From Correlation Structure to Treatment Effect Structure: A New Horizon for Stepped Wedge Designs

Fan Li

https://lifan90.com

Department of Biostatistics
Yale School of Public Health

May 12, 2026

The Fifth International Conference on Stepped Wedge Trial Design

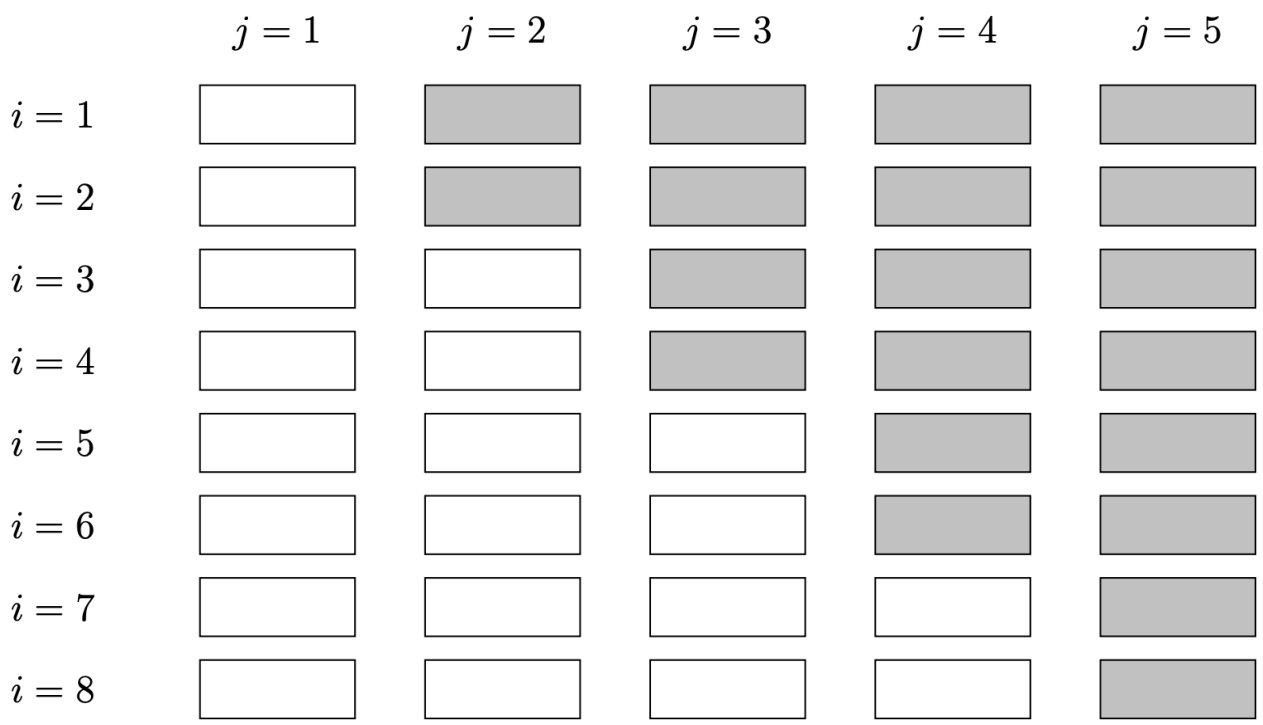
1. Introduction

Introduction

- **SW-CRTs**: The **timing** of crossover is randomized
  - Popularized since introduction
  - Type in "*Stepped wedge cluster randomized trial*" in PubMed

2022	2023	2024	2025	2026
198	220	283	291	88

  - Consensus: more challenging compared to parallel-arm designs
  - **4 broad justifications** (Hemming and Taljaard, 2020 *IJE*)[1]
- **Driving questions**: "*What are essential challenges of this design? What have we learned in the last two decades?*"



[1]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.,* 49(3), 1043-1052

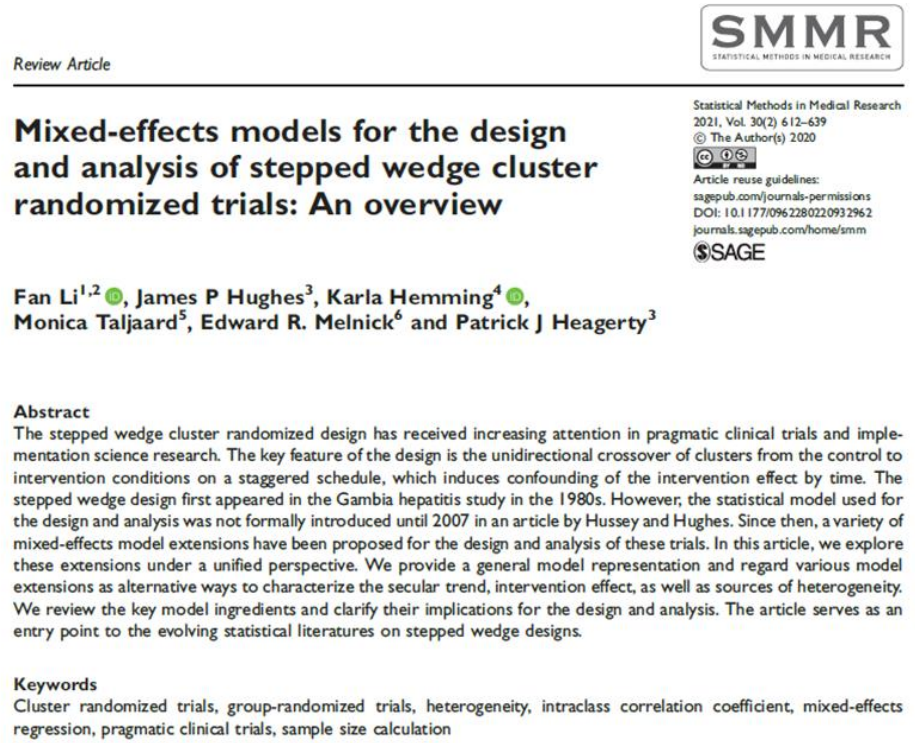
When it comes to the design and analysis...

- Essential task is to estimate the **treatment effects**
- Unique features require more complex considerations on analytical models than those in a parallel-arm CRT
- **Commonly used:** (Generalized) linear mixed model
- Model recipe:[2]

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

  - **Secular trend structure** *(j indexes period)*
  - **Treatment effect structure** *(s indexes timing of treatment)*
  - **Random-effects (correlation) structure** to account for ICCs

    (i indexes cluster and k indexes individual)



SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Review Article

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

Statistical Methods in Medical Research
2021, Vol. 30(2) 612–639

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

[2]**Li, F.,** Hughes, J. P., Hemming, K., Taljaard, M., Melnick, E. R., & Heagerty, P. J. (2021). Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Stat Methods Med Res.,* 30(2), 612-639.

2. Correlation structure

Incomplete summary of correlation structures

Design	Extension	Feature	Example references
Cross-sectional	Nested Exchangeable*	Distinguish between within-period and between-period ICCs	Hooper et al.;[56] Girling and Hemming[38]
	Exponential Decay*	Allow the between-period ICC to decay at an exponential rate over time	Kasza et al.[57] Kasza and Forbes[61]
	Random Intervention	Include random cluster-specific intervention effects, and ICC depends on intervention status	Hughes et al.[55] Hemming et al.[47]
	Random Coefficient	Include random cluster-specific time slopes; ICC tends to be an increasing function of distance in time	Murray et al.[58]
Closed-cohort	Basic	Include cluster-level and subject-level random effects to separate between-individual ICC and within-individual ICC	Baio et al.[65]
	Block Exchangeable*	Include three random effects to distinguish between within-period ICC, between-period ICC, and within-individual ICC	Hooper et al.[56] Girling and Hemming[38]
	Proportional Decay*	Allow the between-period ICC and within-individual ICC to decay over time at the same exponential rate	Li[60]
	Random Intervention	Include random cluster-specific intervention effects, and ICC depends on intervention status	Kasza et al.[27]

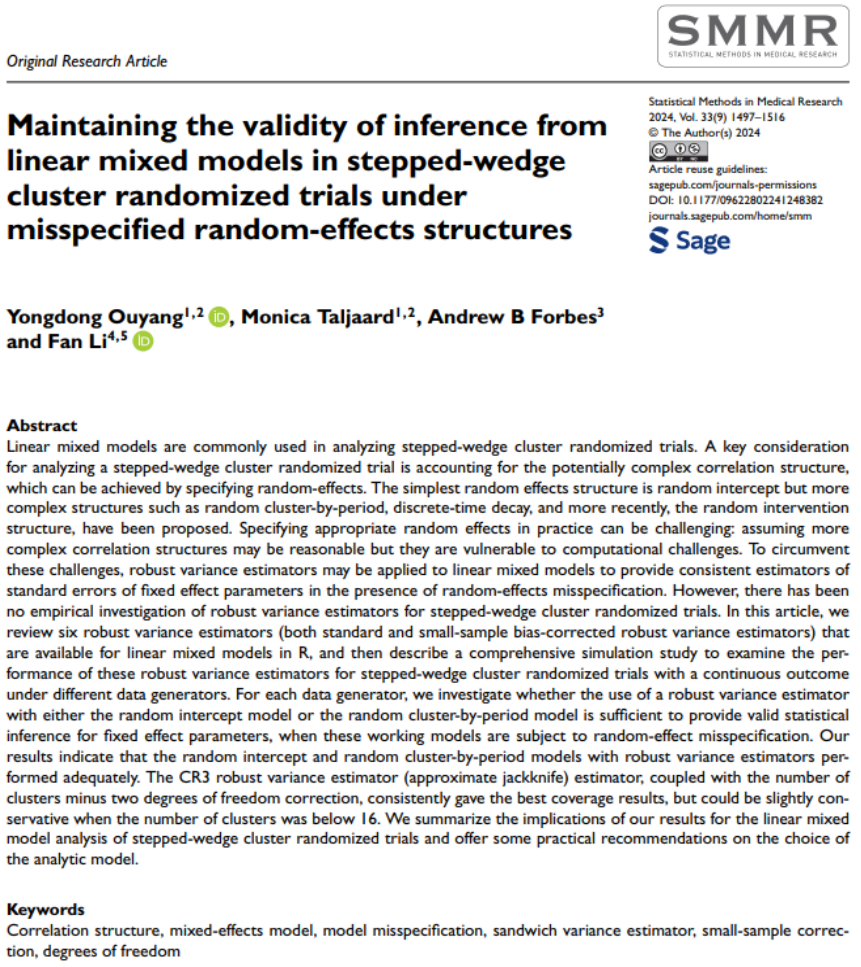
The role of correlation structures for design

- **Conventional wisdom for CRT:** need to get the correlation structure right, otherwise the (model-based) variance expression is **INCORRECT**
- **Critical IMPORTANT in the study planning stage (power and sample size):**
- **Lesson 1**: Under-specification (omitting a necessary decay parameter) results in a smaller model-based variance than necessary[3]
  - Assuming a **simple exchangeable correlation** structure when correlations decay over time can underestimate the treatment-effect variance – sample size too optimistic
  - Assuming a **non-decaying nested exchangeable** structure when a decaying correlation structure is present often underestimates the variance (bias can goes both way) – sample size may be optimistic
- **Lesson 2**: Over-specification (including a decay unnecessarily) leads to only conservative sample sizes

[3]Kasza, J., Hemming, K., Hooper, R., Matthews, J., & Forbes, A. B. (2019). Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Stat Methods Med Res.,* 28(3), 703-716.

