Incomplete Variants of Stepped Wedge Cluster Randomized Designs: Recent Design Innovations and Considerations for Implementation

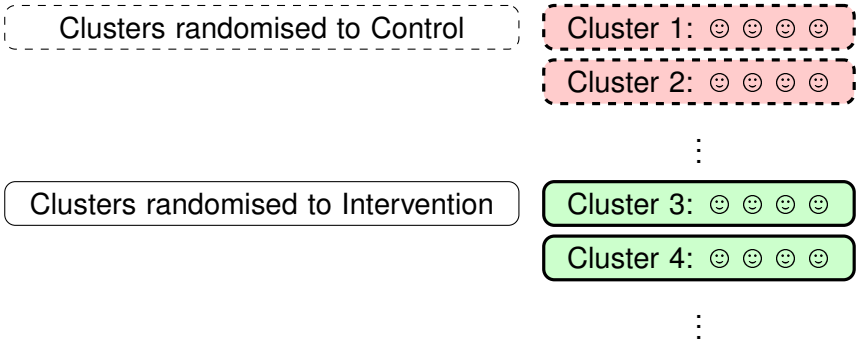
Jessica Kasza
jessica.kasza@monash.edu

Society for Clinical Trials 2025 Conference, Vancouver Canada

Funding acknowledgement

- JK is supported by an NHMRC Investigator Grant (L1), 2033380

Cluster randomised trials



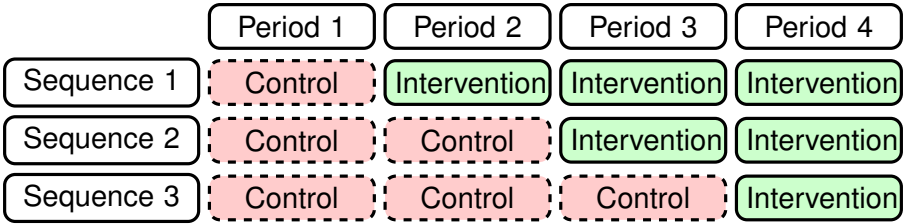
- Intervention at group level may mean that we can't randomise individuals
  - Instead we need to randomise *clusters (groups)* of participants to treatments.
- Clusters could be hospitals, intensive care units, schools, neighbourhoods...
  - The dependence between outcomes for individuals from the same cluster needs to be accounted for.

Sometimes cluster randomised trials are not enough...

We may need more clusters!

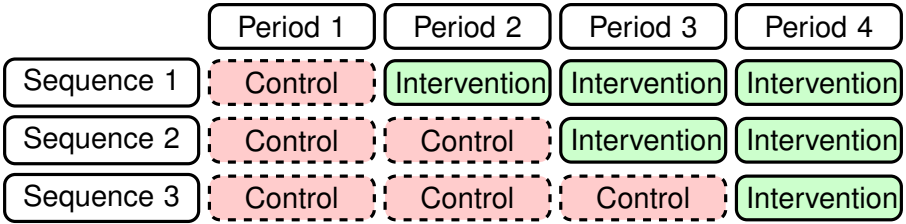
- Clustering reduces the amount of information available about the intervention effect.
- We may need to add more clusters to ensure we have sufficient power to detect intervention effects of interest
- But we might not be able to add more clusters.

The stepped wedge cluster randomised trial design



- Clusters randomised to a *sequence* of treatments.
- Useful when interventions cannot be undone or will be rolled out anyway.
- All clusters know they will receive the intervention (eventually...).

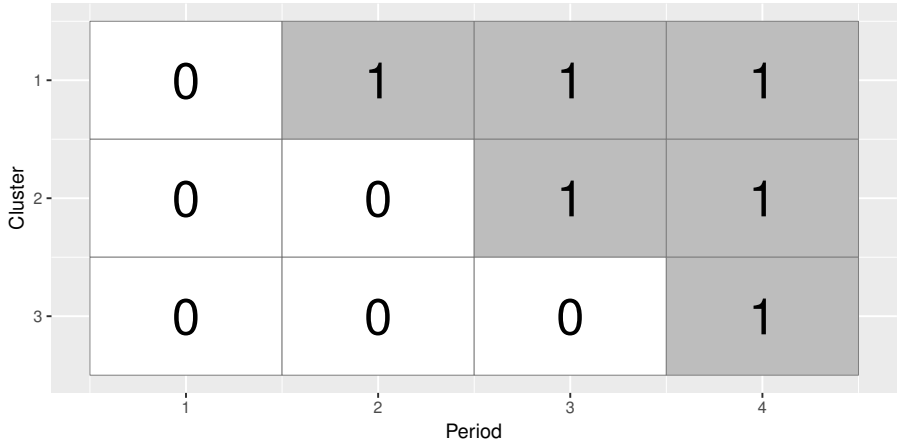
The stepped wedge cluster randomised trial design



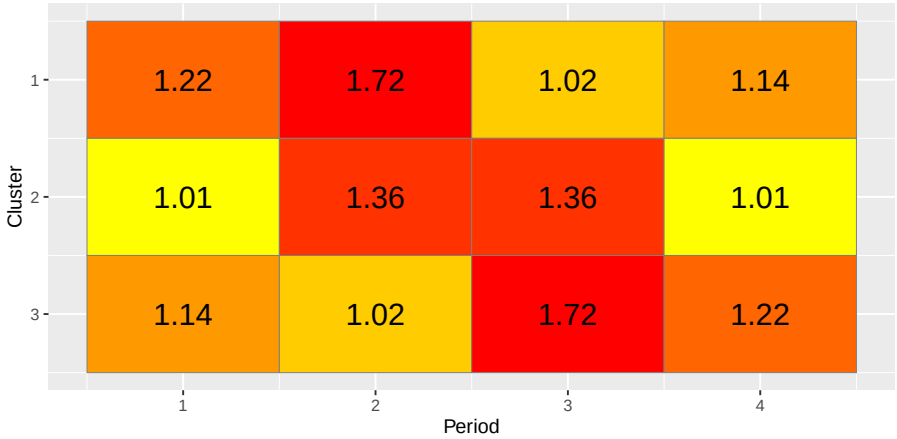
- Clusters randomised to a *sequence* of treatments.
- Useful when interventions cannot be undone or will be rolled out anyway.
- All clusters know they will receive the intervention (eventually...).

But these are burdensome designs!

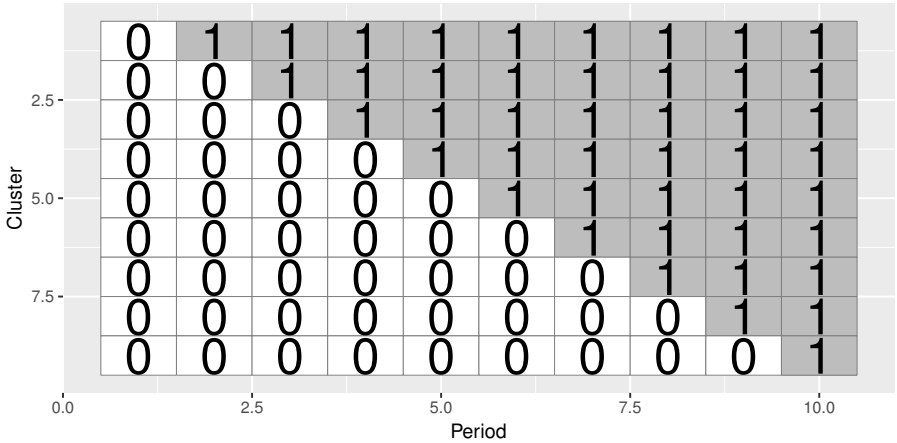
Different cluster-period cells contribute different amounts of information



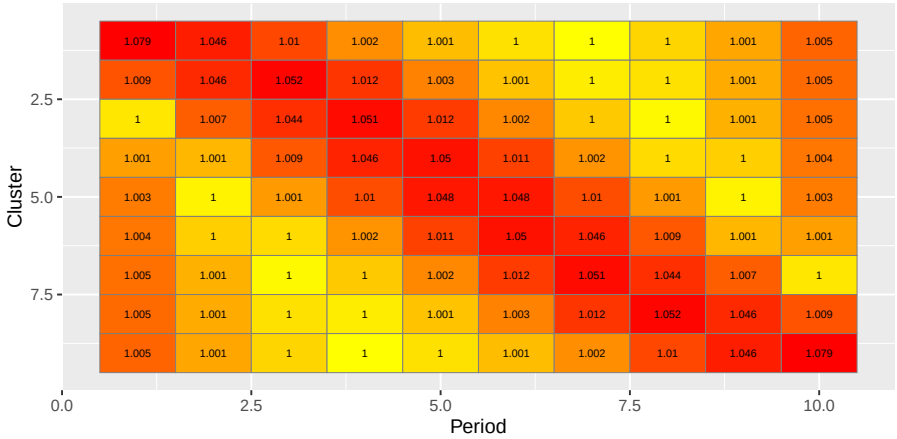
Different cluster-period cells contribute different amounts of information



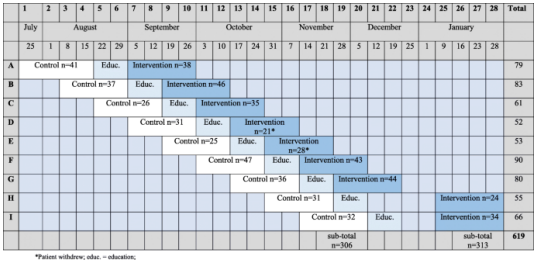
Different cluster-period cells contribute different amounts of information



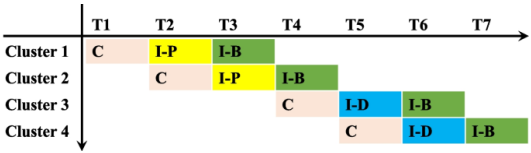
Different cluster-period cells contribute different amounts of information



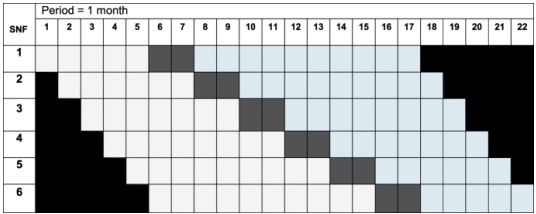
Incomplete variants of the stepped wedge design in practice



Keogh et al, BMC Medicine, 2020.



Zhu et al, BMC Psychiatry, 2023.



Toles et al, Trials, 2021.

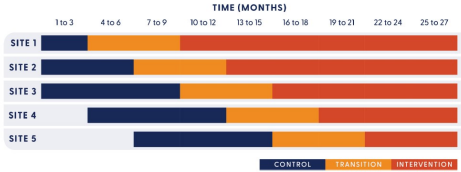


Figure 1 Stepped-wedge design.

Munday et al, BMJ Open, 2025.

This leads to lots of questions!

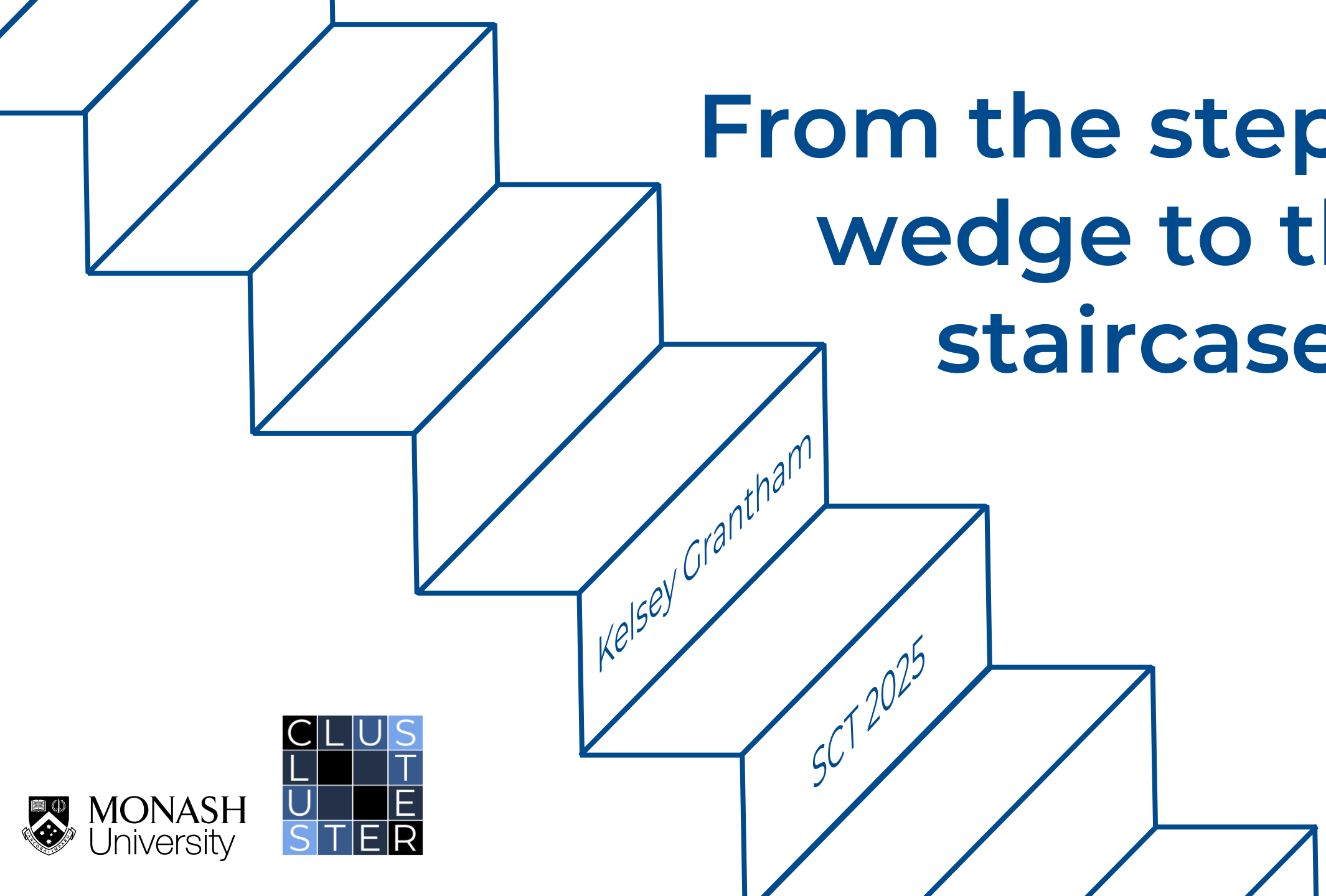
- What sorts of incomplete designs are most beneficial, and when?
- How can we choose the optimal incomplete design for a given setting?
- Does the answer depend on the approach we take to treatment effect estimation?
- How acceptable are incomplete variants of stepped wedge designs to trialists?

Session outline

1. Dr Kelsey Grantham, Monash University
   - From the stepped wedge to the staircase
2. Prof John Preisser, University of North Carolina at Chapel Hill
   - Marginal models, binary outcomes, and incomplete designs
3. Assoc Prof Fan Li, Yale University
   - Demystifying incomplete stepped wedge designs under the working independence assumption
4. Prof Monica Taljaard, University of Ottawa
   - Incomplete designs in practice

Get your questions ready for our panel discussion!

From the stepped wedge to the staircase



MONASH University

CLUSTER

Disclosures

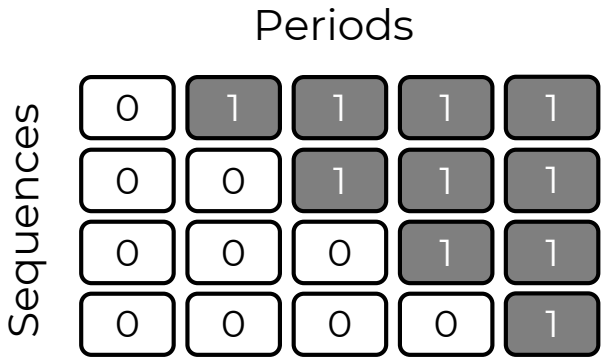
This work was supported by an NHMRC Investigator Grant (L1), 2033380

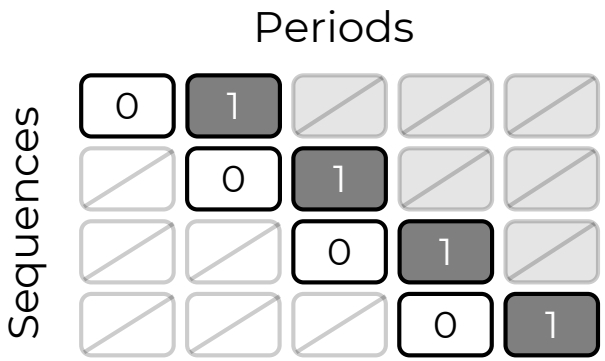
Travel was supported by a Monash SPHPM Southby Travel Grant

Complete vs. incomplete stepped wedge designs

- Complete stepped wedge designs are often too burdensome
  - Clusters must collect data in every period of the trial

- Incomplete stepped wedge designs have one or more cells omitted from the complete stepped wedge
  - No data collection required for omitted cells





0 = Control
1 = Intervention

Incomplete stepped wedge designs are on the rise

Wagner et al (2020)

- Intervention: Education program to improve student self-regulation
- Clusters: Schools in remote Australian Aboriginal communities
- Outcome: Student disruptive behaviour measured by teacher-rated questionnaire
- Motivation for design: “the burden of data collection on participants would be too great in a stepped wedge trial”

	2016						2017					
Study Year and Month	**Apr-May**	**May-Jun**	**End Jun**	**Jul-Sept**	**Sep**	**Nov-Dec**	**Apr-May**	**May-Jun**	**May-Jun**	**Jul-Aug**	**Sep**	**Nov-Dec**
Calendar Time	1		2	3	4	5	6		7	8	9	10
Cluster 1	0		1	1								
Cluster 2	0		0		1	1						
Cluster 3			0		0		0		1	1		
Cluster 4							0		0		1	1

Adapted from Wagner et al (2020)

Wagner B, Latimer J, Adams E, et al. School-based intervention to address self-regulation and executive functioning in children attending primary schools in remote Australian Aboriginal communities. *PLoS One* 2020; 15(6): 1-19.

Obtaining “good” incomplete designs

- “Good” = less burdensome + sufficient statistical power
- One way to obtain such designs is to use the information content of cluster-period cells to guide cell removal
  - Cells near the switch from control to intervention contribute the most information → Retain these cells
  - Cells more distant from the switches typically do not contribute much information → Omit these cells
- Starting with a complete design, repeatedly omit low-information cells until minimum acceptable power is reached (Rezaei-Darzi et al, 2023)
- Key questions: How many low-information cells can be omitted while still maintaining sufficient power? What do these designs tend to look like?

Rezaei-Darzi E, Grantham KL, Forbes AB, Kasza J. The impact of iterative removal of low-information cluster-period cells from a stepped wedge design. *BMC Medical Research Methodology* 2023; 23(1): 160.

