

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



MIND THE GAP NIH WEBINAR SERIES

Stepped Wedge Cluster Randomized Designs for Disease Prevention Research

Presented by

Dr. Monica Taljaard, Ph.D.

Ottawa Hospital Research Institute

MEDICINE: MIND THE GAP
An NIH Office of Disease Prevention Webinar Series



OUTLINE

1. Refresher: Cluster randomized trials (CRTs)
2. What is a stepped wedge cluster randomized trial (SW-CRT)?
3. Key methodological considerations in SW-CRTs
4. Analysis of the SW-CRT
5. Sample size calculation for the SW-CRT
6. Advantages and disadvantages of the SW-CRT

CLUSTER RANDOMIZED TRIALS

- ▶ What is a cluster randomized trial (CRT)?
 - Units of randomization are intact groups (“clusters”) rather than separate individuals
 - Outcomes are observed on multiple individuals within each cluster
- ▶ Key characteristics:
 - Multiple observations from the same cluster usually positively correlated
 - The strength of the correlation can be measured by the Intracluster Correlation Coefficient (ICC)
 - Essential to account for intracluster correlation in both the sample size calculation and analysis

A DEFINITION OF ICC

- ▶ Assume the outcome Y is continuous with variance σ^2
- ▶ The variance σ^2 may be expressed as the sum of two components:

$$\sigma^2 = \sigma_b^2 + \sigma_w^2$$

where

σ_b^2 = between-cluster variance

σ_w^2 = within-cluster variance

- ▶ Then the ICC is defined as

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}; \quad 0 \leq \rho \leq 1$$

QUANTIFYING THE EFFECTS OF CLUSTERING

- ▶ In a standard clinical trial with n individuals randomized to each arm, we have:

$$\text{Var}(\bar{Y}_i) = \frac{\sigma^2}{n}, \quad i = 1, 2$$

- ▶ In a CRT with $n=km$ individuals per arm (where k = number of clusters, and m =number of individuals per cluster), we have:

$$\text{Var}(\bar{Y}_i) = \frac{\sigma^2}{km} [1 + (m-1)\rho]$$

- ▶ The variance inflation factor $1+(m-1)\rho$ is called the “**Design Effect**”
- ▶ Sample size for a CRT may be obtained by multiplying n under individual randomization by the Design Effect (+ any necessary small sample correction)

WHAT IS A STEPPED WEDGE CRT?

- ▶ The stepped wedge cluster randomized trial (SW-CRT)
 - A relatively new type of CRT design
 - Commonly used to evaluate public health, health system and service delivery interventions
 - Many different design variations!
 - Rapid increase in popularity but needs substantial methodological expansion in the next few years

Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*. 2015 Feb 6;350:h391

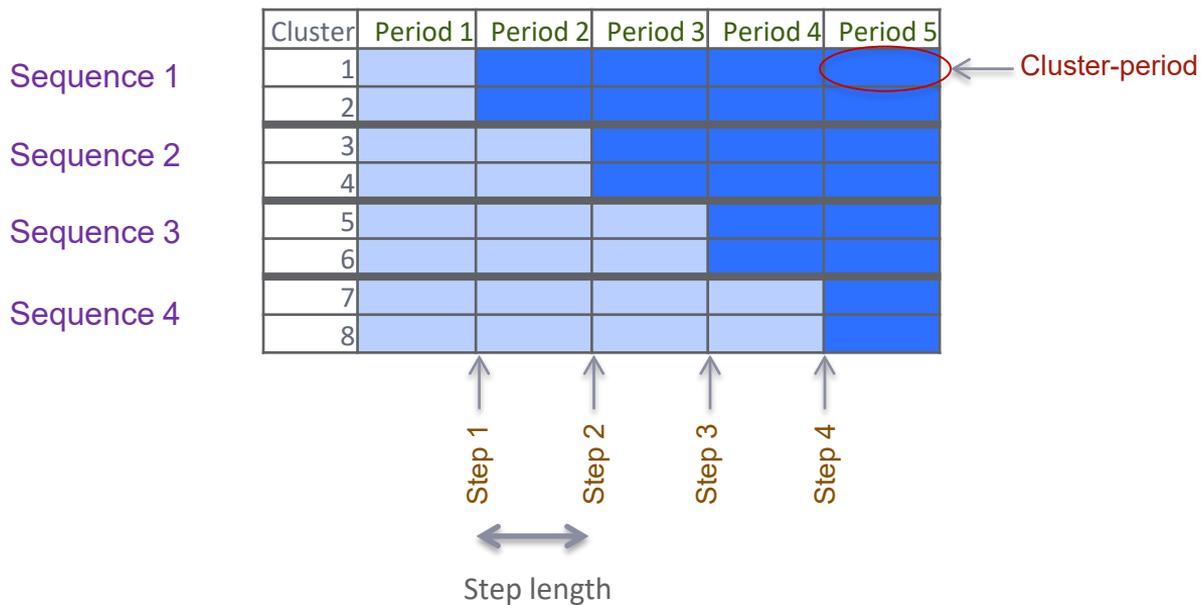
THE “BASIC” STEPPED WEDGE DESIGN



- ▶ Clusters start in control and end in intervention condition
- ▶ (Groups of) clusters cross to intervention sequentially and *in random order*
- ▶ Outcomes are assessed repeatedly in each cluster over time

