *Stat Med*

Nested Exchangeable Correlation Model - Cont'd





- Sample size calculation takes into account both ICCs
  - Same form of variance with HH model, except that

$$\sigma^2_{\text{tot}} = \tau^2_\alpha + \tau^2_\gamma + \sigma^2_\epsilon$$
$$\lambda_1 = 1 + (N-1)\rho_{\text{w}} - N\rho_{\text{b}}$$
$$\lambda_2 = 1 + (N-1)\rho_{\text{w}} + N(J-1)\rho_{\text{b}}$$

- Hooper/Girling model (CAC and CMC) [11,12]
- $\lim_{N\to\infty} \text{var}(\hat{\delta}) \neq 0$

[12] Girling AJ and Hemming K (2016) Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. *Stat Med*

Exponential Decay Model

$$Y_{ijk} = \mu + \beta_j + \delta X_{ij} + \gamma_{ij} + \epsilon_{ijk}$$[13]

- Heterogeneity term $\boldsymbol{R}_{ik}(j,s)'\boldsymbol{\alpha}_i = \gamma_{ij}$
- $\boldsymbol{\gamma}_i = (\gamma_{i1}, \ldots, \gamma_{iJ})' \sim N(0, \tau_\gamma^2 \boldsymbol{M})$

$$\boldsymbol{M} = \begin{pmatrix} 1 & r_0 r & r_0 r^2 & \ldots & r_0 r^{J-1} \\ r_0 r & 1 & r_0 r & \ldots & r_0 r^{J-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_0 r^{J-1} & r_0 r^{J-2} & r_0 r^{J-3} & \ldots & 1 \end{pmatrix}$$

- Allows between-period ICC to decay exponentially

$$\text{corr}[Y_{ijk}(s), Y_{ilm}(s)] = \begin{cases} \rho_w = \tau_\gamma^2/(\tau_\gamma^2 + \sigma_\epsilon^2), & j = l \\ \rho_{b,|j-l|} = \rho_w r^{|j-l|}, & j \neq l. \end{cases}$$

[13] Kasza J et al (2018). Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Stat Methods Med Res*

Exponential Decay Model - Cont'd

- Exponential decay and nested exchangeable correlation models do not have a clear nesting relationship
- Both can reduce to the HH model ($\rho_b = \rho_w$ or $r = 1$)
- var($\hat{\delta}$) does NOT exist in closed form
- From the *design* perspective, estimated sample size can go either direction when the incorrect model is assumed
- From the *analysis* perspective, omitting the decay parameter might lead to an inflated Type I error rate[14]
- The version of continuous-time correlation decay model (Grantham et al. 2020, *Stat Med*)

[14] Kasza J and Forbes AB (2018). Inference for the treatment effect in multiple-period cluster randomised trials when random effect correlation structure is misspecified *Stat Methods Med Res*

Random Intervention Model

$$Y_{ijk} = \mu + \beta_j + (\delta + \nu_i)X_{ij} + \alpha_i + \epsilon_{ijk},$$

where

$$\begin{pmatrix} \alpha_i \\ \nu_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_\alpha^2 & \sigma_{\alpha\nu} \\ \sigma_{\alpha\nu} & \tau_\nu^2 \end{pmatrix}\right],$$

- random cluster-by-treatment interaction (Hughes et al. 2015)[9]
- heterogeneity term $\boldsymbol{R}_{ik}(j,s)'\boldsymbol{\alpha}_i = \alpha_i + \nu_i \mathbb{I}_{[j \geq s]}$
- intervention-condition-specific correlation structures (treatment also affects variance components)
- careful on alternative parameterization to avoid strong assumptions[15]

[15] Hemming K et al (2018) Modeling clustering and treatment effect heterogeneity in parallel and stepped-wedge cluster randomized trials. *Stat Med*