Statistical model

The outcome for the ith $(i = 1, \ldots, m)$ participant measured in period t belonging to the kth $(k = 1, \ldots, K_s)$ cluster allocated to the sth $(s = 1, \ldots, S)$ sequence:

$$Y_{skti} = \beta_t + X_{st}\theta + CP_{skt} + e_{skti}$$

Measured outcome	Time period effect	Treatment effect	Cluster-period effect	Error

$$\boldsymbol{CP}_{sk} \sim N(0, \sigma_C^2 \boldsymbol{R}),$$
$$e_{skti} \sim N(0, \sigma_e^2)$$

Statistical model

$$Y_{skti} = \beta_t + X_{st}\theta + CP_{skt} + e_{skti}$$

- Y_{skti}: Measured outcome
- β_t: Time period effect
- $X_{st}\theta$: Treatment effect
- CP_{skt}: Cluster-period effect
- e_{skti}: Error

$$\boldsymbol{CP}_{sk} \sim N(0, \sigma_C^2 \boldsymbol{R}), \; e_{skti} \sim N(0, \sigma_e^2)$$

where

$$\boldsymbol{R} = \begin{bmatrix} 1 & r & \cdots & r \\ r & 1 & & \vdots \\ \vdots & & \ddots & r \\ r & \cdots & r & 1 \end{bmatrix} \longrightarrow$$

$$corr(Y_{skti}, Y_{skti'}) = \frac{\sigma_C^2}{\sigma_C^2+\sigma_e^2} = \rho,$$

$$corr(Y_{skti}, Y_{skt'i'}) = \frac{\sigma_C^2}{\sigma_C^2+\sigma_e^2} r = \rho r, \quad 0 \le r \le 1$$

- ρ: Within-period intracluster correlation (ICC)
- r: Cluster autocorrelation

Statistical model

$$Y_{skti} = \beta_t + X_{st}\theta + CP_{skt} + e_{skti}$$

Measured outcome	Time period effect	Treatment effect	Cluster-period effect	Error

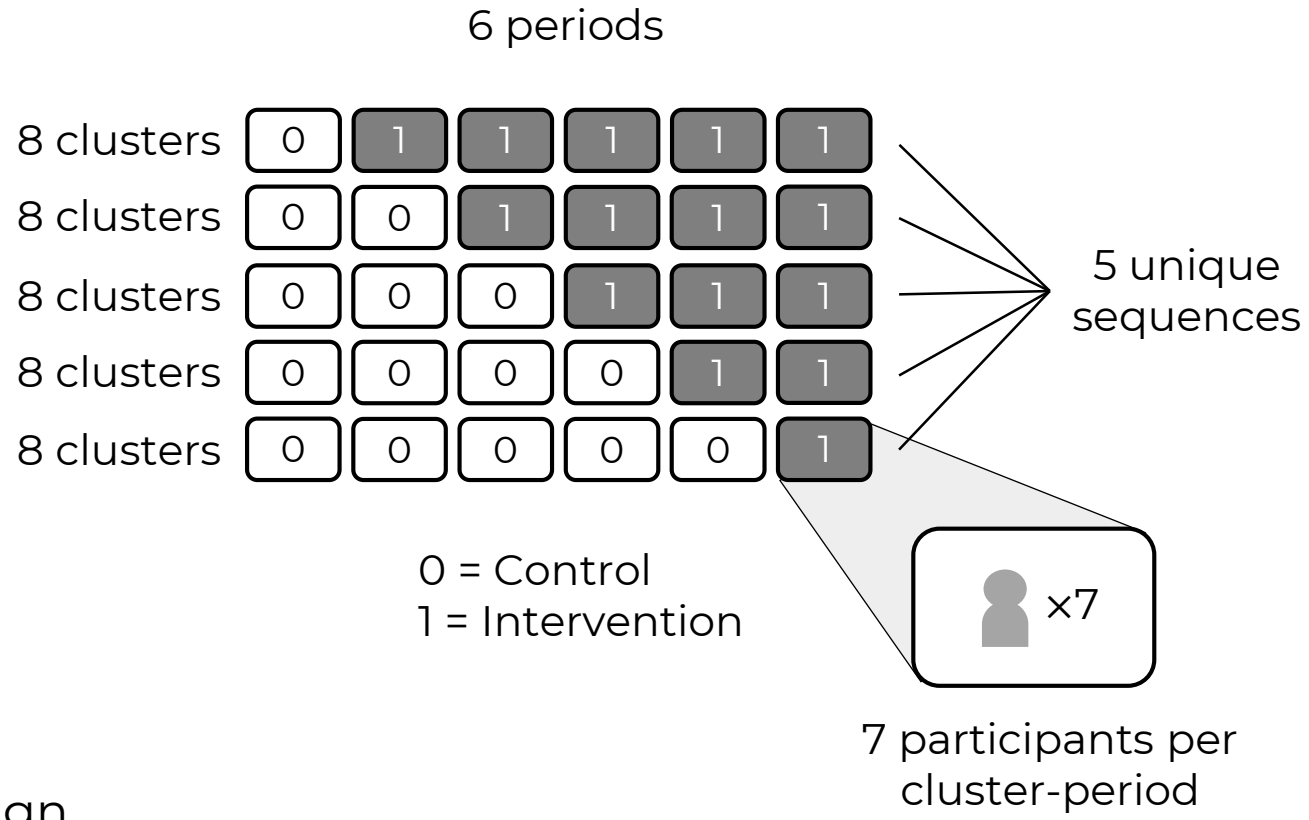
$$corr(Y_{skti}, Y_{skti'}) = \rho, \quad corr(Y_{skti}, Y_{skt'i'}) = \rho r$$

- Interested in $var(\hat{\theta})$, the variance of the treatment effect estimator
  - Where $\hat{\theta}$ is the generalised least squares estimator under this model
- The precision of the treatment effect estimator is: $1/var(\hat{\theta})$

Motivating example: ALLIANCE trial*

Mazza et al (2023)

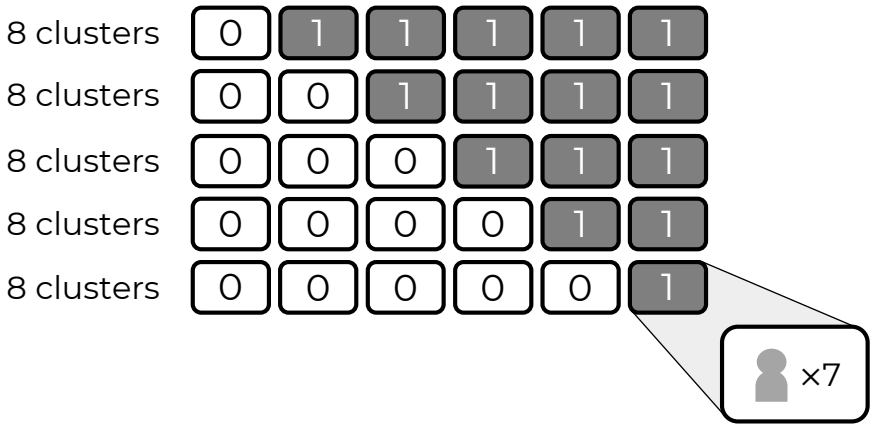
- Population: Women seeking emergency contraception
- Intervention: Provision of contraceptive counselling
- Clusters: Pharmacies
- Outcome: Use of long-acting contraception



*With slight modifications to the trial design

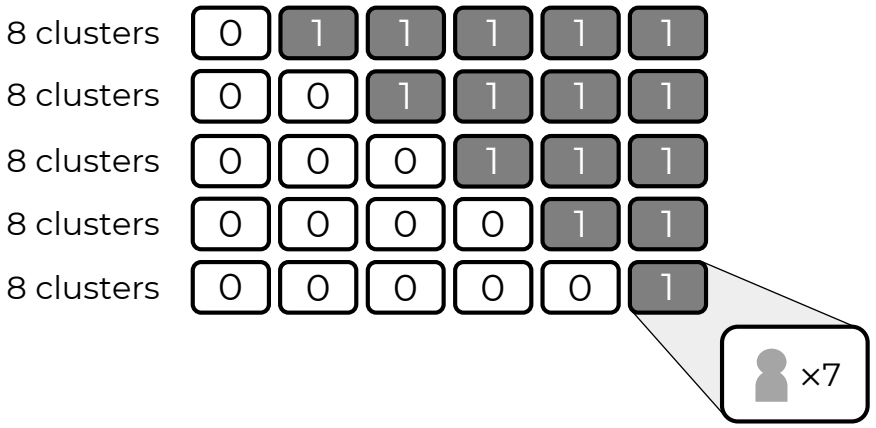
Mazza D, Assifi AR, Hussainy SY, et al. Expanding community pharmacists' scope of practice in relation to contraceptive counselling and referral: A protocol for a pragmatic, stepped-wedge, cluster randomised trial (ALLIANCE). *BMJ Open* 2023; 13(8): e073154.

Motivating example: ALLIANCE trial



- Within-period ICC, $\rho = 0.05$
- Cluster autocorrelation, $r = 0.95$
- Standardised effect size: 0.26

Motivating example: ALLIANCE trial



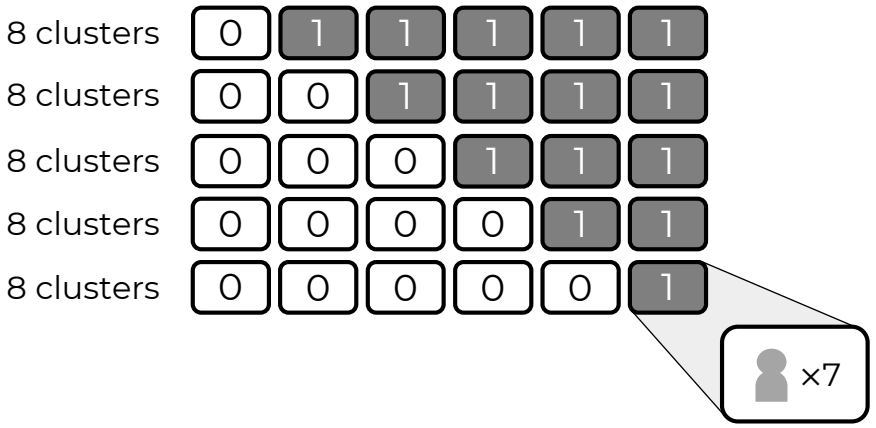
- Within-period ICC, $\rho = 0.05$
- Cluster autocorrelation, $r = 0.95$
- Standardised effect size: 0.26

Information content of cells
(Kasza and Forbes, 2019) :

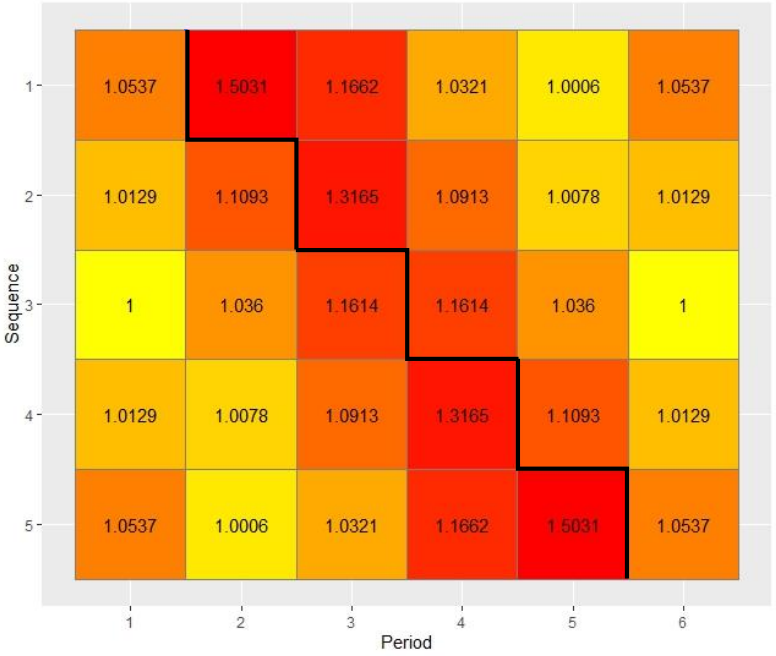
$$\frac{var\left(\hat{\theta}\right) \text{ after omitting the cells}}{var\left(\hat{\theta}\right) \text{ with the cells retained}}$$

Kasza J and Forbes AB. Information content of cluster-period cells in stepped wedge trials. *Biometrics* 2019; 75(1): 144-52.

Motivating example: ALLIANCE trial



- Within-period ICC, $\rho = 0.05$
- Cluster autocorrelation, $r = 0.95$
- Standardised effect size: 0.26



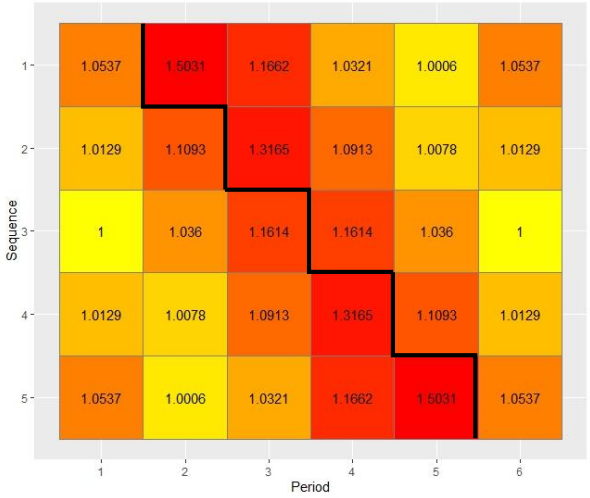
Information content of cells (Kasza and Forbes, 2019) :

$$\frac{var(\hat{\theta}) \text{ after omitting the cells}}{var(\hat{\theta}) \text{ with the cells retained}}$$

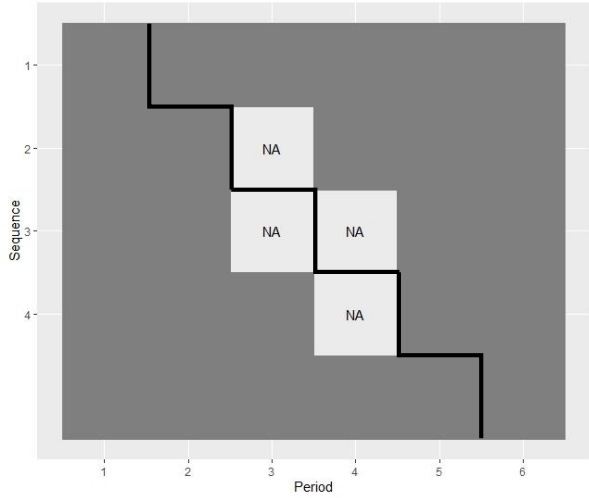
Kasza J and Forbes AB. Information content of cluster-period cells in stepped wedge trials. *Biometrics* 2019; 75(1): 144-52.

Obtaining efficient incomplete designs

Complete



Minimally viable



% of cells retained

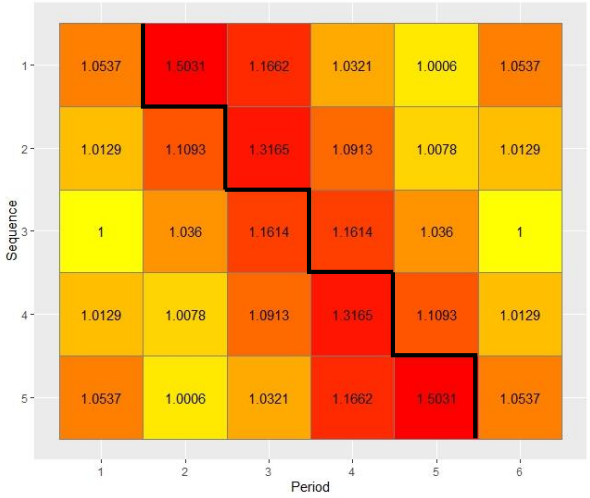
100% 6.7%

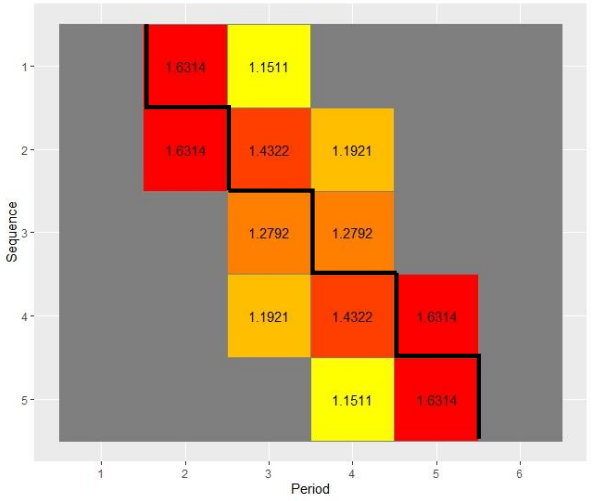
Power (to detect effect size of 0.26)

91.9% 44.7%

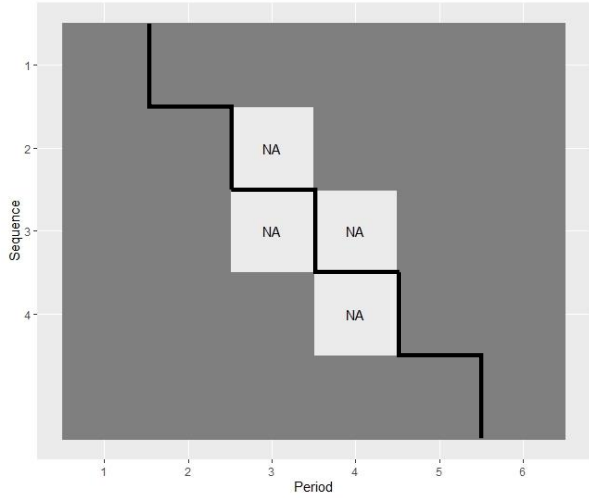
Obtaining efficient incomplete designs

Complete





Minimally viable



% of cells retained

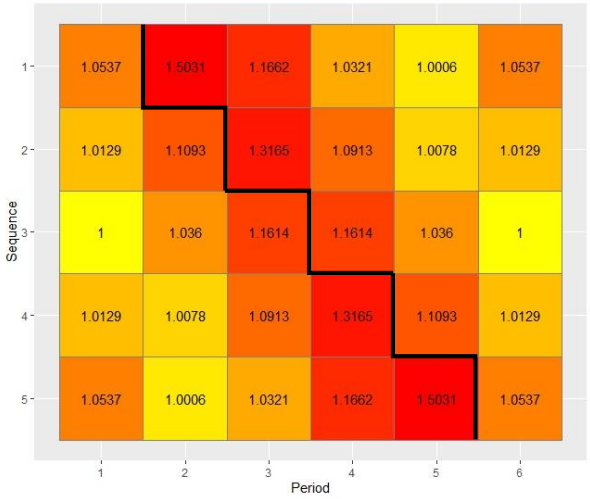
100% 40% 6.7%

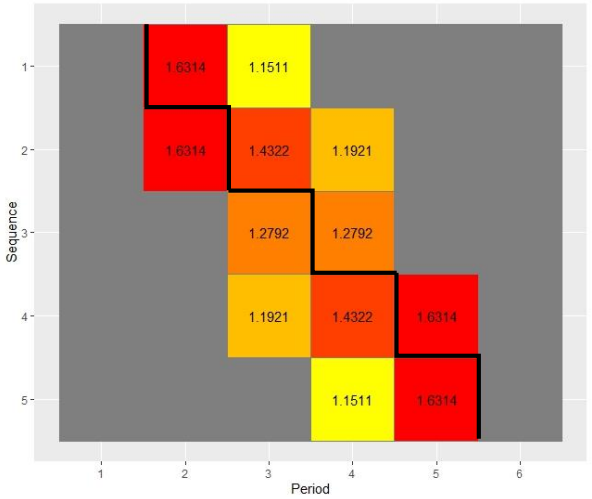
Power (to detect effect size of 0.26)

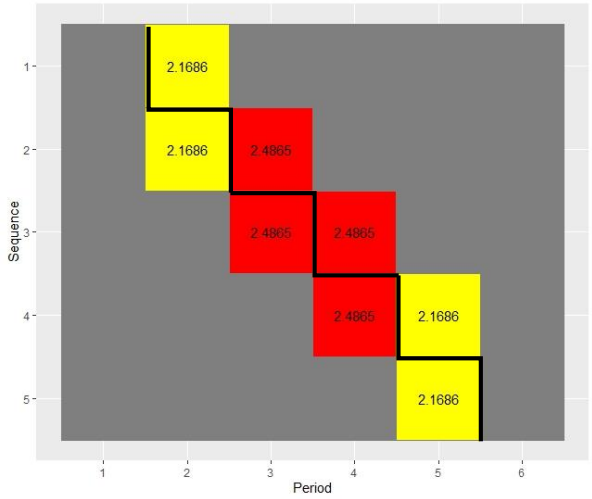
91.9% 86.2% 44.7%

Obtaining efficient incomplete designs

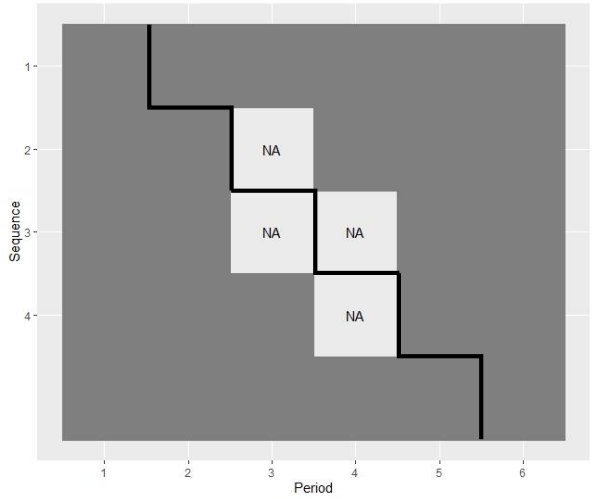
Complete







Minimally viable



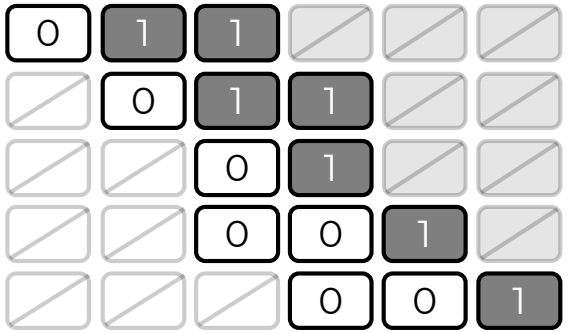
% of cells retained

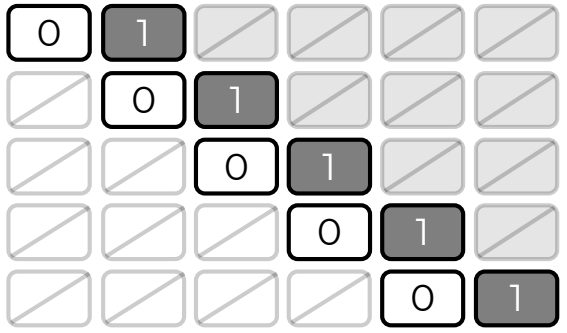
100% 40% 26.7% 6.7%

Power (to detect effect size of 0.26)

91.9% 86.2% 76.8% 44.7%

Staircase designs: Efficient incomplete designs

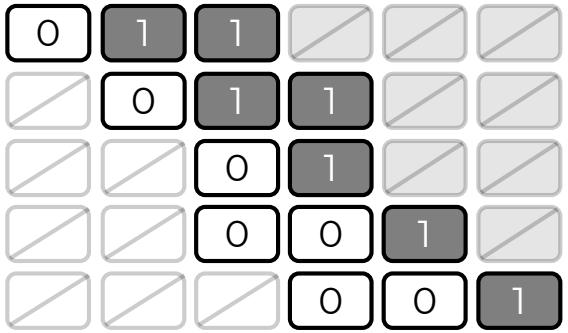


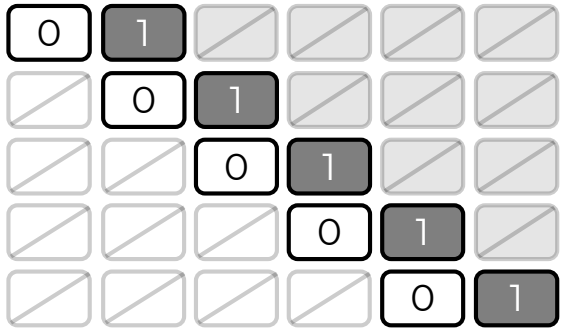


Q: How many low-information cells can be omitted while still maintaining sufficient power?