TERMINOLOGY

- ▶ The basic stepped wedge design

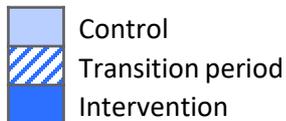


BASIC STEPPED WEDGE WITH TRANSITION PERIOD

- ▶ May need a transition period to allow intervention to be fully implemented (or be able to affect outcomes)

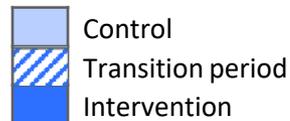
Preferable

Cluster	Period 1	Period 2	Period 3	Period 4	Period 5
1	Control	Transition period	Intervention	Intervention	Intervention
2	Control	Transition period	Intervention	Intervention	Intervention
3	Control	Control	Transition period	Intervention	Intervention
4	Control	Control	Transition period	Intervention	Intervention
5	Control	Control	Control	Transition period	Intervention
6	Control	Control	Control	Transition period	Intervention
7	Control	Control	Control	Control	Transition period
8	Control	Control	Control	Control	Transition period



Not ideal

Cluster	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6
1	Control	Transition period	Intervention	Intervention	Intervention	Intervention
2	Control	Transition period	Intervention	Intervention	Intervention	Intervention
3	Control	Control	Transition period	Intervention	Intervention	Intervention
4	Control	Control	Transition period	Intervention	Intervention	Intervention
5	Control	Control	Control	Transition period	Intervention	Intervention
6	Control	Control	Control	Transition period	Intervention	Intervention
7	Control	Control	Control	Control	Transition period	Intervention
8	Control	Control	Control	Control	Transition period	Intervention



MAIN TYPES OF SW-CRT DESIGNS

- ▶ Repeated cross-sectional design
 - Different individuals are measured each time
- ▶ Cohort design
 - The same individuals are measured each time
 - Closed cohort: no individuals may join during the trial
 - Open cohort: some individuals may leave and others may join during the trial

EXAMPLE 1: MAR TRIAL (CLOSED COHORT)

Open Access

Protocol

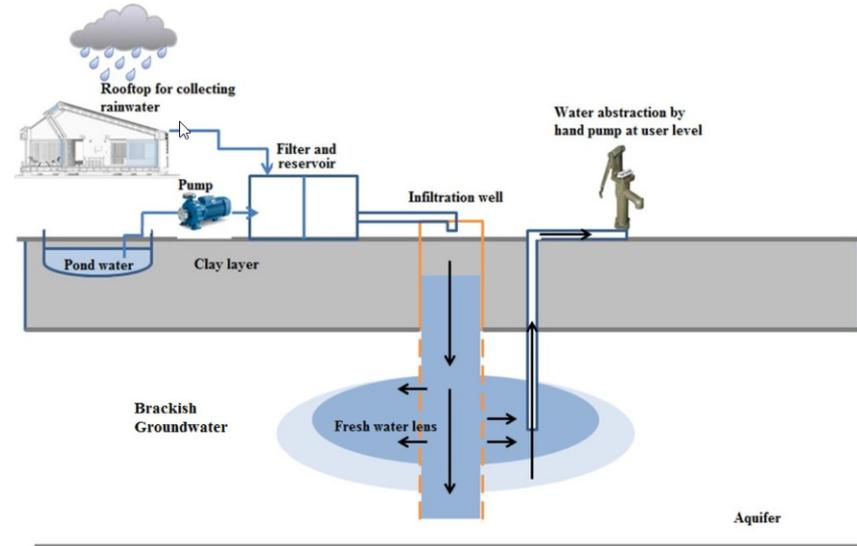
BMJ Open Stepped-wedge cluster-randomised controlled trial to assess the cardiovascular health effects of a managed aquifer recharge initiative to reduce drinking water salinity in southwest coastal Bangladesh: study design and rationale

Abu Mohd Naser,¹ Leanne Unicomb,² Solaiman Doza,² Kazi Matin Ahmed,³ Mahbubur Rahman,² Mohammad Nasir Uddin,² Shamshad B Quraishi,⁴ Shahjada Selim,⁵ Mohammad Shamsudduha,⁶ William Burgess,⁷ Howard H Chang,⁸ Matthew O Gribble,¹ Thomas F Clasen,¹ Stephen P Luby⁹

