Methods: Mind the Gap
Webinar Series

Insights into the Efficiency and Benefits of Staircase Cluster Randomized Trial Designs

Presented by:

Kelsey Grantham, Ph.D.
Monash University

NIH National Institutes of Health
Office of Disease Prevention

Insights into the efficiency and benefits of staircase cluster randomized trial designs

Kelsey Grantham

5 Nov 2024

MONASH University

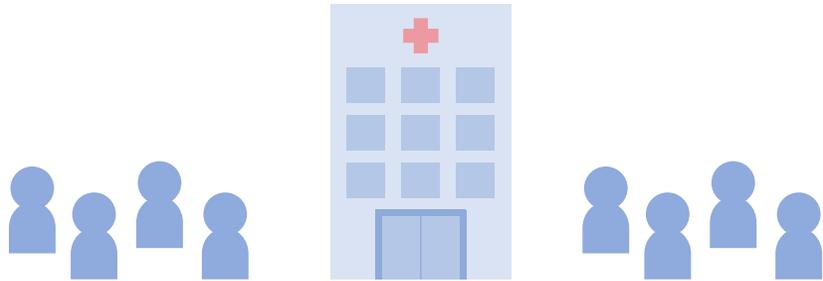CLUSTER

AusTriM

ViCBiostat

Outline

- Intro: What are...
  - ... cluster randomized trials?
  - ... stepped wedge designs?
  - ... staircase designs?

- Underpinning theory for staircase designs

- How efficient are staircase designs compared to stepped wedge designs?

- Which staircase design variants are most efficient?

What is a cluster randomized trial?
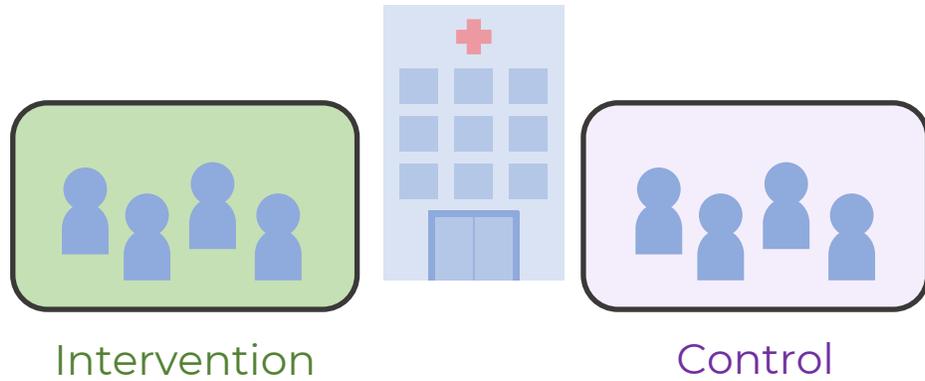
Randomized controlled trial

What is a cluster randomized trial? (cont'd)
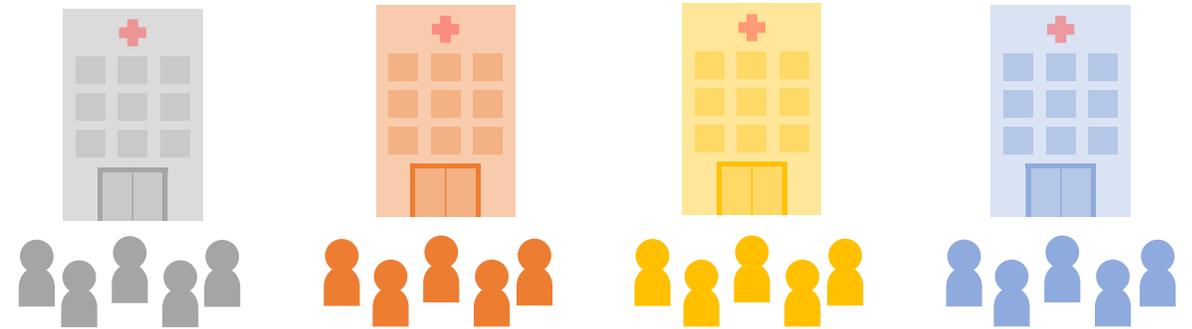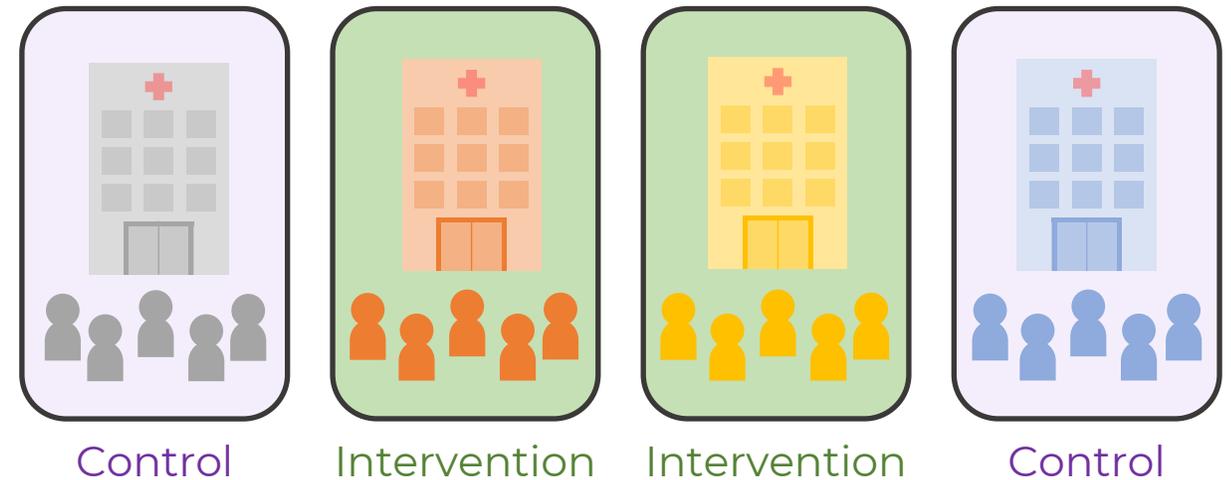
Randomized controlled trial

Intervention

Control

What is a cluster randomized trial? (cont'd)

Randomized controlled trial

Intervention   Control

Cluster randomized trial

What is a cluster randomized trial? (cont'd)
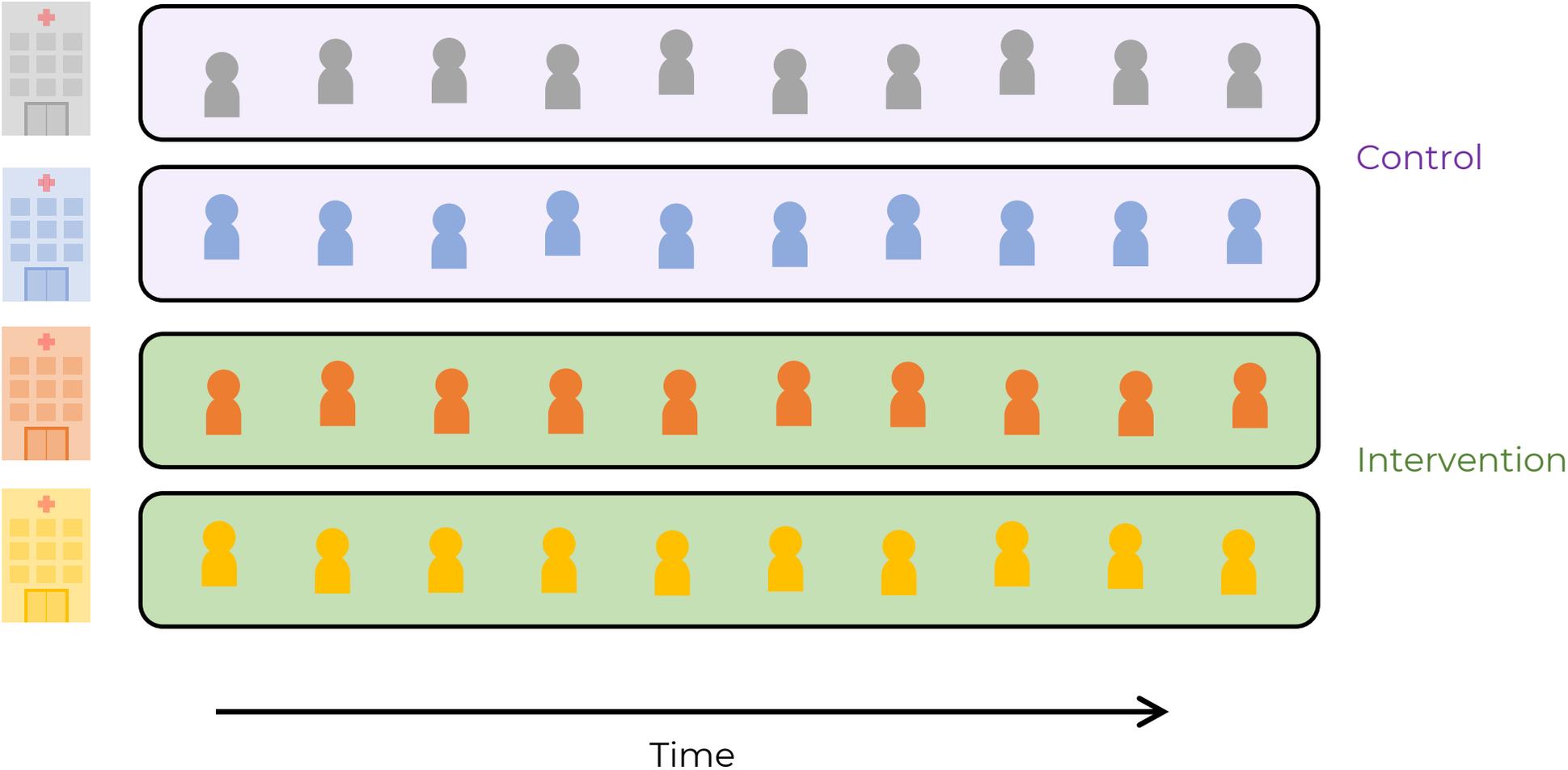
Randomized controlled trial

Intervention

Control
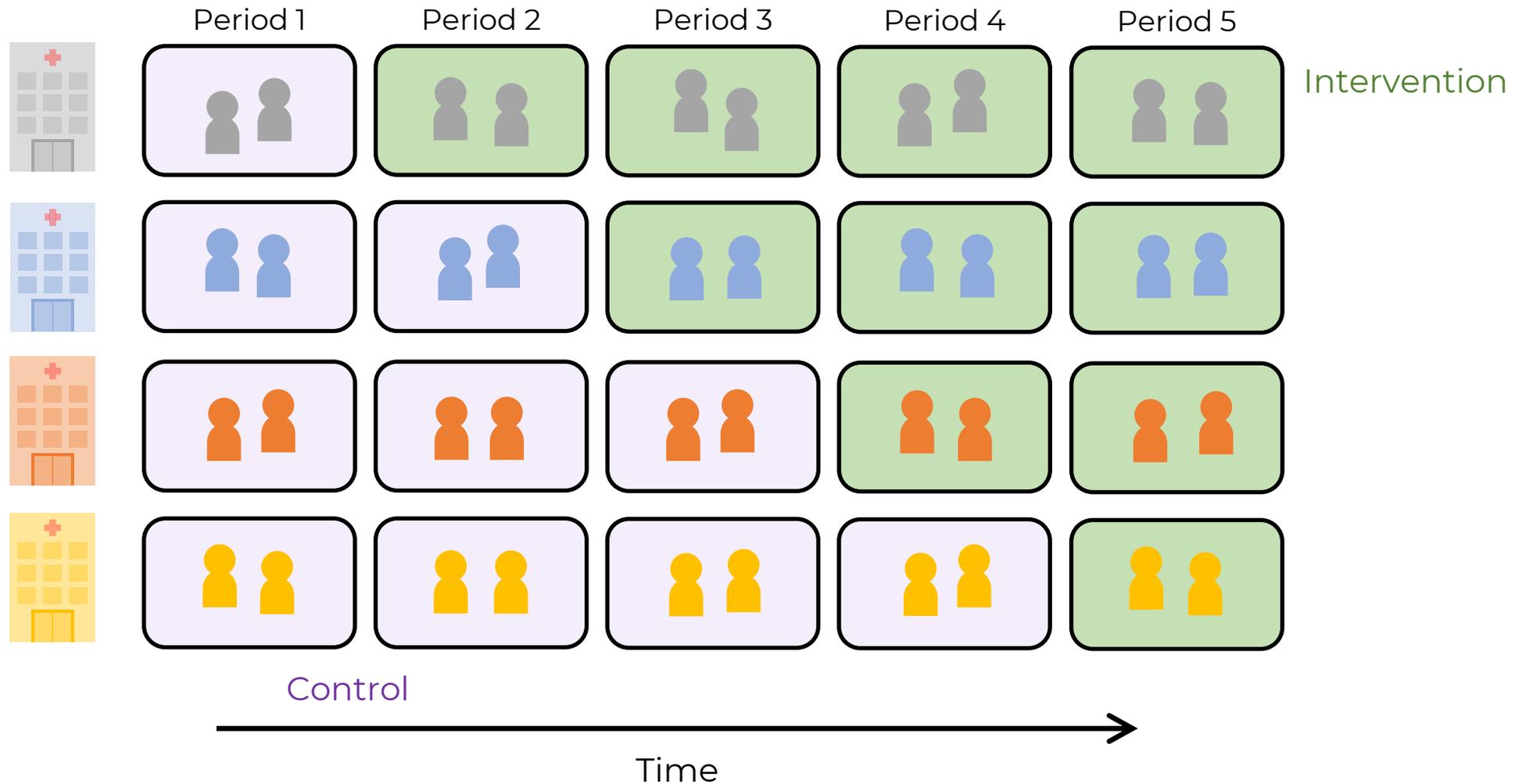
Cluster randomized trial

Control   Intervention   Intervention   Control

Cluster randomized trial designs
Parallel

Control

Intervention

Time

Cluster randomized trial designs (cont'd)
Stepped wedge

Stepped wedge designs

5 periods

8 clusters

0	1	1	1	1
0	1	1	1	1
0	0	1	1	1
0	0	1	1	1
0	0	0	1	1
0	0	0	1	1
0	0	0	0	1
0	0	0	0	1

4 unique sequences

2 clusters per sequence

0 = Control
1 = Intervention

×10

10 participants per cluster-period

Stepped wedge designs (cont'd)