Random Coefficient Model

$$Y_{ijk} = \mu + (\beta + \xi_i) T_j + \delta X_{ij} + \alpha_i + \epsilon_{ijk}.$$

- $T_j = j$ to represent the linear time basis function
- β as the fixed time slope and ξ_i as the random slope

$$\begin{pmatrix} \alpha_i \\ \xi_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_\alpha^2 & \sigma_{\alpha\xi} \\ \sigma_{\alpha\xi} & \tau_\xi^2 \end{pmatrix} \right].$$

- heterogeneity term $\boldsymbol{R}_{ik}(j, s)'\boldsymbol{\alpha}_i = \alpha_i + j\xi_i$
- used in longitudinal parallel CRTs[16]
- ICC structure

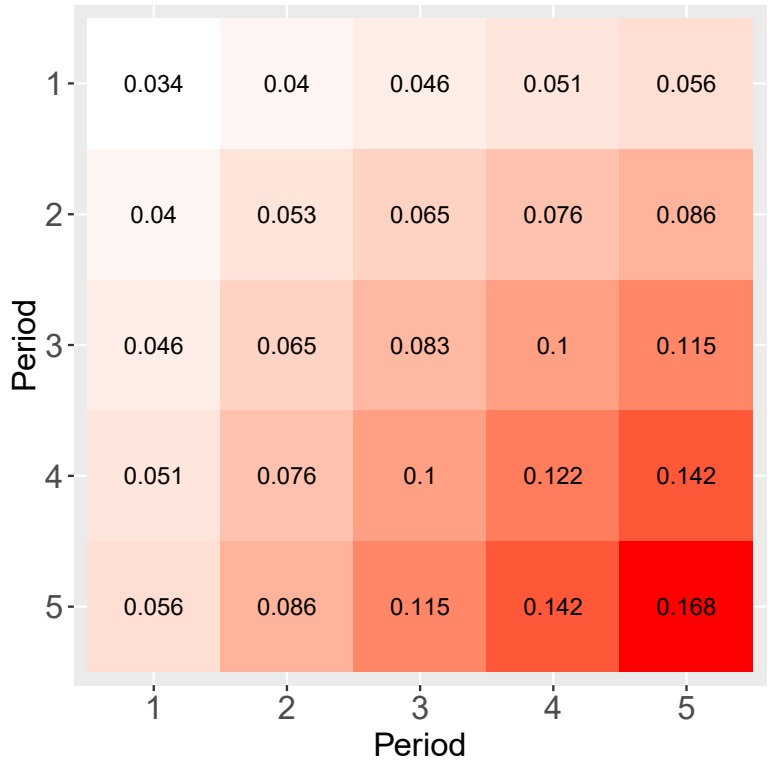
$$\text{corr}[Y_{ijk}(s), Y_{ilm}(s)] = \frac{\tau_\alpha^2 + (j + l)\sigma_{\alpha\xi} + jl\tau_\xi^2}{\sqrt{\tau_\alpha^2 + 2j\sigma_{\alpha\xi} + j^2\tau_\xi^2 + \sigma_\epsilon^2}\sqrt{\tau_\alpha^2 + 2l\sigma_{\alpha\xi} + l^2\tau_\xi^2 + \sigma_\epsilon^2}},$$

[16] Murray DM et al (1998). Analysis of data from group-randomized trials with repeat observations on the same groups. *Stat Med*