Naser AM, et al. *BMJ Open* 2017;**7**:e015205. doi:10.1136/bmjopen-2016-015205

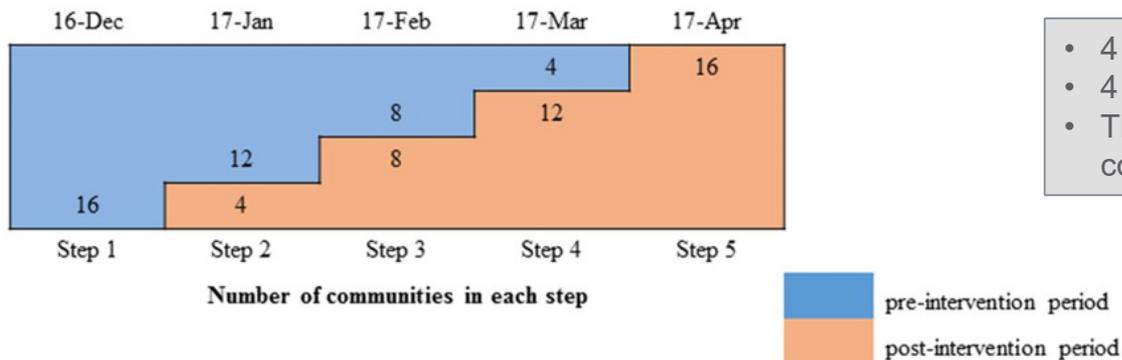
EXAMPLE 1: MAR TRIAL (CLOSED COHORT)

- ▶ **Background:** Groundwater supplies in coastal regions may be contaminated due to saltwater intrusion (can lead to high blood pressure and other adverse effects)
- ▶ **Intervention:** A Managed Aquifer Recharge (MAR) system which can reduce the salt content



EXAMPLE 1: MAR TRIAL (CLOSED COHORT)

- ▶ **Design:** Stepped wedge CRT in 16 communities over 5 months with 4 new communities receiving the system per month
- ▶ **Outcome:** Blood pressures measured on 60 adults per community during 5 monthly visits (recruited before randomization)



- 4 sequences, 5 periods
- 4 communities per sequence
- The same 60 individuals per community measured 5 times

EXAMPLE 2: CROSS-SECTIONAL DESIGN

Poldervaart *et al.* *BMC Cardiovascular Disorders* 2013, **13**:77
<http://www.biomedcentral.com/1471-2261/13/77>



STUDY PROTOCOL

Open Access

The impact of the HEART risk score in the early assessment of patients with acute chest pain: design of a stepped wedge, cluster randomised trial

Judith M Poldervaart, Johannes B Reitsma, Hendrik Koffijberg, Barbra E Backus, A Jacob Six, Pieter A Doevendans and Arno W Hoes

EXAMPLE 2: CROSS-SECTIONAL DESIGN

- ▶ **Background:** It can be difficult to diagnose patients presenting to an emergency department with chest pain. The HEART score is a validated clinical prediction rule which can be used to accurately identify patients with a serious underlying condition.
- ▶ **Objective:** Is the use of the HEART score to guide clinical decision-making for all patients ≥ 18 years presenting with chest pain to the ED safe?
- ▶ **Primary outcome:** Occurrence of major adverse cardiac events within 6 weeks
- ▶ **Design:** Stepped wedge CRT at 10 clinics over 11 months (expect 60 patients presenting per clinic per month)

EXAMPLE 2: CROSS-SECTIONAL DESIGN

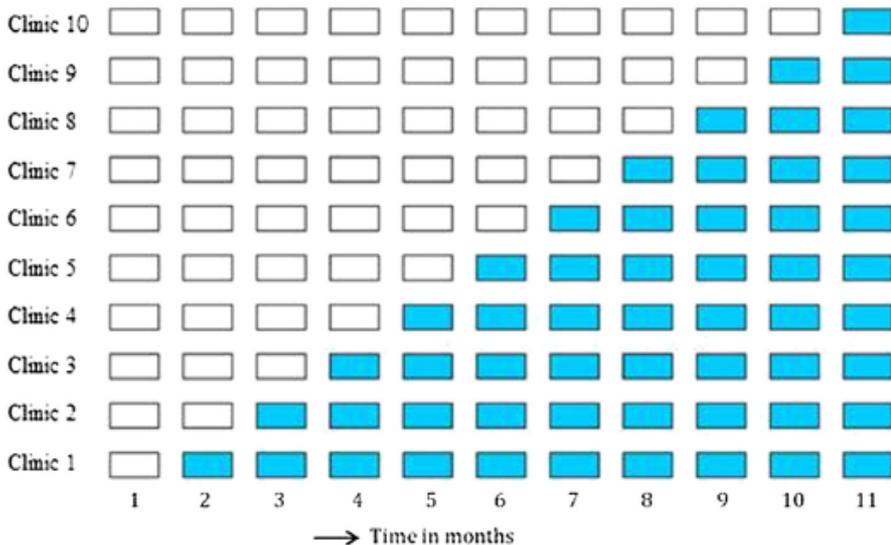


Figure 1 The Stepped Wedge Design for the HEART Impact study.