Rezaei-Darzi E, Grantham KL, Forbes AB, Kasza J. The impact of iterative removal of low-information cluster-period cells from a stepped wedge design. *BMC Medical Research Methodology* 2023; 23(1): 160.

Staircase designs: Efficient incomplete designs



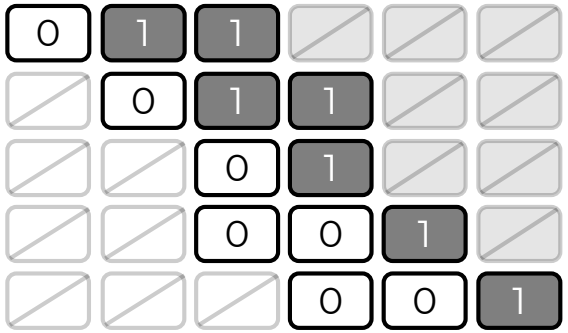


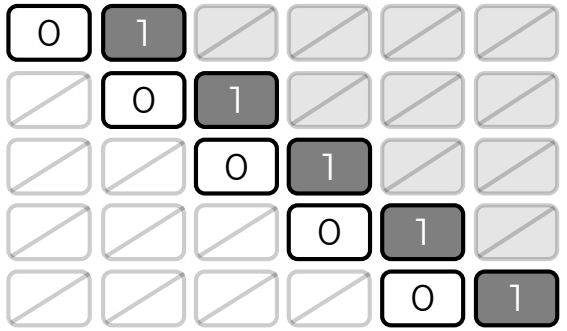
Q: How many low-information cells can be omitted while still maintaining sufficient power?

A: We can often omit over half of the cells and still get a design with adequate power.

Rezaei-Darzi E, Grantham KL, Forbes AB, Kasza J. The impact of iterative removal of low-information cluster-period cells from a stepped wedge design. *BMC Medical Research Methodology* 2023; 23(1): 160.

Staircase designs: Efficient incomplete designs





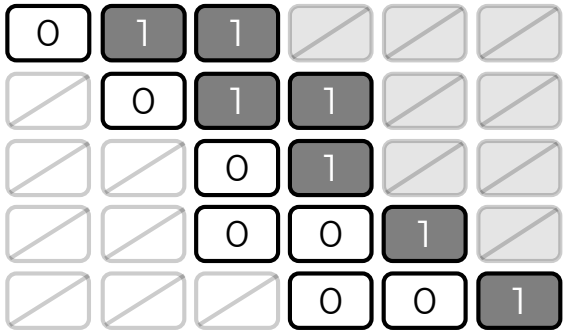
Q: How many low-information cells can be omitted while still maintaining sufficient power?

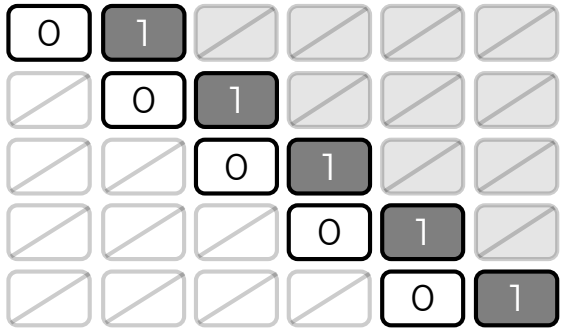
A: We can often omit over half of the cells and still get a design with adequate power.

Q: What do these designs tend to look like?

Rezaei-Darzi E, Grantham KL, Forbes AB, Kasza J. The impact of iterative removal of low-information cluster-period cells from a stepped wedge design. *BMC Medical Research Methodology* 2023; 23(1): 160.

Staircase designs: Efficient incomplete designs





Q: How many low-information cells can be omitted while still maintaining sufficient power?

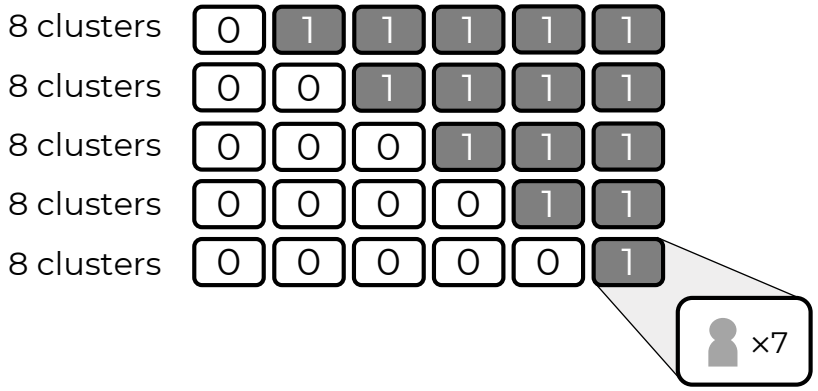
A: We can often omit over half of the cells and still get a design with adequate power.

Q: What do these designs tend to look like?

A: Cells concentrated around the treatment switch ⟶ “Staircase” designs

Rezaei-Darzi E, Grantham KL, Forbes AB, Kasza J. The impact of iterative removal of low-information cluster-period cells from a stepped wedge design. *BMC Medical Research Methodology* 2023; 23(1): 160.

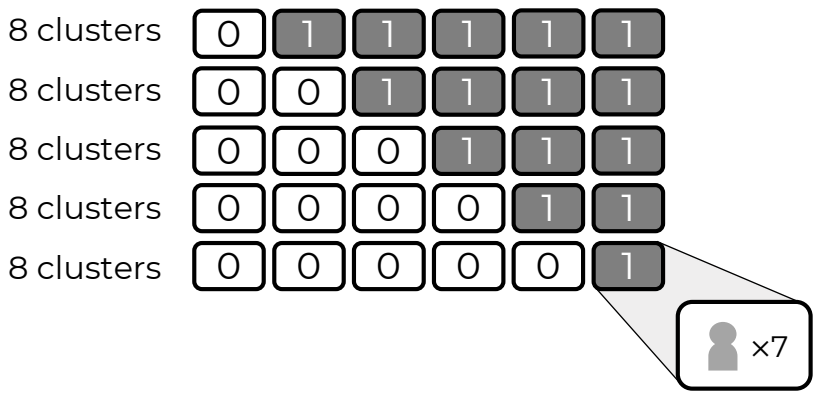
Staircase designs: Even more efficient with modifications

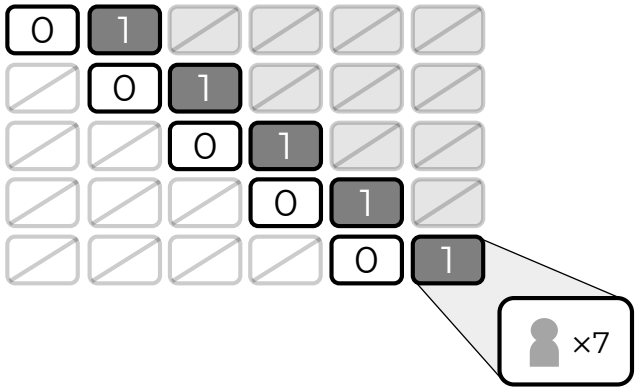


	Complete stepped wedge
Power	91.9%
Total # of participants	1680

Grantham KL, Forbes AB, Hooper R, Kasza J. The relative efficiency of staircase and stepped wedge cluster randomised trial designs. *Statistical Methods in Medical Research* 2025.

Staircase designs: Even more efficient with modifications

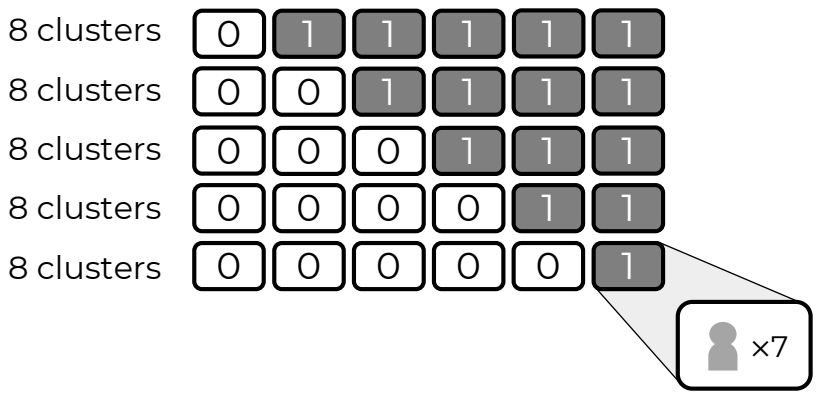


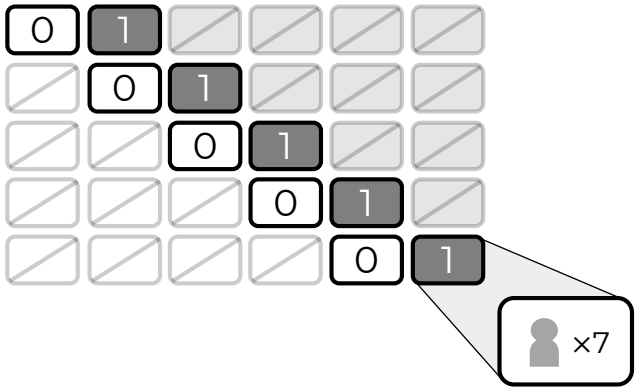


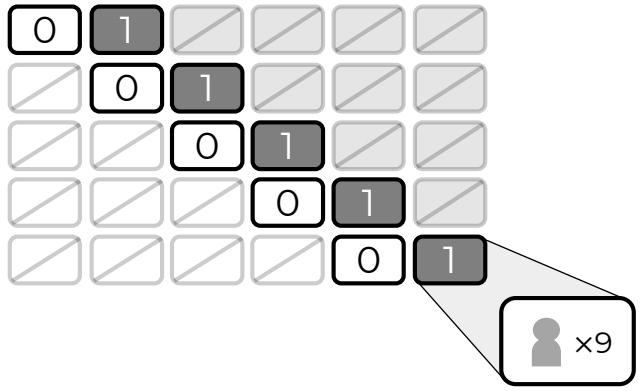
	Complete stepped wedge	Basic staircase
Power	91.9%	76.8%
Total # of participants	1680	560

Grantham KL, Forbes AB, Hooper R, Kasza J. The relative efficiency of staircase and stepped wedge cluster randomised trial designs. *Statistical Methods in Medical Research* 2025.

Staircase designs: Even more efficient with modifications







	Complete stepped wedge	Basic staircase	Basic staircase with larger cluster-period size
Power	91.9%	76.8%	85.4%
Total # of participants	1680	560	720

Grantham KL, Forbes AB, Hooper R, Kasza J. The relative efficiency of staircase and stepped wedge cluster randomised trial designs. *Statistical Methods in Medical Research* 2025.

Key results for staircase designs

- Staircase designs behave differently to stepped wedge designs, statistically speaking (Grantham et al, 2024)
  - Assumed form for time makes a difference
  - Form of treatment effect estimator: primarily uses vertical comparisons
  - Dedicated statistical methodology for sample size/power calculations is required and now available to support their use

Grantham KL, Forbes AB, Hooper R, Kasza J. The staircase cluster randomised trial design: A pragmatic alternative to the stepped wedge. *Statistical Methods in Medical Research* 2024; 33(1): 24-41.

Key results for staircase designs

- Staircase designs behave differently to stepped wedge designs, statistically speaking (Grantham et al, 2024)
  - Assumed form for time makes a difference
  - Form of treatment effect estimator: primarily uses vertical comparisons
  - Dedicated statistical methodology for sample size/power calculations is required and now available to support their use
- Basic staircase designs can be more efficient than complete stepped wedge designs for fewer total participants (Grantham et al, 2025)

Grantham KL, Forbes AB, Hooper R, Kasza J. The staircase cluster randomised trial design: A pragmatic alternative to the stepped wedge. *Statistical Methods in Medical Research* 2024; 33(1): 24-41.

Grantham KL, Forbes AB, Hooper R, Kasza J. The relative efficiency of staircase and stepped wedge cluster randomised trial designs. *Statistical Methods in Medical Research* 2025.

Key results for staircase designs

- Staircase designs behave differently to stepped wedge designs, statistically speaking (Grantham et al, 2024)
  - Assumed form for time makes a difference
  - Form of treatment effect estimator: primarily uses vertical comparisons
  - Dedicated statistical methodology for sample size/power calculations is required and now available to support their use
- Basic staircase designs can be more efficient than complete stepped wedge designs for fewer total participants (Grantham et al, 2025)
- Staircase designs are also highly cost-efficient relative to complete stepped wedge designs (Rezaei-Darzi et al, 2025)

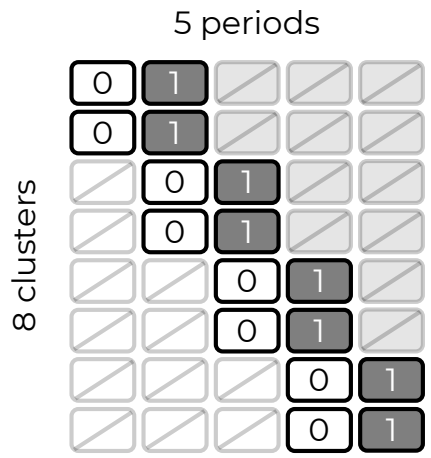
Grantham KL, Forbes AB, Hooper R, Kasza J. The staircase cluster randomised trial design: A pragmatic alternative to the stepped wedge. *Statistical Methods in Medical Research* 2024; 33(1): 24-41.

Grantham KL, Forbes AB, Hooper R, Kasza J. The relative efficiency of staircase and stepped wedge cluster randomised trial designs. *Statistical Methods in Medical Research* 2025.

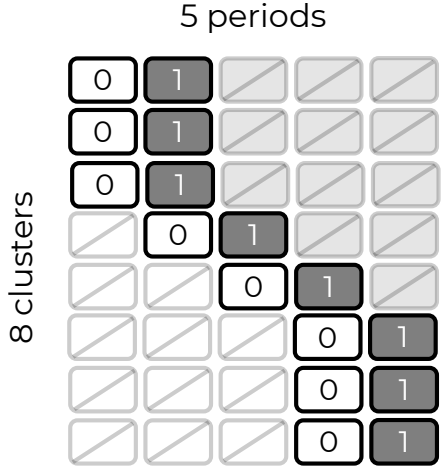
Rezaei-Darzi E, Kasza J, Assifi AR, Mazza D, Forbes AB, Grantham KL. Identifying less burdensome and more cost-efficient incomplete stepped wedge designs for continuous outcomes collected via repeated cross-sections. *Statistics in Medicine* 2025; 44(8-9): e70067.

More work to be done

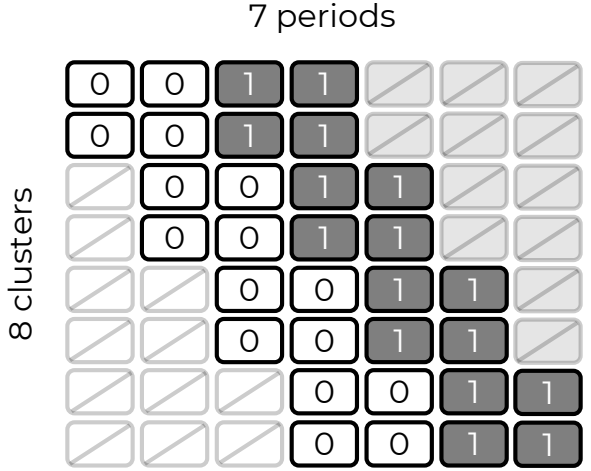
- Many staircase design variants are possible: which are best?
  - More guidance is needed to inform the choice of variant
- Forthcoming work:
  - Optimal allocation of clusters to sequences
  - Balanced versus imbalanced designs
  - How time-varying forms of the treatment effect impact design choice



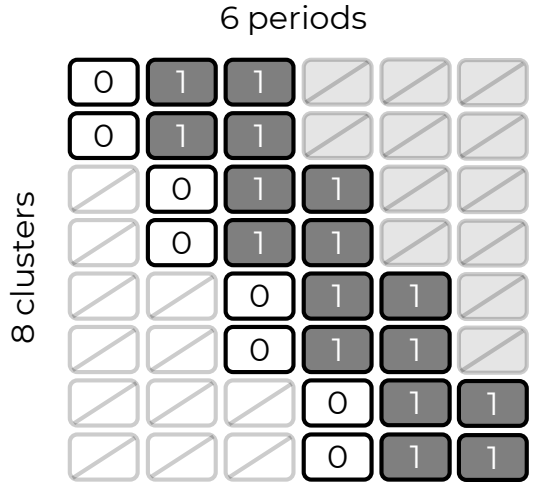
Basic staircase



Unequal allocation of clusters to sequences



Balanced staircase



Imbalanced staircase

Acknowledgements

Jessica Kasza, Monash University, Australia

Ehsan Rezaei-Darzi, Monash University, Australia

Andrew Forbes, Monash University, Australia

Richard Hooper, Queen Mary University of London, UK

Questions?

kelsey.grantham@monash.edu

Marginal Models, Binary Outcomes, and Incomplete designs

John Preisser

Department of Biostatistics
University of North Carolina
at Chapel Hill

May 20, 2025
Vancouver, B.C.

Disclosures

Grant Support: Research in this presentation was funded through a Patient-Centered Outcomes Research Institute (PCORI Award ME-2019C1-16196)

The statements presented in these slides are solely the responsibility of the presenter and do not necessarily represent the views of PCORI®, its Board of Governors or Methodology Committee

J.S.P. has received a stipend for service as a merit reviewer from PCORI®.
J.S.P. did not serve on the PCORI® Merit Review panel that reviewed this project.

Introduction – marginal model analysis with GEE

- The interpretation of the intervention effect in a marginal – or population-averaged (PA) - model is at the population level, in contrast to the cluster-specific level of a generalized linear mixed model.

  Pros: PA models have the same interpretation of the intervention effect under different within-cluster correlation structures.

Introduction – marginal model analysis with GEE

- The interpretation of the intervention effect in a marginal – or population-averaged (PA) - model is at the population level, in contrast to the cluster-specific level of a generalized linear mixed model.

  Pros: PA models have the same interpretation of the intervention effect under different within-cluster correlation structures.

- The usual Generalized Estimating Equations (GEE) procedure requires a large number of clusters (50+) but SW-CRTs often have a small number such that the usual sandwich variance estimator underestimates the true variance.

  Pros: Finite-sample bias-corrected sandwich variance estimators extend GEE to cluster randomized trials (CRTs) with as few as 10 clusters.

Mancl LA, DeRouen TA. A covariance estimator for GEE with improved small-sample properties. *Biometrics*. 2001; 57:126–134.

Statistical Analysis of multi-period CRTs with GEE

- GEE estimates the intra-cluster correlation (ICC), but simple exchangeable correlation does not account for correlation decay in multi-period CRTs.
- Multiple-parameter ICC structures are available for marginal models.
- Prentice's (1988) paired estimating equations approach for correlated binary data flexibly models ICCs and estimates their standard errors.

Prentice RL. Correlated binary regression with covariates specific to each binary observation. *Biometrics*. 1988; 44:1033-1048.

Statistical Analysis of multi-period CRTs with GEE

- GEE estimates the intra-cluster correlation (ICC), but simple exchangeable correlation does not account for correlation decay in multi-period CRTs.
- Multiple-parameter ICC structures are available for marginal models.
- Prentice's (1988) paired estimating equations approach for correlated binary data flexibly models ICCs and estimates their standard errors.
- Matrix-adjusted estimating equations (MAEE) reduces bias of ICC estimates.
- MAEE + bias-corrected variance estimates gives confidence intervals for ICCs with improved coverage (Preisser et al. 2008) to improve reporting of ICCs per CONSORT statements for CRTs and stepped wedge (SW) designs.

Prentice RL. Correlated binary regression with covariates specific to each binary observation. *Biometrics*. 1988; 44:1033-1048.