The role of correlation structures for analysis

- **Cluster-robust variance estimator**
  - Protects against misspecification of the correlation structure[4]
  - **Breaks the conventional wisdom**: DO NOT need to get the correlation structure exactly right for valid inference
- The role of the correlation structure in analysis?
  - Correct specification → more efficient treatment effect estimates
  - When correctly specified: model-based = cluster-robust variance
  - When misspecified: cluster-robust variance remains valid; pay a price in efficiency, not in validity
- ***Foundational methodological work have deepened understanding of correlation structures in SWDs, but other challenges??***



Original Research Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Maintaining the validity of inference from linear mixed models in stepped-wedge cluster randomized trials under misspecified random-effects structures

Statistical Methods in Medical Research
2024, Vol. 33(9) 1497–1516

DOI: 10.1177/09622802241248382
journals.sagepub.com/home/smm
Sage

Yongdong Ouyang[1,2], Monica Taljaard[1,2], Andrew B Forbes[3] and Fan Li[4,5]

Abstract

Linear mixed models are commonly used in analyzing stepped-wedge cluster randomized trials. A key consideration for analyzing a stepped-wedge cluster randomized trial is accounting for the potentially complex correlation structure, which can be achieved by specifying random-effects. The simplest random effects structure is random intercept but more complex structures such as random cluster-by-period, discrete-time decay, and more recently, the random intervention structure, have been proposed. Specifying appropriate random effects in practice can be challenging: assuming more complex correlation structures may be reasonable but they are vulnerable to computational challenges. To circumvent these challenges, robust variance estimators may be applied to linear mixed models to provide consistent estimators of standard errors of fixed effect parameters in the presence of random-effects misspecification. However, there has been no empirical investigation of robust variance estimators for stepped-wedge cluster randomized trials. In this article, we review six robust variance estimators (both standard and small-sample bias-corrected robust variance estimators) that are available for linear mixed models in R, and then describe a comprehensive simulation study to examine the performance of these robust variance estimators for stepped-wedge cluster randomized trials with a continuous outcome under different data generators. For each data generator, we investigate whether the use of a robust variance estimator with either the random intercept model or the random cluster-by-period model is sufficient to provide valid statistical inference for fixed effect parameters, when these working models are subject to random-effect misspecification. Our results indicate that the random intercept and random cluster-by-period models with robust variance estimators performed adequately. The CR3 robust variance estimator (approximate jackknife) estimator, coupled with the number of clusters minus two degrees of freedom correction, consistently gave the best coverage results, but could be slightly conservative when the number of clusters was below 16. We summarize the implications of our results for the linear mixed model analysis of stepped-wedge cluster randomized trials and offer some practical recommendations on the choice of the analytic model.

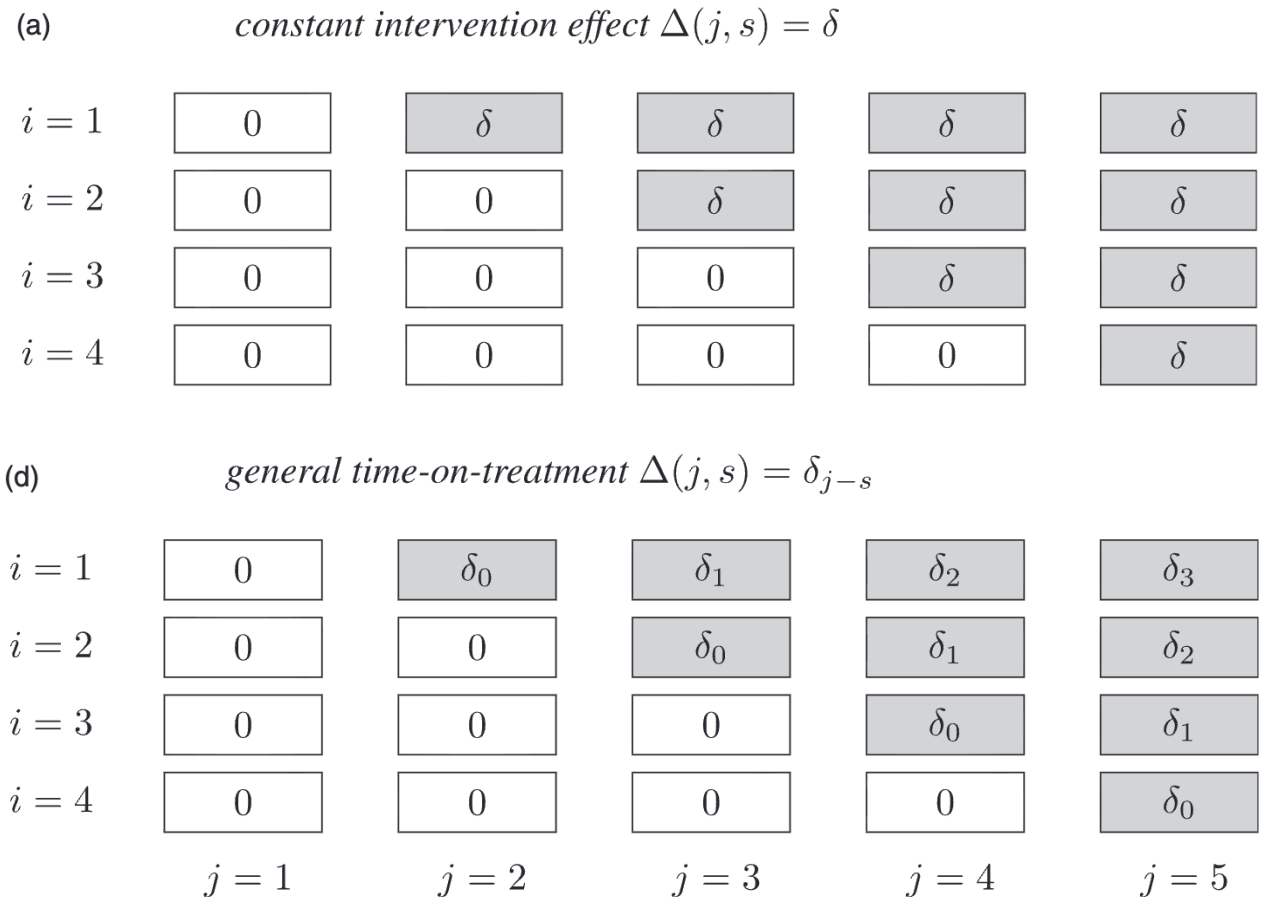
Keywords

Correlation structure, mixed-effects model, model misspecification, sandwich variance estimator, small-sample correction, degrees of freedom

[4]Ouyang, Y., Taljaard, M., Forbes, A. B., & **Li, F.** (2024). Maintaining the validity of inference from linear mixed models in stepped-wedge cluster randomized trials under misspecified random-effects structures. *Stat Methods Med Res.,* 33(9), 1497-1516.

3. Treatment effect structure

What about treatment effect structure?



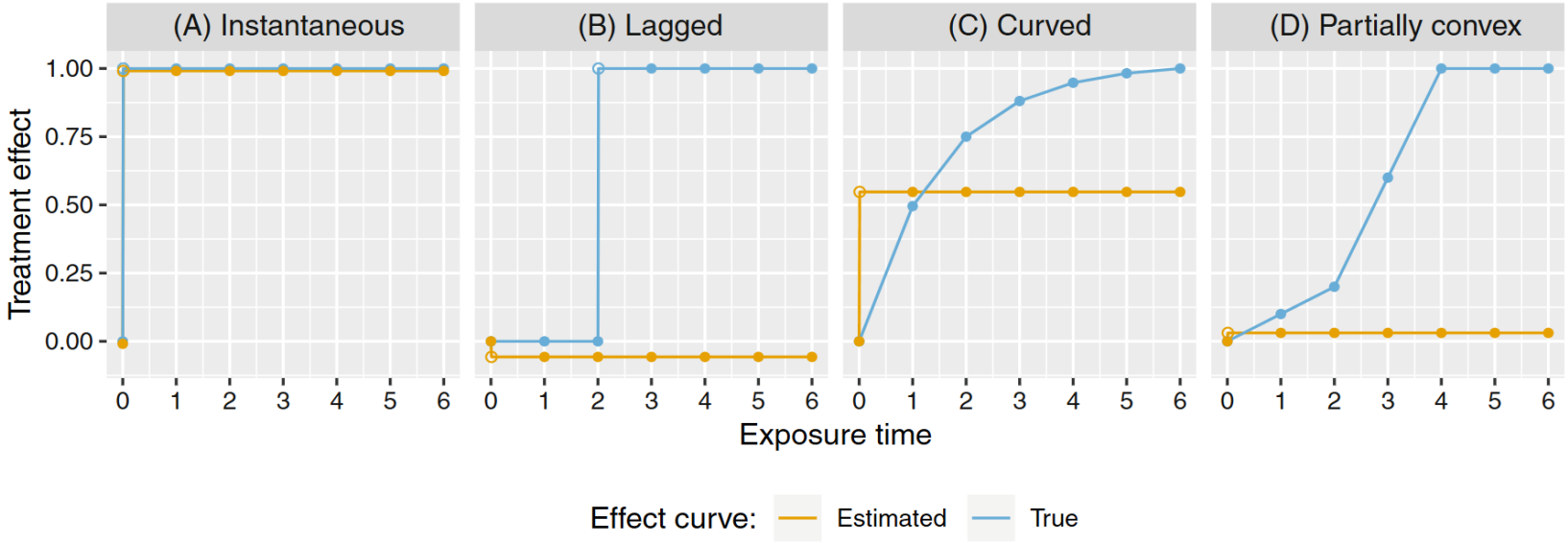
- **Section 3.3 & Figure 2 of Li et al.[2]**
  - 1 page text + 1 page figure v.s. 13 pages of correlation structures
- What is unique about stepped wedge designs?
  - Correlation structure becomes more complex because of "time"
  - But treatment effect structure can also depend on "time"!
  - Two distinct time axes: **calendar time** & **exposure time**
  - This "time" elements sets stepped wedge apart from other longitudinal CRTs

[2]**Li, F.,** Hughes, J. P., Hemming, K., Taljaard, M., Melnick, E. R., & Heagerty, P. J. (2021). Mixed-effects models for the designand analysis of stepped wedge cluster randomized trials: An overview. *Stat Methods Med Res.,* 30(2), 612-639.