- 10 sequences, 11 periods
- 1 clinic per sequence
- 60 different patients per clinic each month
- Outcome measured once on each individual after 6 weeks

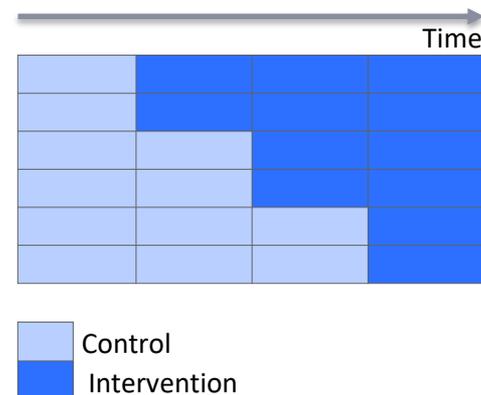
KEY METHODOLOGICAL CONSIDERATIONS

- SW-CRTs have several key characteristics that complicate their design and analysis
- May increase the risks of bias
- Need careful justification for the use of this design
- Need special care in reporting

Hemming, K, Taljaard M, McKenzie J et al. Reporting of The CONSORT extension for Stepped-Wedge Cluster Randomised Trials: Extension of the CONSORT 2010 statement with explanation and elaboration. BMJ. In Press.

(1) CONFOUNDING BY TIME

- ▶ Intervention effect is partially confounded with time
 - Due to staggered implementation, time is correlated with intervention
 - Time may also be correlated with outcome (“secular trend”)
- ▶ Analysis must always adjust for time (even if not significant)



Chen et al. Secular trends and evaluation of complex interventions: the rising tide phenomenon. *BMJ Qual Saf.* 2016 May;25(5):303-10.

(2) CONTAMINATION

- ▶ Increased risk of within-cluster contamination
 - Clusters may implement intervention earlier than planned (they can't wait)
 - Clusters may implement intervention later than planned (difficulties in implementation)

- ▶ As long as contamination is observed and recorded, an “as treated” analysis is possible (but deviates from “Intention-To-Treat”)

Copas AJ e.a. (2015) Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. *Trials*; 16:352

(3) TIME-VARYING INTERVENTION EFFECT

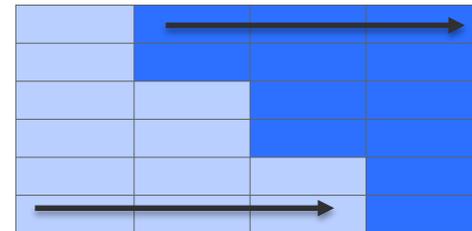
▶ Effect of intervention may vary depending on

- Calendar time

- Seasonal variation, external events

- Time since the intervention was introduced

- Response may increase with more experience
- Response may weaken over time (training is forgotten, decrease in adherence)



▶ An analysis which assumes a constant intervention effect may be biased

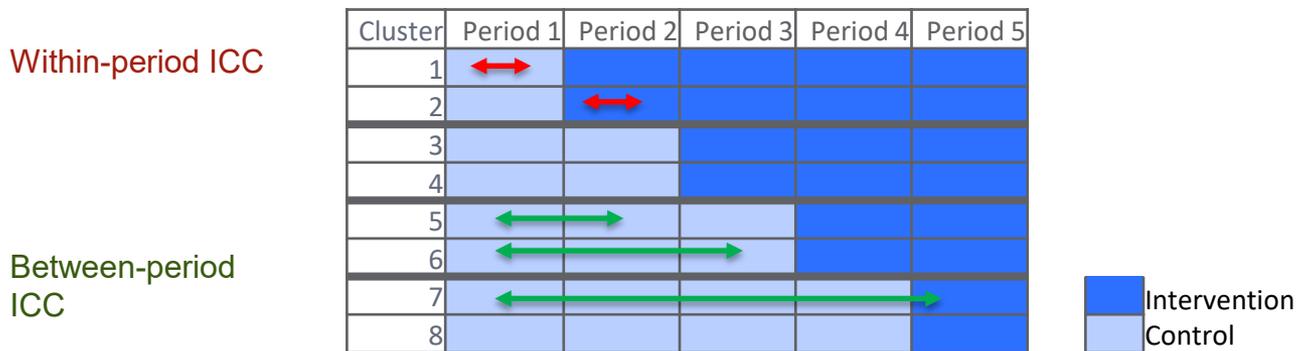
(4) CLUSTER-TREATMENT HETEROGENEITY

- ▶ Treatment effect may vary across clusters
 - Variation in quality of implementation, fidelity, other factors
- ▶ An analysis which assumes a homogeneous intervention effect across clusters may be biased
- ▶ Heterogeneity can reduce power