Preisser JS, Lu B, Qaqish BF. Finite sample adjustments in estimating equations and covariance estimators for intracluster correlations. *Stat Med*. 2008; 27:5764-5785.

Software for GEE+MAEE analysis

- A SAS/IML macro GEEMAEE (Zhang et al. 2023) includes
  - Options for common correlation structures for multi-period ICCs
  - Otherwise, more advanced features are available for flexibly-specified within-cluster correlation models.
- Functions in the geeCRT package in R software (Li et al. 2022)
  - Option for cluster-period mean models for cross-sectional designs
- Both softwares model individual-level data for cross-sectional and cohort multi-period CRTs
- Reliable ICC estimates are needed for accurate power calculation of CRTs!

Li F, Yu H, Rathouz P, Turner E, Preisser J. Marginal modeling of cluster-period means and intra-class correlations in stepped wedge designs with binary outcomes. Biostatistics 2022;23:772-788.

Zhang Y, Preisser JS, Li F, Turner EL, Toles M, Rathouz PJ. GEEMAEE: A SAS macro for the analysis of correlated outcomes based on GEE and finite-sample adjustments with application to cluster randomized trials. Computer Methods and Programs in Biomedicine. 2023 Mar; 230:107362.

Marginal mean models

Under a PA modeling framework, let $\mu_{ijk} = E(Y_{ijk})$ be the marginal mean response for a continuous or categorical outcome Y_{ijk} for the k-th subject at the j-th period in the i-th cluster with categorical periods effects. For a binary outcome and logit link,

$$\text{logit}(\mu_{ijk}) = \beta_j + X_{ij}\delta \qquad (1)$$

where X_{ij}=1 if intervention and X_{ij}=0 if control; δ is the time-adjusted **constant (**or "average") intervention effect, in this case, a log odds ratio. We can also model **incremental** (or fractional) intervention effects, e.g., X_{ij} = {0, 1/4, 2/4, 3/4, 1}..

Instead of categorical period effects, the model with linear period effects:is

$$\text{logit}(\mu_{ijk}) = \beta_0 + \beta_1(j - 1) + X_{ij}\delta \qquad (2)$$

which may be a reasonable choice for CRTs with more periods than clusters.

SOCIETY FOR CLINICAL TRIALS

Common correlation structures for multi-period CRTs

Design	Correlation Structure	$j = j'$	$j \neq j', k \neq k'$	$j \neq j', k = k'$
Cross-Sectional	Simple Exchangeable	α_1	α_1	–
	Nested Exchangeable	α_1	α_2	–
	Exponential Decay	α_0	$\alpha_0 r_0^{\lvert j-j' \rvert}$	–
Cohort	Block Exchangeable	α_1	α_2	α_3
	Proportional Decay	α_0	$\alpha_0 r_0^{\lvert j-j' \rvert}$	$r_0^{\lvert j-j' \rvert}$

- α_0 and α_1 are within period ICC
- α_2 is the between period ICC
- α_3 is the between period ICC for the same participant
- r_0 is the auto-regressive decay rate

General power method for GEE analysis

- GEE power is based on the model-based variance matrix (Rochon, 1998).

$$Cov_{MB}(\hat{\theta}) = \left[\sum_{i=1}^{I} D_i' V_i^{-1} D_i\right]^{-1} \quad (3)$$

where $\theta = (\beta', \delta)$, $D_i = \frac{\partial \mu_i}{\partial \theta'}$ and $V_i = V_i(\alpha)$ is the working covariance matrix.

- The Wald test statistic for the intervention: $\hat{\delta} / \sqrt{Var(\hat{\delta})}$
- Generally, no closed sample size formula for binary outcomes in SW-CRTs.
- A non-simulation, computationally fast power calculation method for SW-CRTs agrees well with simulated power based on GEE with bias-corrected sandwich variance estimators for as few as eight clusters (Li et al. 2018).

Li F, Turner EL, Preisser JS. Sample size determination for GEE Analysis of Stepped Wedge Cluster Randomized Trials. *Biometrics*. 2018; 74(4):1450-1458.

Rochon J. Application of GEE procedures for sample size calculations in repeated measures experiments. *Stat Med*. 1998; 17:1643–1658.

Power software for SW-CRTs designs

- Ouyang et al. (2022) reviewed 18 sample size calculators for SW-CRTs and offered an R Shiny app for users to select a calculator.
- The SAS macro CRTFASTGEEPWR (Zhang et al. 2024) was one of four calculators they featured. It is based on fast GEE power for multi-period CRT designs (Zhang et al. 2023) and applies to complete and incomplete, cross-sectional and cohort SW designs.

Ouyang Y, Li F, Preisser JS, Taljaard M. Sample size calculators for planning stepped-wedge cluster randomized trials: A review and comparison. *Int J Epidemiol*. 2022; 51(6):2000-2013.

Zhang Y, Preisser JS, Turner EL, Rathouz RJ, Toles M, Li F. A general method for calculating power for GEE analysis of complete and incomplete stepped wedge cluster randomized trials. Statistical Methods in Medical Research, 2023; 32:71-87.

Zhang Y, Preisser JS, Li F , Turner EL, Rathouz RJ. %CRTFASTGEEPWR: A SAS macro for power of generalized estimating equations analysis of multi-period cluster randomized trials with application to stepped wedge designs. *Journal of Statistical Software, Code Snippets*, 2024; 108(1), 1-27.

Design Patterns for incomplete multi-period CRTs

Examples of incompleteness in CRT designs:

- Staggered entry or withdrawal of clusters
- Implementation periods of the intervention

Design Patterns for incomplete multi-period CRTs

Examples of incompleteness in CRT designs:

- Staggered entry or withdrawal of clusters
- Implementation periods of the intervention

A Design Pattern (DP) matrix for incompleteness (Hemming et al. 2015).

Example. 3 sequences, 7 periods

$$DP = \begin{pmatrix} 0 & 0 & 2 & 1 & 1 & 2 & 2 \\ 2 & 0 & 0 & 2 & 1 & 1 & 2 \\ 2 & 2 & 0 & 0 & 2 & 1 & 1 \end{pmatrix}$$

0-Control, 1-Intervention, 2-cluster period without data collection

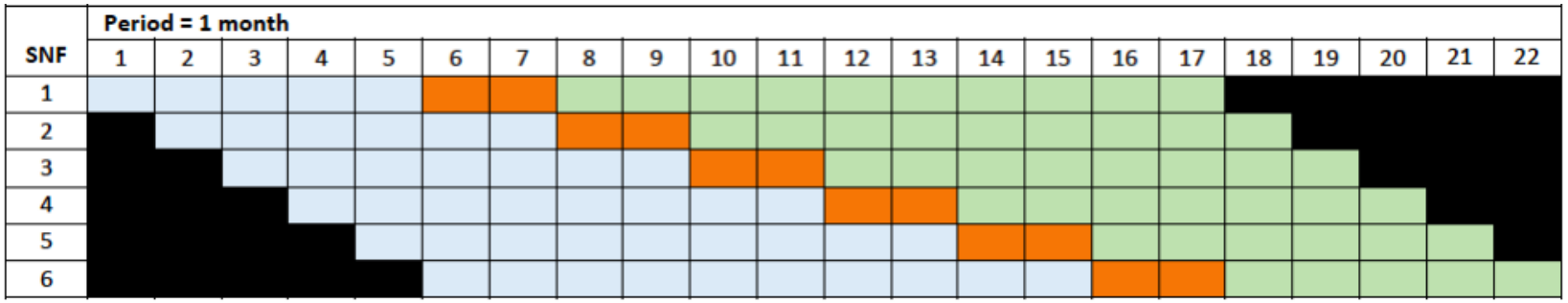
Hemming K, Lilford R, Girling AJ. Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. *Stat Med.* 2015; 34(2):181-96.

Features of the SAS macro **CRTFASTGEEPWR**

- Required inputs include:
  - Design Pattern (DP) and Cluster-period size (CP) matrices
  - Number of clusters in each sequence (M)
  - Specify distribution (binary, Poisson, normal) & link (logit, log, identity)
  - Intervention effect type (AVE, INC, INC_EX)
  - Period_effect_type (categorical vs linear), and beta_period_effects
  - Delta (intervention effect δ), log odds ratio for logistic model
  - Correlation type (NE, ED, BE, PD from Slide 7) and parameter values

- Optional inputs include:
  - Significance level of two-sided test, default is 0.05
  - df_choice (for t-test)
    - 1, df = I – p, where p is number of parameters in mean model
    - 2, df = I - 2

The incomplete stepped wedge design of the Connect-Home Trial (Toles et al 2021)



- Cross-sectional SW-CRT design in six skilled nursing facilities (SNFs) to test an intervention to improve outcomes for rehabilitation patients transitioning to home-based care.
- Intensive staff training intervention required two implementation periods.
- Limited resources led to staggered study entry and early withdrawal of SNFs.
- Planned analysis & power was based on linear period effects.

Toles M, Colón-Emeric C, Hanson LC, Naylor M, Weinberger M, Covington J, Preisser JS. Transitional care from skilled nursing facilities to home: study protocol for a stepped wedge cluster randomized trial. Trials 2021; 22:120.

Example: power calculation for binary outcome in the Connect-Home study

- A binary outcome, 12 clusters
- 2 clusters in each of 6 sequences
- 2 patients in each cluster period
- Linear periods effects, β=(0.85, -0.01)
- Constant intervention effects model with δ = log(0.3) = -1.2, so that the odds of outcome decreased by 70% under the active intervention (OR of 3.3, inverted)
- NE correlation, ICCs = (0.03,0.015)
- Cluster-period size is 2

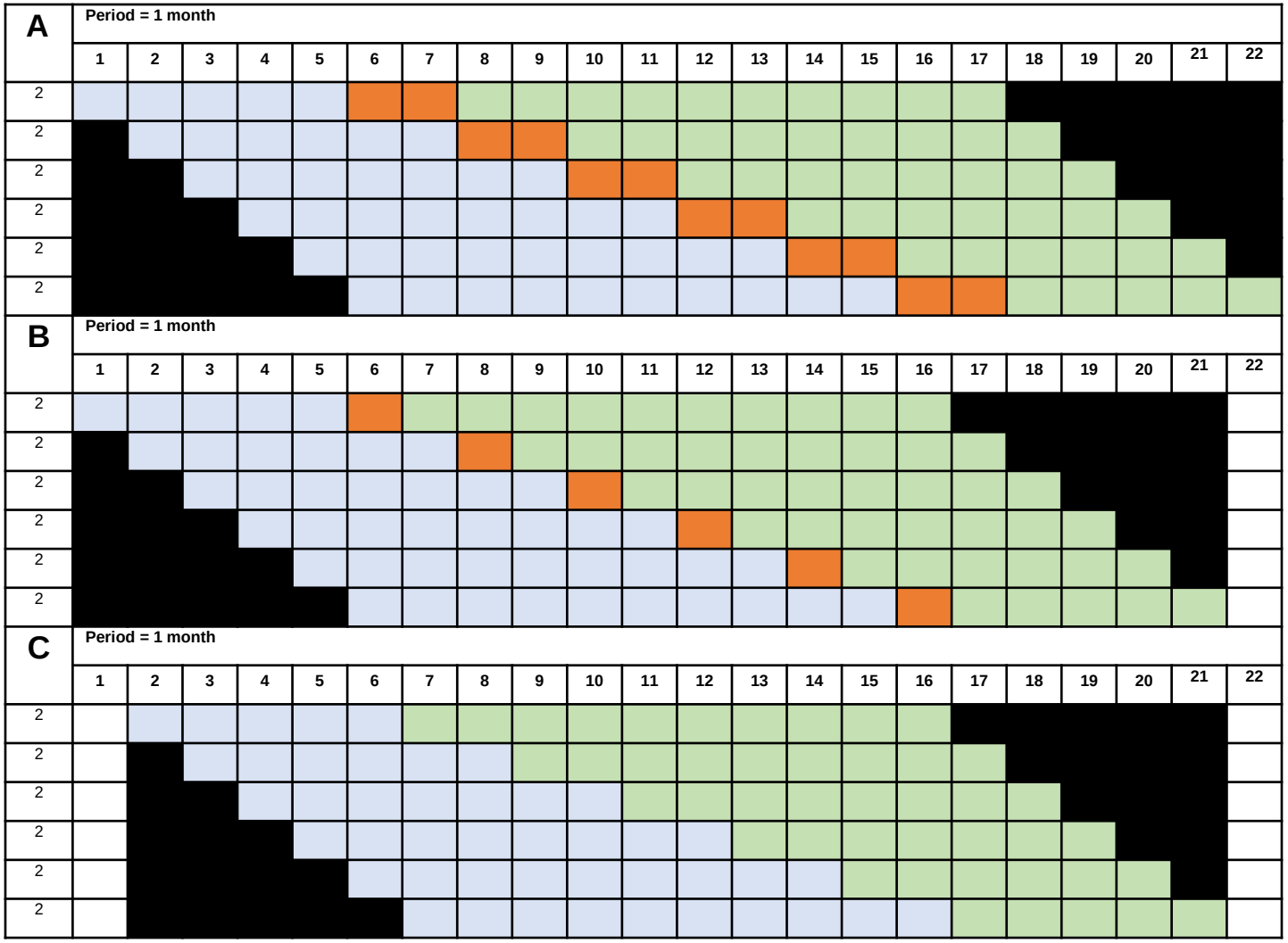
```
%CRTFASTGEEPWR(alpha=0.05, m =%str(J(6,1,2)),
dist = binary, phi=1, period_effect_type=LIN,
beta_period_effects =%str({0.85,-0.01}),
intervention_effect_type=AVE, delta = -1.2,
corr_type = NE, alpha1 = 0.03, alpha2 = 0.015,
CP_size_matrix = %str(
{2 2 2 2 2 0 0 2 2 2 2 2 2 2 2 2 0 0 0 0 0,
 0 2 2 2 2 2 2 0 0 2 2 2 2 2 2 2 2 0 0 0 0,
 0 0 2 2 2 2 2 2 2 0 0 2 2 2 2 2 2 2 0 0 0,
 0 0 0 2 2 2 2 2 2 2 2 0 0 2 2 2 2 2 2 0 0,
 0 0 0 0 2 2 2 2 2 2 2 2 2 0 0 2 2 2 2 2 0,
 0 0 0 0 0 2 2 2 2 2 2 2 2 2 2 0 0 2 2 2 2 2}),
DesignPattern = %str(
{0 0 0 0 0 2 2 1 1 1 1 1 1 1 1 1 2 2 2 2 2,
 2 0 0 0 0 0 0 2 2 1 1 1 1 1 1 1 1 2 2 2 2,
 2 2 0 0 0 0 0 0 0 2 2 1 1 1 1 1 1 1 2 2 2,
 2 2 2 0 0 0 0 0 0 0 0 2 2 1 1 1 1 1 1 2 2,
 2 2 2 2 0 0 0 0 0 0 0 0 0 2 2 1 1 1 1 1 2,
 2 2 2 2 2 0 0 0 0 0 0 0 0 0 2 2 1 1 1 1 1}));
```

T	S	clusters	df	theta	totaln	Dist	Link	stddel	zpower	tpower
22	6	12	9	0.85	360	BINARY	LOGIT	3.1568	0.8843	0.8029
				-0.01						
				-1.2						

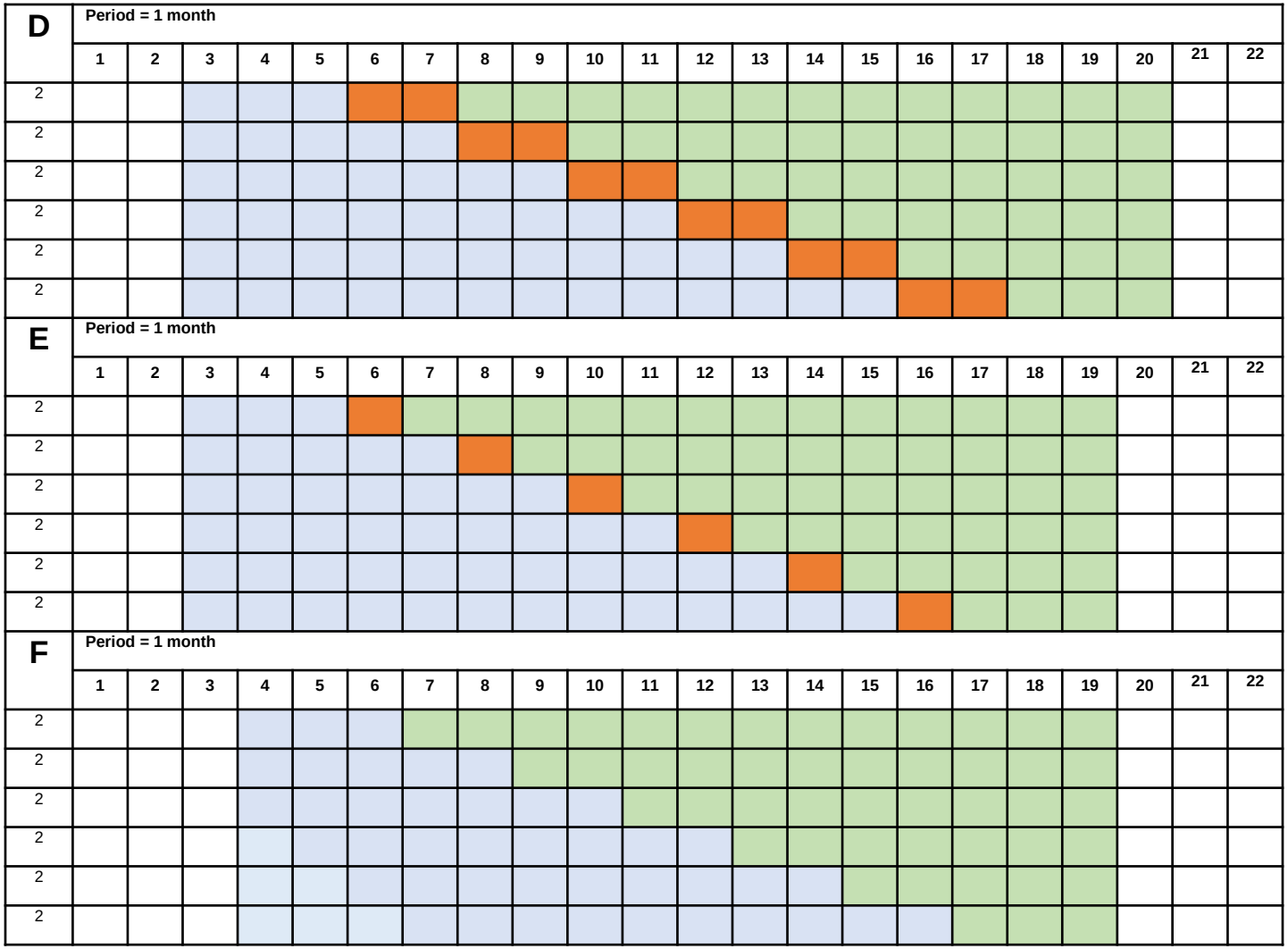
What is the impact of implementation periods and staggered entry and withdrawal on statistical power?

- How does the power of the Connect-Home study ("Design A") compare to other designs with comparable features?
  - 0, 1 or 2 implementation periods
  - with or without staggered entry and start.
- All 6 designs have a total sample size of 360 patients.
- Separately evaluate Constant and Incremental Intervention effects models.

SW Designs with staggered start & end



SW Designs without staggered start & end



Power of incomplete SW designs: constant intervention effect

- A binary outcome
- 12 clusters , 6 sequences,
- Cluster-period size of 2
- Total of 360 patients
- Linear periods effects
- Constant intervention effects model

  $\delta = \log(0.3) = -1.2$
- NE correlation ICCs = (0.03,0.015)

Introduction of implementation periods substantially reduces power, while use of staggering has little effect.

	Staggered Start & End			
Number of Implementation Periods		Yes		No
2	A	80	D	80
1	B	86	E	86
0	C	91	F	91

Power of incomplete SW designs: incremental intervention effect

- A binary outcome
- 12 clusters , 6 sequences
- Total of 360 patients
- Linear periods effects
- Incremental intervention effects model
- $X_{ij} = \{0, 1/10, \ldots, 9/10, 1\}$

  $\delta = \log(0.09) = -2.4$
- NE correlation ICCs = (0.03,0.015)

Introduction of implementation periods and staggering reduces power to similar degrees.

	Staggered Start & End			
Number of Implementation Periods		Yes		No
2	A	89	D	94
1	B	91	E	95
0	C	93	F	96

Discussion

- Power results for a binary outcome suggest that incomplete designs suffer power loss relative to complete designs, and in ways that may not be generalizable.

- GEE with finite-sample bias corrections can be used to reliably fit marginal models to stepped wedge and other CRTs.

- GEE software for power and analysis for designing and analyzing multi-period CRTs applies to parallel, cluster crossover, and stepped wedge designs.

Thank YOU.