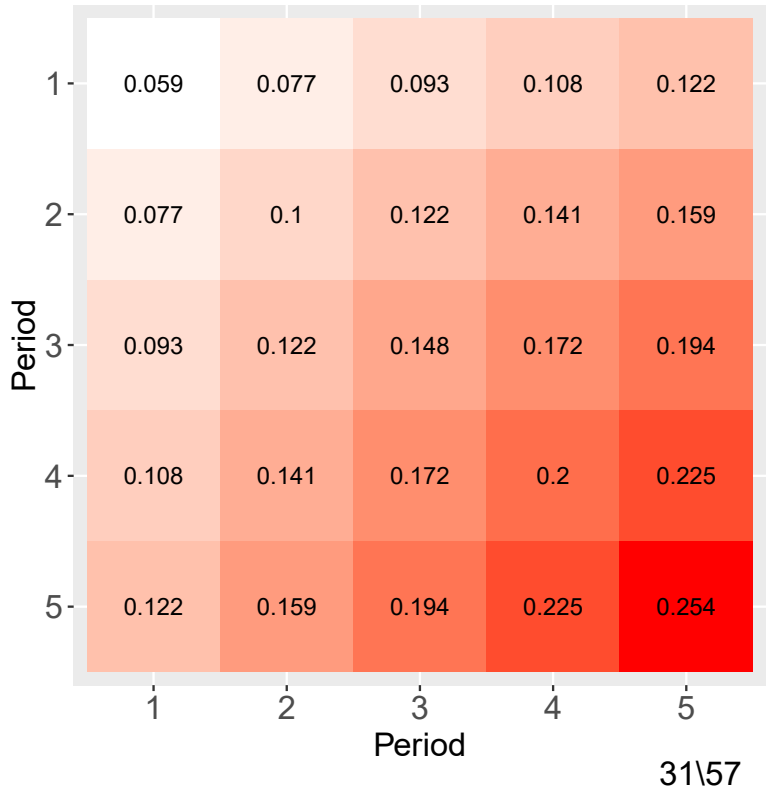
Random Coefficient Model - Cont'd

- Even this basic form remains to be studied more
- Could imply unique correlation structure opposed to exponential decay

(a) Zero covariance $\sigma_{\alpha\xi} = 0$



(b) Positive covariance $\sigma_{\alpha\xi} = 0.5$



Modeling Heterogeneity in CC Designs

- Example extensions to the Hussey and Hughes model closed-cohort designs; all models assume a continuous outcome and an identity link function[17]

Basic	Include cluster-level and subject-level random effects to separate between-individual ICC and within-individual ICC	Baio et al. (2015)[62]
Block Exchangeable*	Include three random effects to distinguish between within-period ICC, between-period ICC, and within-individual ICC	Hooper et al. (2016)[56] Girling and Hemming (2016)[38]
Proportional Decay*	Allow the between-period ICC and within-individual ICC to decay over time at the same exponential rate	Li (2019)[60]
Random Intervention	Include random cluster-specific intervention effects, and ICC depends on intervention status	Kasza et al. (2019)[27]

[17] Li F. et al (2020) Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Stat Methods Med Res*

Basic Model Extending Hussey and Hughes

$$Y_{ijk} = \mu + \beta_j + \delta X_{ij} + \alpha_i + \phi_{ik} + \epsilon_{ijk}^{18} \quad (1)$$

- $\phi_{ik} \sim N(0, \tau_\phi^2)$ random effect for the repeated measures
- heterogeneity term $\boldsymbol{R}_{ik}(j,s)'\boldsymbol{\alpha}_i = \alpha_i + \phi_{ik}$
- implies a *nested exchangeable* correlation structure

$$\text{corr}[Y_{ijk}(s), Y_{ilm}(s)] = \begin{cases} \rho_{\text{a}} = (\tau_\alpha^2 + \tau_\phi^2)/(\tau_\alpha^2 + \tau_\phi^2 + \sigma_\epsilon^2), & k = m \\ \rho_{\text{d}} = \tau_\alpha^2/(\tau_\alpha^2 + \tau_\phi^2 + \sigma_\epsilon^2), & k \neq m \end{cases}$$

- closed-form variance shares the same form of HH variance (with changes in total variance and eigenvalues)
- $\lim_{N\to\infty} \text{var}(\hat{\delta}) = 0$!!