A fundamental rethinking

- Almost all prior efforts have focused **a constant treatment effect** model
  - Default, conventional practice
  - assumes that the full effect of the treatment is reached **immediately** and remains **constant**

- Kenny et al. (2022)[5] proved that, in the presence of treatment effect heterogeneity across exposure time, the resulting estimator from **a constant treatment effect** model does NOT necessarily represent the average of exposure-time specific treatment effect

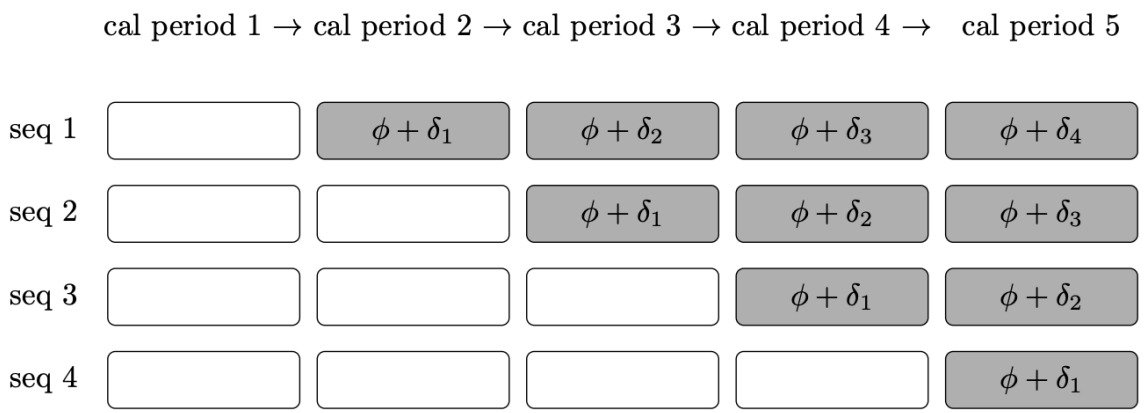


[5]Kenny, A., Voldal, E. C., Xia, F., Heagerty, P. J., & Hughes, J. P. (2022). Analysis of stepped wedge cluster randomized trials in the presence of a time-varying treatment effect. *Stat in Med.,* 41(22), 4311-4339.

Fixed or random treatment effects across exposure time?

- A random effect model formulation accounting for exposure-time treatment effect heterogeneity[6]
- More efficient by estimating a single average treatment effect
- Facilitates a permutation test to examine the existence of such heterogeneity
- Empirically, the time-averaged estimates under this model tend to be closer to the constant effect model

Random-effects model formulation:

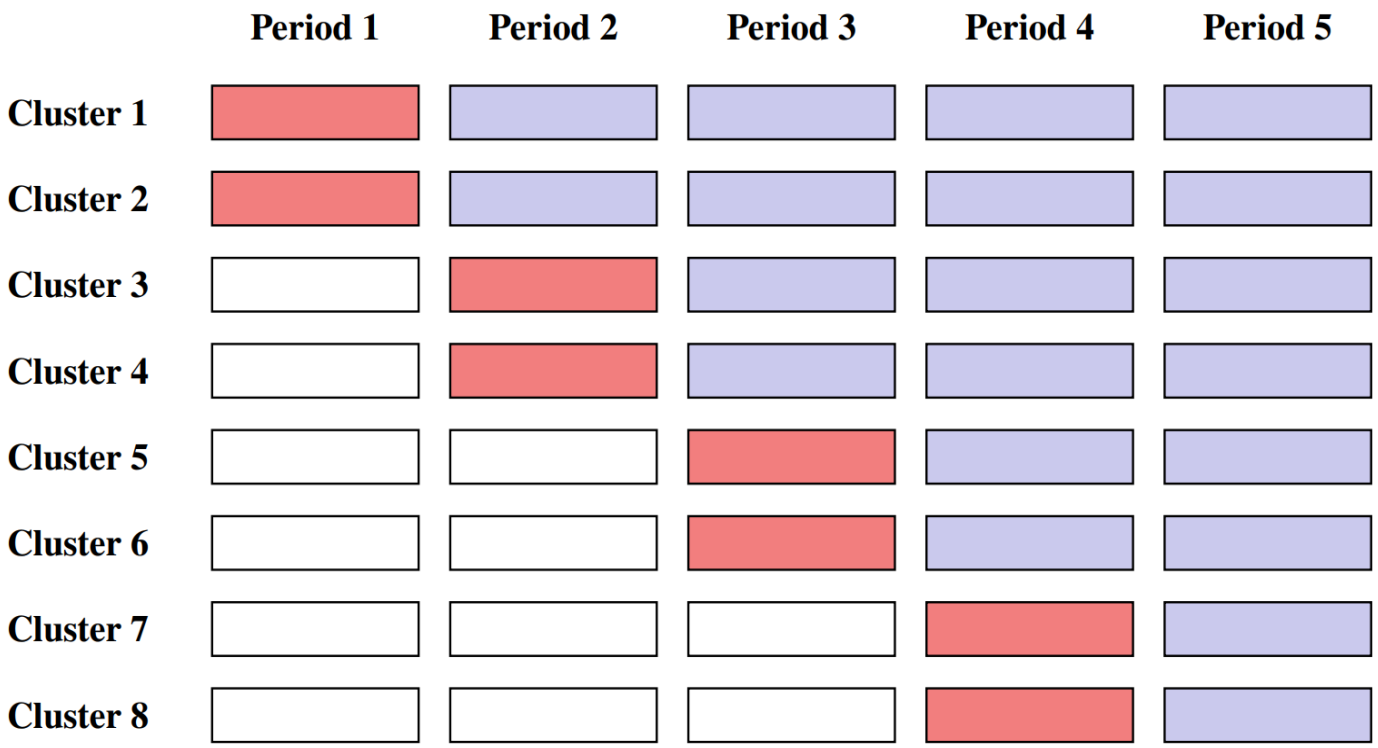


[6]Maleyeff, L., **Li, F.**, Haneuse, S., & Wang, R. (2023). Assessing exposure-time treatment effect heterogeneity in stepped-wedge cluster randomized trials. *Biometrics,* 79(3), 2551-2564.

3.1 What happens when exposure time equals -1?

What if clusters know that the intervention is coming?

- **Anticipation effect!**
  - Potential for cluster members to **change behavior prior to intervention adoption** simply because they have foreknowledge of the intervention timing
- **Empirical examples in other field**:
  - Students study harder expecting a new teaching method
  - Behavior shifts before expected insurance price changes
- **Although blinding is recommended**
  - Complete blinding is **impractical** from time to time



A roadmap for a full investigation

- **Gap:**
  - Almost no literature on the **anticipation effect** in SW-CRTs
- **Objectives:**
  - Systematically examine linear mixed model estimators when **anticipation, exposure-time heterogeneity, or both** are present[7]
  - Derive **closed-form bias formulas** for the 6 unstudied misspecified scenarios

Index	True Data Generating Process			Working Specification			Scenario	References
	Model	Anticipation	TVE	Model	Anticipation	TVE		
1	HH	×	×	HH	×	×	Correct	Hussey and Hughes (2007)
2	HH	×	×	HH-ANT	✓	×	Over-specification	this article
3	HH	×	×	ETI	×	✓	Over-specification	this article
4	HH	×	×	ETI-ANT	✓	✓	Over-specification	this article
5	HH-ANT	✓	×	HH	×	×	Under-specification	this article
6	HH-ANT	✓	×	HH-ANT	✓	×	Correct	this article
7	HH-ANT	✓	×	ETI	×	✓	Mismatch	this article
8	HH-ANT	✓	×	ETI-ANT	✓	✓	Over-specification	this article
9	ETI	×	✓	HH	×	×	Under-specification	Kenny et al. (2022)
10	ETI	×	✓	HH-ANT	✓	×	Mismatch	this article
11	ETI	×	✓	ETI	×	✓	Correct	this article
12	ETI	×	✓	ETI-ANT	✓	✓	Over-specification	this article
13	ETI-ANT	✓	✓	HH	×	×	Under-specification	this article
14	ETI-ANT	✓	✓	HH-ANT	✓	×	Under-specification	this article
15	ETI-ANT	✓	✓	ETI	×	✓	Under-specification	this article
16	ETI-ANT	✓	✓	ETI-ANT	✓	✓	Correct	this article

[7]Wang, H., Chen, X., Courtright, K. R., Halpern, S. D., Harhay, M. O., Taljaard, M., & **Li, F.** (2026). On anticipation effect in stepped wedge cluster randomized trials. *Stat in Med.,* 45(3-5), e70380

A roadmap for a full investigation

- **Gap:**
  - Almost no literature on the **anticipation effect** in SW-CRTs
- **Objectives:**
  - Systematically examine linear mixed model estimators when **anticipation, exposure-time heterogeneity, or both** are present
  - Derive **closed-form bias formulas** for the 6 unstudied misspecified scenarios

Index	True Data Generating Process			Working Specification			Scenario	References
	Model	Anticipation	TVE	Model	Anticipation	TVE		
2	HH	×	×	HH-ANT	✓	×	Over-specification	this article
3	HH	×	×	ETI	×	✓	Over-specification	this article
4	HH	×	×	ETI-ANT	✓	✓	Over-specification	this article
5	HH-ANT	✓	×	HH	×	×	Under-specification	this article
7	HH-ANT	✓	×	ETI	×	✓	Mismatch	this article
8	HH-ANT	✓	×	ETI-ANT	✓	✓	Over-specification	this article
9	ETI	×	✓	HH	×	×	Under-specification	Kenny et al. (2022)
10	ETI	×	✓	HH-ANT	✓	×	Mismatch	this article
12	ETI	×	✓	ETI-ANT	✓	✓	Over-specification	this article
13	ETI-ANT	✓	✓	HH	×	×	Under-specification	this article
14	ETI-ANT	✓	✓	HH-ANT	✓	×	Under-specification	this article
15	ETI-ANT	✓	✓	ETI	×	✓	Under-specification	this article

A roadmap for a full investigation

- **Gap:**
  - Almost no literature on the **anticipation effect** in SW-CRTs
- **Objectives:**
  - Systematically examine linear mixed model estimators when **anticipation, exposure-time heterogeneity, or both** are present
  - Derive **closed-form bias formulas** for the 6 unstudied misspecified scenarios

Index	True Data Generating Process			Working Specification			Scenario	References
	Model	Anticipation	TVE	Model	Anticipation	TVE		
5	HH-ANT	✓	×	HH	×	×	Under-specification	this article
7	HH-ANT	✓	×	ETI	×	✓	Mismatch	this article
9	ETI	×	✓	HH	×	×	Under-specification	Kenny et al. (2022)
10	ETI	×	✓	HH-ANT	✓	×	Mismatch	this article
13	ETI-ANT	✓	✓	HH	×	×	Under-specification	this article
14	ETI-ANT	✓	✓	HH-ANT	✓	×	Under-specification	this article
15	ETI-ANT	✓	✓	ETI	×	✓	Under-specification	this article

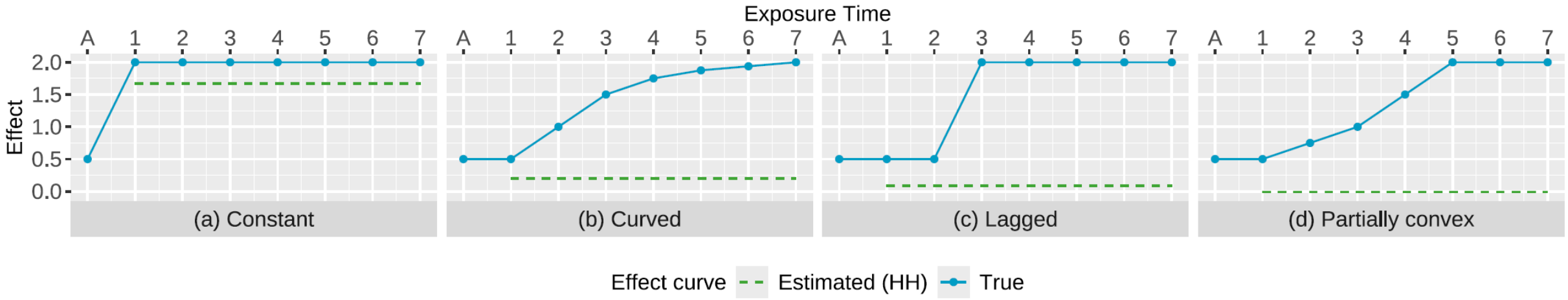
16 combinations of true × working models => Four scenarios