5 periods

8 clusters

0	1	1	1	1
0	1	1	1	1
0	0	1	1	1
0	0	1	1	1
0	0	0	1	1
0	0	0	1	1
0	0	0	0	1
0	0	0	0	1

- Appealing
  - All clusters eventually receive the intervention
  - Gradual rollout of the intervention across clusters

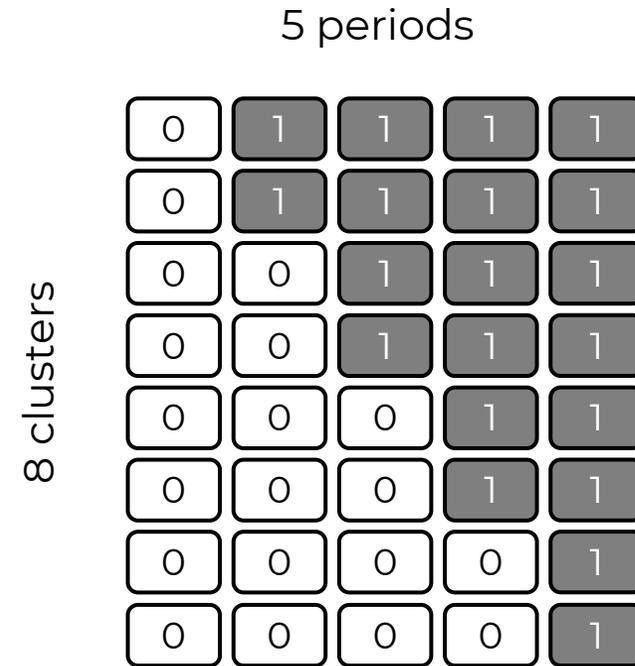- More efficient than parallel designs
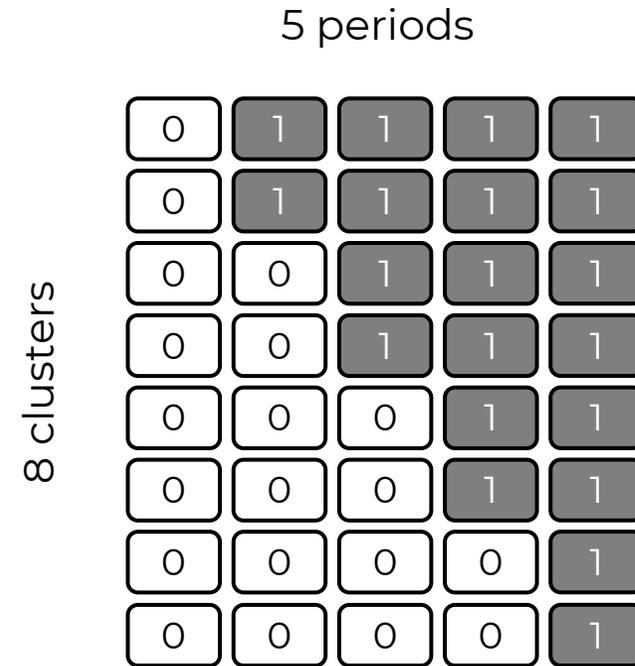 - Make use of both between- and within-cluster comparisons

Stepped wedge designs (cont'd)

- Appealing
  - All clusters eventually receive the intervention
  - Gradual rollout of the intervention across clusters
- More efficient than parallel designs
  - Make use of both between- and within-cluster comparisons
- Can be burdensome and expensive
  - All clusters must contribute data
    in *all* periods of the trial

5 periods

8 clusters

0	1	1	1	1
0	1	1	1	1
0	0	1	1	1
0	0	1	1	1
0	0	0	1	1
0	0	0	1	1
0	0	0	0	1
0	0	0	0	1

Stepped wedge designs (cont'd)

5 periods

8 clusters

0	1	1	1	1
0	1	1	1	1
0	0	1	1	1
0	0	1	1	1
0	0	0	1	1
0	0	0	1	1
0	0	0	0	1
0	0	0	0	1

- Appealing
  - All clusters eventually receive the intervention
  - Gradual rollout of the intervention across clusters
- More efficient than parallel designs
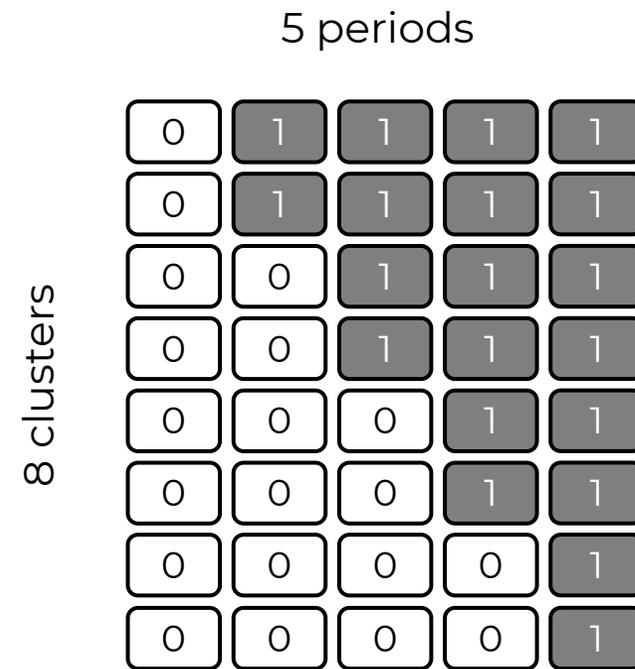 - Make use of both between- and within-cluster comparisons
- Can be burdensome and expensive
  - All clusters must contribute data in *all* periods of the trial

→
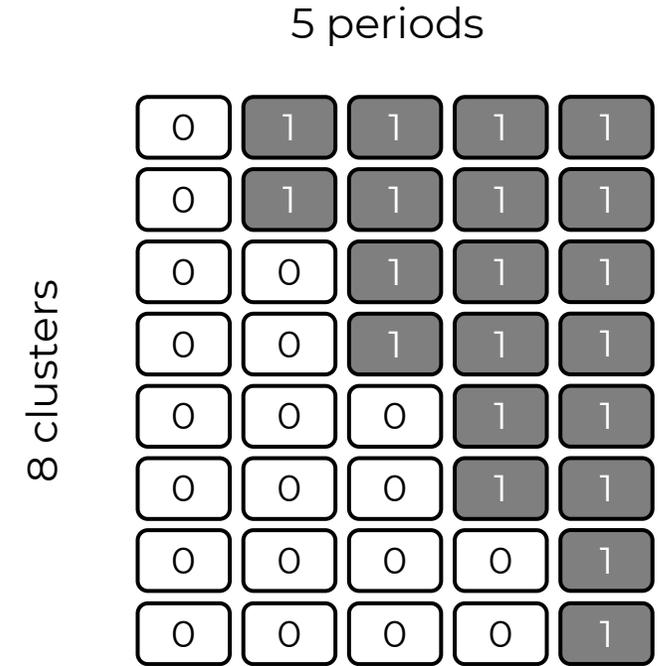
- Infeasible
  - Trial cannot be conducted, intervention cannot be tested
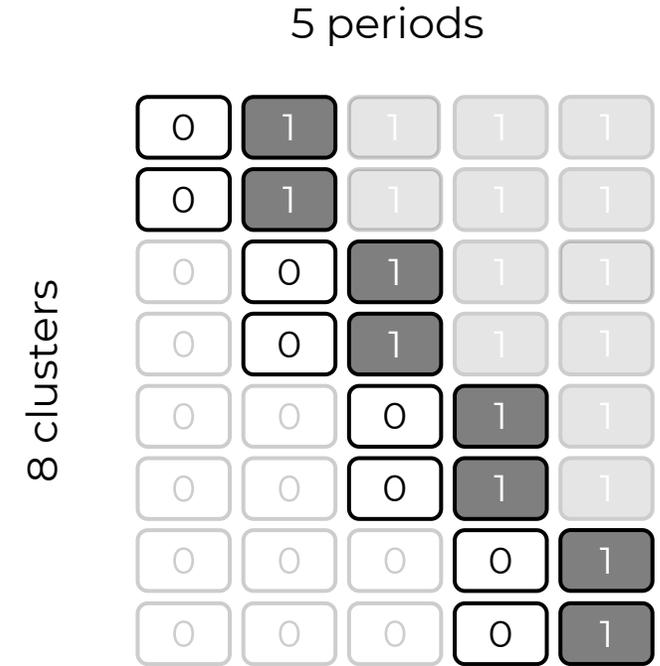
Incomplete stepped wedge designs

- Certain cluster-period cells in a stepped wedge contribute more information towards estimation of the treatment effect than others (Kasza and Forbes, 2019)

5 periods

8 clusters

0	1	1	1	1
0	1	1	1	1
0	0	1	1	1
0	0	1	1	1
0	0	0	1	1
0	0	0	1	1
0	0	0	0	1
0	0	0	0	1

Kasza J and Forbes AB. Information content of cluster-period cells in stepped wedge trials. *Biometrics* 2019; 75(1):144-152.

Incomplete stepped wedge designs (cont'd)

- ## Certain cluster-period cells in a stepped wedge contribute more information towards estimation of the treatment effect than others (Kasza and Forbes, 2019)

  - ### Information-rich cluster-periods: Near the treatment switch

  - ### Information-poor cluster-periods: Distant from the treatment switch

5 periods

8 clusters

Kasza J and Forbes AB. Information content of cluster-period cells in stepped wedge trials. *Biometrics* 2019; 75(1):144-152.

Incomplete stepped wedge designs are on the rise

Keogh et al (2020)

- Clusters: Hospital wards

- Intervention: Peripheral intravenous catheter (PIVC) flushing education program

- Outcome: All-cause PIVC failure

- Motivation for design:
  - Staggered rollout of intervention
  - Minimising 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**

Keogh S, Shelverton C, Flynn J, et al. Implementation and evaluation of short peripheral intravenous catheter flushing guidelines: a stepped wedge cluster randomised trial. *BMC Med* 2020; 18(1): 1-11.

Incomplete stepped wedge designs are on the rise (cont'd)

Wagner et al (2020)

- Clusters: Schools in remote Aboriginal communities
- Intervention: Education program to improve student self-regulation
- Outcome: Teacher-rated questionnaire measuring student disruptive behaviour
- Motivation for design: "the burden of data collection on participants would be too great in a stepped wedge trial"

Study Year and Month	2016						2017					
	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.

Incomplete stepped wedge designs are on the rise (cont'd)

Kim et al (2021)

- Clusters: Nursing homes

- Intervention: Integrated care management model

- Outcome: Quality of care measure

- Motivation for design:
  - Nursing homes incentivized to participate due to eventually receive the intervention
  - Staggered rollout of intervention
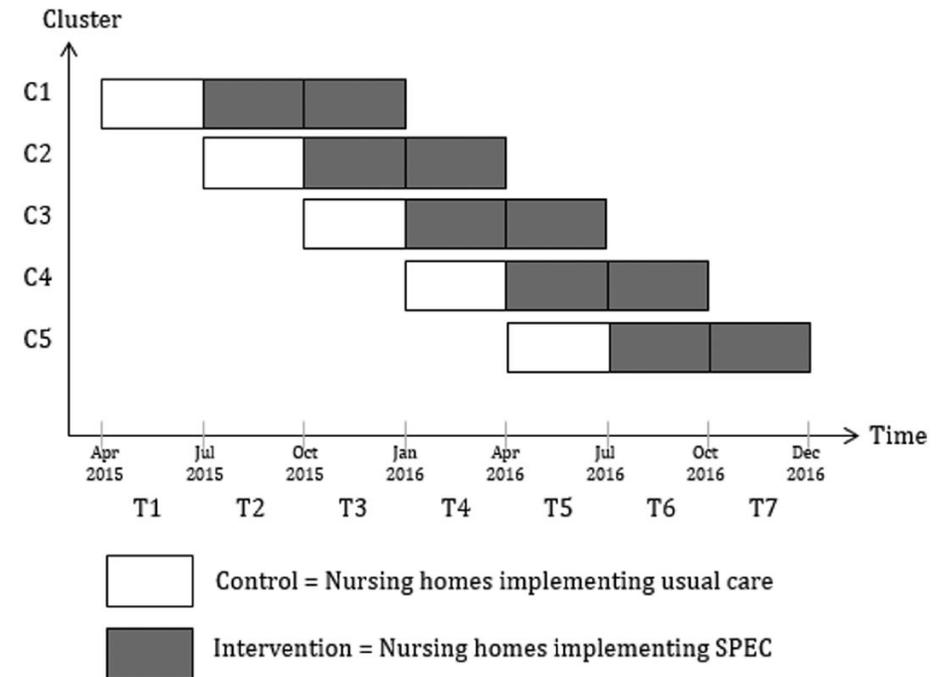
Kim H, Jung Y, Kim G, et al. Effectiveness of a Technology-Enhanced Integrated Care Model for Frail Older People: A Stepped-Wedge Cluster Randomized Trial in Nursing Homes. *The Gerontologist* 2021; 61(3): 460-469.