[18] Baio G et al. (2015) Sample size calculation for a stepped wedge trial. *Trials*

Block Exchangeable Correlation Model

$$Y_{ijk} = \mu + \beta_j + \delta X_{ij} + \alpha_i + \gamma_{ij} + \phi_{ik} + \epsilon_{ijk},$$

- $\phi_{ik} \sim N(0, \tau_\phi^2)$ random effect for the repeated measures
- Hooper/Girling model
- heterogeneity $\boldsymbol{R}_{ik}(j,s)'\boldsymbol{\alpha}_i = \alpha_i + \gamma_{ij} + \phi_{ik}$
- correlation structure with three ICCs
  - *within-period* and *between-period* ICC different individuals
  - *(between-period) within-individual* ICC for repeated measures (ρ_a)
- Same form of variance with HH model, except that

$$\begin{aligned}
\sigma^2_{\text{tot}} &= \tau^2_\alpha + \tau^2_\gamma + \tau^2_\phi + \sigma^2_\epsilon, \\
\lambda_1 &= 1 + (N-1)(\rho_\mathrm{w} - \rho_\mathrm{b}) - \rho_\mathrm{a}, \\
\lambda_2 &= 1 + (N-1)\rho_\mathrm{w} + (J-1)(N-1)\rho_\mathrm{b} + (J-1)\rho_\mathrm{a}.
\end{aligned}$$

- $\lim_{N\to\infty} \text{var}(\hat{\delta}) \neq 0$

Block Exchangeable Correlation Model - Cont'd

- Common choice of models in closed-cohort designs
- Closed-form variance allows us to confirm[19]
  - within-period correlation $\rho_w \uparrow$, power $\downarrow$ (traditional ICC)
  - between-period correlation $\rho_b \uparrow$, power $\uparrow$
  - within-individual correlation $\rho_a \uparrow$, power $\uparrow$
- Only ρ_w is mostly likely to be found in the literature, few published ρ_b, ρ_a
- Conservatively small values for ρ_b, ρ_a will NOT underpower the study

[19] Li F et al. (2018) Sample size determination for GEE analyses of stepped wedge cluster randomized trials. *Biometrics*

Proportional Decay Model

$$Y_{ijk} = \mu + \beta_j + \delta X_{ij} + \gamma_{ij} + \epsilon_{ijk},$$

- Mimicking the exponential decay model with $\boldsymbol{R}_{ik}(j, s)'\boldsymbol{\alpha}_i = \gamma_{ij}$
- Further assume a similar autoregressive structure for residual errors of the kth person in cluster i as

$$\boldsymbol{\epsilon}_{ik} = (\epsilon_{i1k}, \ldots, \epsilon_{iJk})' \sim N(0, \sigma_\epsilon^2 \boldsymbol{M}), \qquad \boldsymbol{\epsilon}_{ik} \perp \boldsymbol{\epsilon}_{im}, \quad k \neq m$$

- Implies a *proportional decay* correlation structure

$$\text{corr}[Y_{ijk}(s), Y_{ilm}(s)] = \begin{cases} \rho_{\text{w}} = \tau_\gamma^2/(\tau_\gamma^2 + \sigma_\epsilon^2), & j = l, k \neq m, \\ \rho_{\text{a},|j-l|} = r^{|j-l|}, & j \neq l, k = m, \\ \rho_{\text{b},|j-l|} = \rho_{\text{w}} r^{|j-l|}, & j \neq l, k \neq m, \end{cases}$$

- ρ_{w} and $\rho_{\text{b},|j-l|}$ are the within-period and between-period ICCs
- $\rho_{\text{a},|j-l|}$ the within-individual ICC that decays exponentially

Proportional Decay Model - Cont'd

- Originally studied under marginal model[20]
- Assumption: same decay rate r applies to both the within-individual ICC and the between-period ICC for different individuals
- Separability of correlation matrix allow us to obtain

$$\text{var}(\hat{\delta}) = \frac{(\sigma^2_{\text{tot}}/N)I(1-r^2)\{1+(N-1)\rho_{\text{w}}\}}{(IU-W)(1+r^2)-2(IP-Q)r},$$