True Model \ Working Model	HH	HH-ANT	ETI	ETI-ANT
HH	Correct	Overspecified	Overspecified	Overspecified
HH-ANT	Underspecified	Correct	Mismatch	Overspecified
ETI	Underspecified	Mismatch	Correct	Overspecified
ETI-ANT	Underspecified	Underspecified	Underspecified	Correct

- **Correct (4)**
  - **No bias**
- **Overspecified (5)**
  - **No bias, efficiency loss**
- **Underspecified/Mismatch (7)**
  - **We derive 6 new bias formulas**

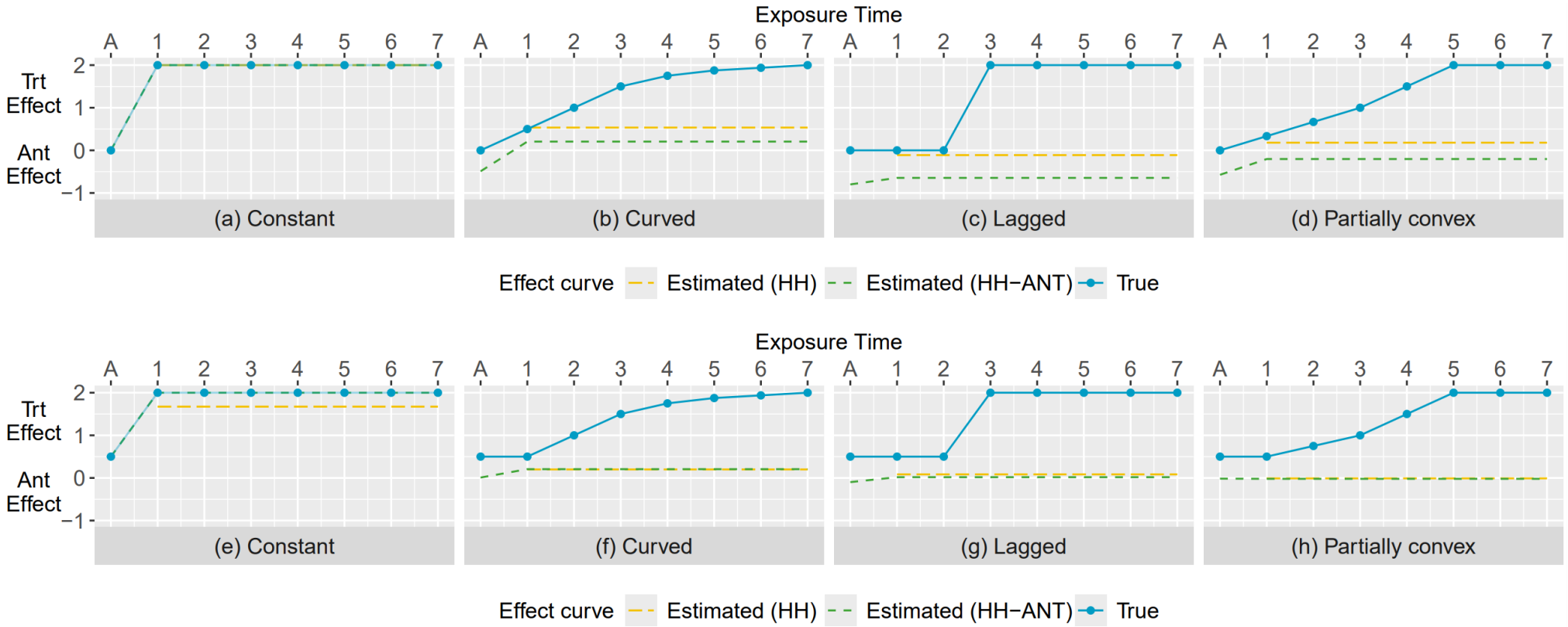
From the analysis perspective: an example

- The true data generating process has the **anticipation effect**
- However, the working model does not account for it
  - Lead to **Biased** estimates
  - Estimated curves **lie entirely below** the true effect curve



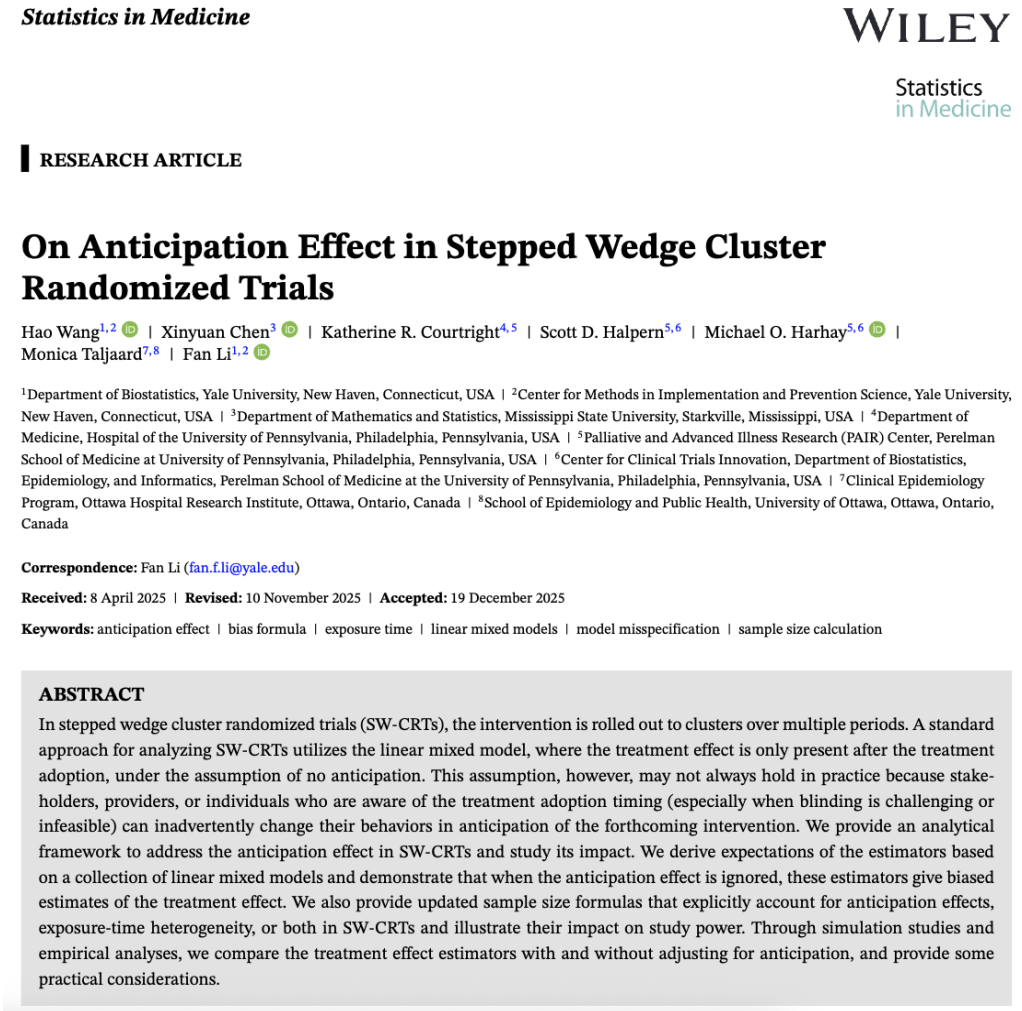
When the treatment effect structure is misspecified

- Working models: **HH and HH-ANT models**
  - HH-ANT may give **more underestimations** compared to HH in a **mismatched** scenario



Implications from anticipation effects

- Calls for **greater awareness** of anticipation that was historically neglected[7]
- There are risks of **not** accounting for the anticipation effect in SW-CRTs when they exist
- Always try to minimize the anticipation through **careful trial designs – another justification for transition periods!!**
- **However, when blinding is infeasible**
  - Anticipation effects can be accounted for in the analysis
  - One can view this effect as the coefficient when the treatment indicator is set to -1



Statistics in Medicine

WILEY

Statistics in Medicine

RESEARCH ARTICLE

On Anticipation Effect in Stepped Wedge Cluster Randomized Trials

Hao Wang[1,2] | Xinyuan Chen[3] | Katherine R. Courtright[4,5] | Scott D. Halpern[5,6] | Michael O. Harhay[5,6] | Monica Taljaard[7,8] | Fan Li[1,2]

[1]Department of Biostatistics, Yale University, New Haven, Connecticut, USA | [2]Center for Methods in Implementation and Prevention Science, Yale University, New Haven, Connecticut, USA | [3]Department of Mathematics and Statistics, Mississippi State University, Starkville, Mississippi, USA | [4]Department of Medicine, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania, USA | [5]Palliative and Advanced Illness Research (PAIR) Center, Perelman School of Medicine at University of Pennsylvania, Philadelphia, Pennsylvania, USA | [6]Center for Clinical Trials Innovation, Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA | [7]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada | [8]School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada

Correspondence: Fan Li (fan.f.li@yale.edu)

Received: 8 April 2025 | **Revised:** 10 November 2025 | **Accepted:** 19 December 2025

Keywords: anticipation effect | bias formula | exposure time | linear mixed models | model misspecification | sample size calculation

ABSTRACT

In stepped wedge cluster randomized trials (SW-CRTs), the intervention is rolled out to clusters over multiple periods. A standard approach for analyzing SW-CRTs utilizes the linear mixed model, where the treatment effect is only present after the treatment adoption, under the assumption of no anticipation. This assumption, however, may not always hold in practice because stakeholders, providers, or individuals who are aware of the treatment adoption timing (especially when blinding is challenging or infeasible) can inadvertently change their behaviors in anticipation of the forthcoming intervention. We provide an analytical framework to address the anticipation effect in SW-CRTs and study its impact. We derive expectations of the estimators based on a collection of linear mixed models and demonstrate that when the anticipation effect is ignored, these estimators give biased estimates of the treatment effect. We also provide updated sample size formulas that explicitly account for anticipation effects, exposure-time heterogeneity, or both in SW-CRTs and illustrate their impact on study power. Through simulation studies and empirical analyses, we compare the treatment effect estimators with and without adjusting for anticipation, and provide some practical considerations.