Staircase designs

- ## Appealing
  - Clusters receive the intervention sooner upon commencing data collection
  - Gradual rollout of the intervention across clusters
- ## Less burdensome
  - Clusters contribute data in *a limited number of periods*

5 periods
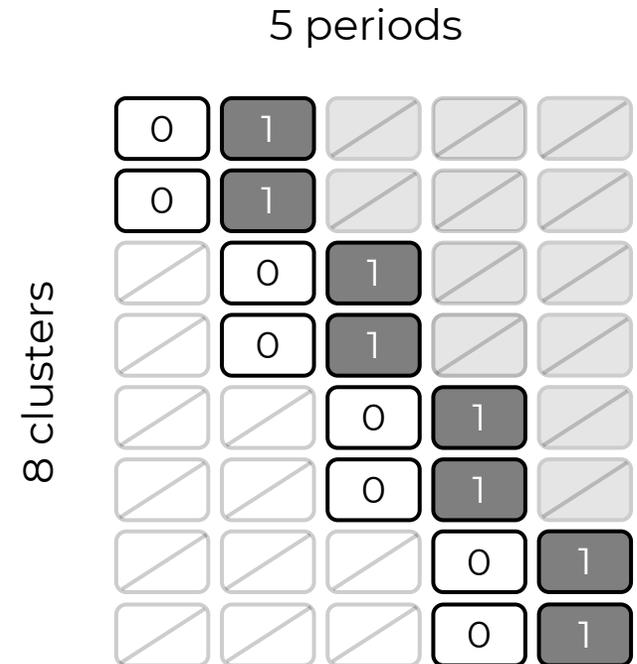
8 clusters

Staircase designs (cont'd)

- Appealing
  - Clusters receive the intervention sooner upon commencing data collection
  - Gradual rollout of the intervention across clusters
- Less burdensome
  - Clusters contribute data in *a limited number of periods*
- How efficient compared to stepped wedge?
- Which staircase design variants are most efficient?

5 periods

8 clusters

0	1			
0	1			
	0	1		
	0	1		
		0	1	
		0	1	
			0	1
			0	1
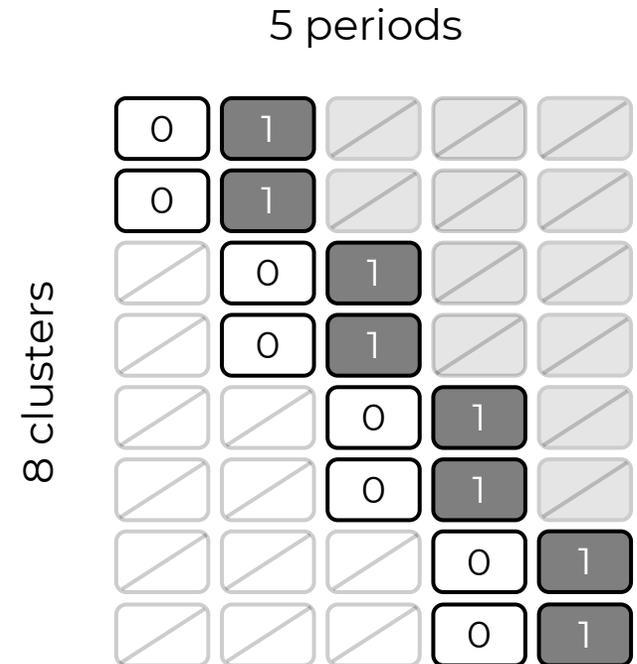
Many staircase design variants are possible
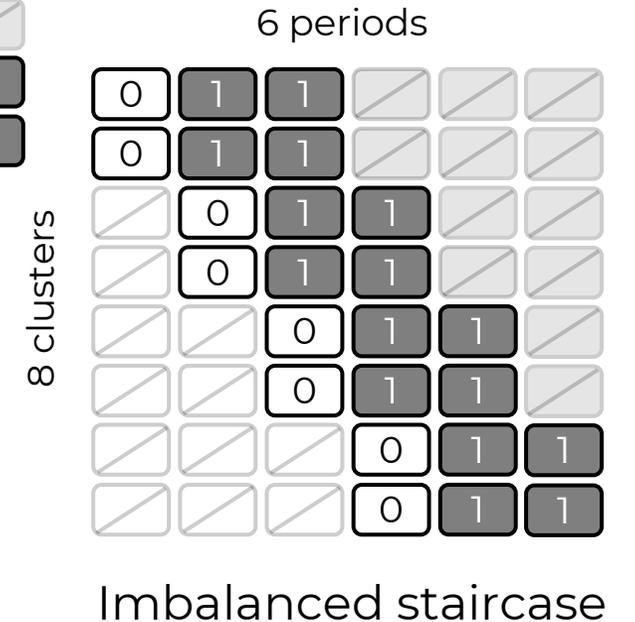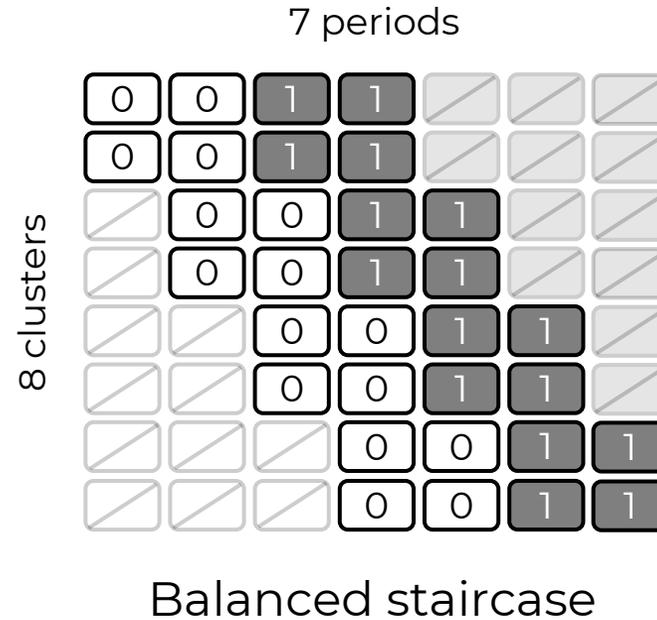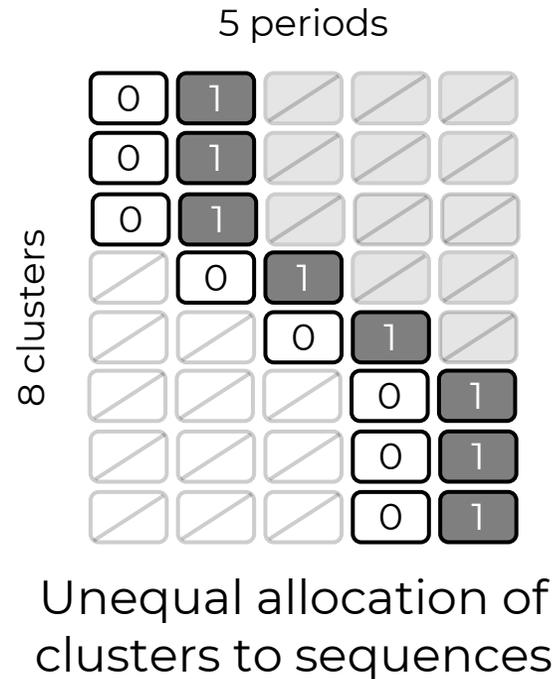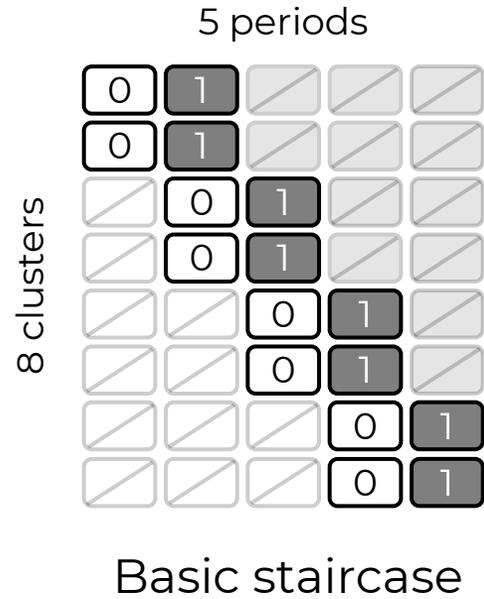
5 periods

8 clusters

Basic staircase

5 periods

8 clusters

Unequal allocation of clusters to sequences

7 periods

8 clusters

Balanced staircase

6 periods

8 clusters

Imbalanced staircase

Outline

- Intro: What are...
  - ... cluster randomized trials?
  - ... stepped wedge designs?
  - ... staircase designs?

- Underpinning theory for staircase designs

- How efficient are staircase designs compared to stepped wedge designs?

- Which staircase design variants are most efficient?

Underpinning theory for staircase designs

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.

Statistical model: individual level

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:

$$Y_{skti} = \mathbf{Z}_{st}\boldsymbol{\beta} + X_{st}\theta + CP_{skt} + e_{skti}$$

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

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

Statistical model: individual level (cont'd)

$$Y_{skti} = \boldsymbol{Z}_{st}\boldsymbol{\beta} + X_{st}\theta + CP_{skt} + e_{skti}$$

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

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

Within-period intracluster correlation (ICC)

where

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

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

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

Cluster autocorrelation

Statistical model: cluster-period mean level

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

$$Y_{skt} = \boldsymbol{Z}_{st}\boldsymbol{\beta} + X_{st}\theta + CP_{skt} + e_{skt}$$

| Mean measured outcome* | Time period effect | Treatment effect | Cluster-period effect | Error |

*Averaged over all participants' outcomes in the cluster-period

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

$$e_{skt} \sim N\left(0, \frac{\sigma_e^2}{m}\right)$$

Statistical model: correlation b/w cluster-period means

$$corr(Y_{skt}, Y_{skt'}) = \frac{\sigma_C^2 r}{\sigma_C^2 + \sigma_e^2/m} = \frac{m\rho}{1+(m-1)\rho} r = \psi, \quad 0 \leq \psi \leq 1$$

Cluster-period size	Within-period ICC	Cluster autocorrelation

Correlation matrix for a cluster:
$$\begin{bmatrix} 1 & \psi & \cdots & \psi \\ \psi & 1 & . & \vdots \\ \vdots & . & \ddots & \psi \\ \psi & \cdots & \psi & 1 \end{bmatrix}$$

Correlation between cluster-period means

- $\psi = \dfrac{m\rho}{1+(m-1)\rho}\, r$ increases with increasing
    - Cluster-period size, m (measuring more participants)
    - Within-period ICC, ρ (greater similarity between participants)
    - Cluster autocorrelation, r (less waning of similarity over time)
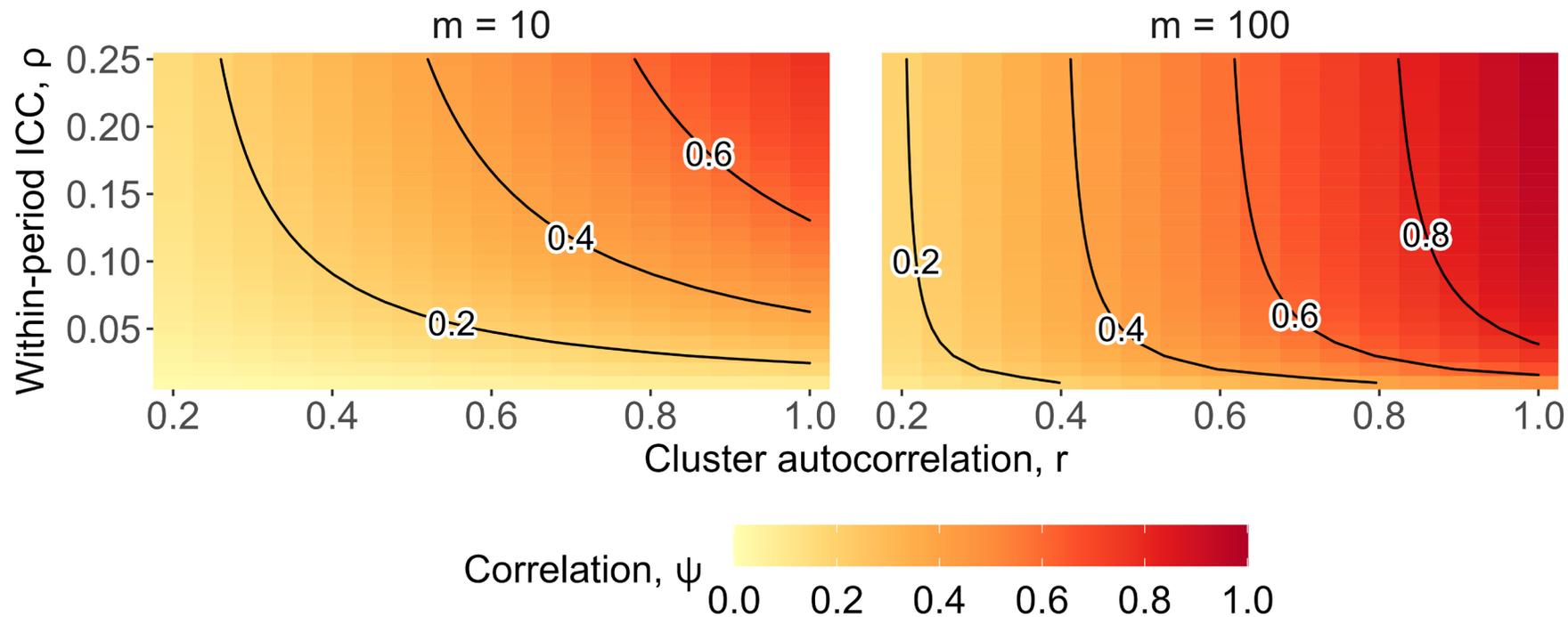
General model: cluster-period mean level

$$Y_{skt} \quad = \quad \mathbf{Z}_{st}\boldsymbol{\beta} \quad + \quad X_{st}\theta \quad + \quad CP_{skt} \quad + \quad e_{skt}$$

Mean measured outcome	Time period effect	Treatment effect	Cluster-period effect	Error

$$corr(Y_{skt}, Y_{skt'}) = \frac{m\rho}{1+(m-1)\rho}r = \psi, \quad 0 \leq \psi \leq 1$$

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

Variance of the treatment effect estimator
Staircase designs

$$var(\hat{\theta})_{\text{SC}} = \frac{1}{SK}\left[X^T V_*^{-1} X - \frac{1}{S} X^T V_*^{-1} \sum_{s=1}^{S} Z_s \left(\sum_{s=1}^{S} Z_s^T V_*^{-1} Z_s \right)^{-1} \sum_{s=1}^{S} Z_s^T V_*^{-1} X \right]^{-1}$$

Number of clusters per sequence

Treatment sequence

Covariance matrix for a cluster

Time period design matrix

Variance of the treatment effect estimator (cont'd)
Basic staircase designs, categorical period effects

$$var(\hat{\theta})_{SC,cat} = \frac{2var(Y_{skt})(1-\psi)^2}{K\left[S(1-\psi) - \sqrt{1-\psi^2}\frac{\left(1+\sqrt{1-\psi^2}\right)^S - \psi^S}{\left(1+\sqrt{1-\psi^2}\right)^S + \psi^S}\right]}$$

where $var(Y_{skt}) = \frac{1+(m-1)\rho}{m}$ and $\psi = \frac{m\rho}{1+(m-1)\rho}r$

Within-period ICC, ρ	↑	Variance	↑ *	*most* of the time!
Cluster autocorrelation, r	↑	Variance	↓	less correlation decay
Cluster-period size,	↑	Variance	↓	but tapers off quickly

How efficient are staircase designs compared to stepped wedge designs?