Hughes JP, Granston TS, Heagerty PJ. (2015) On the design and analysis of stepped wedge trials. Contemporary Clinical Trials. 45(Pt A):55-60

(5) COMPLEX CORRELATIONS

- ▶ Repeated measures on same clusters (and possibly same participants)
- ▶ Need to account for within-period ICCs as well as between-period ICCs
- ▶ Bias can be introduced by misspecifying the correlation structure



ANALYSIS OF THE SW-CRT

- ▶ Many possible methods of analysis, e.g.
 - Cluster-level methods vs. individual-level methods
 - Within-cluster (“Horizontal” approaches) vs. Between-cluster comparisons (“Vertical approaches”)
- ▶ Currently no consensus over best method
- ▶ Focusing here on ONE possible method: combination of within-cluster and between-cluster information using General(ized) Linear Mixed Model (GLMM)

Davey C, Hargreaves J, Thompson JA, Copas AJ, Beard E, Lewis JJ, Fielding KL. Analysis and reporting of stepped wedge randomised controlled trials: synthesis and critical appraisal of published studies, 2010 to 2014. *Trials* 2015 Aug 17;16:358.

THE SIMPLEST “DISCRETE TIME” GLMM

- ▶ MODEL 1: Hussey and Hughes (2007) proposed a mixed-effects regression approach for the cross-sectional SW-CRT design:

$$Y_{ijl} = \mu + \beta_j + \theta X_{ij} + u_i + e_{ijl} \quad \text{where}$$
$$u_i \sim N(0, \sigma_u^2); \quad e_{ijl} \sim N(0, \sigma_e^2)$$