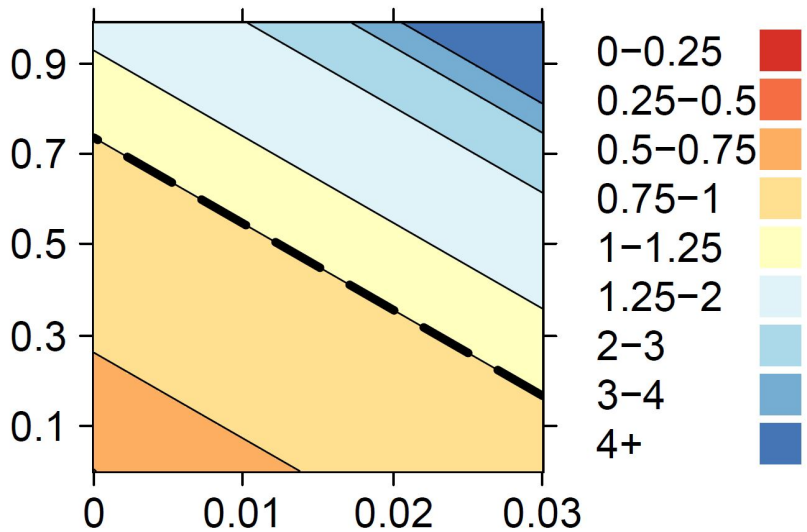
where $U = \sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}$, $W = \sum_{j=1}^{J}(\sum_{i=1}^{I} X_{ij})^2$, $P = \sum_{i=1}^{I}\sum_{j=1}^{J-1} X_{ij}X_{i,j+1}$, $Q = \sum_{j=1}^{J-1}(\sum_{i=1}^{I} X_{ij})(\sum_{i=1}^{I} X_{i,j+1})$ are new design constants

- Parabolic relationship between $\text{var}(\hat{\delta})$ and decay

[20] Li F (2020). Design and analysis considerations for cohort stepped wedge cluster randomized trials with a decay correlation structure. *Stat Med*

Proportional Decay Model - Cont'd

- Sample size calculation may be sensitive to correlation assumptions, and can go either direction[20]
  - recall nested exchangeable versus exponential decay (cross-sectional)
  - block exchangeable (BE) versus proportional decay (PD) (closed-cohort)



- contour of $\text{var}^{\text{PD}}(\hat{\delta})/\text{var}^{\text{BE}}(\hat{\delta})$
- common $\rho_w = 0.03$
- $r = 0.5$ from PD
- x-axis ρ_b from BE
- y-axis ρ_a from BE

[20] Li F (2020). Design and analysis considerations for cohort stepped wedge cluster randomized trials with a decay correlation structure. *Stat Med*

Random Intervention Model

$$Y_{ijk} = \mu + \beta_j + (\delta + \nu_i)X_{ij} + \gamma_{ij} + \phi_{ik} + \epsilon_{ijk},$$[21]

- $\phi_{ik} \sim N(0, \tau^2_\phi)$ for repeated measures, γ_{ij} is the cluster-period-specific random deviation from the group average, as in the exponential decay model
- ν_i is the cluster-specific random intervention effect
- $\boldsymbol{R}_{ik}(j,s)'\boldsymbol{\alpha}_i = \gamma_{ij} + \phi_{ik} + \nu_i \mathbb{I}_{[j \geq s]}$
- implies eight different ICC parameters
- generalizes exponential decay and random intervention models under CS design, but does not nest proportional decay model

[21] Kasza J et al. (2019) Information content of stepped wedge designs when treatment effect heterogeneity and/or implementation periods are present. *Stat Med*

Modeling Heterogeneity in OC Designs

- Open-cohort (OC) design can be considered as a mix of a cross-sectional design and a closed-cohort design
- N_{ij} individuals included during period j in cluster i
- Exists an *overlapping* number $\left(0 \leq n_i(j,l) \leq \min\{N_{ij}, N_{il}\}\right)$ of individuals for period j and period l in cluster i, depending on the degree of cohort openness
- Notation generalizes that of the previous two designs
  - cross-sectional: $n_i(j,l) = 0$ for all j and l (maximum degree of openness)
  - closed-cohort: $n_i(j,l) = N_{ij} = N_{il}$ for all j and l (minimum degree of openness).