Grantham KL, Forbes AB, Hooper R, Kasza J. The relative efficiency of staircase and stepped wedge cluster randomised trial designs. doi:10.26180/24636114.v1.

Design comparisons

Stepped wedge
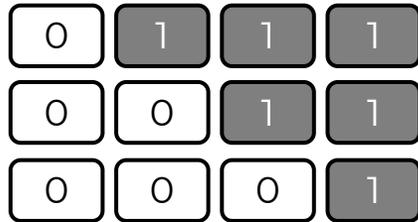
0	1	1	1
0	0	1	1
0	0	0	1

- S sequences
- K clusters per sequence
- m participants per cluster-period
- $S(S+1)Km$ total participants

Design comparisons (cont'd)

Stepped wedge

0	1	1	1
0	0	1	1
0	0	0	1

Embedded staircase

0	1		
	0	1	
		0	1

- S sequences
- K clusters per sequence
- m participants per cluster-period
- $S(S+1)Km$ total participants

- S sequences
- K clusters per sequence
- m participants per cluster-period
- $2SKm$ total participants

Design comparisons (cont'd)

Stepped wedge
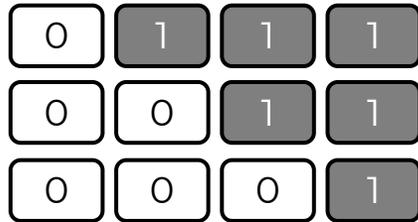
Embedded staircase
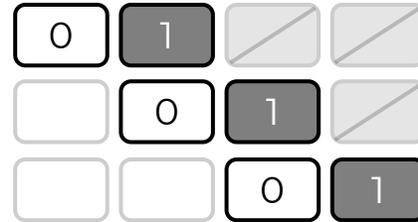
Extended staircase

- S sequences
- K clusters per sequence
- m participants per cluster-period
- $S(S+1)Km$ total participants

- S sequences
- K clusters per sequence
- m participants per cluster-period
- $2SKm$ total participants
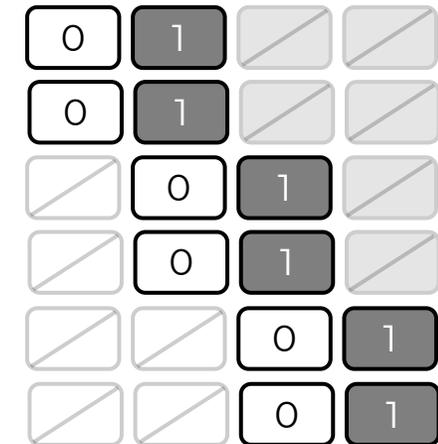
- S sequences
- $\frac{(S+1)K}{2}$ clusters per sequence
- m participants per cluster-period
- $S(S+1)Km$ total participants

Relative efficiency
Stepped wedge vs. staircase

$$\text{Relative efficiency} := \frac{prec(\hat{\theta})_{SC}}{prec(\hat{\theta})_{SW}}$$

$> 1 \longrightarrow$ Staircase is more precise than stepped wedge

Relative efficiency (cont'd)
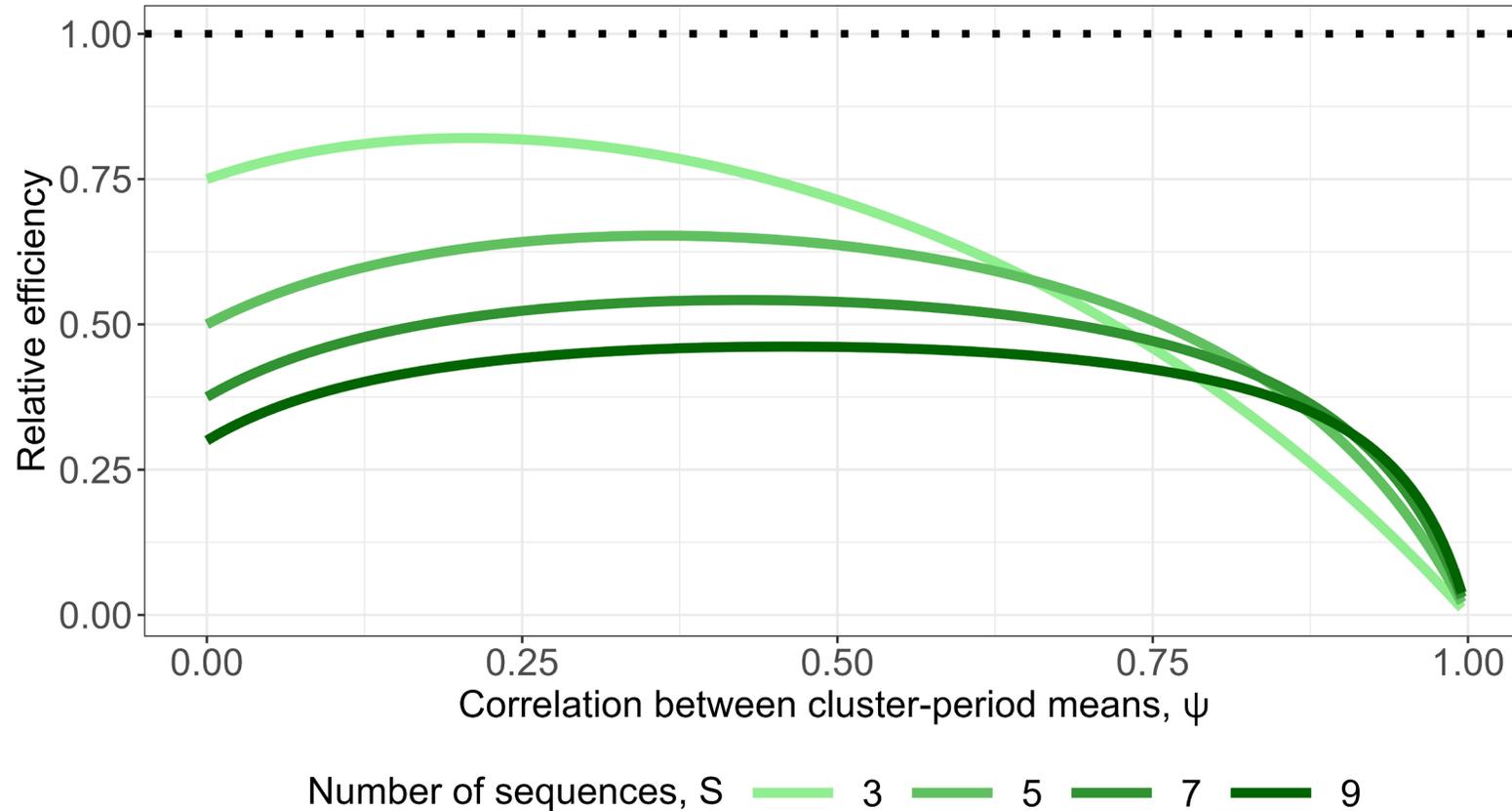
Stepped wedge vs. embedded staircase

Relative efficiency (cont'd)
Stepped wedge vs. extended staircase

Intuition behind relative efficiency results

- The treatment effect estimators (and their precision) for stepped wedge and embedded basic staircase designs are most similar when ψ is small (i.e. small m, ρ, r)

  - Matthews and Forbes (2017) showed that the stepped wedge estimator takes the form of a:

    - "vertical" estimator (comparisons within columns/periods) when ψ and S are small
    - "horizontal" estimator (comparisons within rows/clusters) when ψ and S are large

  - Grantham et al (2024) showed that the basic staircase estimator resembles the stepped wedge "vertical" estimator

Matthews JNS and Forbes AB. Stepped wedge designs: insights from a design of experiments perspective. *Statistics in Medicine* 2017; 36(24):3772-3790.

Which staircase design variants are most efficient?

Which staircase variants are most efficient?

- In which trial settings are particular staircase design variants a good choice?

- Develop guidance to choose between different staircase design variants:
  - Unequal versus equal allocation of clusters to sequences
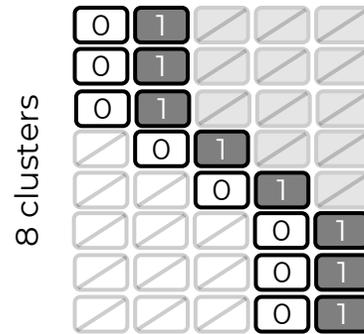
Unequal allocation
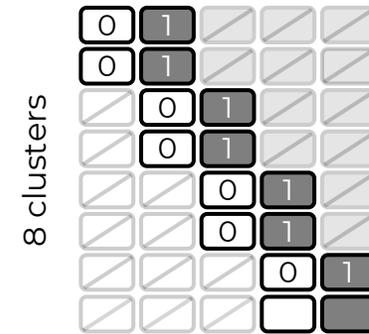
Equal allocation

  - Imbalanced versus balanced staircase designs
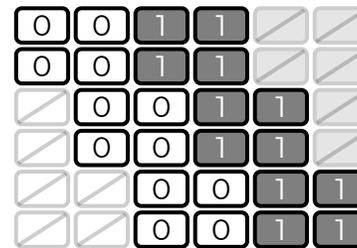
Balanced staircase
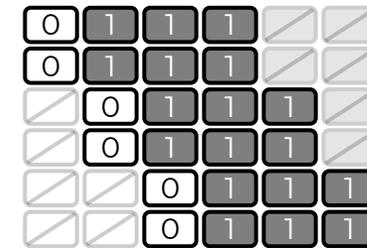
Imbalanced staircase

Optimal allocation of clusters to sequences
4-sequence basic staircase design

N clusters $\times$

p_1 $= K_1$ clusters to sequence 1:

p_2 $= K_2$ clusters to sequence 2:

p_2 $= K_2$ clusters to sequence 3:

p_1 $= K_1$ clusters to sequence 4:

Optimal allocation of clusters to sequences (cont'd)
4-sequence basic staircase design + Constraints

N clusters ×

p_1 $\quad$ = K_1 clusters to sequence 1:

p_2 $\quad$ = K_2 clusters to sequence 2:

p_2 $\quad$ = K_2 clusters to sequence 3:

p_1 $\quad$ = K_1 clusters to sequence 4:

where

- $2p_1 + 2p_2 = 1$

- $p_1, p_2 \geq \dfrac{1}{N}$ $\longrightarrow$ At least one cluster per sequence

Optimal allocation of clusters to sequences (cont'd)
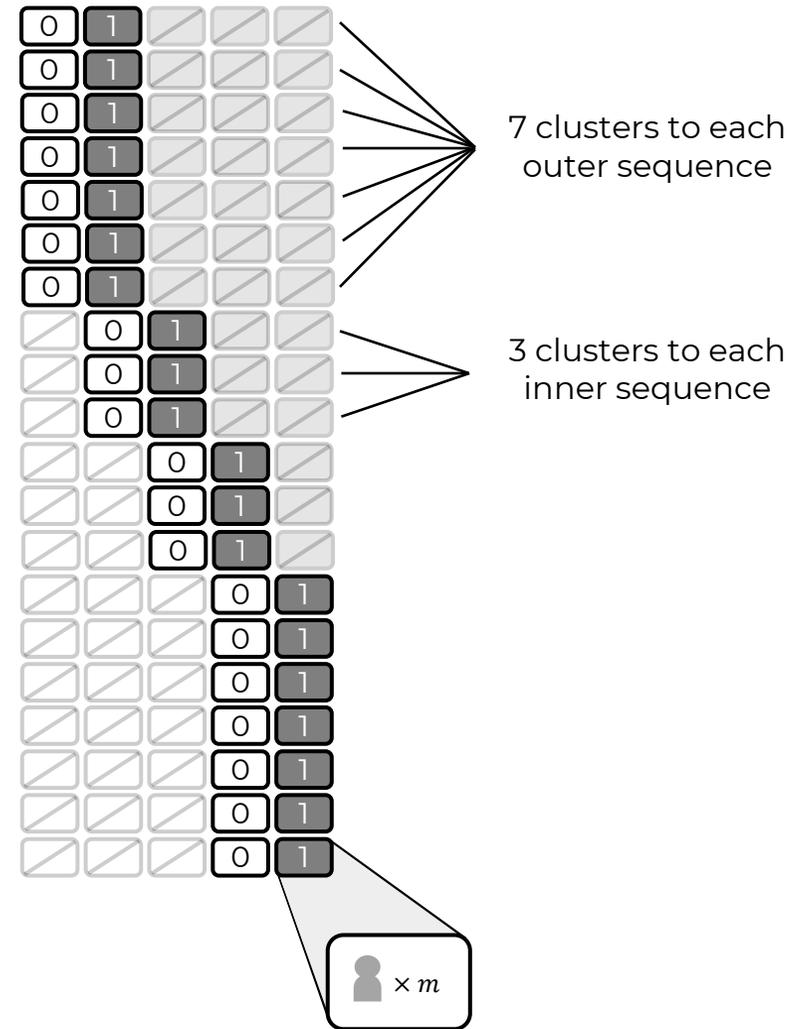## 4-sequence basic staircase design + Constraints (cont'd)