Link for SAS macro codes:

http://www.bios.unc.edu/~preisser/personal/crtfastgeepwr/

I would like to acknowledge Ying Zhang and Di Hu for contributions to development of the macro CRTFASTGEEPWR

Disclosures

Grant Support: Research in this presentation was funded through a Patient-Centered Outcomes Research Institute (PCORI Award ME-2019C1-16196)

The statements presented in these slides are solely the responsibility of the presenter and do not necessarily represent the views of PCORI®, its Board of Governors or Methodology Committee

J.S.P. has received a stipend for service as a merit reviewer from PCORI®.

J.S.P. did not serve on the PCORI® Merit Review panel that reviewed this project.

Demystifying incomplete stepped wedge designs under the working independence assumption

Fan Li

Department of Biostatistics
Yale School of Public Health

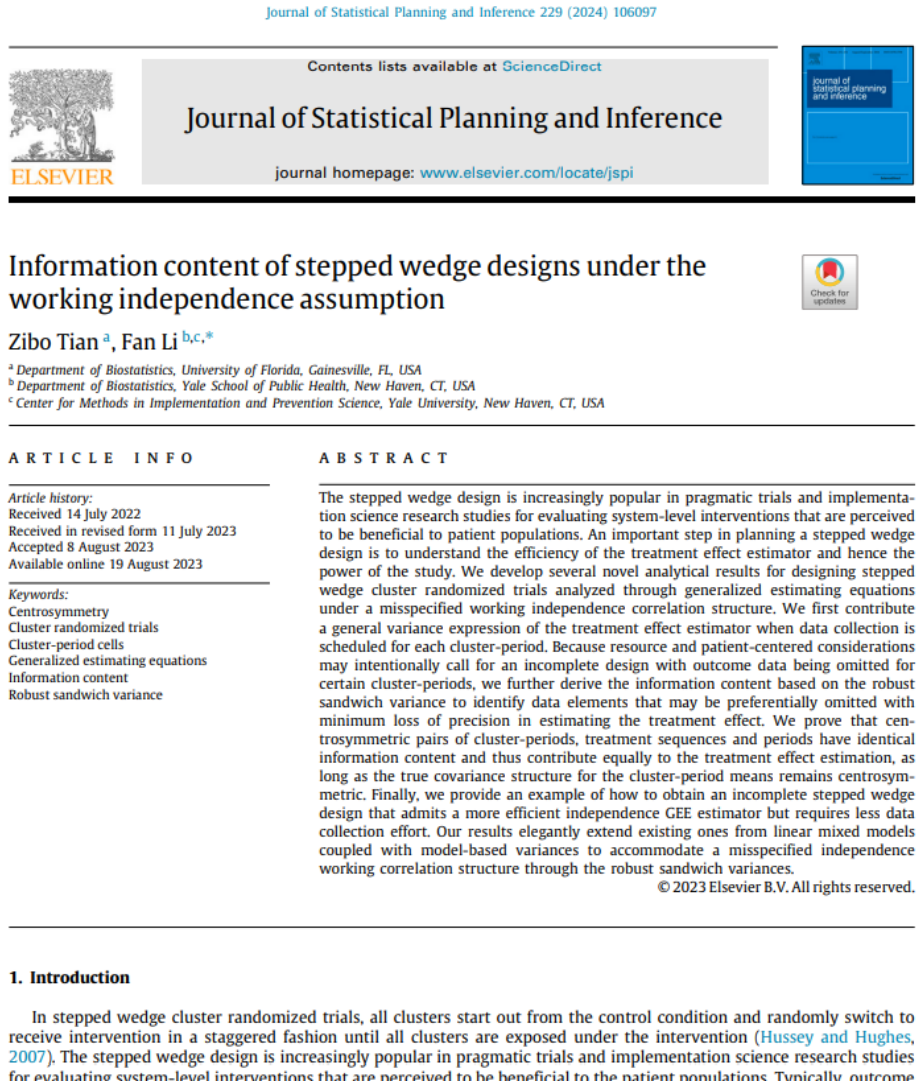
https://lifan90.com/

The Society for Clinical Trials (SCT) 46th Annual Meeting
May 20, 2025

Acknowledgement

- Research in this article was funded through a Patient-Centered Outcomes Research Institute® (PCORI® Award ME-2019C1-16196)

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



Journal of Statistical Planning and Inference 229 (2024) 106097

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Information content of stepped wedge designs under the working independence assumption

Zibo Tian [a], Fan Li [b,c,*]

[a] Department of Biostatistics, University of Florida, Gainesville, FL, USA
[b] Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA
[c] Center for Methods in Implementation and Prevention Science, Yale University, New Haven, CT, USA

ARTICLE INFO

Article history:
Received 14 July 2022
Received in revised form 11 July 2023
Accepted 8 August 2023
Available online 19 August 2023

Keywords:
Centrosymmetry
Cluster randomized trials
Cluster-period cells
Generalized estimating equations
Information content
Robust sandwich variance

ABSTRACT

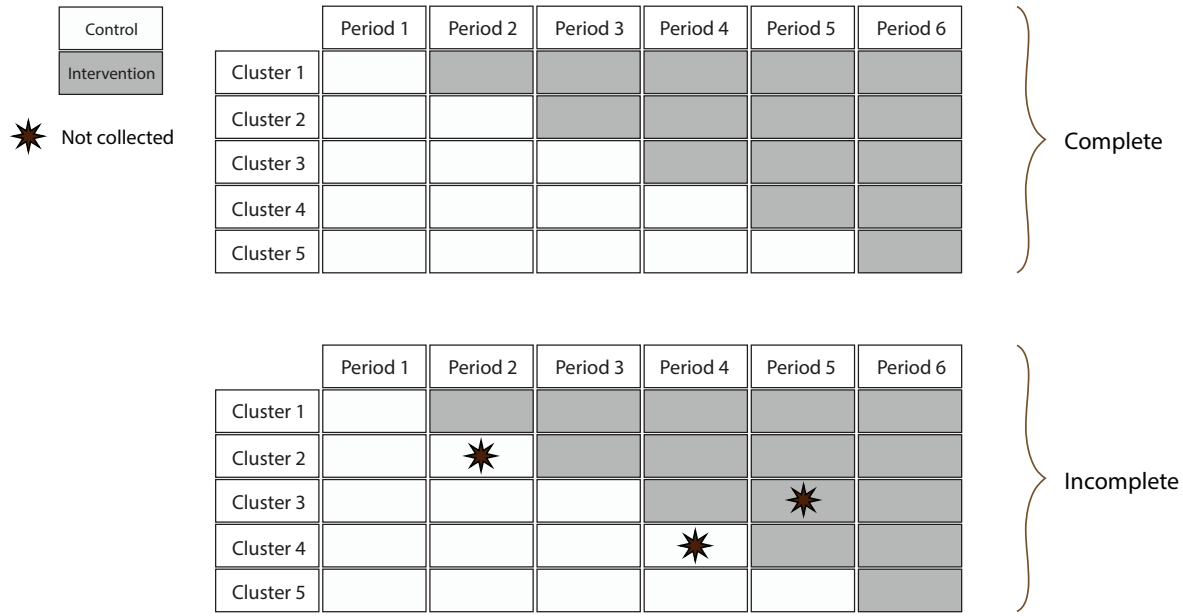
The stepped wedge design is increasingly popular in pragmatic trials and implementation science research studies for evaluating system-level interventions that are perceived to be beneficial to patient populations. An important step in planning a stepped wedge design is to understand the efficiency of the treatment effect estimator and hence the power of the study. We develop several novel analytical results for designing stepped wedge cluster randomized trials analyzed through generalized estimating equations under a misspecified working independence correlation structure. We first contribute a general variance expression of the treatment effect estimator when data collection is scheduled for each cluster-period. Because resource and patient-centered considerations may intentionally call for an incomplete design with outcome data being omitted for certain cluster-periods, we further derive the information content based on the robust sandwich variance to identify data elements that may be preferentially omitted with minimum loss of precision in estimating the treatment effect. We prove that centrosymmetric pairs of cluster-periods, treatment sequences and periods have identical information content and thus contribute equally to the treatment effect estimation, as long as the true covariance structure for the cluster-period means remains centrosymmetric. Finally, we provide an example of how to obtain an incomplete stepped wedge design that admits a more efficient independence GEE estimator but requires less data collection effort. Our results elegantly extend existing ones from linear mixed models coupled with model-based variances to accommodate a misspecified independence working correlation structure through the robust sandwich variances.

1. Introduction

In stepped wedge cluster randomized trials, all clusters start out from the control condition and randomly switch to receive intervention in a staggered fashion until all clusters are exposed under the intervention (Hussey and Hughes, 2007). The stepped wedge design is increasingly popular in pragmatic trials and implementation science research studies for evaluating system-level interventions that are perceived to be beneficial to the patient populations. Typically, outcome

Incomplete stepped wedge designs

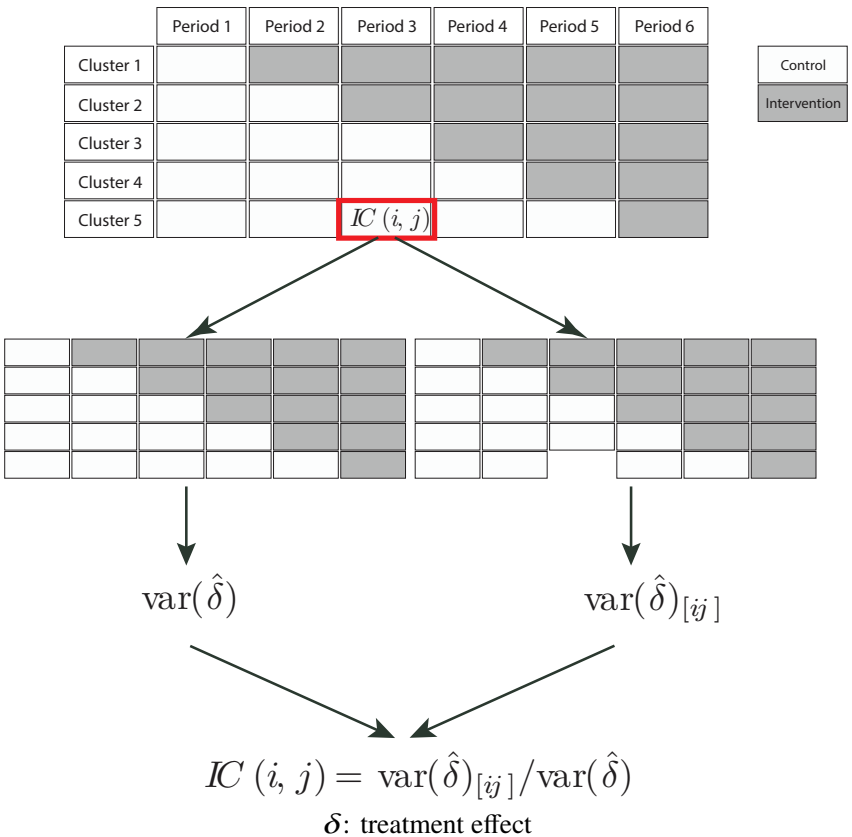
- **Complete design**: outcome data collection scheduled for all cluster-periods
- **Incomplete design**: outcome data collection omitted for certain cluster-periods due to logistical, and resource considerations



- What is an optimal incomplete design?
  - helpful to identify cells, sequences, and periods that contribute the most or least amount of information to the estimation of treatment effect

Information content

Information content (IC) of a cell: ratio of the variance of the treatment effect estimator when cell omitted, to that under the complete design.[1]



Under linear mixed models

1. $IC \geq 1$: omission does not reduce the variance of treatment effect estimator (decrease study power).
2. Centrosymmetric covariance structures correspond to centrosymmetric patterns in IC of cluster-period cells, treatment sequences, and periods.
3. Omitting cluster-period cells closest to the treatment switching results in the greatest reduction of precision for estimating the treatment effect parameter.
4. Outer sequences and inner periods often carry the highest IC.

[1] Jessica Kasza and Andrew B Forbes. "Information content of cluster–period cells in stepped wedge trials". In: *Biometrics* 75.1 (2019), pp. 144–152.

Objective

- In general, IC can depend on
  - the true correlation structure of the data generating process
  - the type of outcome (continuous, binary or count)[2]
  - the choice of analytic model, especially when assumed correlation structure ≠ truth (but this is much less known)
- **Objective**: Demystify IC when the analysis proceeds with marginal models (GEE) under the working independence assumption
- Why working independence?
  - computationally convenient, often the practically feasible approach for complex analyses (e.g., survival, quantile)
  - more robust to model misspecification (when targeting counterfactual estimands)
  - general theoretical interest

[2] Fan Li et al. "Generalizing the information content for stepped wedge designs: A marginal modeling approach". In: *Scandinavian Journal of Statistics* 50.3 (2023), pp. 1048–1067.

Recap of marginal mean model

Let Y_{ijl} denote the continuous outcome for individual $l = 1, \ldots, N$ in cluster $i = 1, \ldots, I$ and period $j = 1, \ldots, J$. The linear marginal mean model:

$$\mu_{ij} = \beta_j + X_{ij}\delta,$$

- β_j: secular trend in period j
- $X_{ij} \in \{0, 1\}$: treatment indicator
- δ: constant time-adjusted treatment effect

True intracluster correlation matrix follows a symmetric block structure:

$$\text{corr}(\boldsymbol{Y}_i) = (\boldsymbol{B}_1 - \boldsymbol{B}_2) \otimes \boldsymbol{I}_N + \boldsymbol{B}_2 \otimes \boldsymbol{J}_N$$

where $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ encode within- and between-subject correlation. Four typical examples: (1) nested exchangeable, (2) exponential decay, (3) block exchangeable, (4) double decay.

A closer look at independence GEE

The independence GEE leads to a vertical estimator with overlap propensity score weights

$$\hat{\delta} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\left\{(1-p_j)X_{ij} - p_j(1-X_{ij})\right\}\overline{Y}_{ij}}{\sum_{i'=1}^{I}\sum_{j'=1}^{J} p_{j'}(1-p_{j'})} = \sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}\overline{Y}_{ij}, \qquad (1)$$

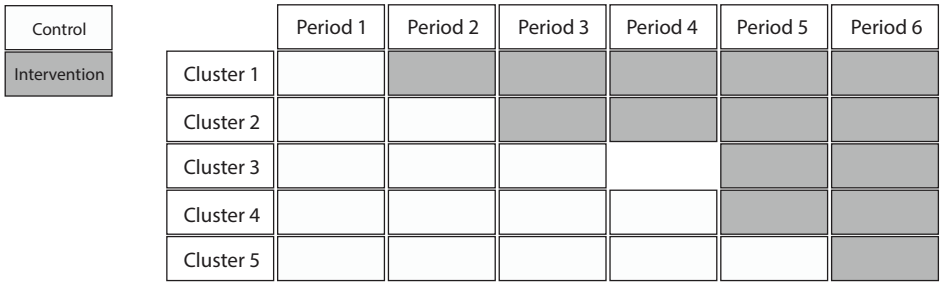
where $p_j = \frac{1}{I}\sum_{i=1}^{I} X_{ij}$ is the proportion of treated clusters in period j.

Owing to the simple structure in (1)

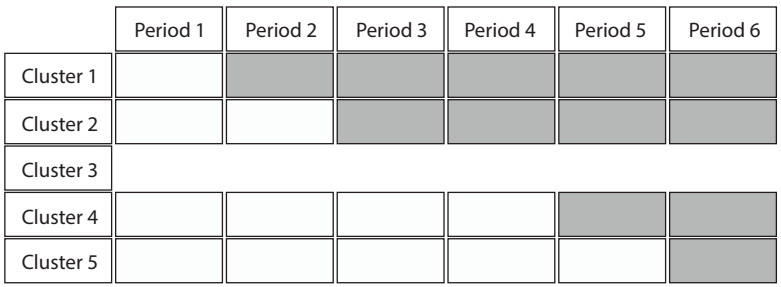
- IC for cells in first and last periods are automatically zero
- IC for each cells driven by cluster-period means and the weights
- $\text{var}(\widehat{\delta})$ in closed form for all symmetric block structure (including decay) to facilitate sample size estimation

Scope of analytical investigation

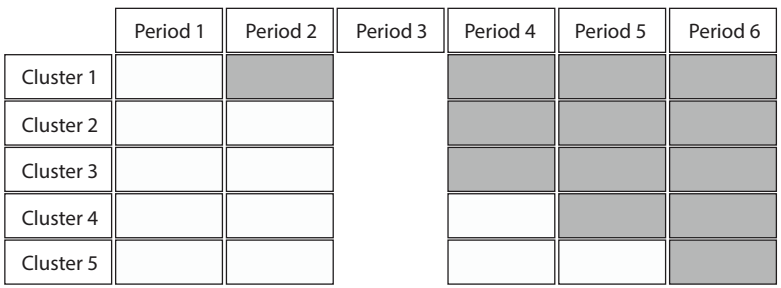
We provide scalar expressions for 3 types of omissions[3]



$$\rightarrow \quad IC\,(i,\,j) = \mathrm{var}(\widehat{\delta})_{[ij\,]}/\mathrm{var}(\widehat{\delta})$$



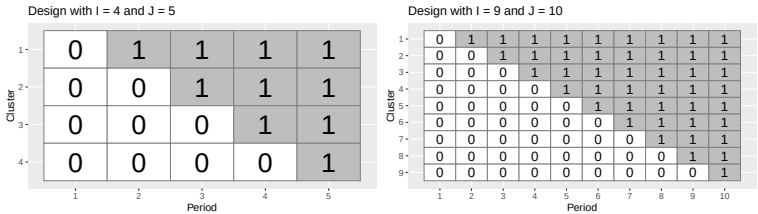
$$\rightarrow \quad IC\,(i,\,\bullet) = \mathrm{var}(\widehat{\delta})_{[i\bullet]}/\mathrm{var}(\widehat{\delta})$$



$$\rightarrow \quad IC\,(\bullet,j) = \mathrm{var}(\widehat{\delta})_{[\bullet j]}/\mathrm{var}(\widehat{\delta})$$

[3] Zibo Tian and Fan Li. "Information content of stepped wedge designs under the working independence assumption". In: *Journal of statistical planning and inference* 229 (2024), p. 106097.

Scope of numerical evidence

Design configuration	**Correlation Structures and Parameters**
Design diagram: 	**(i) Simple Exchangeable (Cross-sectional)** $r_{jj} = \alpha_0 = 0.05$ $r_{jj'} = r^*_{jj'} = \alpha_1 = 0.05$ **(ii) Exponential Decay (Cross-sectional)** $r_{jj} = \alpha_0 = 0.05$ $r_{jj'} = r^*_{jj'} = \alpha_0 \cdot \rho^{\lvert j-j' \rvert}$ $\rho = 0.95$
Cluster-period size set $N = 100$. **Design settings:** Both *cross-sectional* and *closed-cohort* trials included.	**(iii) Block Exchangeable (Closed-cohort)** $r_{jj} = \alpha_0 = 0.05$ $r_{jj'} = \alpha_1 = 0.05$ $r^*_{jj'} = \alpha_2 = 0.7$ **(iv) Double Decay (Closed-cohort)** $r_{jj} = \alpha_0 = 0.05$ $r_{jj'} = \alpha_0 \cdot \rho_0^{\lvert j-j' \rvert}, \rho_0 = 0.95$ $r^*_{jj'} = \rho_1^{\lvert j-j' \rvert}, \rho_1 = 0.7$

Overall pattern

Theorem 1 (Rotational symmetric property)

When all clusters have a common true covariance structure $\boldsymbol{W} = cov(\overline{\boldsymbol{Y}}_i)$ *that is centrosymmetric, then for a standard stepped wedge design,*

$$IC(i,j) = IC(I+1-i, J+1-j)$$
$$IC(i,\bullet) = IC(I+1-i,\bullet),$$
$$IC(\bullet,j) = IC(\bullet, J+1-j).$$

That is, centrosymmetric (invariant to row + column reversion) pairs of cluster-period cells, symmetric pairs of treatment sequences and periods contribute the same amount of information in estimating δ *when using an independence GEE estimator.*