Modeling Heterogeneity in OC Designs - Cont'd

- In principle, the models developed for the closed-cohort design can still be used to represent the outcome trajectory in the open-cohort design
- Caveat is the repeated measures are only available for individuals included in more than one period
- For example, the block exchangeable model still applies

$$Y_{ijk} = \mu + \beta_j + \delta X_{ij} + \alpha_i + \gamma_{ij} + \phi_{ik} + \epsilon_{ijk},$$

- The implied within-cluster correlation matrix is neither *nested exchangeable* nor *block exchangeable*, but becomes a blend of these two

Blended Block Correlation Structure

Each block represents a given cluster-period or between two cluster-periods, and $J = 3$. In the open-cohort design, we assume only one individual is followed through all periods, and a new individual will be supplemented in each period.

Nested exchangeable structure

$$\begin{pmatrix} 1 & \rho_w & \rho_b & \rho_b & \rho_b & \rho_b \\ \rho_w & 1 & \rho_b & \rho_b & \rho_b & \rho_b \\ \rho_b & \rho_b & 1 & \rho_w & \rho_b & \rho_b \\ \rho_b & \rho_b & \rho_w & 1 & \rho_b & \rho_b \\ \rho_b & \rho_b & \rho_b & \rho_b & 1 & \rho_w \\ \rho_b & \rho_b & \rho_b & \rho_b & \rho_w & 1 \end{pmatrix}$$

Block exchangeable structure

$$\begin{pmatrix} 1 & \rho_w & \rho_a & \rho_b & \rho_a & \rho_b \\ \rho_w & 1 & \rho_b & \rho_a & \rho_b & \rho_a \\ \rho_a & \rho_b & 1 & \rho_w & \rho_a & \rho_b \\ \rho_b & \rho_a & \rho_w & 1 & \rho_b & \rho_a \\ \rho_a & \rho_b & \rho_a & \rho_b & 1 & \rho_w \\ \rho_b & \rho_a & \rho_b & \rho_a & \rho_w & 1 \end{pmatrix}$$

Blended exchangeable structure

$$\begin{pmatrix} 1 & \rho_w & \rho_a & \rho_b & \rho_a & \rho_b \\ \rho_w & 1 & \rho_b & \rho_b & \rho_b & \rho_b \\ \rho_a & \rho_b & 1 & \rho_w & \rho_a & \rho_b \\ \rho_b & \rho_b & \rho_w & 1 & \rho_b & \rho_b \\ \rho_a & \rho_b & \rho_a & \rho_b & 1 & \rho_w \\ \rho_b & \rho_b & \rho_b & \rho_b & \rho_w & 1 \end{pmatrix}$$

- All three matrices have the same diagonal block

Blended Block Correlation Structure

Each block represents a given cluster-period or between two cluster-periods, and $J = 3$. In the open-cohort design, we assume only one individual is followed through all periods, and a new individual will be supplemented in each period.