N clusters ×

p_1 $= K_1$ clusters to sequence 1:

p_2 $= K_2$ clusters to sequence 2:

p_2 $= K_2$ clusters to sequence 3:

p_1 $= K_1$ clusters to sequence 4:

where

- $2p_1 + 2p_2 = 1$

- $p_1, p_2 \geq \dfrac{1}{N}$ ⟶   At least one cluster per sequence

Goal: Find p_1 and p_2 that give the most precise estimate of the treatment effect (maximize $prec(\hat{\theta})$ subject to constraints)

Optimal allocation of clusters to sequences (cont'd)
4-sequence basic staircase, assuming categorical period effects

- Optimal proportions depend on ψ
  - Small ψ: Larger proportion of clusters allocated to inner sequences
  - Large ψ: Larger proportion of clusters allocated to outer sequences
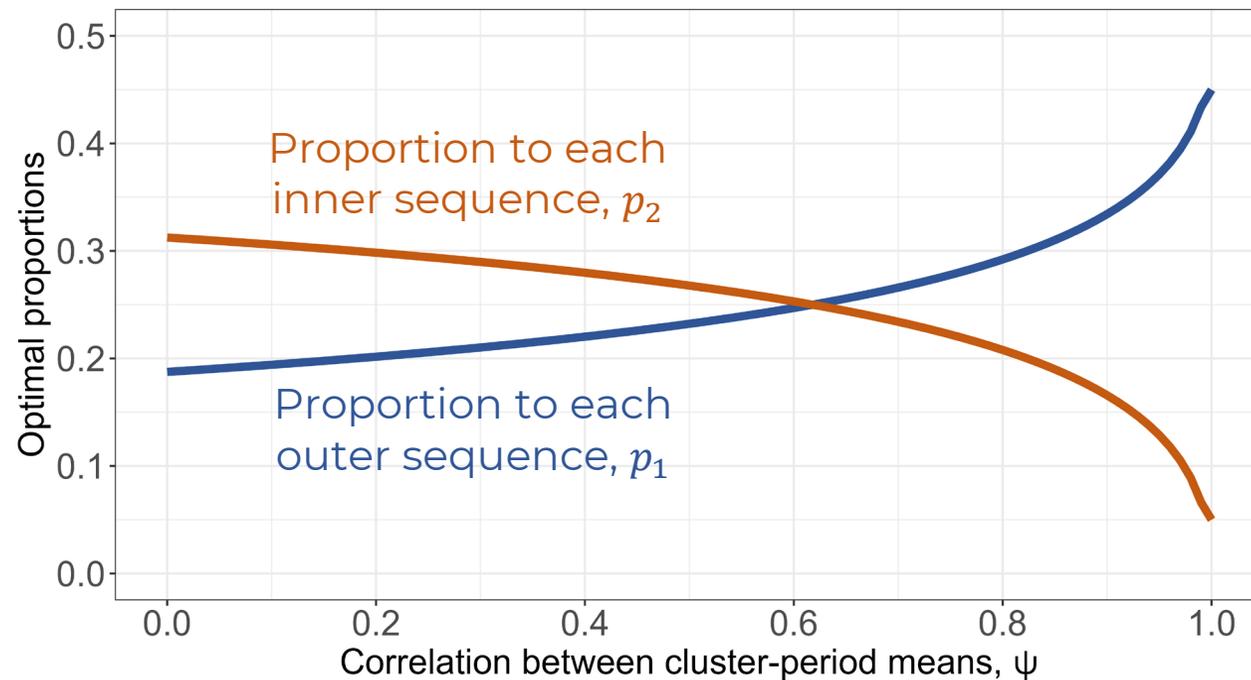
Optimal allocation of clusters to sequences (cont'd)
4-sequence basic staircase, assuming categorical period effects (cont'd)

- $N = 20$ clusters, small ψ:
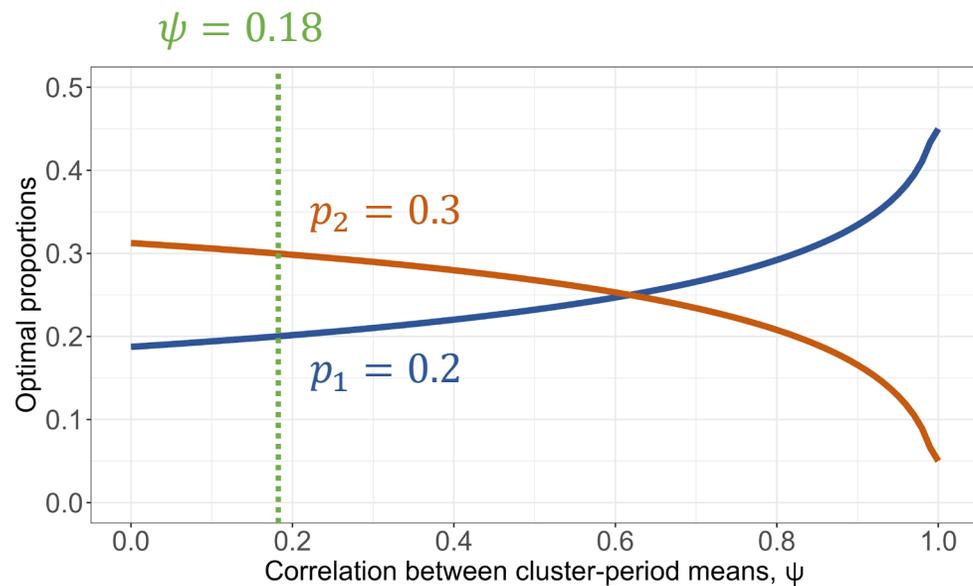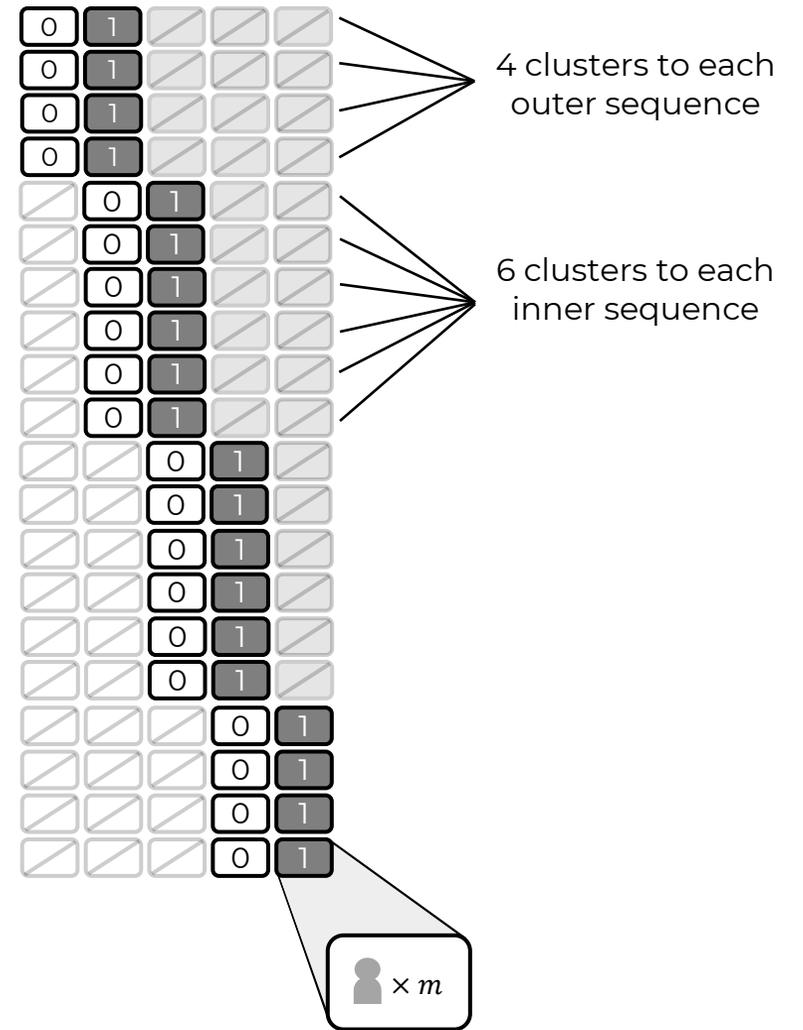
4 clusters to each outer sequence

6 clusters to each inner sequence

Optimal design

$\times m$

Optimal allocation of clusters to sequences (cont'd)

4-sequence basic staircase, assuming categorical period effects (cont'd)

- $N = 20$ clusters, large ψ:

7 clusters to each outer sequence

3 clusters to each inner sequence

Optimal design

$\psi = 0.93$

$p_1 = 0.35$

$p_2 = 0.15$

$\times m$
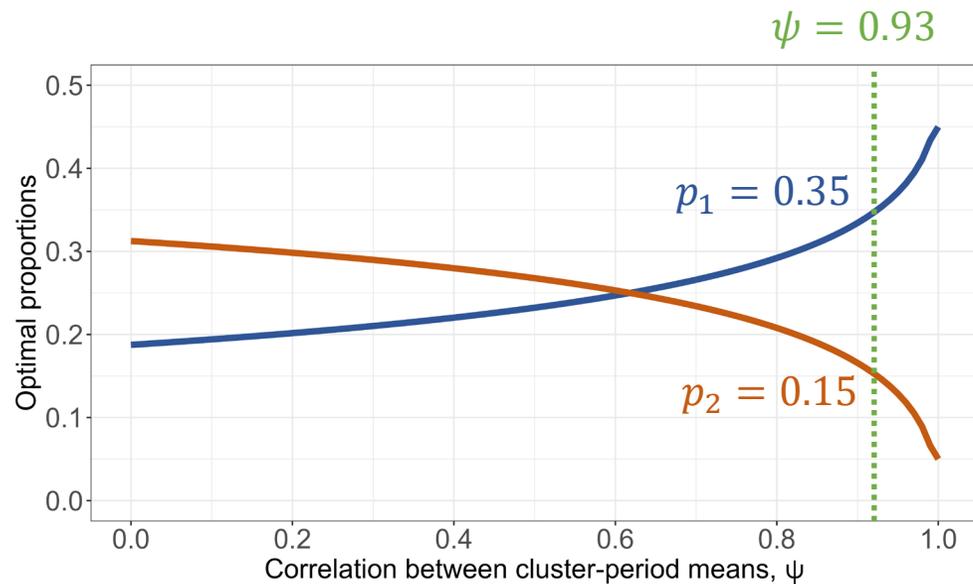
Relative efficiency
Optimal vs. equal allocation

Relative efficiency $:= \dfrac{prec(\hat{\theta})_{\text{optimal}}}{prec(\hat{\theta})_{\text{equal}}}$

> 1 $\longrightarrow$ Optimal allocation design is more precise than equal allocation design

Relative efficiency: Optimal vs. equal allocation design
4-sequence basic staircase, assuming categorical period effects
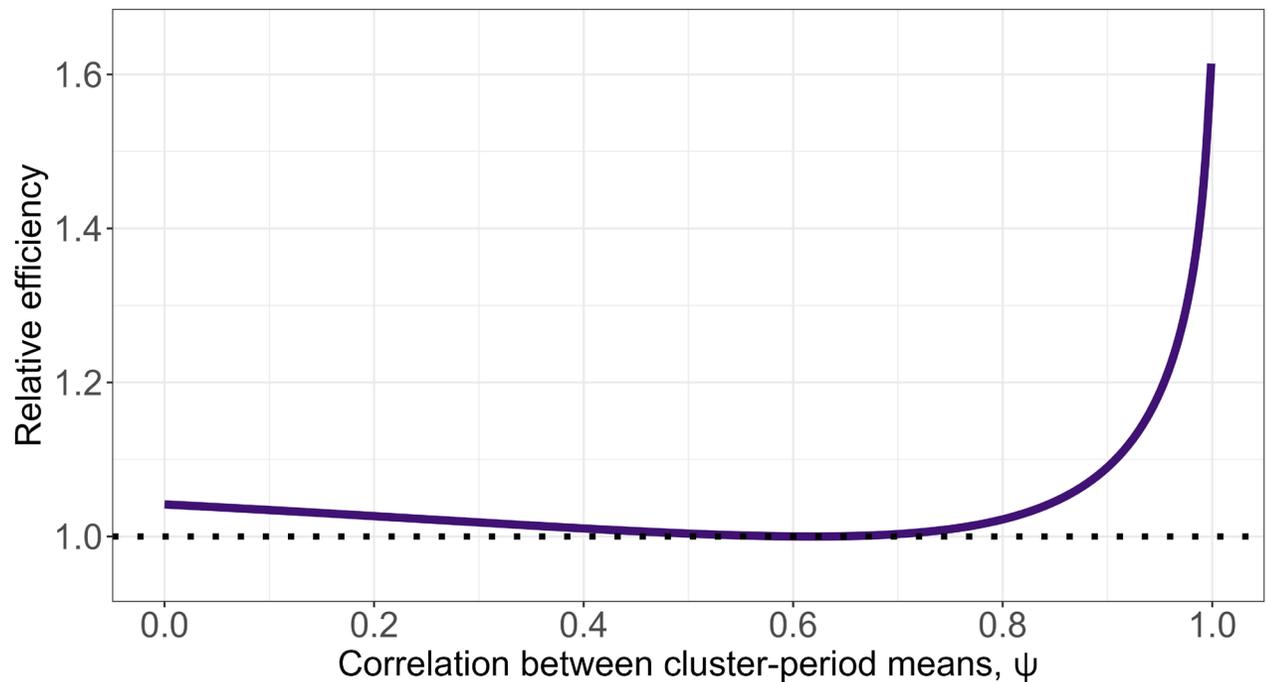
Optimal design is slightly more precise