Fixed intervention effect

β_j = fixed effect of time

X_{ij} = 1 if intervention, 0 if control

θ = intervention effect

u_i = random effect for cluster

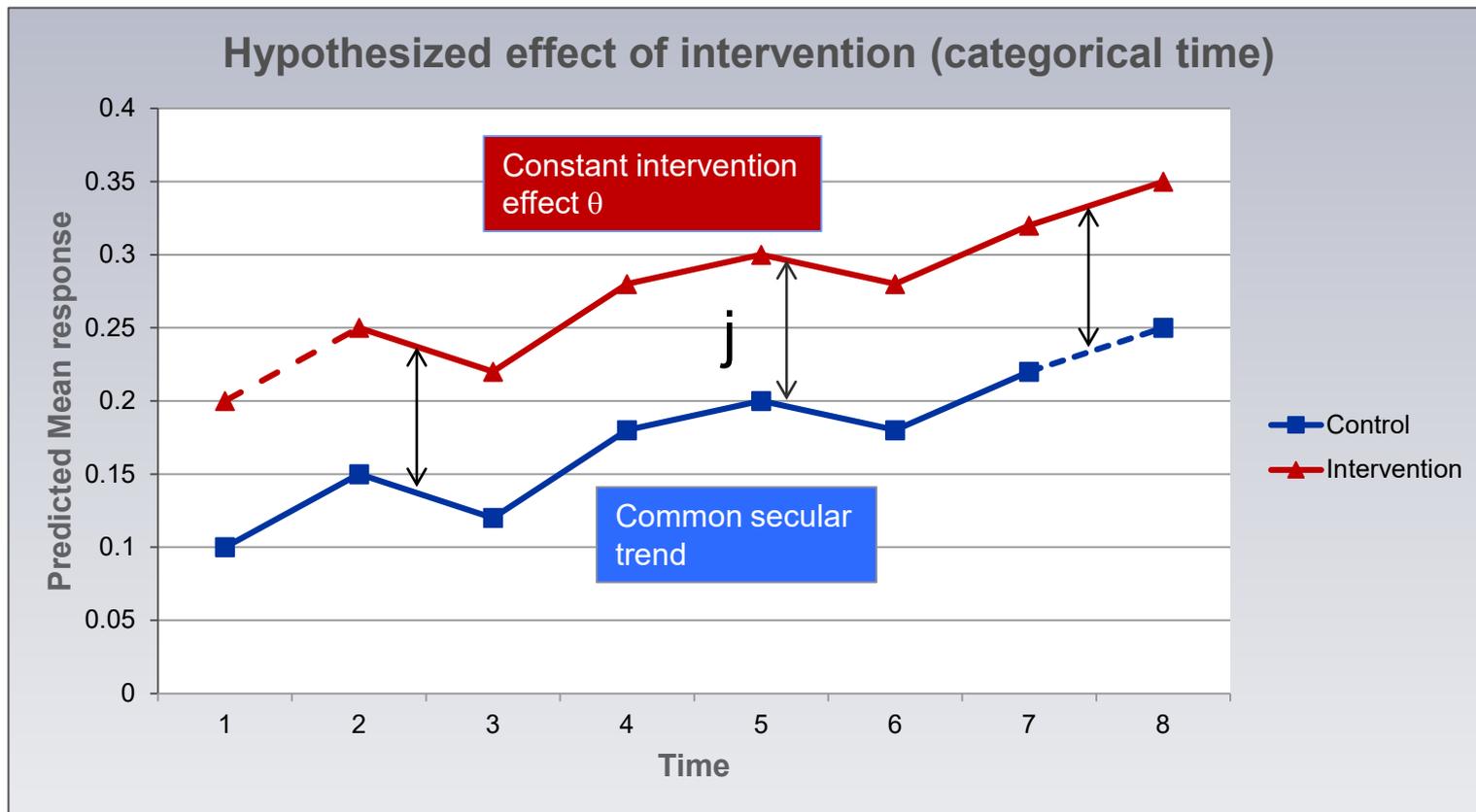
e_{ijl} = residual

$i = 1, \dots, k$ clusters

$j = 1, \dots, T$ periods

$l = 1, \dots, m$ individuals

NOTE THE MODEL ASSUMPTIONS (1)



NOTE THE MODEL ASSUMPTIONS (2)

- ▶ Single random intercept for cluster - implies a constant ICC over time (within-period ICC = between-period ICC)

ρ	ρ	ρ	ρ
	ρ	ρ	ρ
		ρ	ρ
			ρ

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

- ▶ A **decay** in the strength of the correlation over time is more likely

TWO ALTERNATIVE DISCRETE TIME “DECAY” MODELS

- ▶ Two alternative discrete time models have been proposed that allow the correlations to decay to a lesser or greater extent over time
 - MODEL 2: Hooper et al. (2016)
 - MODEL 3: Kasza & Forbes (2017)

- ▶ Summary of models:

Model	Within-period ICC	Between-period ICC
1. Hussey & Hughes	ρ	ρ
2. Hooper	ρ	$r \times \rho$
3. Kasza & Forbes	ρ	$r^{ t-s } \times \rho$

- $0 \leq r \leq 1$ is the decay parameter

TWO ALTERNATIVE “DECAY” MODELS (HOOPER)

- ▶ MODEL 2: Hooper model allows the between-period ICC to be $<$ within-period ICC, but allows no further decay:

ρ	$r\rho$	$r\rho$	$r\rho$
	ρ	$r\rho$	$r\rho$
		ρ	$r\rho$
			ρ

- ▶ This model is easy to fit (see SAS code in supplementary slides)
- ▶ We have simple design effects for easy sample size calculation

Hooper R, Teerenstra S, de Hoop E, Eldridge S. (2016) Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. *Statistics in Medicine* 35(26):4718-4728

TWO ALTERNATIVE “DECAY” MODELS (KASZA)

- ▶ MODEL 3: Kasza & Forbes model allows the between-period ICCs to decay exponentially:

ρ	$r\rho$	$r^2\rho$	$r^3\rho$
	ρ	$r\rho$	$r^2\rho$
		ρ	$r\rho$
			ρ

- ▶ This model is more difficult to fit (only possible in SAS – see code in supplementary slides)
- ▶ We have no simple design effects for it

Kasza J, Hemming K, Hooper R, Matthews JNS, Forbes AB et al. (2017) Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Stat Methods Med Res*. DOI: 10.1177/0962280217734981

DETAILS FOR HOOPER MODEL

- ▶ MODEL 2: Discrete time mixed effects regression model for the cross-sectional SW-CRT design allowing a different between-period ICC:

$$Y_{ijl} = \mu + \beta_j + \theta X_{ij} + u_i + (u\tau)_{ij} + e_{ijl};$$
$$u_i \sim N(0, \sigma_u^2); \quad (u\tau)_{ij} \sim N(0, \sigma_{u\tau}^2); \quad e_{ijl} \sim N(0, \sigma_e^2)$$

β_j = fixed effect of time

X_{ij} = 1 if intervention, 0 if control

θ = intervention effect

u_i = random effect for cluster

$(u\tau)_{ij}$ = random time effect for cluster

e_{ijl} = residual

$i = 1, \dots, k$ clusters

$j = 1, \dots, T$ periods

$l = 1, \dots, m$ individuals

WITHIN AND BETWEEN-PERIOD ICCs (HOOPER)

- ▶ Within-period ICC: between two individuals in the same cluster and same period

$$wpICC = \rho_0 = \frac{\sigma_u^2 + \sigma_{u\tau}^2}{\sigma_u^2 + \sigma_{u\tau}^2 + \sigma_e^2};$$

- ▶ Between-period ICC: between two individuals in the same cluster but different periods

$$bpICC = \rho_1 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_{u\tau}^2 + \sigma_e^2}$$

- $bpICC \leq wpICC$

THE CLUSTER AUTOCORRELATION COEFFICIENT

- ▶ The ratio of the between-period and within-period ICCs is called the “Cluster Autocorrelation Coefficient” (CAC)
- ▶ CAC measures the extent of the correlation decay (e.g., CAC=0.8 implies a 20% decay in the correlation)

$$CAC = \frac{bpICC}{wpICC} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_{ur}^2} = \omega$$