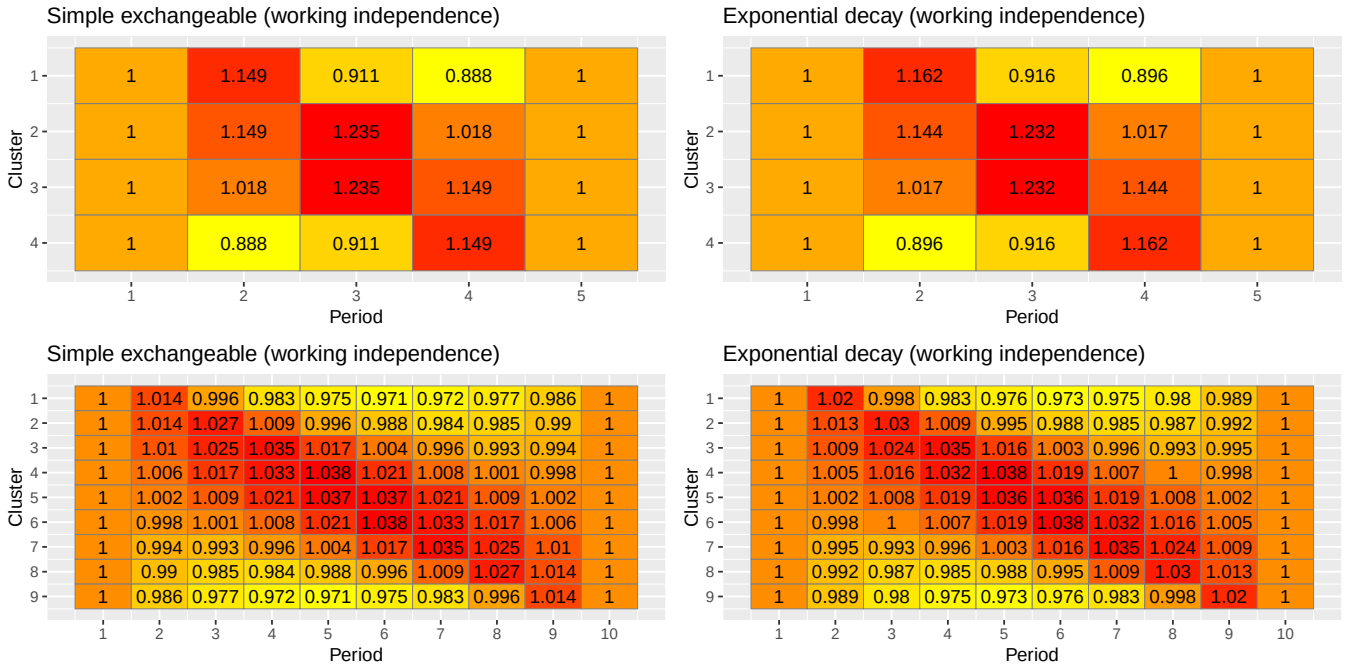
- omitting data from cluster-period (i, j) leads to the same variance inflation as omitting its centrosymmetric counterpart $(I+1-i, J+1-j)$; similarly for sequences/periods
- applicable to: (1) nested exchangeable, (2) exponential decay, (3) block exchangeable, (4) double decay.
- **same** as Kasza and Forbes (2018)

Expressions of $IC(i, j)$, $IC(i, \bullet)$, $IC(\bullet, j)$

- We derived an analytical expression for $IC(i, j)$, and confirmed the following:
  - $IC(i, 1) = IC(i, J) = 1$: first and last periods contribute no information under standard stepped wedge designs (**expected** as a result of vertical comparison)
  - counter-intuitively, $IC(i, j) < 1$ is possible — omitting a cell can even reduce the variance for estimating δ!
  - a main difference from Kasza and Forbes (2018), and arises because (1) independence GEE is not a fully-efficient estimator, and (2) the pattern is based on the robust sandwich variance rather than the model-based variance
- We also derived analytical expressions for $IC(i, \bullet)$ and $IC(\bullet, j)$
  - variance of $\widehat{\delta}$ computed using the Moore–Penrose generalized inverse, as β_j becomes unidentifiable when omitting period j
  - IC generally exceeds 1, except for $IC(\bullet, 1) = IC(\bullet, J) = 1$ (**expected**)

Numerical demonstration for $IC(i, j)$

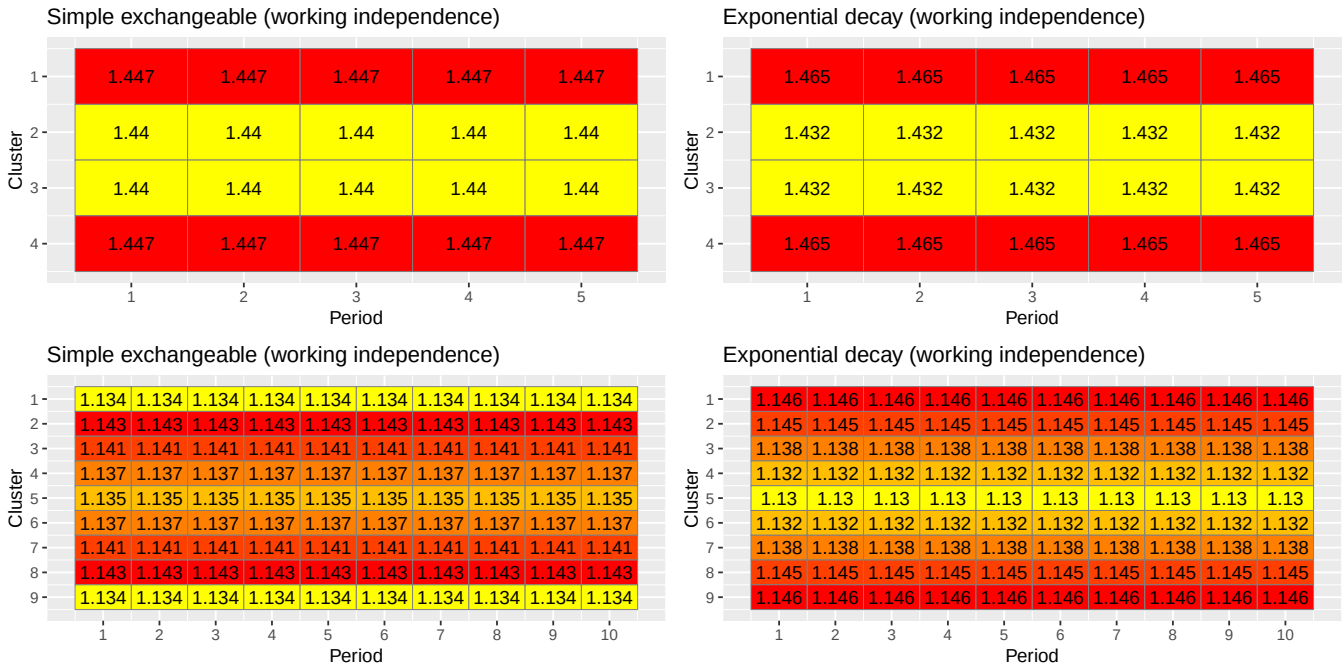
Cross-sectional design



- Centrosymmetric cell pairs have equal $IC(i, j)$
- Cells near treatment switching always contribute the highest amount of information (somewhat different from Kasza and Forbes (2018))
- Cells farthest away from treatment switching has $IC(i, j) < 1$

Numerical demonstration for $IC(i, \bullet)$

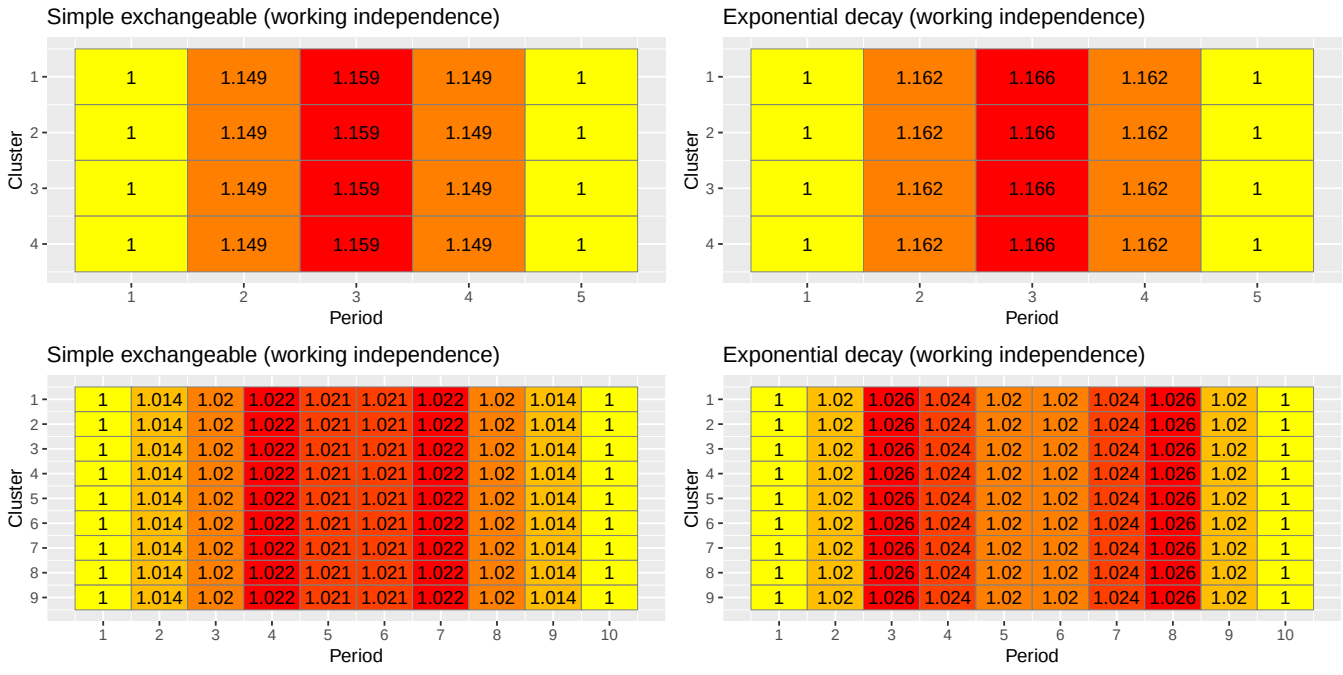
Cross-sectional design



- centrosymmertric
- 4-sequence, 5-period design: the **first/last sequences** contain the most information
- 9-sequence, 10-period design: with **simple exchangeable**, the **second and second-to-last sequences** are most informative; under **exponential decay**, outer sequences are most informative

Numerical demonstration for $IC(\bullet, j)$

Cross-sectional design



- centrosymmetric
- 4-sequence, 5-period design: **middle period** contains the most information
- 9-sequence, 10-period design: **middle periods** are the most informative
- highly similar patterns under closed-cohort designs

Efficient incomplete designs under working independence?

- Consider a stepwise elimination algorithm to systematically remove cells with $IC(i, j) < 1$
- Core algorithm same as Rezaei-Darzi et al[4] but for a different purpose

(i) **Initialization:** Let $\mathcal{A}_0 = \emptyset$ denote the set of omitted cells. Compute information content for all cluster-period cells.

(ii) **Cell selection:** At step r, define $\mathcal{B}_r$ as the set of cells with minimum information content not exceeding one, based on the current design:

$$\mathcal{B}_r = \arg\min_{(i,j)} \left\{ IC_{[\mathcal{A}_{r-1}]}(i, j) : IC_{[\mathcal{A}_{r-1}]}(i, j) \leq 1 \right\}$$

where $IC_{[\mathcal{A}]}(i, j) = \text{var}(\widehat{\delta})_{[ij \cup \mathcal{A}]} / \text{var}(\widehat{\delta})_{[\mathcal{A}]}$ denotes the updated information content conditional on already omitted set $\mathcal{A}$.

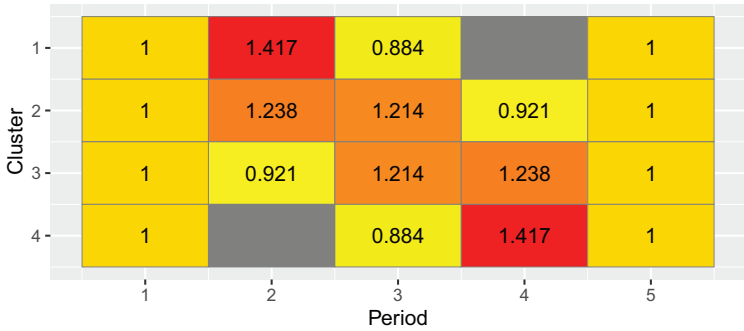
(iii) **Update:** Define $\mathcal{A}_r = \mathcal{A}_{r-1} \cup \mathcal{B}_r$, recompute $IC_{[\mathcal{A}_r]}(i, j)$ for all remaining cells.

(iv) **Repeat:** Iterate until all remaining cells satisfy $IC_{[\mathcal{A}_r]}(i, j) > 1$ or $\mathcal{B}_r = \emptyset$.

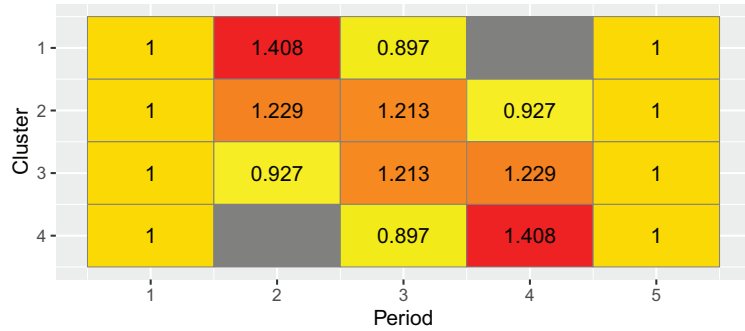
[4] Ehsan Rezaei-Darzi et al. "The impact of iterative removal of low-information cluster-period cells from a stepped wedge design". In: *BMC Medical Research Methodology* 23.1 (2023), p. 160.

Example 1: 4-sequence 5-period design

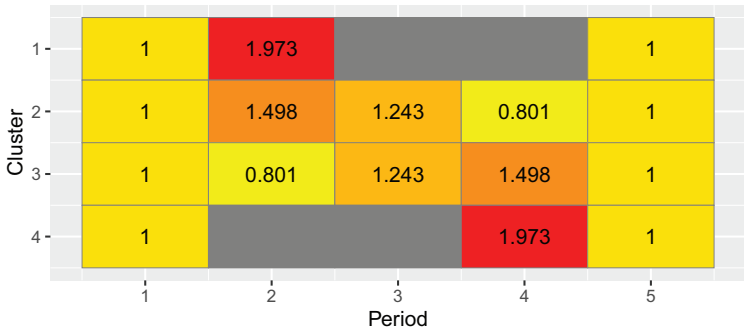
Simple exchangeable (step 1)
Variance of treatment effect estimator: 0.0337



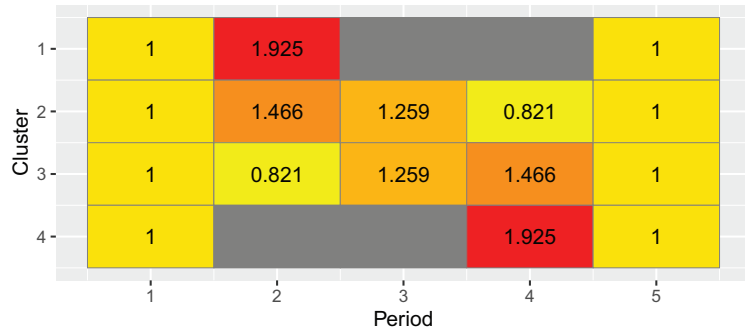
Exponential decay (step 1)
Variance of treatment effect estimator: 0.0334



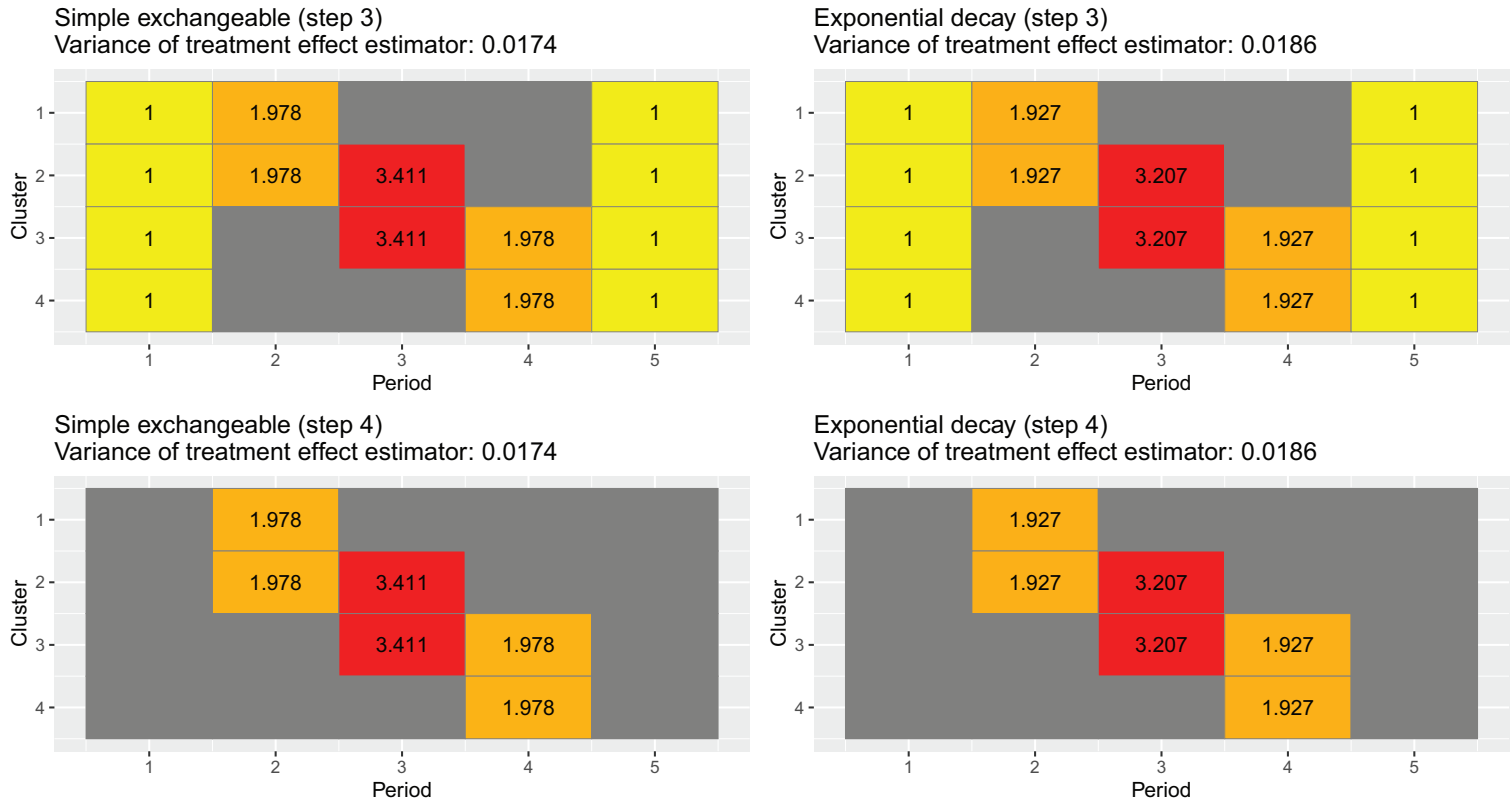
Simple exchangeable (step 2)
Variance of treatment effect estimator: 0.0258



Exponential decay (step 2)
Variance of treatment effect estimator: 0.0265

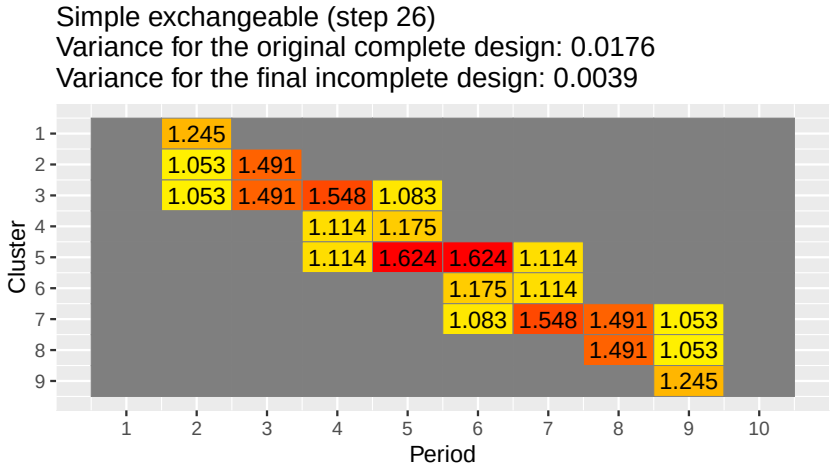


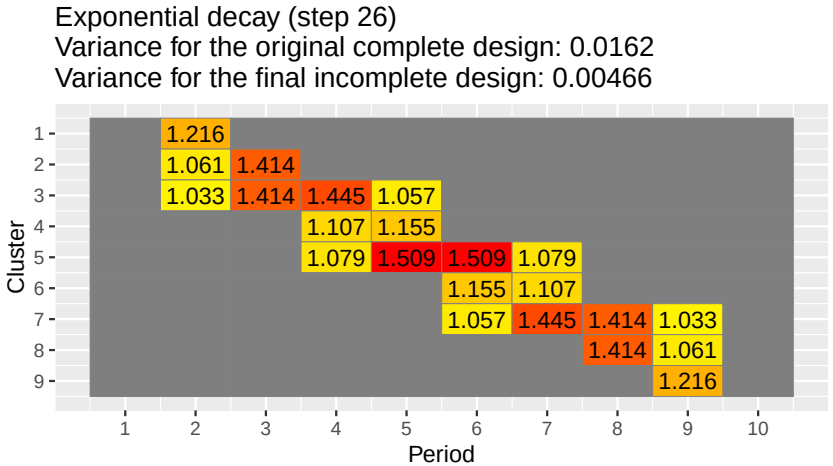
Example 1: 4-sequence 5-period design - cont'd



- stop in 4 steps and the remaining cells are located near the treatment switching points $\Longrightarrow$ staircase or dog-leg design
- robust sandwich variance does not increase as the step counter increases
- the final incomplete design admits a more efficient independence GEE estimator for δ, but requires a much lighter data collection schedule.