[7]Wang, H., Chen, X., Courtright, K. R., Halpern, S. D., Harhay, M. O., Taljaard, M., & **Li, F.** (2026). On anticipation effect in stepped wedge cluster randomized trials. *Stat in Med.,* 45(3-5), e70380

3.2 Is the treatment effect structure confined to exposure time?

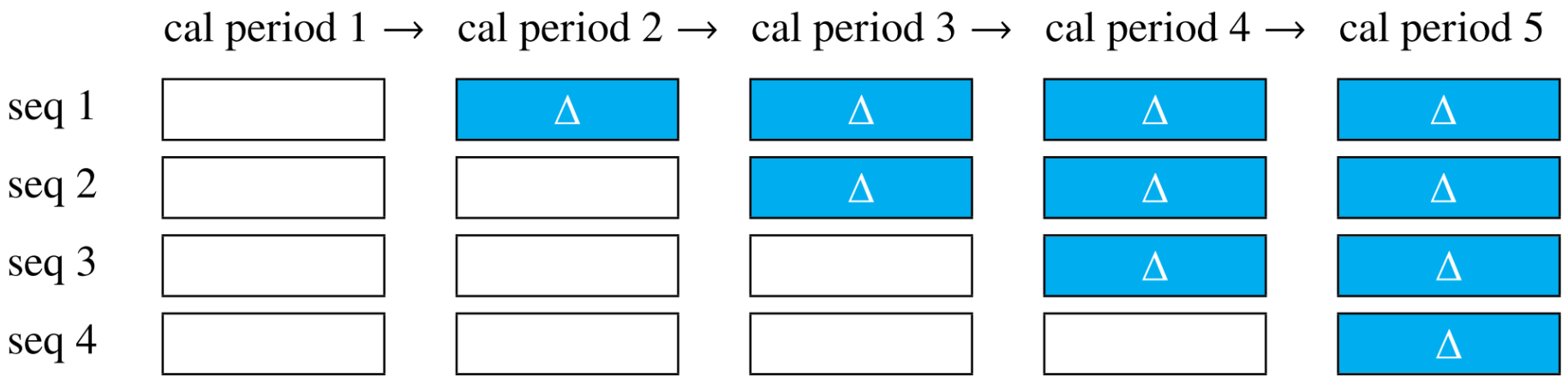
Are there more to the treatment effect structure?

- **In parallel-arm CRTs**
  - Treatment coefficient from a linear mixed model converges to the average treatment effect, under arbitrary model misspecification (model robustness)[8]
- But in SW-CRTs, $\theta(j, s)$ is allowed to depend on two distinct time axes - **calendar time** (j) and **exposure time** (s)
  - Exposure time measures the duration of the intervention for a cluster
  - Learning effect over **exposure time**
  - Seasonal impact over **calendar time**
- ***What if covariates are also included in the model? What is correlation structure gets misspecified under non-constant treatment effect structure?***

[8]Wang, B., Harhay, M. O., Tong, J., Small, D. S., Morris, T. P., & **Li, F.** (2026). On the mixed-model analysis of covariance in cluster-randomized trials. *Stat. Sci.,* 41(1), 49-68

Case 1: Constant treatment effect

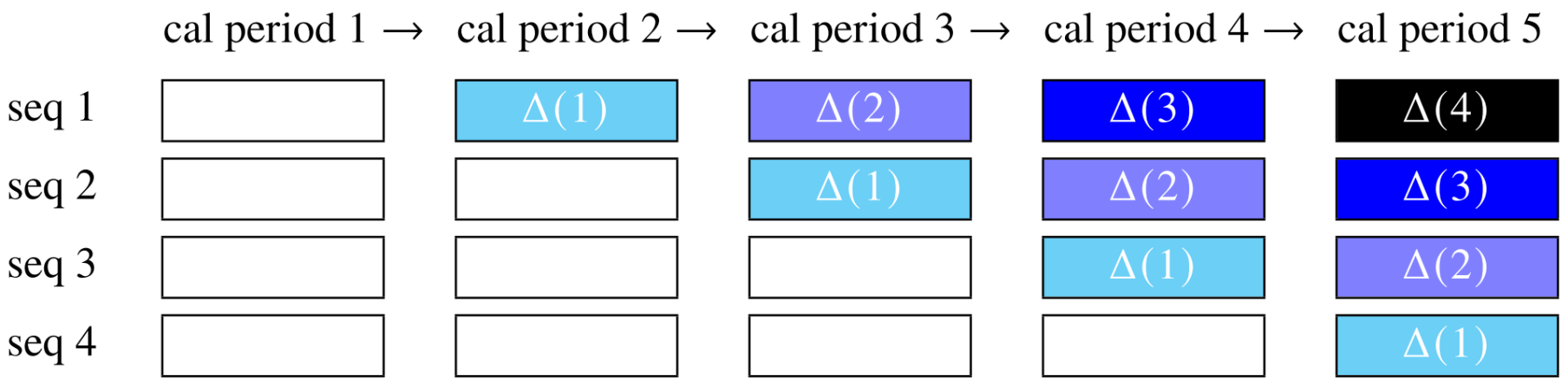
- Assume $\Delta_j(d)$ is constant for all j and d, i.e., $\Delta_j(d) \equiv \Delta$



  - The majority of the existing literature is built on this assumption
  - **Standard practice**

Case 2: Duration-specific treatment effect

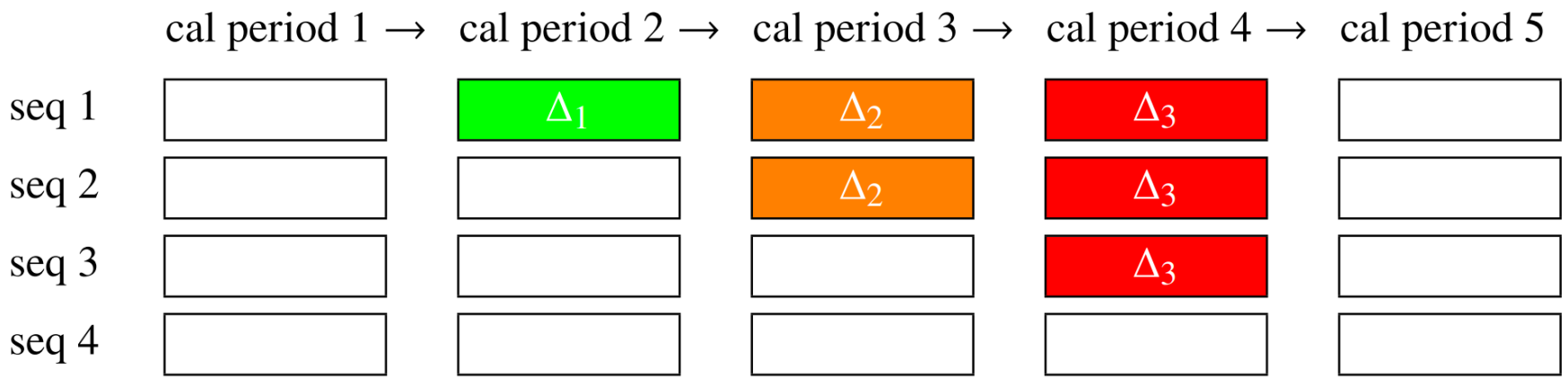
- Consider $\Delta_j(d)$ to be constant across j, but vary by d, i.e., $\Delta_j(d) = \Delta(d)$



  - Ignoring duration effects (exposure-time heterogeneity) can lead to erroneous conclusions

Case 3: Period-specific treatment effect

- Allow the treatment effect to vary by the period of measurement, but not the duration of treatment, i.e., $\Delta_j(d) = \Delta_j$

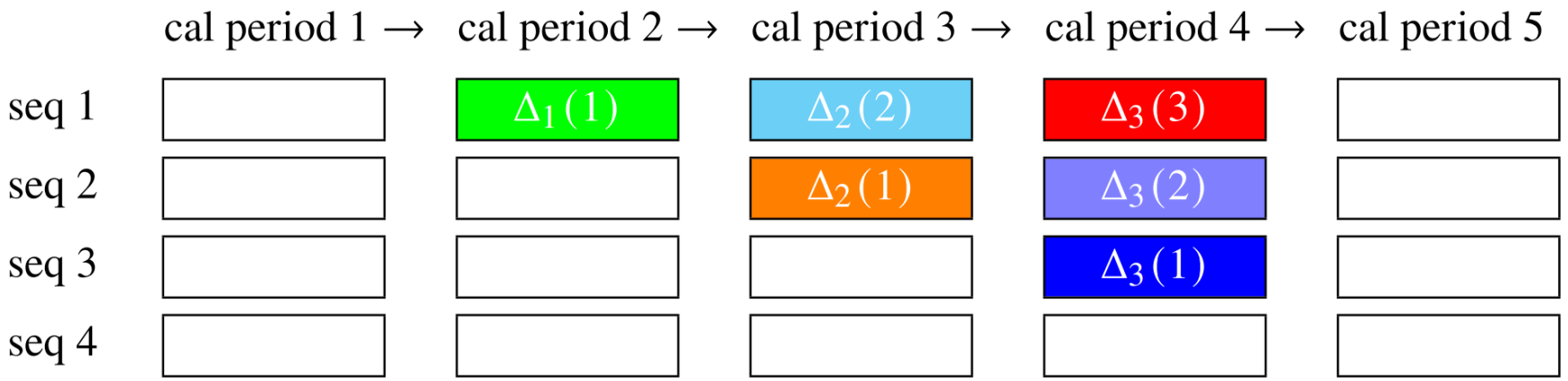


  - Treatment effect is seasonal, or is affected by external events (pandemic or other concurrent intervention) that are correlated with the calendar time
  - Δ_J not defined because $Y_{ijk}(0)$ is truncated by design
  - bias due to ignoring exposure time, however, should be more concerning that bias due to ignoring calendar time in treatment effect structure [9]

[9] Lee, K. M., Turner, E. L., & Kenny, A. (2025). Analysis of stepped-wedge cluster randomized trials when treatment effects vary by exposure time or calendar time. *Stat in Med.,* 44(20-22), e70256.

Case 4: Saturated treatment effect

- The finest set of estimands



  - Make the least assumptions, include the most parameters to estimate, but not always easy to interpret
  - Typical summary estimands in Cases 2-4 as an overall effect measure:

$$\Delta^{D\text{-avg}} = J^{-1} \sum_{d=1}^{J} \Delta(d)$$
$$\Delta^{P\text{-avg}} = (J-1)^{-1} \sum_{j=1}^{J-1} \Delta_j$$
$$\Delta^{S\text{-avg}} = \{(J-1)J\}^{-1} 2 \sum_{j=1}^{J-1} \sum_{d=1}^{j} \Delta_j(d)$$

Linear mixed model

- Consider the following working linear mixed model (LMM)

$$Y_{ijk} = \underbrace{\beta_{0j}}_{\text{secular}} + \underbrace{\boldsymbol{\beta}_{\boldsymbol{X}}^{\top} \boldsymbol{X}_{ik}}_{\text{covariates}} + TE_{ij} + RE_{ij} + \varepsilon_{ijk},$$

TABLE 1 Summary of estimands, model specifications, and estimators when a working linear mixed model is considered.