- ▶ CAC is an important parameter in sample size calculation
 - Note: CAC=1 implies between-period ICC = within-period ICC (Hussey & Hughes model)

IMPLICATIONS OF MISSPECIFIED CORRELATION

- ▶ Under-specification (omitting a necessary decay) results in bias of the variance of the treatment effect estimate
 - Assuming Model 1 (CAC=1) when Model 3 holds will underestimate variance (p-values too small; CI too narrow)
 - Assuming Model 2 when Model 3 holds will *usually* underestimate variance
 - Impact depends on strength of correlation decay, within-period ICC and cluster period size (m)
- ▶ Over-specification (including a decay unnecessarily) does not lead to bias

Kasza J & Forbes A. Estimating variance components in multiple-period cluster randomised trials when random effect correlation structure is misspecified. *Statistical Methods in Medical Research*. Under review.

ADDITIONAL CONSIDERATIONS FOR ANALYSIS

- ▶ Highly recommended considering accounting for:
 - Cluster treatment heterogeneity
 - Include random cluster by treatment effect
 - Time-varying intervention effect
 - Include intervention by time interaction (either calendar time or time on treatment)
- ▶ Pre-specify how delays/transition periods will be handled
 - Omit observations during transition period
 - Analyze observations as unexposed under Intent-To-Treat
 - Analyze observations as a fraction of full effect

SAMPLE SIZE CALCULATION FOR THE SW-CRT

- ▶ Can be based on
 - Simulation
 - Analytical formula
 - Design effect
- ▶ Here, illustrate the simplest approach which uses design effects based on Hooper model
- ▶ (No design effects available for Kasza model but see R Shiny App – slide 42)

CALCULATION OF THE REQUIRED NUMBER OF CLUSTERS

► Five steps:

1. Calculate **total** required sample size assuming individual randomization N_{ind}
2. Multiply by design effect due to clustering $Deff_c = 1 + (m - 1)r0$
3. Multiply by design effect due to repeated assessment $Deff_t$ (see next slide)
4. Divide by cluster size per period (m) to determine **total** required number of clusters (k)

$$k = \frac{N_{ind} \times Deff_c \times Deff_t}{m}$$

5. May need to round up to multiple of number of steps

- Hooper R, Bourke L. Cluster randomised trials with repeated cross sections: alternatives to parallel group designs. *BMJ*. 2015 Jun 8;350:h2925
- Hooper R et al. (2016) Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. *Statistics in Medicine* 35(26):4718-4728

DESIGN EFFECT DUE TO REPEATED ASSESSMENT

- ▶ Function of number of sequences t and the correlation between cluster means at two different times R :

$$Deff_t = \frac{3t(1-R)(1+tR)}{(t^2-1)(2+tR)}$$

- ▶ R is defined, for cross-sectional and cohort designs, respectively as:

$$R = \frac{m\rho_0\omega}{1+(m-1)\rho_0} \quad R = \frac{m\rho_0\omega + (1-\rho_0)\tau}{1+(m-1)\rho_0}$$

where ω is the Cluster Autocorrelation Coefficient (CAC) and τ is the Individual Autocorrelation Coefficient (IAC)

EMPIRICAL ESTIMATES FOR DESIGN PARAMETERS

- ▶ Calculations critically depend on the assumed correlation structure (no between-period decay)
- ▶ Obtaining empirical estimates for the within-period ICC and CAC is challenging
 - Ideally need raw longitudinal data with the correct period length (e.g., historical, routinely collected data)
 - If no prior information, consider assuming CAC between 0.6 to 0.8
 - Essential to examine sensitivity to a range of assumed values

WORKED EXAMPLE: MAR TRIAL (BANGLADESH)

- ▶ How many communities are required for the MAR trial? (Basic SW with a closed cohort design, continuous outcome)
- ▶ Sample size parameters:
 - 80% power, $\alpha = 0.05$
 - $t = 4$ sequences
 - $m=60$ residents per community per period
 - Standard deviation $\sigma = 20$ mmHg
 - Target difference = 3 mmHg
 - $wplCC = 0.05$, $CAC=0.7$, $IAC=0.9$

WORKED EXAMPLE: MAR TRIAL (BANGLADESH)

- ▶ Calculate total sample size required under individual randomization: $N_{ind} = 1396$
- ▶ Calculate design effect due to clustering: $Deff_c = 1 + (m - 1)\rho_0 = 3.95$
- ▶ Calculate R for cohort design:

$$R = \frac{m\rho_0\omega + (1 - \rho_0)\tau}{1 + (m - 1)\rho_0} = 0.75$$

- ▶ Calculate design effect due to time:

$$Deff_t = \frac{3t(1 - R)(1 + tR)}{(t^2 - 1)(2 + tR)} = 0.16$$

- ▶ Calculate required number of clusters: $k = \frac{N_{ind} \times Deff_c \times Deff_t}{m} = 14.8$
- ▶ Round to a multiple of the number of sequences: 16