Example 2: 9-sequence 10-period design



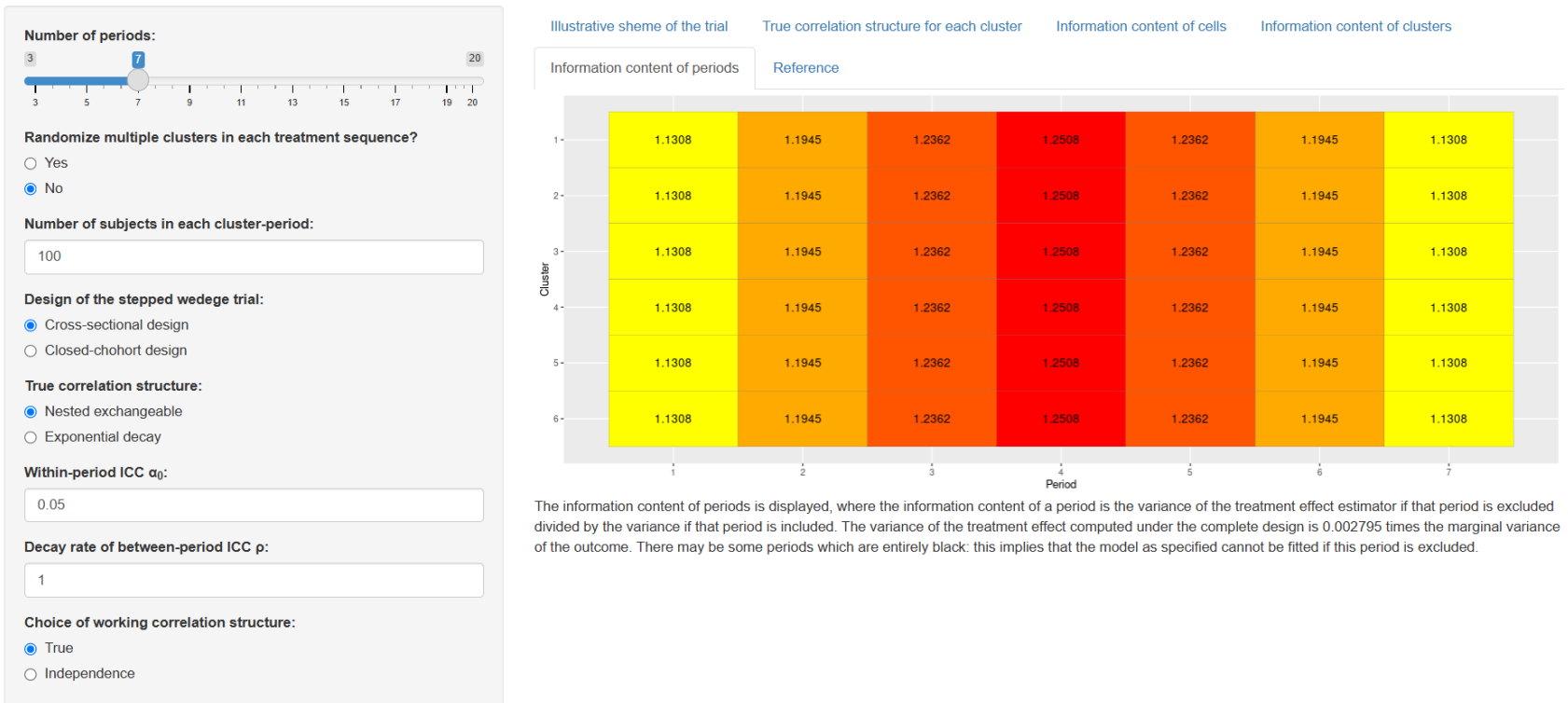


- final step after 26 elimitation steps
- still lead to some form of the staircase design, with a much smaller variance[5]
- interestingly, staircase designs happen to be the "optimal design" justified by independence GEE

[5] Kelsey L Grantham et al. "The staircase cluster randomised trial design: a pragmatic alternative to the stepped wedge". In: *Statistical Methods in Medical Research* 33.1 (2024), pp. 24–41.

Software

Alternative choices of the design parameters of the stepped wedge trials (e.g., cluster-period size N, number of treatment sequences, and number of periods) can be investigated in our R Shiny App available at:
https://information-content-binary-count.shinyapps.io/IC-OLS/



Discussion

- **Central message**: information content and an efficient incomplete stepped wedge design depend on the choice of analytic model
  - independence GEE + sandwich variance exhibit somewhat different patterns from more complex GEE + model-based variance
  - not to recommend independence GEE for all application, but rather to explore the properties of vertical estimators
  - staircase designs come out as optimally efficient designs under independence GEE $\Longrightarrow$ a somewhat surprising result, a paradox
    - insights into $\text{var}(\widehat{\delta}) = \text{Depth} \times \text{Overlap}$ for incomplete designs under OLS (Andrew Forbes @ Monash)
  - regardless of analytic models, cells near treatment switching often carry the largest amount of information and should be retained from an efficiency point of view
- **Caveat**: these patterns may change with a binary/count outcome, and when the treatment effect depends on exposure time (future research)

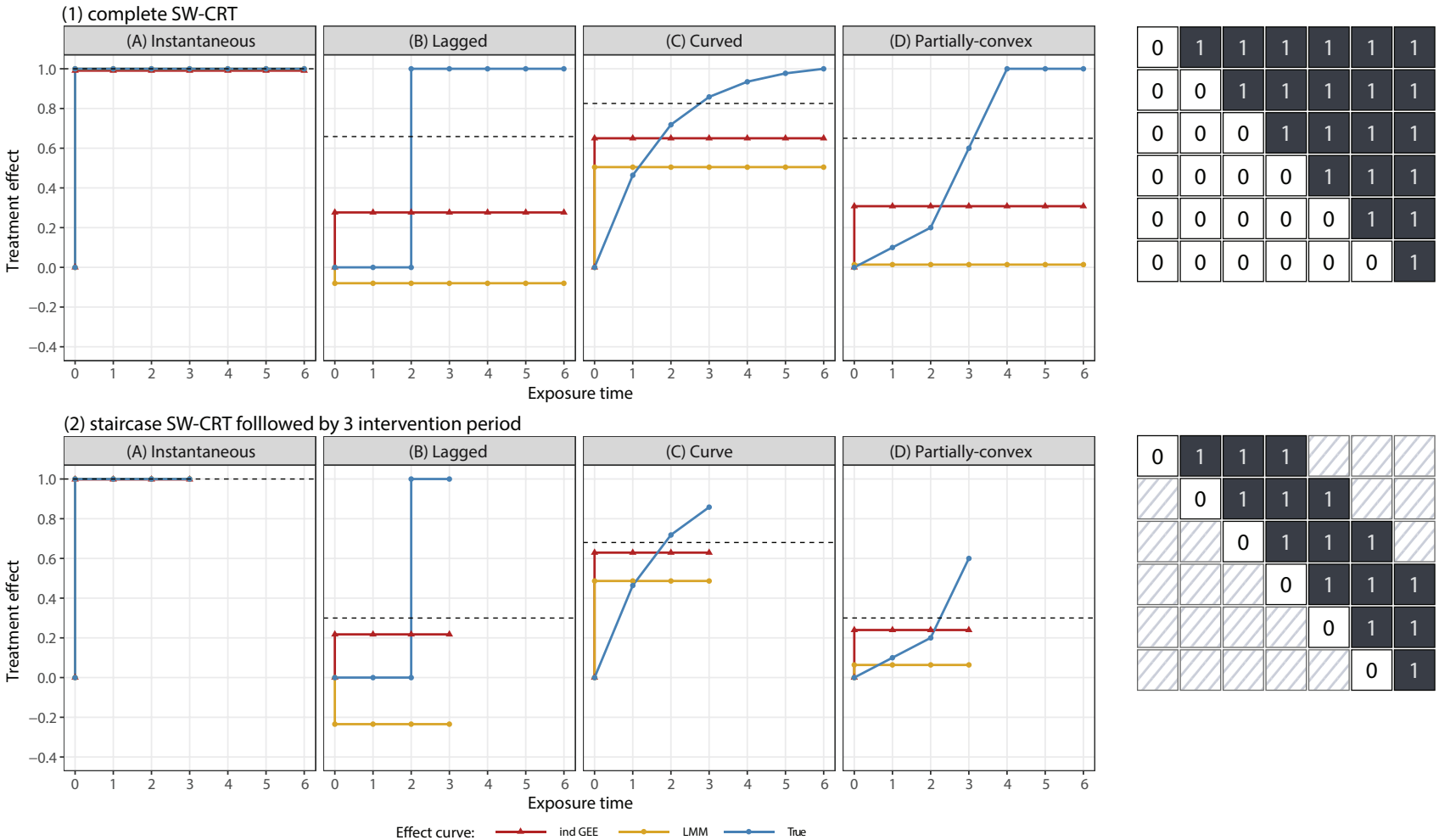
Extra: Discussion

- It all boil down to WEIGHTING
- **Explicit weighting**: e.g., cluster size weights, missing data weights
- **Implicit weighting** (focus of this work)
  - weighting implicit from fitting linear mixed models or independence GEE
  - weighting implicit from an incomplete design
- The implicit weighting under a complete design

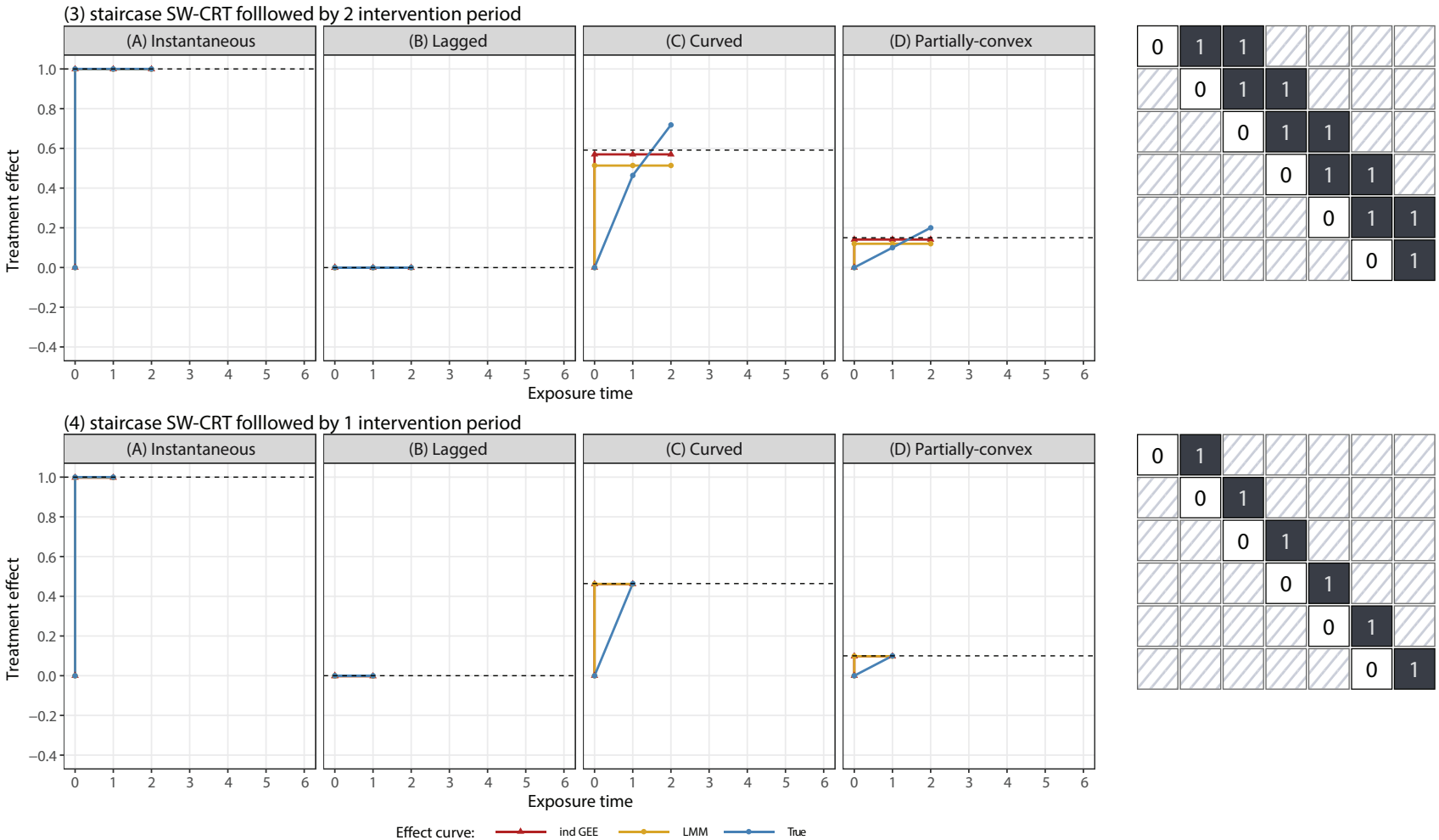
	Weights as a function of?
Linear mixed model	cluster-period size and ICCs
Independence GEE	propensity score

- The implicit weighting changes under a different incomplete design! $\Longrightarrow$ the more closer we get to a minimum staircase design
  - linear mixed model ($\Downarrow$) $\approx$ independence GEE ($\Uparrow$)

Extra: Bias due to ignoring time-varying treatment effect



Extra: Bias due to ignoring time-varying treatment effect



Incomplete Stepped Wedge Cluster Randomized Designs in Practice

Monica Taljaard
Senior Scientist, OHRI
Professor, School of Epidemiology and Public Health, University of Ottawa

Society for Clinical Trials 2025

Disclosures

- No relevant disclosures

Outline

01

Review the current landscape of published Stepped Wedge Cluster Randomized Trials (SW-CRTs)

02

Identify potential barriers to adoption of incomplete SW-CRT variants in practice

03

Describe my anecdotal experience with a planned SW-CRT

Current landscape of SW-CRTs

- What are typical characteristics of published SW-CRTs?
- What do incomplete designs tend to look like?
- Why are incomplete designs being used?

Source data

- Systematic review of SW-CRTs published 2016-2022 (N=160)

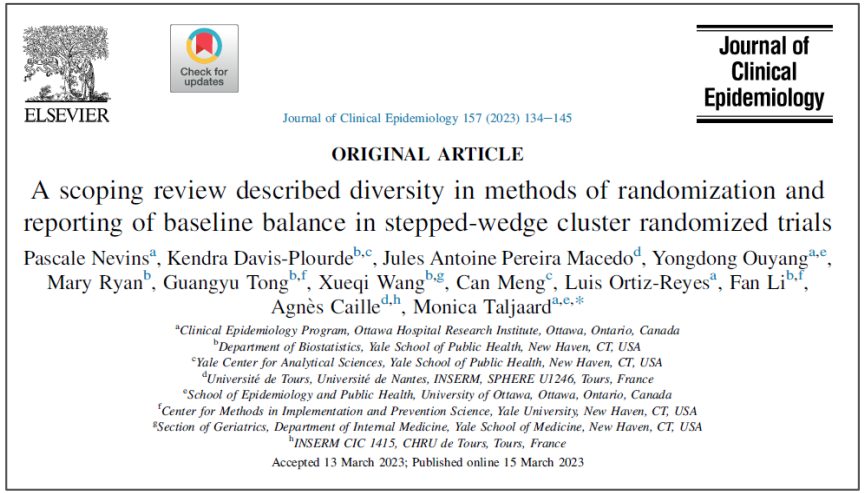
Article

CLINICAL TRIALS

Clinical Trials
2024, Vol. 21(2) 199–210
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/17407745231208397
journals.sagepub.com/home/ctj
Sage

Adherence to key recommendations for design and analysis of stepped-wedge cluster randomized trials: A review of trials published 2016–2022

Pascale Nevins[1], Mary Ryan[2], Kendra Davis-Plourde[2,3], Yongdong Ouyang[1,4], Jules Antoine Pereira Macedo[5], Can Meng[2,3], Guangyu Tong[2,6], Xueqi Wang[2,7], Luis Ortiz-Reyes[1], Agnès Caille[5,8], Fan Li[2,6] and Monica Taljaard[1,4]



ELSEVIER

Check for updates

Journal of Clinical Epidemiology

Journal of Clinical Epidemiology 157 (2023) 134–145

ORIGINAL ARTICLE

A scoping review described diversity in methods of randomization and reporting of baseline balance in stepped-wedge cluster randomized trials

Pascale Nevins[a], Kendra Davis-Plourde[b,c], Jules Antoine Pereira Macedo[d], Yongdong Ouyang[a,e], Mary Ryan[b], Guangyu Tong[b,f], Xueqi Wang[b,g], Can Meng[c], Luis Ortiz-Reyes[a], Fan Li[b,f], Agnès Caille[d,h], Monica Taljaard[a,e,*]

[a]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada
[b]Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA
[c]Yale Center for Analytical Sciences, Yale School of Public Health, New Haven, CT, USA
[d]Université de Tours, Université de Nantes, INSERM, SPHERE U1246, Tours, France
[e]School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada
[f]Center for Methods in Implementation and Prevention Science, Yale University, New Haven, CT, USA
[g]Section of Geriatrics, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA
[h]INSERM CIC 1415, CHRU de Tours, Tours, France

Accepted 13 March 2023; Published online 15 March 2023

Trial characteristics

	N=160
Cross-sectional design	76%
# sequences (Median, Q1-Q3)	5 (4-7)
# clusters (Median, Q1-Q3)	11 (8-18)
# participants	2,724 (643-14,734)
Active recruitment of participants	50%
Exclusively routinely collected data	6%
Incomplete design (including transition periods)	28%
Continuous primary outcome	26%

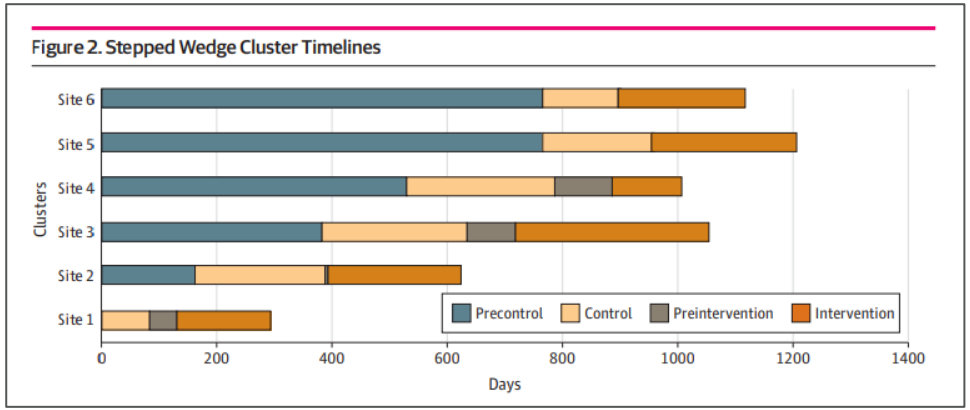
What do incomplete designs look like?

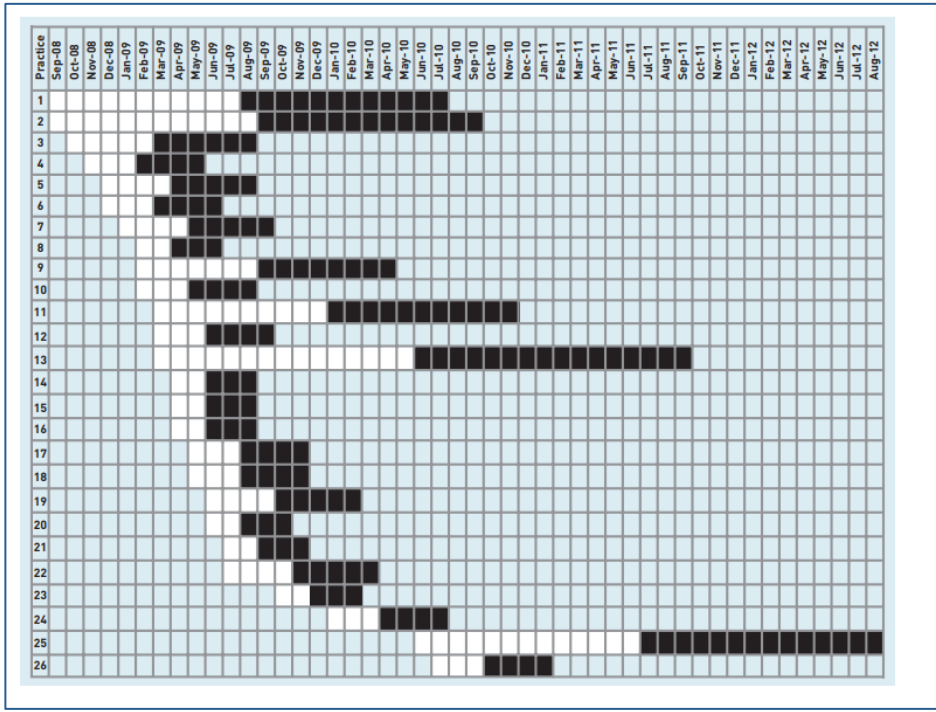
- Reviewed design diagrams, when available
- Identified three broad groups
- (Ignoring for now, designs with only transition periods)

Group 1

- Ad hoc pattern
  - Based on anticipated logistical considerations (e.g., major holidays)
  - Post-hoc, based on recruitment or implementation challenges
  - Reflect misunderstandings of the SW-CRT

Group 1: Examples



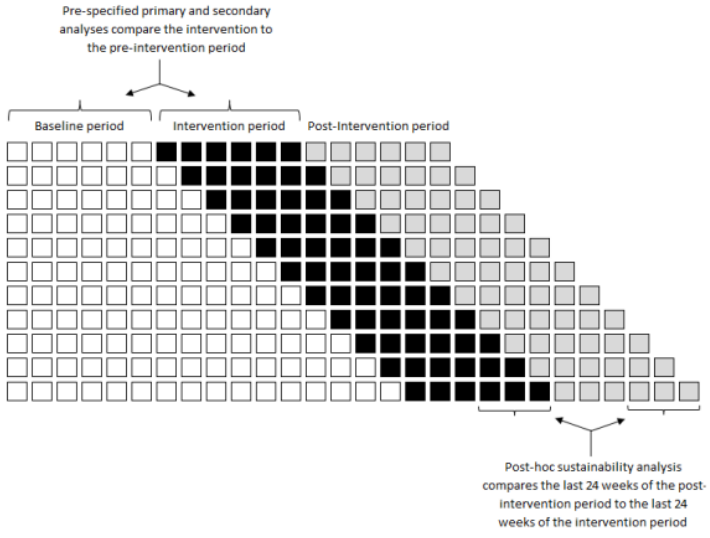


Group 2

- Eliminate control periods (staggered at start, aligned at end)
- Eliminate intervention periods (aligned at start, staggered at end)
- Different trial duration across sequences

Group 2: Examples

Figure S1: Illustration of the stepped wedge design as implemented in the DQIP trial. Each row represents practices randomised to the same start date and each box represents an eight weekly measurement of the primary outcome.