Treatment effect structure	TE_{ij} in model (3)	Point estimator	Variance estimator
Constant	$\beta_Z I\{Z_i \leq j\}$	$\widehat{\beta}_Z$	$\widehat{V}$
Duration-specific	$\sum_{d=1}^{j-1} \beta_{Zd} I\{Z_i = j-d+1\}$	$\widehat{\boldsymbol{\beta}}_Z^D = (\widehat{\beta}_{Z1}, \ldots, \widehat{\beta}_{ZJ})^{\top}$	$\widehat{\mathbf{V}}^D$
Period-specific	$\beta_{jZ} I\{Z_i \leq j\}$	$\widehat{\boldsymbol{\beta}}_Z^P = (\widehat{\beta}_{1Z}, \ldots, \widehat{\beta}_{J-1,Z})^{\top}$	$\widehat{\mathbf{V}}^P$
Saturated	$\sum_{d=1}^{j-1} \beta_{jZd} I\{Z_i = j-d+1\}$	$\widehat{\boldsymbol{\beta}}_Z^S = (\widehat{\beta}_{1Z1}, \ldots, \widehat{\beta}_{J-1,Z,J-1})^{\top}$	$\widehat{\mathbf{V}}^S$

- (1) $RE_{ij} = 0$,
- (2) exchangeable: $RE_{ij} = \alpha_i$ with $\alpha_i \sim N(0, \tau^2)$,
- (3) nested exchangeable: $RE_{ij} = \alpha_i + \gamma_{ij}$ with $\alpha_i \sim N(0, \tau^2)$ and $\gamma_{ij} \sim N(0, \kappa^2)$
- Cluster-robust sandwich variance estimator

What this work is about

Volume 80, Issue 4
December 2024

JOURNAL ARTICLE

How to achieve model-robust inference in stepped wedge trials with model-based methods?

Bingkai Wang, Xueqi Wang, Fan Li

Biometrics, Volume 80, Issue 4, December 2024, ujae123,
https://doi.org/10.1093/biomtc/ujae123

- **Gap:** little theory on design-based properties in SW-CRTs, and a model includes many moving parts
- **Objectives:**
  - Articulate a **potential outcomes framework** to define estimands that allow for treatment effect heterogeneity across calendar time and/or exposure time
  - Adapt **linear mixed models** and **GEE marginal models** (in published paper) to construct consistent estimators that are robust to working model misspecification
- Offer insights on what stepped wedge randomization can offer and cannot offer to enhance its causal inference rigor[10]

[10]Wang, B., Wang, X., & **Li, F.** (2024). How to achieve model-robust inference in stepped wedge trials with model-based methods? *Biometrics,* 45(3-5), e70380

What does stepped wedge randomization offer?

- **Take-away:** randomization guarantees consistency in working linear mixed models to causal estimands
  - As long as the assumed treatment effect structure **is correct**;
  - Even if **all other aspects of the working model are wrong**
- **What does randomization offer in a SW-CRT?**
  - Randomization offer robustness to misspecification of **covariate effects**, **random-effects structure** and **error structure**
  - But need to have the right **treatment effect structure**!
  - An additional complexity that is usually **not a concern** in standard parallel-arm CRTs

Result 1. (Model-robust property of LMM)

Under mild assumptions, (a) assuming a constant treatment effect,

$$\widehat{V}^{-1/2}(\widehat{\beta}_Z - \Delta) \xrightarrow{d} N(0,1);$$

(b) assuming duration-specific treatment effects,

$$\left(\widehat{\mathbf{V}}^D\right)^{-1/2}\left(\widehat{\boldsymbol{\beta}}_Z^D - \boldsymbol{\Delta}^D\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_J);$$

(c) assuming period-specific treatment effects,

$$\left(\widehat{\mathbf{V}}^P\right)^{-1/2}\left(\widehat{\boldsymbol{\beta}}_Z^P - \boldsymbol{\Delta}^P\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{J-1});$$

(d) assuming saturated treatment effects,

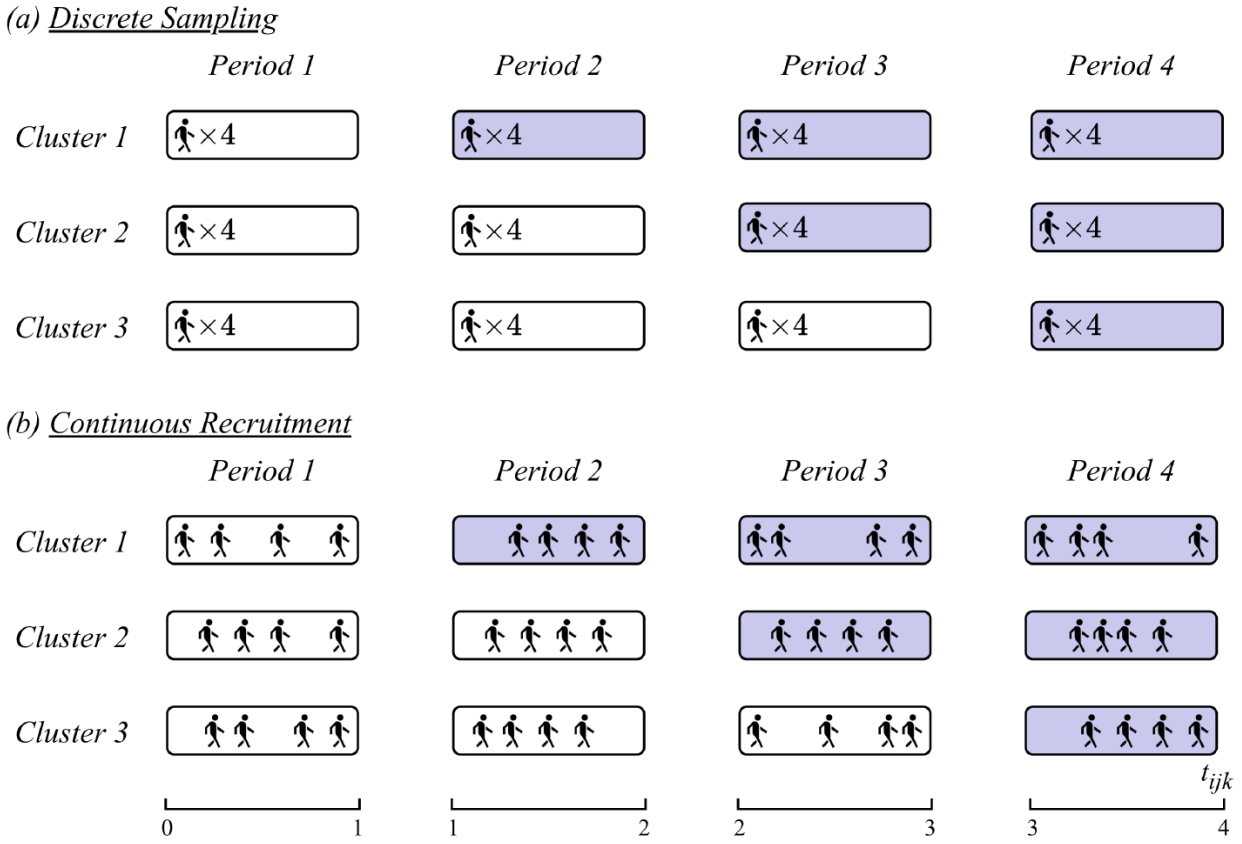
$$\left(\widehat{\mathbf{V}}^S\right)^{-1/2}\left(\widehat{\boldsymbol{\beta}}_Z^S - \boldsymbol{\Delta}^S\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{J(J-1)/2}),$$

where $\mathbf{I}_q$ *is the* $q \times q$ *identity matrix for a positive integer* q.

4. Double complexity: when time becomes continuous

Discrete sampling vs continuous recruitment?

- **Discrete sampling designs**
  - Individuals are recruited at **fixed** time points (typically once per period)
- **Continuous recruitment designs**
  - Individuals to enter the trial at **arbitrary** times within each period
  - Some strong references include (Hooper and Copas, 2019 *J. Clin. Epidemiol.*; Hooper et al., 2020 *BMC MRM*; Hooper et al., 2024 *Clinical Trials*)
- A recent systematical review included 160 SW-CRTs[11]
  - 76.3% were cross-sectional designs; among these, 95.1% implemented **continuous recruitment**
  - In sharp contrast to conventional **discrete-time analysis**



[11]Nevins, P., Ryan, M., Davis-Plourde, K., Ouyang, Y., ... **Li, F.**, Taljaard, M. (2024). Adherence to key recommendations for design and analysis of stepped-wedge cluster randomized trials: A review of trials published 2016-2022. *Clinical Trials,* 21(2), 199-210.

What do we know already?

- **Design methods** for SW-CRTs with continuous recruitment
  - Ignoring continuous recruitment leads to an underestimation of the required sample size (Grantham et al., 2019 *SIM*)
  - Efficient designs for SW-CRTs with continuous recruitment (Hooper et al., 2020 *BMC MRM*; Hooper et al., 2024 *Clinical Trials*)
- **Analysis methods** for SW-CRTs with model misspecification
  - LMMs are robust against misspecification of covariate effects, the correlation structure, and the error structure, as long as the intervention effect structure is correctly specified (Wang et al., 2024 *Biometrics*)
  - Robust variance estimator (RVE) can provide nominal coverage for the intervention effect under LMMs even when the correlation structure is misspecified (Ouyang., 2024 *SMMR*)
- Limited attention to analyzing SW-CRTs with **continuous recruitment**

What do we want to know more?

- **Gap: Lack of evidence** that discrete-time analysis of continuous recruitment designs delivers credible result
  - True data-generating process includes **continuous recruitment**
  - Working LMMs are still **discrete-time**
- **Objectives:**
  - Examine the behaviors of LMM estimators with **model misspecification**
  - How to draw **model-robust inference** when analyzing such SW-CRTs?
  - Implications for **trial planning** in the case of continuous recruitment
- Special considerations are needed when evaluating treatment impacts in SW-CRTs with continuous recruitment

Envisioning the data-generating process under continuous recruitment

- **Recruitment time** $t_{ijk} \in (j - 1, j]$
  - Individual k from cluster i in period j
- **Continuous period effects** $T(t_{ijk})$
  - $T(t_{ijk})$ function that describes the underlying **continuous effect** of time on expected outcome in terms of recruitment time t_{ijk}
- **Continuous-time decay (CTD) correlation structure** (Grantham et al. 2019 *SIM*; Hooper et al., 2024 *Clinical Trials*)
  - $\text{Cov}(Y_{ijk}, Y_{ij'k'}) = \sigma_\gamma^2 r^{|t^{ijk} - t^{ij'k'}|}$
  - In contrast, $\text{Cov}(Y_{ijk}, Y_{ij'k'}) = \sigma_\gamma^2 r^{|j-j'|}$ under the DTD structure (Kasza et al. 2017 *SMMR*)
- We still focus on **discrete intervention effect structure**

Continuous recruitment patterns

- Quantify the recruitment pattern
  - In terms of **cluster-period-specific enrollment density**
  - i.e., distribution of individual enrollment timing in each cluster-period
- Three basic patterns:
  - Uniform (U) Pattern: $t_{ijk} \sim \text{Uniform}(j-1, j]$
  - Normal Pattern: $t_{ijk} \sim \text{N}(0, 1)$
  - Exponential Pattern: $t_{ijk} \sim \text{Exp}(0, 1.5)$
- More complex patterns:
  - Cluster (C) mixed pattern: draw one of the three distributions with probability 1/3 for each **cluster**
  - Cluster-period (CP) mixed pattern: draw one of the three distributions with probability 1/3 for each **cluster-period**