Optimal design is much more precise

Optimal 4-sequence basic staircase
Assuming categorical period effects

- Recall: $\psi = \frac{m\rho}{1+(m-1)\rho}r$ is a function of cluster-period size, m, within-period ICC, ρ, and cluster autocorrelation, r
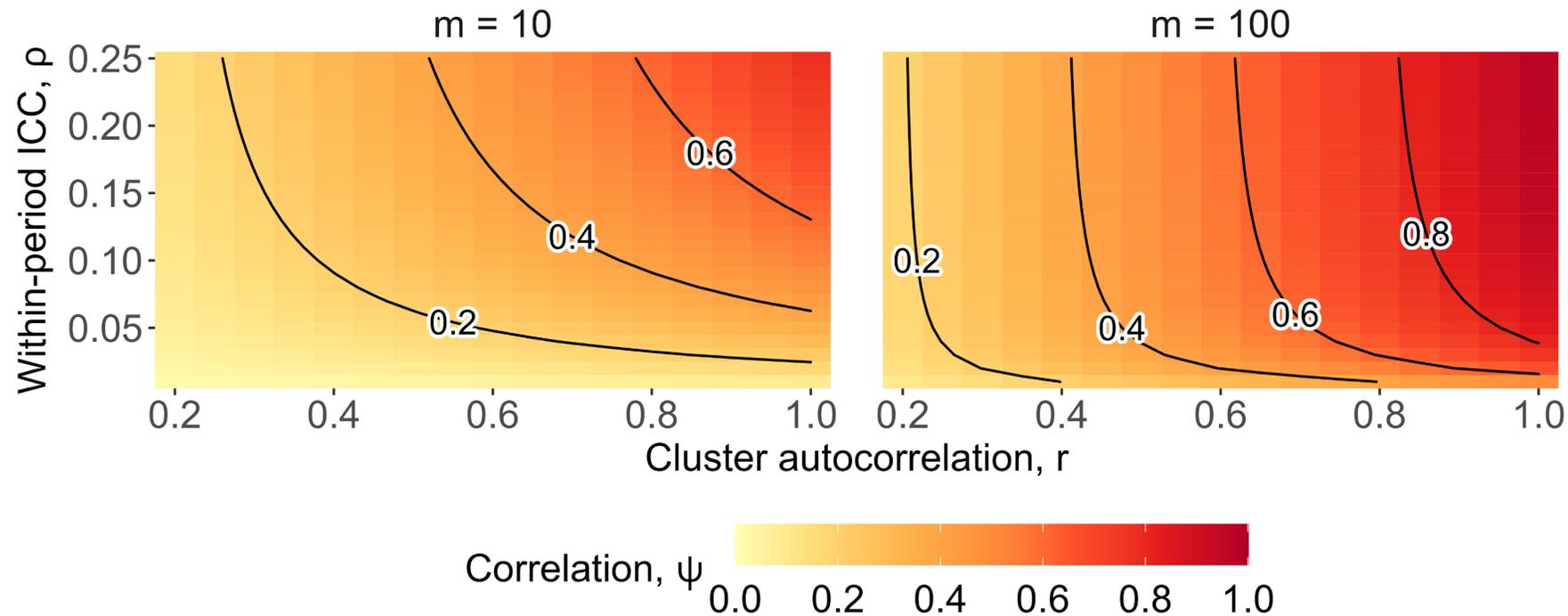
Optimal 4-sequence basic staircase (cont'd)
Assuming categorical period effects (cont'd)

- Recall: $\psi = \frac{m\rho}{1+(m-1)\rho}r$ is a function of cluster-period size, m, within-period ICC, ρ, and cluster autocorrelation, r

Large m, large r, moderate $\rho \rightarrow$ large ψ
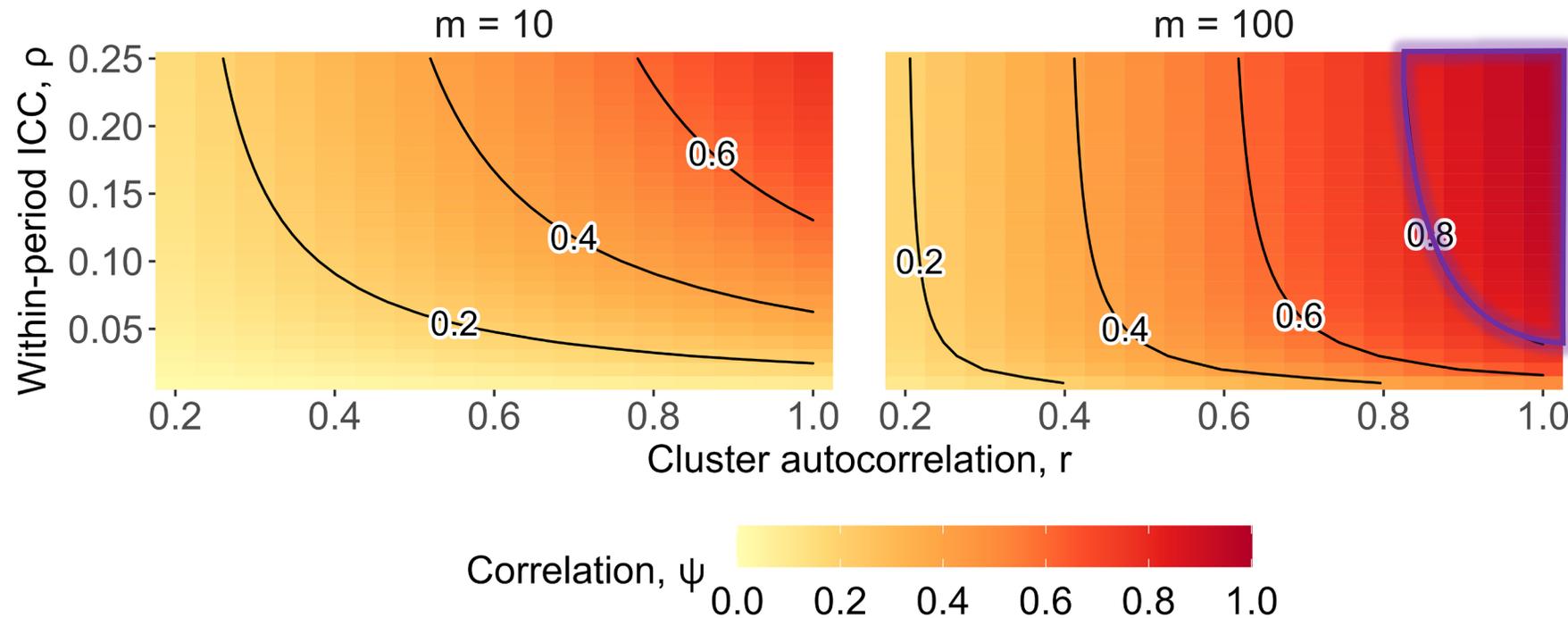
Optimal 6-sequence basic staircase
Assuming categorical period effects
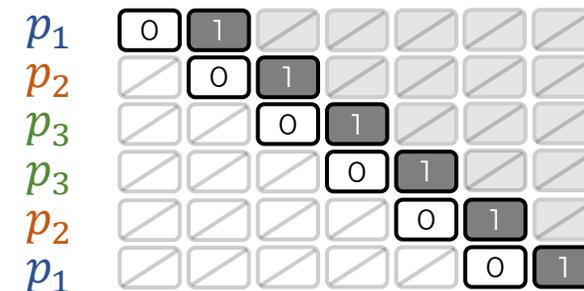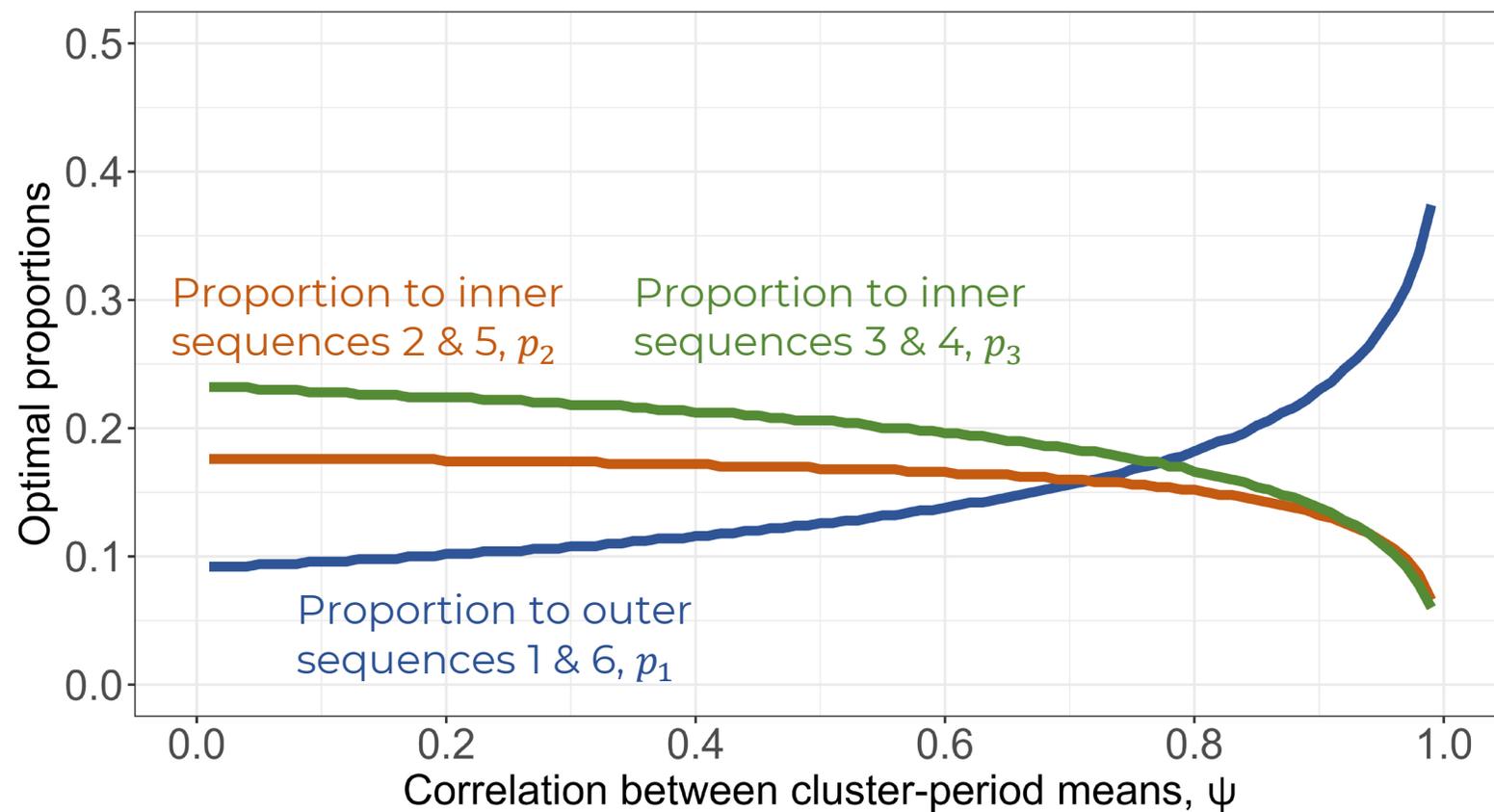
Which staircase variants are most efficient?

- In which trial settings are particular staircase design variants a good choice?

- Develop guidance to choose between different staircase design variants:
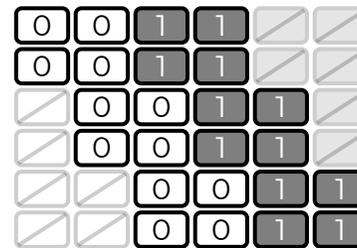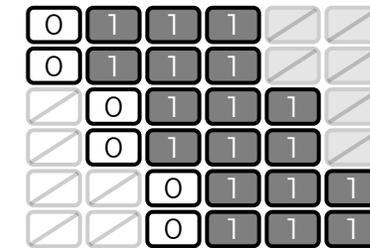 - Unequal versus equal allocation of clusters to sequences
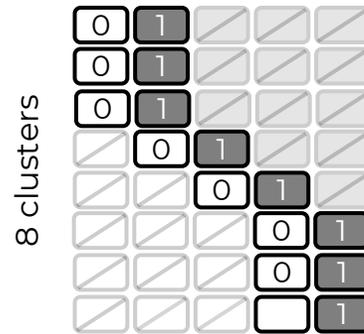
Unequal allocation
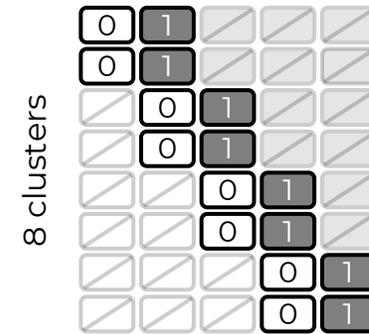
Equal allocation

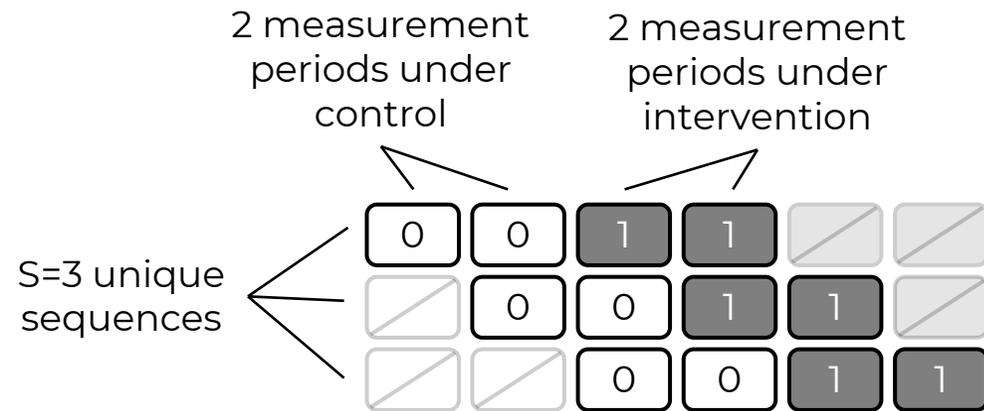 - Imbalanced versus balanced staircase designs