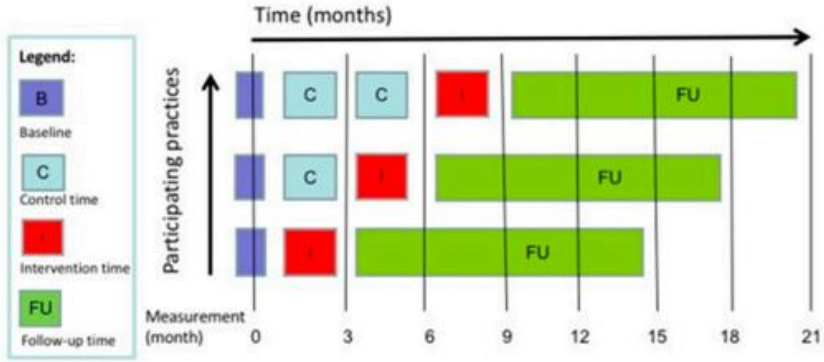
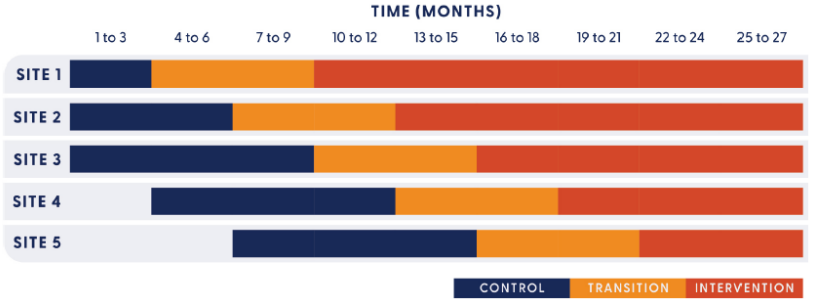
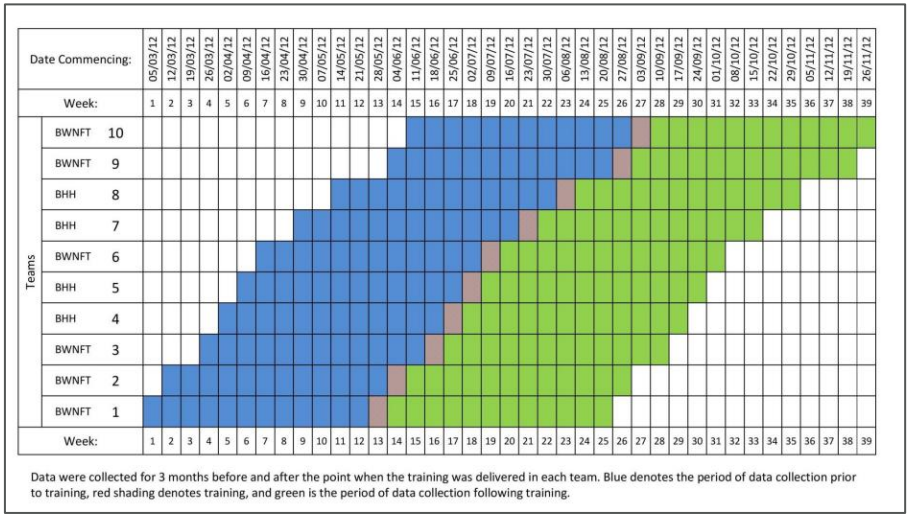


Figure 1. Flowchart stepped wedge design.

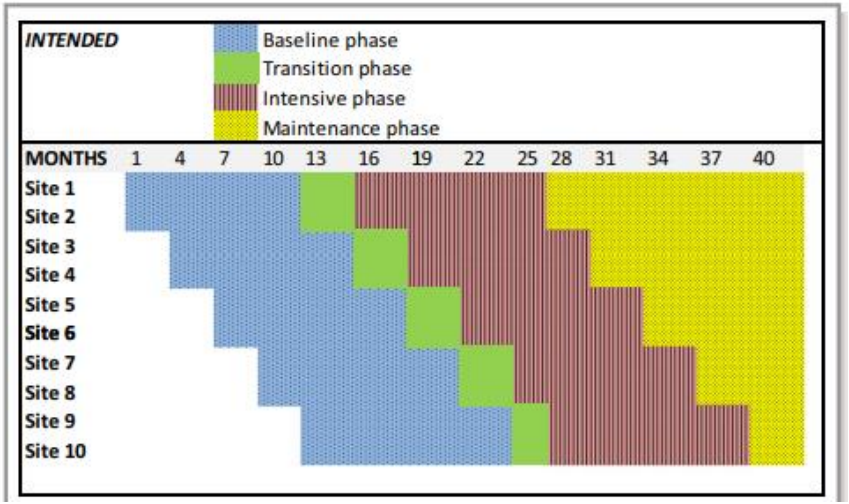


Group 3

- Staircase patterns
  - Staggered at start and end
  - Equal or unequal number of before & after periods
  - Same trial duration across sequences



Data were collected for 3 months before and after the point when the training was delivered in each team. Blue denotes the period of data collection prior to training, red shading denotes training, and green is the period of data collection following training.



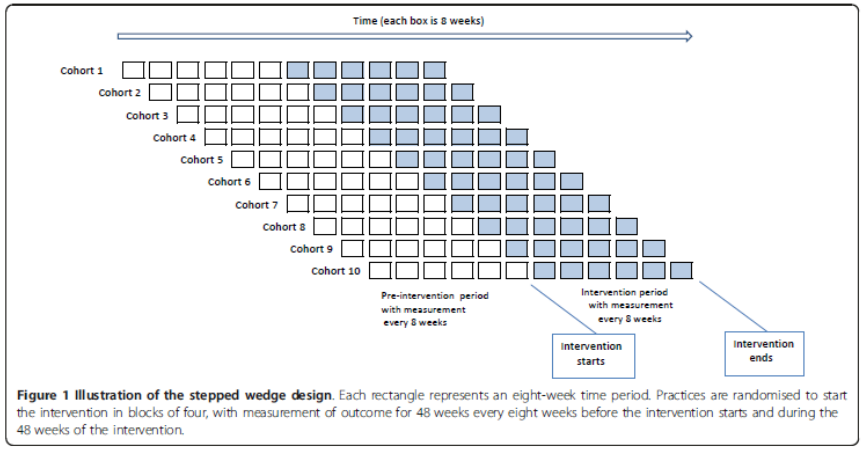


Figure 1 Illustration of the stepped wedge design. Each rectangle represents an eight-week time period. Practices are randomised to start the intervention in blocks of four, with measurement of outcome for 48 weeks every eight weeks before the intervention starts and during the 48 weeks of the intervention.

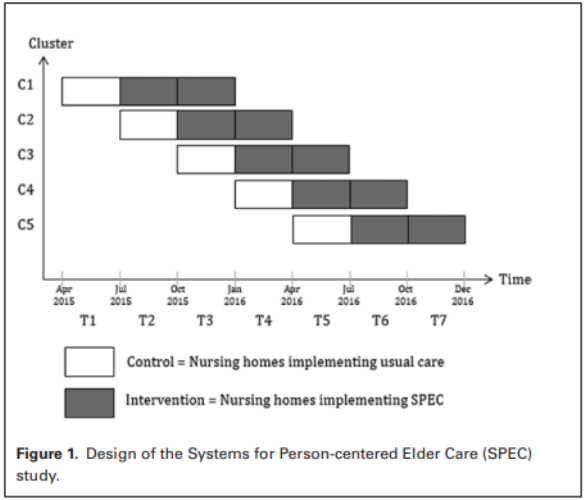


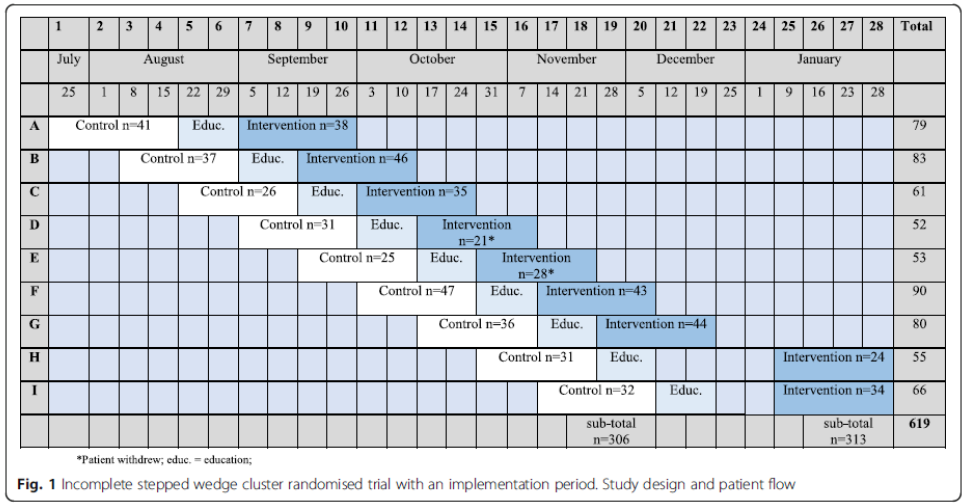
Figure 1. Design of the Systems for Person-centered Elder Care (SPEC) study.

Why are incomplete designs being used?

- Clear rationale seldom provided
- Seems to be mostly based on logistics
- Not based on information content of cells

Example

- **Keogh et al.** “Wards had equal exposure periods… which helped avoid the potentially confounding impact of different levels of exposure to the intervention… as well as minimizing the measurement burden”



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	Total
	July	August					September				October					November				December				January					
	25	1	8	15	22	29	5	12	19	26	3	10	17	24	31	7	14	21	28	5	12	19	25	1	9	16	23	28	
A	Control n=41				Educ.		Intervention n=38																						79
B			Control n=37				Educ.		Intervention n=46																				83
C					Control n=26				Educ.		Intervention n=35																		61
D							Control n=31				Educ.		Intervention n=21*																52
E									Control n=25				Educ.		Intervention n=28*														53
F											Control n=47				Educ.		Intervention n=43												90
G													Control n=36				Educ.		Intervention n=44										80
H															Control n=31				Educ.						Intervention n=24				55
I																	Control n=32				Educ.				Intervention n=34				66
																		sub-total n=306								sub-total n=313			619

*Patient withdrew; educ. = education;

Fig. 1 Incomplete stepped wedge cluster randomised trial with an implementation period. Study design and patient flow

Outline

01

Review the current landscape of published SW-CRTs

02

Identify potential barriers to adoption of incomplete SW-CRT variants in practice

03

Describe my anecdotal experience with a planned SW-CRT

Potential barriers 1: Knowledge

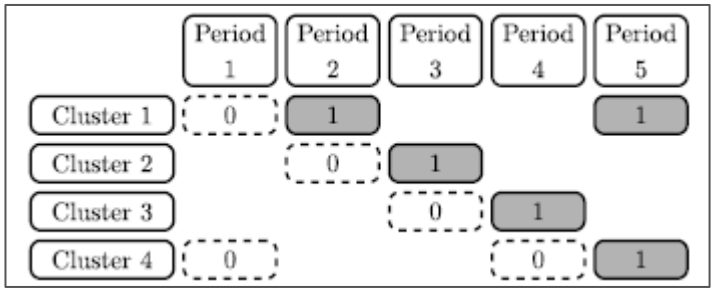
- Lack of awareness / expertise
- Lack of guidance on how to choose a particular design
  - Based on logistics? Based on power? What if they conflict?

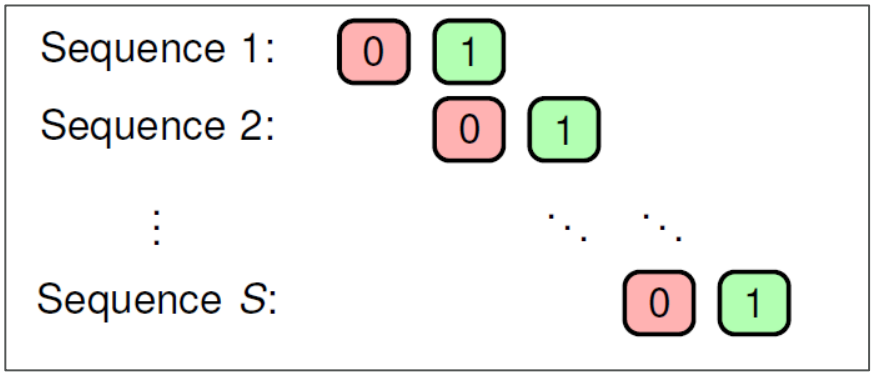
Potential barriers 2: Power

- Information content of cells depends on the assumed correlation structure and postulated ICCs – which are usually unknown
- Concerns about potential loss of information (secondary outcomes, subgroup analyses, exploratory analyses)

Potential barriers 3: Bias

- SW-CRTs are confounded with period effects: are staircase designs much better than before and after designs?
- Analytical approaches depend on several model assumptions: are staircase designs more vulnerable to violation of model assumptions?





Potential barriers 4: Other

- Reduced ability to assess learning or decay effects
- Less flexibility in dealing with implementation and recruitment challenges – may only have “one chance”
- Potentially a long gap between randomization and trial initiation for some clusters – increased risk of attrition?

Outline

01

Review the current landscape of published SW-CRTs

02

Identify potential barriers to adoption of incomplete SW-CRT variants in practice

03

Describe my anecdotal experience with a planned SW-CRT

IRIS trial

- **Background**: Each year in the USA, syncope is responsible for over 1.2M ED visits and over $2.4B in healthcare costs; the validated Canadian Syncope Risk Score (CSRS) can be used to safely discharge low risk patients
- **Objective**: Evaluate implementation of Clinical Decision Support using the CSRS
- **Primary outcome:** Risk-concordant disposition (high risk patient admitted; low risk patient not admitted)

Sample size constraints

- Ideally want to have a 3-year trial (maximize grant $)
- Maximum of 9 independent ED clusters available
- Step lengths of 3 months
- Expect an average of 170 patients per cluster-period

Sample size parameters

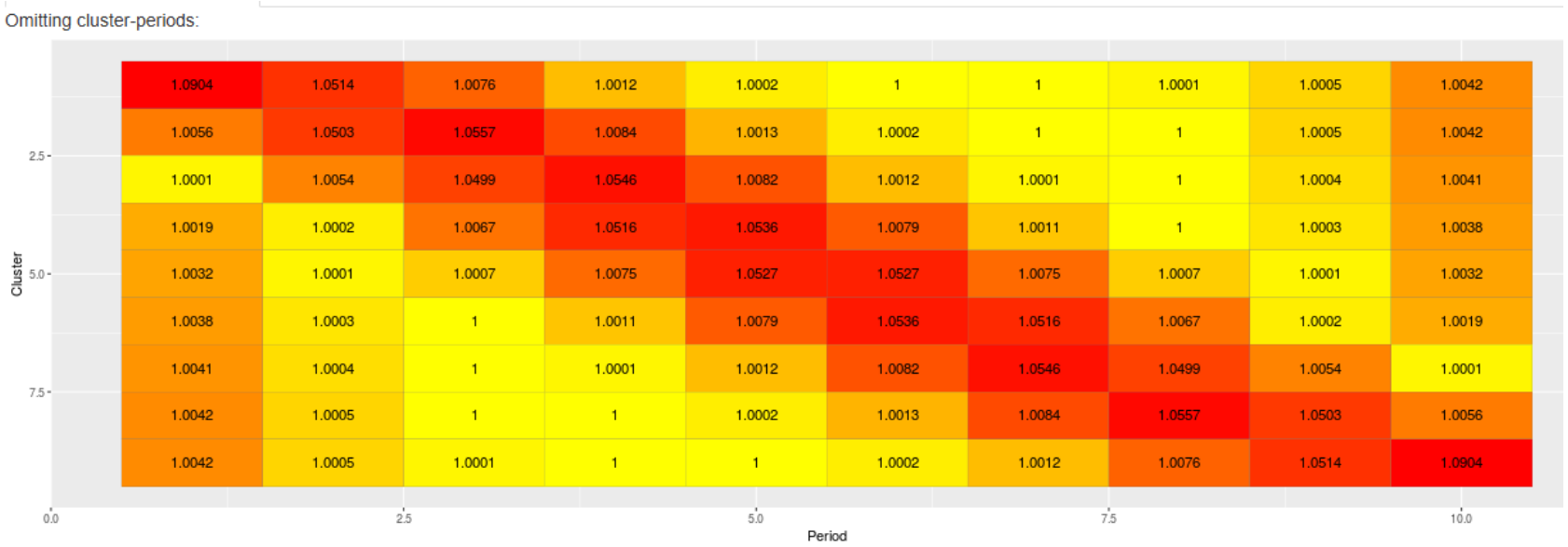
- Control proportion: 0.75
- Effect size: 10% absolute increase
- Within-period ICC: 0.05
- CAC: 0.95 (exponential decay)

Complete SW-CRT design

- Estimated power: 99%

	Quarters									
Clusters	1	2	3	4	5	6	7	8	9	10
1	0	1	1	1	1	1	1	1	1	1
2	0	0	1	1	1	1	1	1	1	1
3	0	0	0	1	1	1	1	1	1	1
4	0	0	0	0	1	1	1	1	1	1
5	0	0	0	0	0	1	1	1	1	1
6	0	0	0	0	0	0	1	1	1	1
7	0	0	0	0	0	0	0	1	1	1
8	0	0	0	0	0	0	0	0	1	1
9	0	0	0	0	0	0	0	0	0	1

Information content



https://monash-biostat.shinyapps.io/informationcontent/

Incomplete SW-CRT design

- Estimated power: 99%

	Quarters									
Clusters	1	2	3	4	5	6	7	8	9	10
1	0	1	1	1					1	1
2	0	0	1	1	1				1	1
3		0	0	1	1	1			1	1
4	0		0	0	1	1	1		1	1
5	0		0	0	0	1	1	1		1
6	0	0		0	0	0	1	1		1
7	0	0			0	0	0	1	1	
8	0	0				0	0	0	1	1
9	0	0					0	0	0	1

Staircase SW-CRT design

- Estimated power: 99%

	Quarters										
Clusters	1	2	3	4	5	6	7	8	9	10	11
1	0	1	1								
2		0	1	1							
3			0	1	1						
4				0	1	1					
5					0	1	1				
6						0	1	1			
7							0	1	1		
8								0	1	1	
9									0	1	1

Ultimate decision

- Complete design with new primary outcome (overall admissions)
- Estimated power 92%

	Quarters									
Clusters	1	2	3	4	5	6	7	8	9	10
1	0	1	1	1	1	1	1	1	1	1
2	0	0	1	1	1	1	1	1	1	1
3	0	0	0	1	1	1	1	1	1	1
4	0	0	0	0	1	1	1	1	1	1
5	0	0	0	0	0	1	1	1	1	1
6	0	0	0	0	0	0	1	1	1	1
7	0	0	0	0	0	0	0	1	1	1
8	0	0	0	0	0	0	0	0	1	1
9	0	0	0	0	0	0	0	0	0	1

Conclusions

- Although theoretically, ethically, and intuitively appealing, I don’t (yet) routinely recommend incomplete designs in consultations
- Need more methodological development
- Logistics - rather than information content - may drive adoption of a particular incomplete variant
- Maybe consider incomplete variant as a nested substudy (e.g., for patient-reported outcomes)