Nested exchangeable structure

$$\begin{pmatrix} 1 & \rho_w & \rho_b & \rho_b & \rho_b & \rho_b \\ \rho_w & 1 & \rho_b & \rho_b & \rho_b & \rho_b \\ \rho_b & \rho_b & 1 & \rho_w & \rho_b & \rho_b \\ \rho_b & \rho_b & \rho_w & 1 & \rho_b & \rho_b \\ \rho_b & \rho_b & \rho_b & \rho_b & 1 & \rho_w \\ \rho_b & \rho_b & \rho_b & \rho_b & \rho_w & 1 \end{pmatrix}$$

Block exchangeable structure

$$\begin{pmatrix} 1 & \rho_w & \rho_a & \rho_b & \rho_a & \rho_b \\ \rho_w & 1 & \rho_b & \rho_a & \rho_b & \rho_a \\ \rho_a & \rho_b & 1 & \rho_w & \rho_a & \rho_b \\ \rho_b & \rho_a & \rho_w & 1 & \rho_b & \rho_a \\ \rho_a & \rho_b & \rho_a & \rho_b & 1 & \rho_w \\ \rho_b & \rho_a & \rho_b & \rho_a & \rho_w & 1 \end{pmatrix}$$

Blended exchangeable structure

$$\begin{pmatrix} 1 & \rho_w & \rho_a & \rho_b & \rho_a & \rho_b \\ \rho_w & 1 & \rho_b & \rho_b & \rho_b & \rho_b \\ \rho_a & \rho_b & 1 & \rho_w & \rho_a & \rho_b \\ \rho_b & \rho_b & \rho_w & 1 & \rho_b & \rho_b \\ \rho_a & \rho_b & \rho_a & \rho_b & 1 & \rho_w \\ \rho_b & \rho_b & \rho_b & \rho_b & \rho_w & 1 \end{pmatrix}$$

- All three matrices have the same diagonal block
- Difference in off-diagonal blocks determined by overlapping number of individuals and hence degree of cohort openness

Design Considerations

- Assuming same cluster-period sizes and constant attrition rate, Kasza et al. derived the closed-form variance for power calculation [22]
- The attrition rate reflects the degree of openness — represents continuum between cross-sectional and closed-cohort designs
  - facilitates efficiency comparisons between these two designs
  - closed-cohort design is usually at least as efficient as the cross-sectional design as long as $\rho_a \geq \rho_b$
  - the reverse when $\rho_a < \rho_b$ (not plausible under mixed models)
- Extensions to the correlation decay[22]

[22] Kasza J et al. (2020) Sample size and power calculations for open cohort longitudinal cluster randomized trials. *Stat Med*

Considerations for Binary Outcomes

- Not as many article on binary outcomes, which are nonetheless common as primary endpoints
- From design perspective
  - Zhou et al.[23] provided a maximum likelihood approach for power calculation with binary outcomes
  - extending the HH model to estimate risk difference
  - SAS and R package forthcoming `swdpwr`
  - can be accessed via `https://publichealth.yale.edu/cmips/research/software/swdpwr/`
- Important message is that linear mixed model approximation may not be accurate for power calculation with binary outcomes

[23] Zhou X et al. (2020) A maximum likelihood approach to power calculations for stepped wedge designs of binary outcomes. *Biostatistics*

Considerations for Binary Outcomes - Cont'd

- Limited investigations on generalized linear mixed models with more complex random-effects structure
- An exception is Thompson et al.[24], compared
  - Hussey and Hughes model
  - Nested exchangeable model (Hooper/Girling model)
  - Random intervention model
- The logistic nested exchangeable correlation model

$$\text{logit}(\mu_{ij}) = \mu + \beta_j + \delta X_{ij} + \alpha_i + \gamma_{ij}, \quad \alpha_i \sim N(0, \tau_\alpha^2), \quad \gamma_{ij} \sim N(0, \tau_\gamma^2)$$

  had more robust performance in terms of bias and type I error rates across a number of data generating processes

- Careful on the interpretation of δ

[24] Thompson et al. (2017) Bias and inference from misspecified mixed-effect models in stepped wedge trial analysis. *Stat Med*