Balanced staircase

Imbalanced staircase

Imbalanced vs. balanced designs
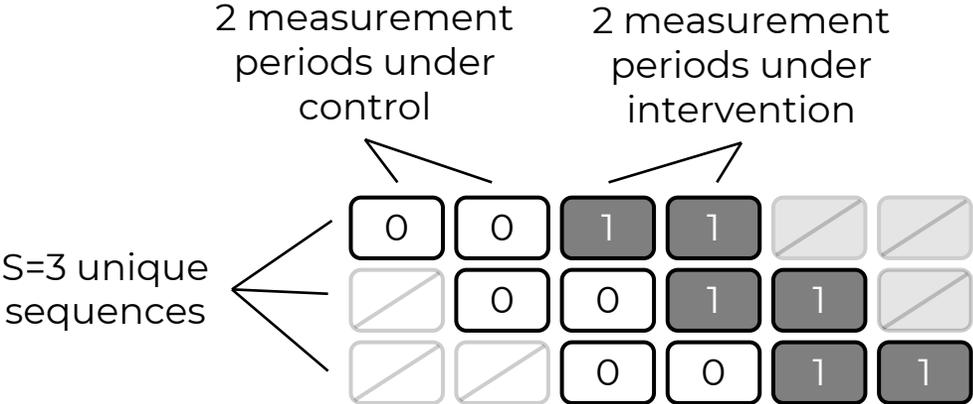## 3 sequences, 4 measurement periods in each sequence

Balanced staircase design

2 measurement periods under control

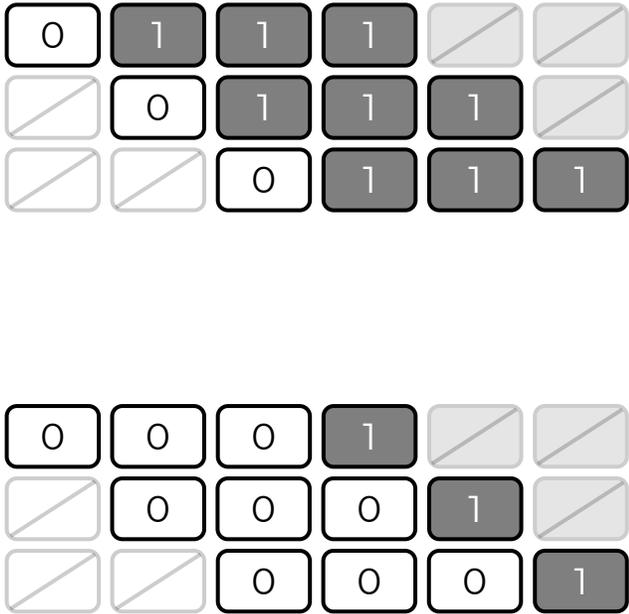2 measurement periods under intervention

S=3 unique sequences

Imbalanced vs. balanced designs (cont'd)
3 sequences, 4 measurement periods in each sequence (cont'd)

Balanced staircase design

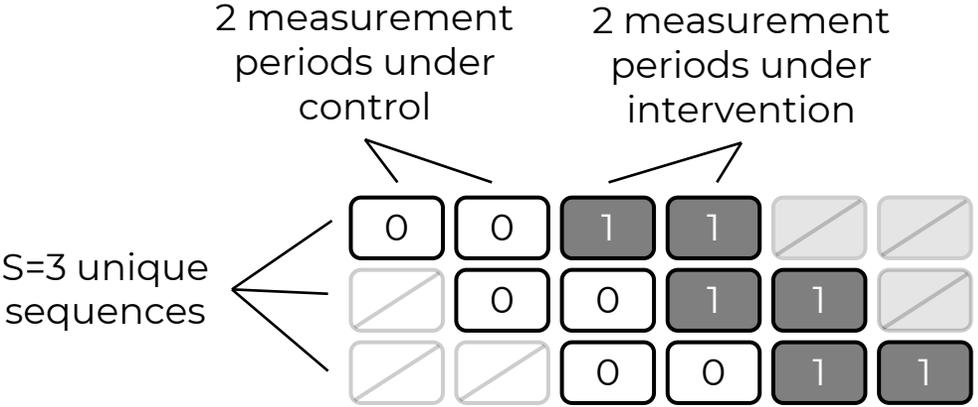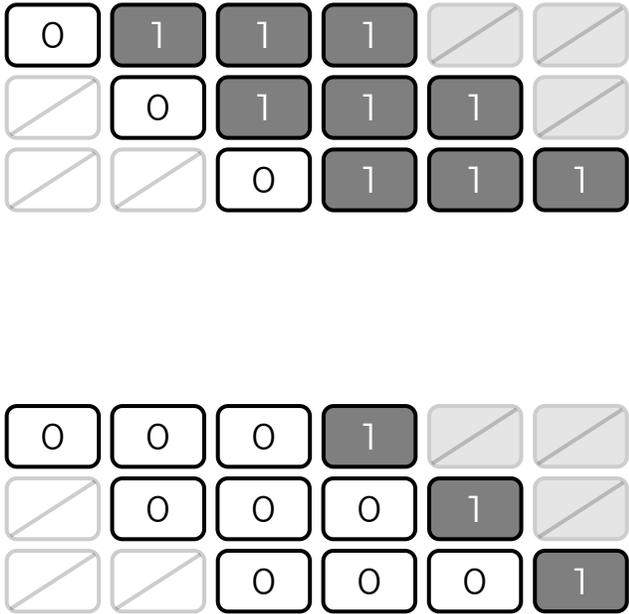Imbalanced staircase designs

2 measurement periods under control

2 measurement periods under intervention

S=3 unique sequences

Imbalanced vs. balanced designs (cont'd)
3 sequences, 4 measurement periods in each sequence (cont'd)

Balanced staircase design

2 measurement periods under control

2 measurement periods under intervention

S=3 unique sequences

Imbalanced staircase designs

Equally precise
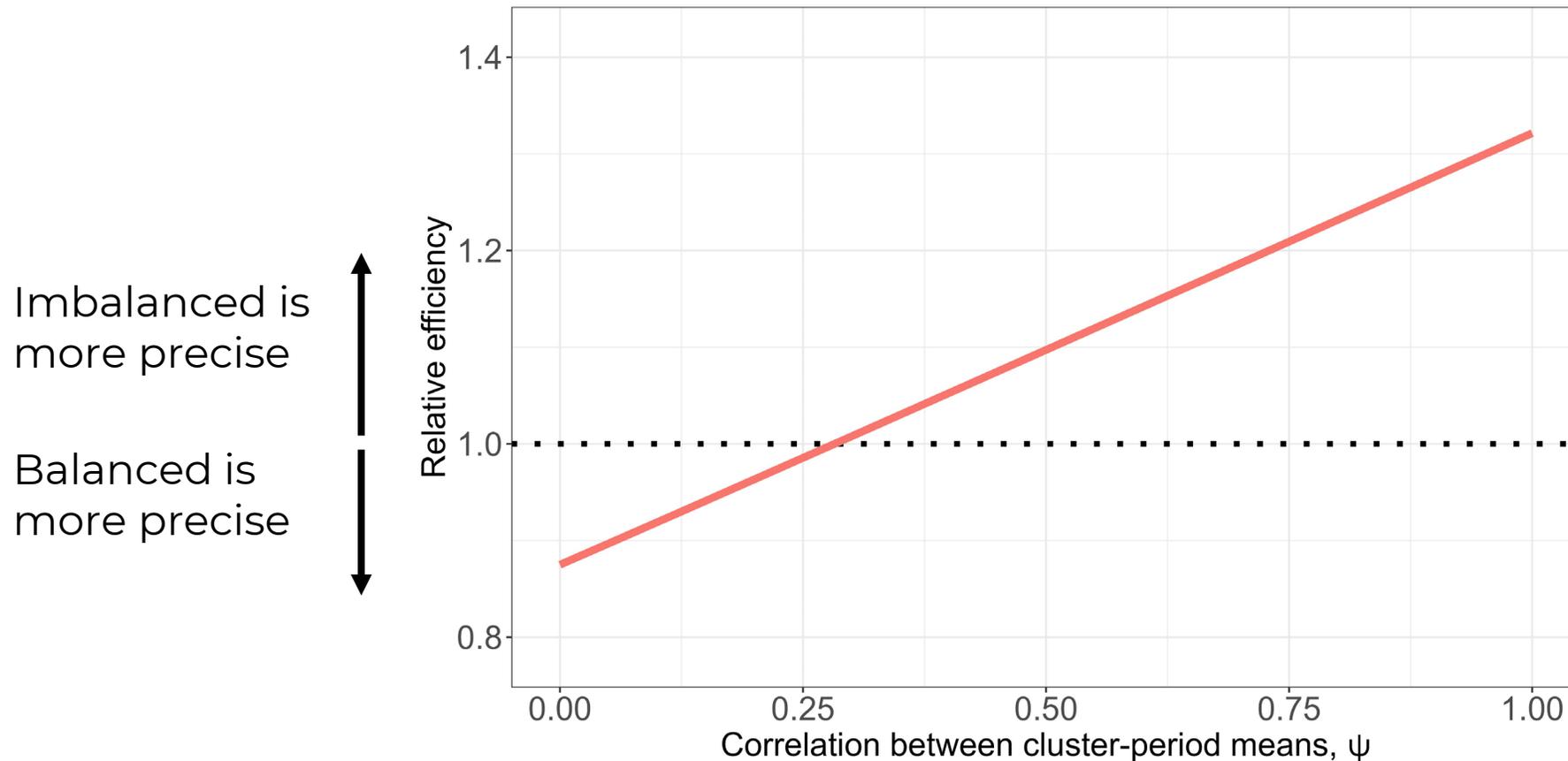
Relative efficiency: Imbalanced vs. balanced

Relative efficiency $:= \dfrac{prec(\hat{\theta})_{\text{imbalanced}}}{prec(\hat{\theta})_{\text{balanced}}}$

$> 1 \quad \longrightarrow \quad$ Imbalanced staircase is more precise than balanced staircase design

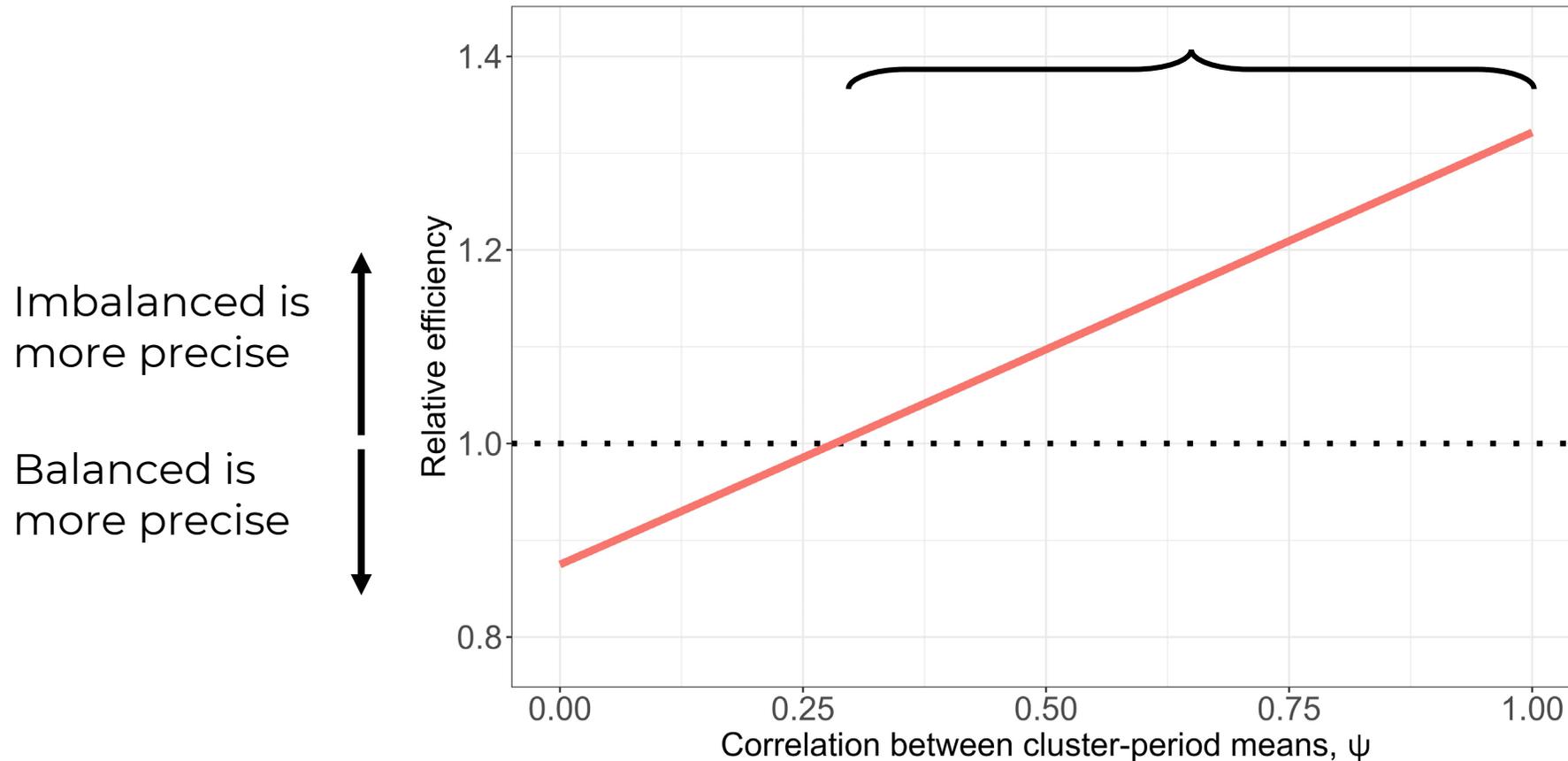# Relative efficiency: Imbalanced vs. balanced (cont'd)
3 sequences, 4 measurement periods in each sequence