Figure 1: Cluster Mixed

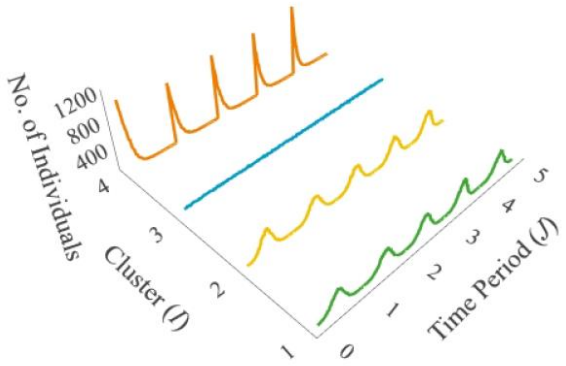


Figure 2: Cluster-Period Mixed

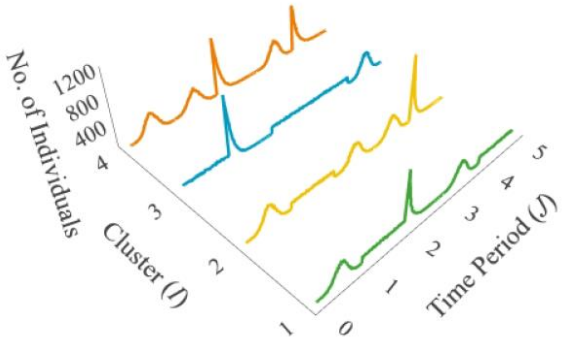
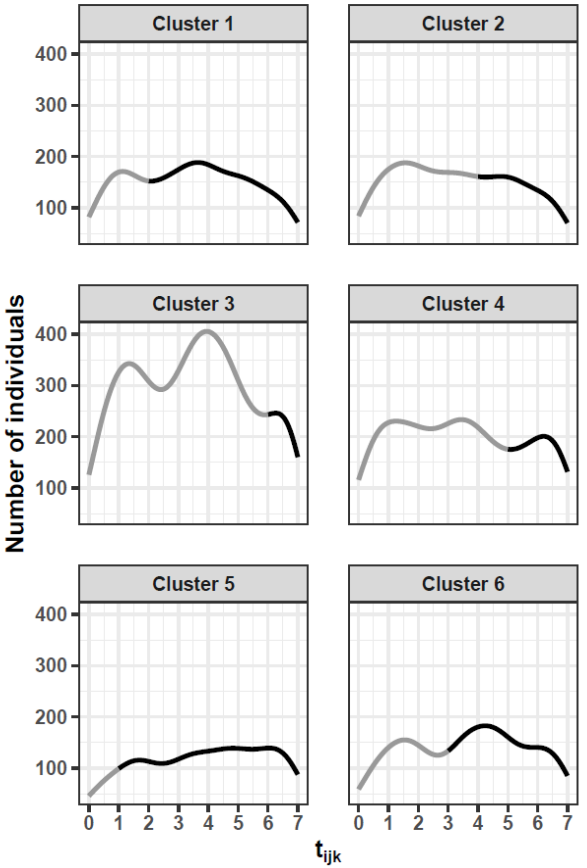


Figure 3: Empirical Patterns
(Haines et al., 2017 PLoS Medicine)



Impact of continuous period effects

- How **continuous period effects** in the true data-generating process affects the performance of discrete-time LMMs?

- **True data-generating process:**

$$Y_{ijk} = \mu + \underbrace{\beta_j}_{\text{Discrete}} + (\delta + v_i)Z_{ij} + \underbrace{\gamma_{i,t_{ijk}}}_{\text{CTD}} + \epsilon_{ijk}, \quad (1)$$

$$Y_{ijk} = \mu + \underbrace{T(t_{ijk})}_{\text{Continuous}} + (\delta + v_i)Z_{ij} + \underbrace{\gamma_{i,t_{ijk}}}_{\text{CTD}} + \epsilon_{ijk}, \quad (2)$$

where $v_i \sim \mathcal{N}(0, \tau_v^2)$ is the random intervention (RI) effect

- **Working models: discrete-time** LMMs

Correlation Structure	Model Specification
Simple Exchangeable (EXCH)	$Y_{ijk} = \mu + \beta_j + \delta Z_{ij} + \alpha_i + \epsilon_{ijk}$
Nested Exchangeable (NE)	$Y_{ijk} = \mu + \beta_j + \delta Z_{ij} + \alpha_i + \gamma_{ij} + \epsilon_{ijk}$
Discrete-Time Decay (DTD)	$Y_{ijk} = \mu + \beta_j + \delta Z_{ij} + \gamma_{ij} + \epsilon_{ijk}$

Impact of continuous period effects - cont'd

- 32 clusters, 5 periods, and 50 individuals per cluster-period
  - Within-period ICC under control: 0.01
  - Within-period ICC under treatment: 0.1
  - CAC = 0.5 and $\delta = 0$

Period Effect	Pattern	Model	Bias	$\text{sd}(\hat{\delta})$	$\mathbb{V}_{\text{Naive}}(\hat{\delta})$	$\mathbb{C}_{\text{Naive}}(\hat{\delta})$	$\mathbb{V}_{\text{RVE}}(\hat{\delta})$	$\mathbb{C}_{\text{RVE}}(\hat{\delta})$	$\mathbb{V}^{\text{MD}}_{\text{RVE}}(\hat{\delta})$	$\mathbb{C}^{\text{MD}}_{\text{RVE}}(\hat{\delta})$
Discrete	CP	EXCH	0.0005	0.0727	0.0398	**73.60**	0.0681	**93.20**	0.0727	**95.00**
		NE	0.0010	0.0731	0.0593	**89.55**	0.0681	**93.15**	0.0727	**95.20**
		DTD	0.0011	0.0727	0.0568	**88.55**	0.0676	**93.95**	0.0722	**95.05**
Continuous	CP	EXCH	0.0007	0.0741	0.0405	**73.90**	0.0699	**93.70**	0.0745	**95.45**
		NE	0.0015	0.0742	0.0614	**89.75**	0.0696	**93.60**	0.0743	**95.25**
		DTD	0.0015	0.0741	0.0591	**89.15**	0.0694	**93.80**	0.0740	**95.30**

- **Main findings:**
  - Estimates are **consistent** for the true intervention effect
  - Model-based variance estimator leads to an **under-coverage**
  - RVE with the Mancl and DeRouen (MD) correction leads to the **nominal coverage**

Impact of recruitment patterns

Period Effect	Pattern	Model	Bias	$\text{sd}(\hat{\delta})$	$\mathbb{V}_{\text{Naive}}(\hat{\delta})$	$\mathbb{C}_{\text{Naive}}(\hat{\delta})$	$\mathbb{V}_{\text{RVE}}(\hat{\delta})$	$\mathbb{C}_{\text{RVE}}(\hat{\delta})$	$\mathbb{V}^{\text{MD}}_{\text{RVE}}(\hat{\delta})$	$\mathbb{C}^{\text{MD}}_{\text{RVE}}(\hat{\delta})$
Continuous	CP	EXCH	0.0007	0.0741	0.0405	**73.90**	0.0699	**93.70**	0.0745	**95.45**
		NE	0.0015	0.0742	0.0614	**89.75**	0.0696	**93.60**	0.0743	**95.25**
		DTD	0.0015	0.0741	0.0591	**89.15**	0.0694	**93.80**	0.0740	**95.30**
	C	EXCH	0.0011	0.0726	0.0405	**73.80**	0.0689	**94.65**	0.0735	**95.85**
		NE	0.0006	0.0731	0.0609	**91.15**	0.0688	**94.15**	0.0734	**95.55**
		DTD	0.0009	0.0721	0.0582	**90.45**	0.0683	**94.10**	0.0729	**95.70**
	U	EXCH	0.0005	0.0719	0.0405	**74.35**	0.0683	**94.10**	0.0728	**95.50**
		NE	0.0003	0.0723	0.0595	**90.35**	0.0681	**93.40**	0.0726	**95.15**
		DTD	-0.0001	0.0721	0.0572	**88.85**	0.0677	**93.90**	0.0722	**95.00**

- **Main findings:**
  - Recruitment patterns have **almost no influence** on the performance of discrete-time LMMs or their variance estimators
  - Empirical standard error of the treatment effect estimator under the cluster-period mixed pattern is **slightly higher** compared to the uniform and the cluster mixed patterns

Impact of intervention-dependent recruitment

- **Recruitment sizes** depend on the intervention
  - Balanced recruitment (50 for both control and treatment)
  - Moderately unbalanced (25 for control; 75 for treatment)
  - Severely unbalanced (10 for control; 90 for treatment)

K_{ij}	Pattern	Model	Bias	$\text{sd}(\hat{\delta})$	$\mathbb{V}_{\text{Naive}}(\hat{\delta})$	$\mathbb{C}_{\text{Naive}}(\hat{\delta})$	$\mathbb{V}_{\text{RVE}}(\hat{\delta})$	$\mathbb{C}_{\text{RVE}}(\hat{\delta})$	$\mathbb{V}^{\text{MD}}_{\text{RVE}}(\hat{\delta})$	$\mathbb{C}^{\text{MD}}_{\text{RVE}}(\hat{\delta})$
50 + 50	CP + CP	EXCH	-0.0001	0.0432	0.0234	**72.40**	0.0417	**94.00**	0.0425	**94.55**
		NE	-0.0003	0.0432	0.0356	**88.90**	0.0416	**93.95**	0.0425	**94.30**
		DTD	-0.0003	0.0432	0.0342	**88.20**	0.0415	**94.00**	0.0424	**94.65**
25 + 75	CP + CP	EXCH	-0.0059	0.0452	0.0258	**74.45**	0.0440	**94.15**	0.0450	**94.65**
		NE	-0.0057	0.0436	0.0383	**91.20**	0.0422	**93.60**	0.0431	**94.15**
		DTD	-0.0059	0.0437	0.0368	**90.65**	0.0426	**94.10**	0.0435	**94.45**
10 + 90	CP + CP	EXCH	0.0144	0.0515	0.0331	**78.55**	0.0513	**94.05**	0.0523	**94.35**
		NE	-0.0144	0.0481	0.0431	**90.85**	0.0473	**93.40**	0.0483	**93.90**
		DTD	-0.0145	0.0483	0.0432	**91.50**	0.0478	**93.50**	0.0488	**94.25**

- **Main findings:**
  - Intervention effect estimator is **consistent**
  - RVEs **achieve the nominal coverage**

Impact of intervention-dependent recruitment - cont'd

- **Recruitment patterns** depend on the intervention
  - Control: **uniform** patterns
  - Treatment: **cluster-period mixed** pattern

K_{ij}	Pattern	Model	Bias	$\text{sd}(\hat{\delta})$	$\mathbb{V}_{\text{Naive}}(\hat{\delta})$	$\mathbb{C}_{\text{Naive}}(\hat{\delta})$	$\mathbb{V}_{\text{RVE}}(\hat{\delta})$	$\mathbb{C}_{\text{RVE}}(\hat{\delta})$	$\mathbb{V}^{\text{MD}}_{\text{RVE}}(\hat{\delta})$	$\mathbb{C}^{\text{MD}}_{\text{RVE}}(\hat{\delta})$
50 + 50	U + CP	EXCH	-0.0297	0.0426	0.0234	**62.30**	0.0412	**88.55**	0.0421	**89.45**
		NE	-0.0294	0.0429	0.0354	**82.20**	0.0412	**88.55**	0.0421	**89.45**
		DTD	-0.0293	0.0424	0.0339	**80.05**	0.0410	**88.70**	0.0419	**89.70**
25 + 75	U + CP	EXCH	-0.0307	0.0445	0.0258	**64.10**	0.0439	**89.25**	0.0448	**90.15**
		NE	-0.0322	0.0432	0.0382	**84.00**	0.0419	**87.75**	0.0428	**88.30**
		DTD	-0.0324	0.0437	0.0367	**82.30**	0.0423	**87.85**	0.0432	**88.45**
10 + 90	U + CP	EXCH	-0.0274	0.0520	0.0331	**73.70**	0.0511	**91.80**	0.0522	**92.45**
		NE	-0.0291	0.0482	0.0430	**87.65**	0.0473	**90.35**	0.0482	**91.05**
		DTD	-0.0297	0.0487	0.0432	**87.45**	0.0477	**90.80**	0.0487	**91.30**

- **Main findings:**
  - Intervention effect estimator **carries bias**
  - RVEs **fail to achieve nominal coverage**
    - Mainly caused by the biased estimate of the intervention effect

Main takeaways

- From the **analysis** perspective[12]
  - Discrete-time LMMs can yield **consistent** intervention effect estimates
  - Model-based variance estimators **underestimate** empirical variance
  - The **RVE with the MD correction** can ensure the validity of inference
  - **Exception:** recruitment patterns change systematically between control and intervention period
- From the **trial planning** perspective
  - Documenting **recruitment patterns** during the trial
  - **Privacy issue:** when exact enrollment dates may identify participants, cluster-period-specific enrollment densities provide an acceptable alternative

Can discrete-time analyses be trusted for stepped wedge trials with continuous recruitment?