Estimation and Inference

- Fitting variants of mixed effects models have become standard in common software
  - `proc mixed`, `glimmix` or `hpmixed` (SAS)
  - `nlme` or `lme4` (R)
- Flexible choice of readily-available complex random effects structure with linear mixed model compared to generalized linear mixed model
  - provide intervention effect parameter estimate
  - estimate variance component, but need additional step to compute ICC (simple for continuous outcomes, not as much for binary)

Estimation and Inference - Cont'd

- Permutation inference has gained traction for accurate type I error rate control
- General idea is to obtain the reference distribution of a given test statistic by permuting the intervention sequences across clusters
  - requires exchangeability across permuted intervention sequences under the null
- Other recent permutation methods discussed in Li et al.[25]

[25] Li F. et al (2020) Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Stat Methods Med Res*

4. Concluding Remarks

On Mixed Model Variants

- Review variants of mixed models for stepped wedge cluster randomized trials under a unified perspective

$$g[\mu_{ijk}(s)] = \underbrace{\boldsymbol{F}^0(j)'\boldsymbol{\beta}}_{\text{secular trend}} + \underbrace{F_i^1(j,s)\Delta(j,s)}_{\text{intervention effect}} + \underbrace{\boldsymbol{R}_{ik}(j,s)'\boldsymbol{\alpha}_i}_{\text{heterogeneity}}.$$

- Majority of models assumed categorical time effect and a scalar intervention effect (convenient for sample size estimation)
- Current literature devoted to continuous outcomes and variations of random-effects structure
- Relatively limited literature on binary or count outcomes

Choices of Models

- Generally a difficult question with no uniform solution
- Can depend on the accuracy of characterizing the outcome trajectories
- Consistent choice of models in design and analysis stage through pre-specification
- Design stage – prior information/pilot data/sensitivity analysis
- Analysis stage –
  - robust analysis
  - information criteria?[26]
  - open question

[26] Murray DM et al (1998). Analysis of data from group-randomized trials with repeat observations on the same groups. *Stat Med*

Alternatives: Marginal Models

- Marginal models
  - separate mean and correlation models
  - population-averaged interpretation
  - inference robust to correlation specification (# cluster *I* large)
- Recent literature studying sample size and finite-sample behaviour of GEE estimators, e.g[27]
- Directly estimating correlations instead of variance components
- Software available, but more need to be developed

[27] Li F et al. (2018) Sample size determination for GEE analyses of stepped wedge cluster randomized trials. *Biometrics*

Revisit CONSORT extension to SW-CRTs

- CONSORT item 7a: Sample size[28]
  - *Extension for SW-CRTs – . . . Method of calculation and relevant parameters with sufficient detail so the calculation can be replicated. Assumptions made about correlations between outcomes of participants from the same cluster*.
- CONSORT item 12a & b: Statistical methods
  - *Extension for SW-CRTs – . . . Statistical methods used to compare treatment conditions for primary and secondary outcomes including how time effects, clustering, and repeated measures were taken into account.*.
- Important to explicitly describe model variants

[28] Hemming K (2018) Reporting of stepped wedge cluster randomised trials: Extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ*

More on CONSORT extension to SW-CRTs

- CONSORT item 17a: Outcomes and estimation
  - *Extension for SW-CRTs – For each primary and secondary outcome, results for each treatment condition, and the estimated effect size and its precision; any correlations (or covariances) and time effects estimated in the analysis outcome*
- With either choice of models, reporting ICC or variance components is highly recommended as this can be informative for the planning of future trials
- In particular for the correlation decay and random intervention models, to facilitate the design of trials based on these more recent extensions

Thank you for listening!

- questions are welcome either during the webinar or `fan.f.li@yale.edu`

Back up slide: Other Important Aspects Not Mentioned

- Using baseline covariates?
  - constrained randomization (design)
  - improve power (analysis)
- Multiple layers of clustering
- Addressing “missing data"
  - complete designs, missing outcomes (e.g. closed-cohort)
  - incomplete designs
- Other interesting and important questions to be solved