Relative efficiency: Imbalanced vs. balanced (cont'd)
3 sequences, 4 measurement periods in each sequence (cont'd)
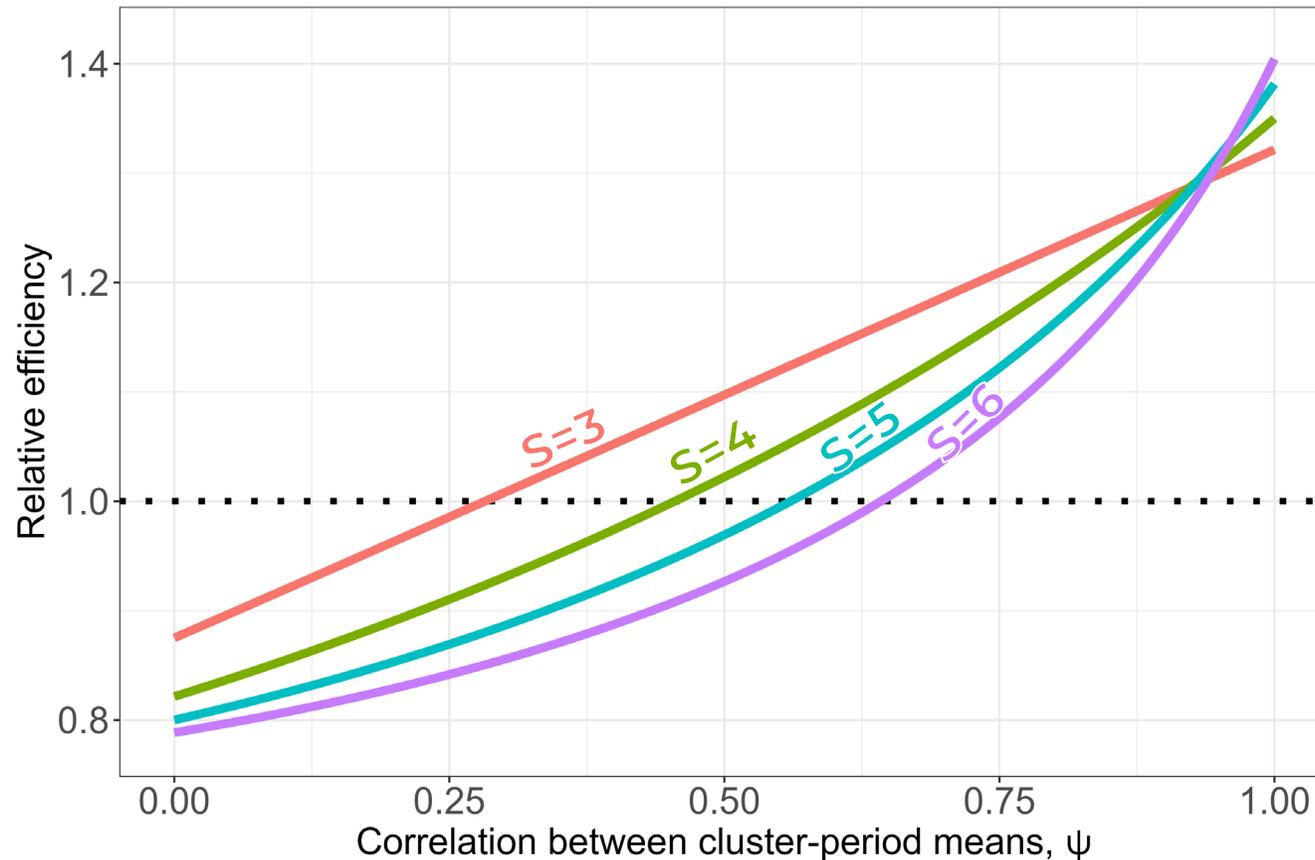
Imbalanced more precise
than balanced when: $0.282 < \psi$

Imbalanced is
more precise

Balanced is
more precise

Relative efficiency

Correlation between cluster-period means, ψ

Relative efficiency: Imbalanced vs. balanced (cont'd)
S sequences, 4 measurement periods in each sequence

Relative efficiency: Imbalanced vs. balanced (cont'd)
3 sequences, 4 measurement periods in each sequence

Relative efficiency: Imbalanced vs. balanced (cont'd)
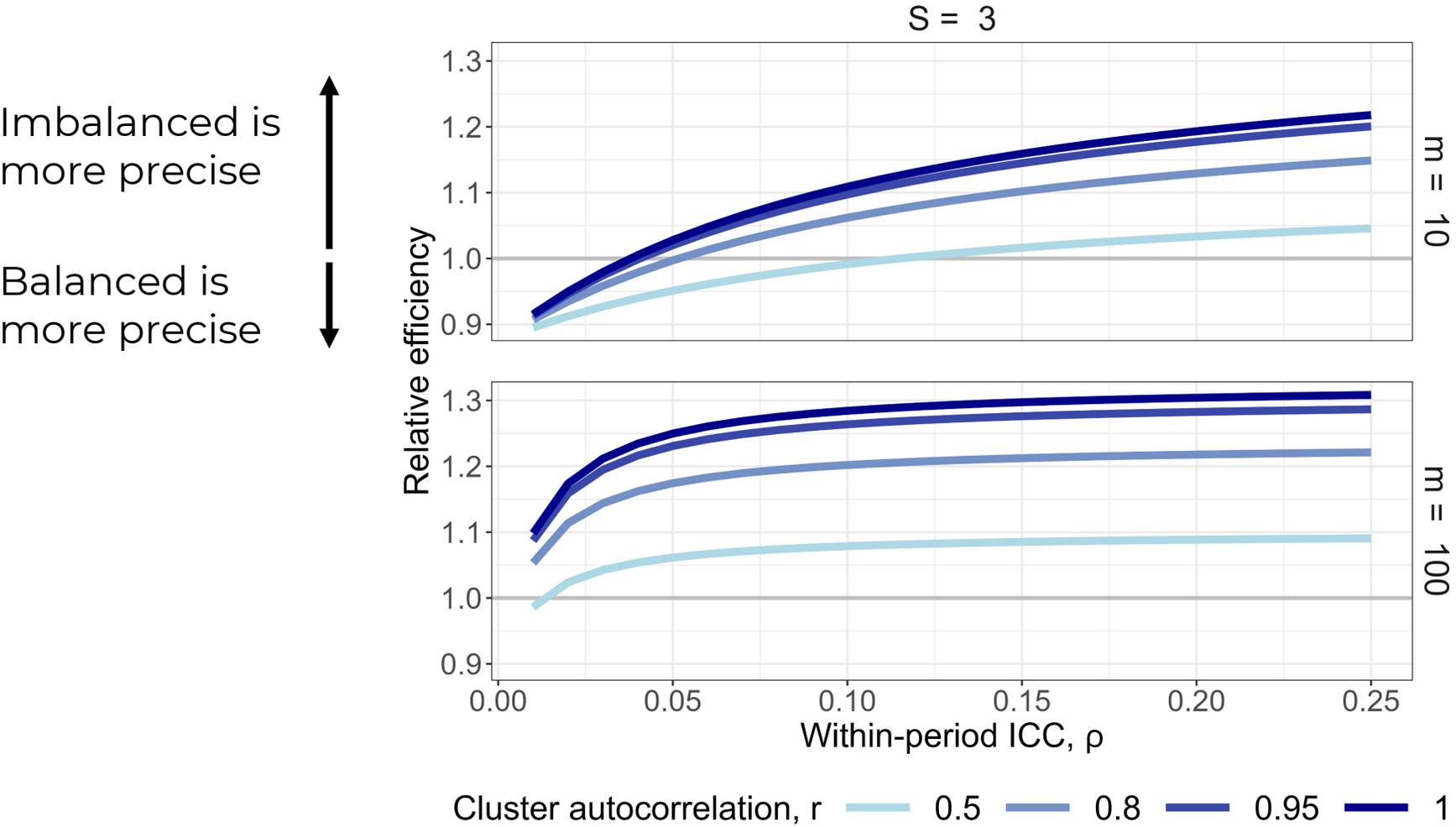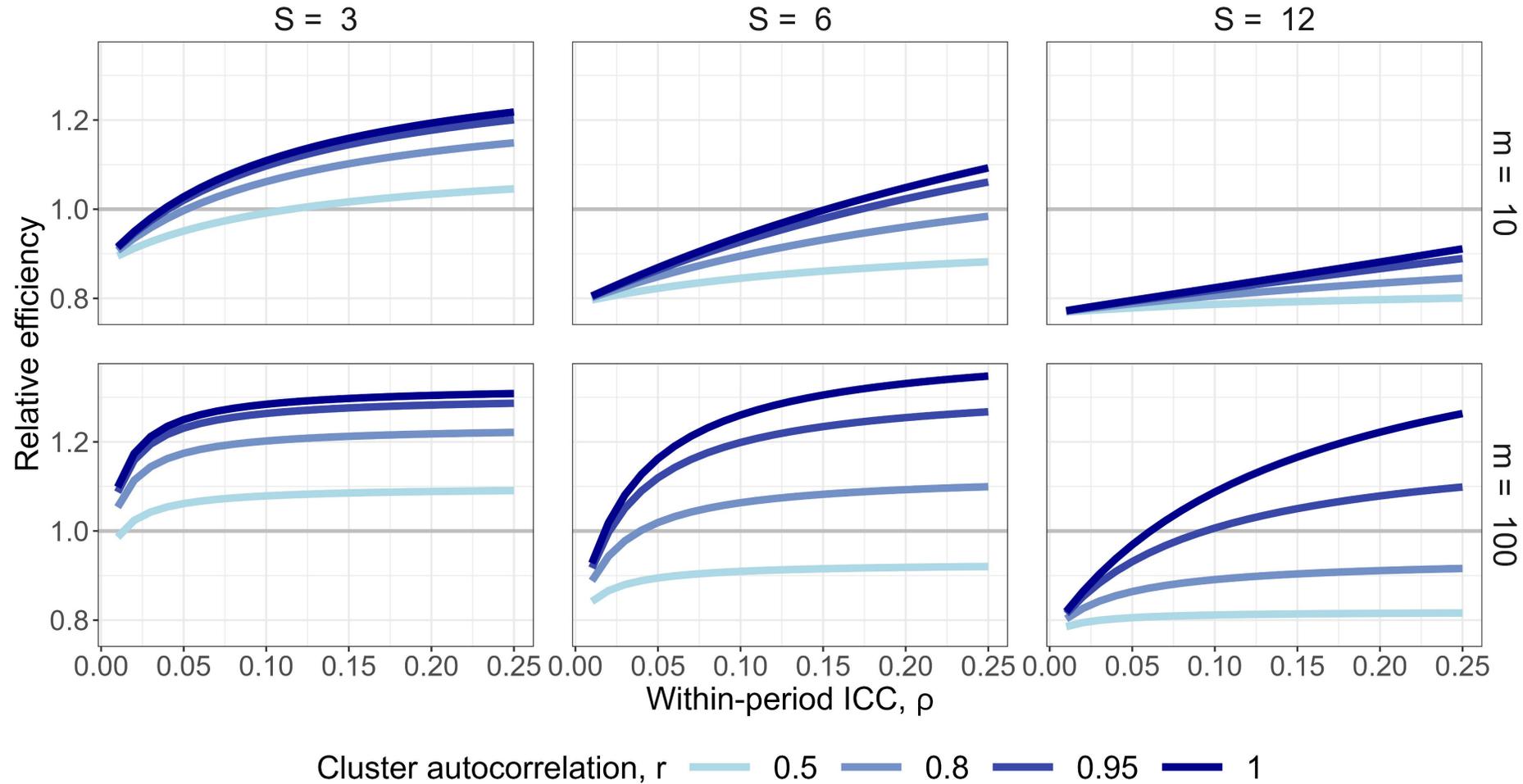## S sequences, 4 measurement periods in each sequence

Overall summary

- Staircase designs are pragmatic and less burdensome alternatives to stepped wedge designs

- Underpinning statistical theory supports accurate sample size and power calculations

- There is no definitive "best" staircase design: design efficiency depends heavily on the trial setting
  - For *certain* trial settings, basic staircase designs can be more efficient than stepped wedge designs, but some rearrangement of measurements is required
  - Optimal unequal allocation of clusters is much more efficient than equal allocation in some settings, but only slightly more efficient in others
  - Balanced staircase designs are more efficient in some settings, imbalanced designs are more efficient in others

Design efficiency summary

Trial setting		
	Small ψ	**Large ψ**
Cluster-period sizes	Smaller	Larger
Similarity between participant outcomes	Less similarity	More similarity
Waning of similarity over time	More waning	Less waning
Design choice: Which is more efficient?		
Embedded basic staircase vs. stepped wedge	Stepped wedge	Stepped wedge
Extended basic staircase vs. stepped wedge	Staircase	Stepped wedge
Basic staircase with optimal unequal vs. equal allocation of clusters	Unequal allocation with more clusters to inner sequences	Unequal allocation with more clusters to outer sequences
Balanced vs. imbalanced staircase	Balanced staircase	Imbalanced staircase

Ongoing and future work on staircase designs

- Cost-efficiency of incomplete stepped wedge designs, including staircase designs

- Robustness of inference for the treatment effect to misspecification of the within-cluster correlation structure and to cluster dropout

- Inference with small number of clusters

- Batched staircase designs

Acknowledgments

Andrew Forbes, Monash University, Australia

Richard Hooper, Queen Mary University of London, UK

Jessica Kasza, Monash University, Australia

Questions?

kelsey.grantham@monash.edu

Methods: Mind the Gap Webinar Series

- Updates on Research Methods Resources (researchmethodsresources.nih.gov)
  - Time-varying intervention effects in Stepped Wedge Designs
  - Group Regression Discontinuity Design section with background, key references, and a sample size calculator.