Hao Wang[1,2], **Guangyu Tong**[1,2,3], **Heather Allore**[1,4], **Monica Taljaard**[5,6], **and Fan Li**[1,2,3,*]

Abstract

In stepped wedge cluster randomized trials (SW-CRTs), interventions are sequentially rolled out to clusters over multiple periods. It is common practice to analyze SW-CRTs using discrete-time linear mixed models, in which measurements are considered to be taken at discrete time points. However, a recent systematic review found that 95.1% of cross-sectional SW-CRTs recruit individuals continuously over time. Despite the high prevalence of designs with continuous recruitment, there has been limited guidance on how to draw model-robust inference when analyzing such SW-CRTs. In this article, we investigate through simulations the implications of using discrete-time linear mixed models in the case of continuous recruitment designs with a continuous outcome. First, in the data-generating process, we characterize continuous recruitment with a continuous-time exponential decay correlation structure in the presence or absence of a continuous period effect, addressing scenarios both with and without a random or exposure-time-dependent intervention effect. Then, we analyze the simulated data under three popular discrete-time working correlation structures: simple exchangeable, nested exchangeable, and discrete-time exponential decay, with a robust sandwich variance estimator. Our results demonstrate that discrete-time analysis often yields minimum bias, and the robust variance estimator with the Mancl and DeRouen correction consistently achieves nominal coverage and type I error rate. One important exception occurs when recruitment patterns vary systematically between control and intervention periods, where discrete-time analysis leads to slightly biased estimates. Finally, we illustrate these findings by reanalyzing a concluded SW-CRT.

Keywords

Cluster Randomized Trials, Continuous-Time Decay, Linear Mixed Models, Model Misspecification, Robust Sandwich Variance, Recruitment Pattern

[12]Wang, H., Tong, G., Allore, H., Taljaard, M., & **Li, F.** (2025). Can discrete-time analyses be trusted for stepped wedge trials with continuous recruitment? *arXiv:2511.18731*

5. Discussion

Final reflections

- SW-CRTs are more complicated also because of **treatment effect structure**
  - Needs to be recognized
  - Although no consensus on best practice, secondary analysis is recommended
- Stepped wedge designs still offer a solid causal inference basis, but one needs to sort out the treatment effect structure
  - **Anticipation effect:** justify transition periods & incomplete designs
  - **Continuous-time recruitment:** the reality of stepped wedge designs, we need better recommendations
- More importantly, all these complexities should consist of reasons **not to consider** stepped wedge design unless you have to (**4 broad justifications**)[1]
- **Acknowledgement:** reference not inclusive, due to the massive literature already

[1]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.,* 49(3), 1043-1052.

Published literature on treatment effect structure (~2016 - 2019)

ELSEVIER

Contemporary Clinical Trials

Volume 28, Issue 2, February 2007, Pages 182-191

Design and analysis of stepped wedge cluster randomized trials

Michael A. Hussey [a], James P. Hughes [b]

Statistics in Medicine

Research Article | Full Access

Sample size calculation for stepped wedge and other longitudinal cluster randomised trials

Richard Hooper, Steven Teerenstra, Esther de Hoop, Sandra Eldridge

Biometrics

Issues | Advance Articles | Submit | Purchase | Alerts | About

JOURNAL ARTICLE

Sample Size Determination for GEE Analyses of Stepped Wedge Cluster Randomized Trials

Fan Li, Elizabeth L. Turner, John S. Preisser

Statistics in Medicine

RESEARCH ARTICLE | Full Access

Accounting for a decaying correlation structure in cluster randomized trials with continuous recruitment

Kelsey L. Grantham, Jessica Kasza, Stephane Heritier, Karla Hemming, Andrew B. Forbes

Statistical Methods in Medical Research

Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials

J Kasza, K Hemming, R Hooper, JNS Matthews, and AB Forbes on behalf of the ANZICS Centre for Outcomes & Resource Evaluation (CORE) Committee -3 View all

Statistics in Medicine

RESEARCH ARTICLE | Full Access

Design and analysis considerations for cohort stepped wedge cluster randomized trials with a decay correlation structure

Fan Li

Statistics in Medicine

TUTORIAL IN BIOSTATISTICS | Open Access

Analysis of Stepped-Wedge Cluster Randomized Trials: A Tutorial Using Marginal Models

Elizabeth L. Turner, John S. Preisser, Ying Zhang, Xueqi Wang, Mark Toles, Samuel Cykert, Fan Li, Paul J. Rathouz

International Journal of Epidemiology

Issues | More Content | Submit | Purchase | Alerts | About

JOURNAL ARTICLE

Estimating intra-cluster correlation coefficients for planning longitudinal cluster randomized trials: a tutorial

Yongdong Ouyang, Karla Hemming, Fan Li, Monica Taljaard

Statistical Methods in Medical Research

Maintaining the validity of inference from linear mixed models in stepped-wedge cluster randomized trials under misspecified random-effects structures

Yongdong Ouyang, Monica Taljaard, Andrew B Forbes, and Fan Li

Published literature on treatment effect structure (~2021 - current)

Statistics in Medicine

RESEARCH ARTICLE | Full Access

Analysis of stepped wedge cluster randomized trials in the presence of a time-varying treatment effect

Avi Kenny, Emily C. Voldal, Fan Xia, Patrick J. Heagerty, James P. Hughes

Scandinavian Journal of Statistics Theory and Applications

ORIGINAL ARTICLE | Full Access

Model-assisted analysis of covariance estimators for stepped wedge cluster randomized experiments

Xinyuan Chen, Fan Li

Biometrics

Issues | Advance Articles | Submit | Purchase | Alerts | About

Assessing Exposure-Time Treatment Effect Heterogeneity in Stepped-Wedge Cluster Randomized Trials FREE

Lara Maleyeff, Fan Li, Sebastien Haneuse, Rui Wang

Biometrics

Issues | Advance Articles | Submit | Purchase | Alerts | About

JOURNAL ARTICLE

How to achieve model-robust inference in stepped wedge trials with model-based methods?

Bingkai Wang, Xueqi Wang, Fan Li

Statistics in Medicine

RESEARCH ARTICLE | Full Access

On Anticipation Effect in Stepped Wedge Cluster Randomized Trials

Hao Wang, Xinyuan Chen, Katherine R. Courtright, Scott D. Halpern, Michael O. Harhay, Monica Taljaard, Fan Li

Statistics in Medicine

RESEARCH ARTICLE | Open Access

Analysis of Stepped-Wedge Cluster Randomized Trials When Treatment Effects Vary by Exposure Time or Calendar Time

Kenneth M. Lee, Elizabeth L. Turner, Avi Kenny

Journal of the Royal Statistical Society Series A: Statistics in Society

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS

Model-robust standardization in stepped wedge cluster randomized trials Get access

Xi Fang, Xueqi Wang, Patrick J Heagerty, Bingkai Wang, Fan Li

Statistics in Medicine

RESEARCH ARTICLE | Full Access

Cluster randomized trial designs for modeling time-varying intervention effects

Kenneth Menglin Lee, Yin Bun Cheung

Statistics in Medicine

RESEARCH ARTICLE | Open Access

Analysis of Cohort Stepped Wedge Cluster-Randomized Trials With Nonignorable Dropout via Joint Modeling

Alessandro Gasparini, Michael J. Crowther, Emiel O. Hoogendijk, Fan Li, Michael O. Harhay

A Forthcoming *CRT-Estimands Framework*

ELSEVIER

Journal of Clinical Epidemiology

Volume 189, January 2026, 112015

Original Research

A scoping review identified additional considerations for defining estimands in cluster randomized trials

Dongquan Bi [a], Andrew Copas [a], Fan Li [b c d], Brennan C. Kahan [a]

Home > Trials > Article

Development of a consensus extension of the estimands framework for cluster randomised trials (CRT-estimands): results from an international Delphi study

Research | Open access | Published: 18 December 2025

Volume 27, article number 70, (2026) Cite this article

You have full access to this open access article

Download PDF | Save article

Brennan C. Kahan, Melanie Bahti, Dongquan Bi, Frank Bretz, Gary S. Collins, Andrew Copas, Michael O. Harhay, Fan Li & Catherine L. Auriemma

The CRT-Estimands Framework: a consensus-based extension of the ICH E9(R1) addendum for cluster randomised trials

Brennan C Kahan[1], Frank Bretz[2,3], Andrew Copas[1], Michael O Harhay[1,4], Gary S Collins[5], Melanie Bahti[6], Dongquan Bi[1], Marion K Campbell[7], Katherine R Courtright[6], Timothy Feeney[8,9], Andrew B Forbes[10], Amit X Garg[11], Scott D Halpern[4,6], Patrick J Heagerty[12], Karla Hemming[13], Sally Hopewell[14], Jim Hughes[12], David M Murray[15], Joseph S Ross[16], Derek Stewart[17], Monica Taljaard[18,19], Obioha C Ukoumunne[20], Catherine L Auriemma[6], Fan Li[21]

article forthcoming at *BMJ*

Summary points

- To interpret and evaluate a randomised trial accurately, readers need a clear understanding of the research question being addressed (i.e., the estimand)
- The CRT-Estimands Framework provides guidance on defining estimands for cluster randomised trials
- The CRT-Estimands Framework contains five attributes that should be described when defining estimands for cluster randomised trials, an expanded table of points to consider when defining estimands, and guidance describing how estimands can be used to align statistical analysis methods and sensitivity analyses with key research objectives
- Trial investigators should use the CRT-Estimands Framework when defining estimands for cluster randomised trials to ensure trial objectives are transparent

Acknowledgement

Many of the published work supported by a Patient-Centered Outcomes Research Institute Award® (PCORI® Award ME-2022C2- 27676).

The statements presented are solely the responsibility of the authors and do not necessarily represent the official views of PCORI®, its Board of Governors, or the Methodology Committee.

This work in this presentation credits to my amazing trainees (e.g. Hao Wang) and faculty collaborators (e.g., Bingkai Wang, PhD), whose leadership made it possible.

Thank you, and Q&A