MAR TRIAL: COMPARE SAMPLE SIZES, SW VS. PARALLEL

Parallel arm

	Month	
Cluster	1	
1		
...		
...		
...		
...		
K		

K=92
N=5520

Parallel before & after repeated measures

	Month				
Cluster	1	2	3	4	5
1					
...					
...					
...					
...					
K					

K=24
N=1440

Parallel arm before and after

	Month	
Cluster	1	2
1		
...		
...		
K		

K=42
N=2520

Stepped wedge

	Month				
Steps	1	2	3	4	5
1					
2					
3					
4					

K=16
N=960

SAMPLE SIZE RESOURCES

- ▶ R package 'swCRTdesign' <http://faculty.washington.edu/jphughes/pubs.html>
 - Allows for fractional treatment indicator, incomplete designs, cluster treatment heterogeneity (but not correlation decay)
- ▶ R-Shiny (Hemming & Kasza) <https://clusterrcts.shinyapps.io/rshinyapp/>
 - Includes parallel arm longitudinal, stepped wedge, and cross-over designs
 - Continuous, binary or count outcomes
 - Repeated cross-sectional and cohort designs
 - Equal or unequal allocation
 - Complete or incomplete designs (but not fractional treatment indicator)
 - Adjustments for cluster size variability
 - Allows for correlation decay and cluster treatment heterogeneity

ADVANTAGES AND DISADVANTAGES OF SW-CRTs

ADVANTAGES

- ▶ All clusters receive the intervention during the study
 - Easier to recruit clusters if all will receive intervention
 - Or it may be a stakeholder *requirement* - in this situation, the SW-CRT is a stronger study design than an uncontrolled before and after evaluation
- ▶ Usually require fewer clusters than parallel arm design (unless ICC and cluster period sizes small)
- ▶ Delivery of intervention can be spread out over time (e.g., by having only one cluster cross each time) (although also possible with parallel arm design)

DISADVANTAGES

1. Can be difficult to separate the effect of the intervention from the effect of secular trends
2. All participating clusters must be recruited upfront (so they can be randomized)
 - Doesn't scale easily if you want to add clusters
3. Can be logistically challenging to ensure all clusters are ready to implement on schedule
4. Can increase the data collection burden (unless routinely collected outcomes)
5. Can take longer to complete the study
 - May increase the risk of clusters dropping out
 - May increase the risk of contamination or external events influencing outcomes
6. More complicated to analyze and interpret results (requires many assumptions)
 - May not work well if intervention does not have an immediate effect
 - May not work well if intervention effect might change over time

CONCLUSIONS

- ▶ The SW-CRT is a novel type of cluster randomized design that is rapidly increasing in popularity
- ▶ It is more complicated to design and analyze and may increase the risks of bias
- ▶ There is no consensus over its design and analysis
- ▶ Needs substantial methodological development in the next years
- ▶ Need to carefully consider whether adoption of the design is justified

- de Hoop E, van der Tweel I, van der Graaf R, Moons KG, et al. (2015) The need to balance merits and limitations from different disciplines when considering the stepped wedge cluster randomized trial design. *BMC Med Res Methodol.*; 15:93.
- Hargreaves JR, Copas AJ, Beard E, et al. (2015) Five questions to consider before conducting a stepped wedge trial. *Trials*; 16:350

KEY REFERENCES

- ▶ Hussey MA & Hughes JP (2007) Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials*; 28:182-191
- ▶ Hughes JP, Granston TS, Heagerty PJ. (2015) On the design and analysis of stepped wedge trials. *Contemporary Clinical Trials*. 45(Pt A):55-60
- ▶ Hooper R, Bourke L (2015) Cluster randomised trials with repeated cross sections: alternatives to parallel group designs. *BMJ*. Jun 8;350:h2925
- ▶ Hemming K, Lilford R, Girling AJ. (2015) Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple level designs. *Statist. Med*; 34(2):181-196
- ▶ Copas AJ e.a. (2015) Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. *Trials*; 16:352
- ▶ Girling AJ and Hemming K. (2016) Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. *Statist Med* 35(13):2149-66
- ▶ Hooper R, Teerenstra S, de Hoop E, Eldridge S. (2016) Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. *Statist Med* 35(26):4718-4728
- ▶ Thompson JA et al. (2017) The optimal design of stepped wedge trials with equal allocation to sequences and a comparison to other trial designs. *Clin Trials*. Dec;14(6):639-647.
- ▶ Kasza J, Hemming K, Hooper R, Matthews JNS, Forbes AB et al. (2017) Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Stat Methods Med Res*. DOI: 10.1177/0962280217734981
- ▶ Thompson JA, Davey C, Fielding K, Hargreaves JR, Hayes RJ. (2018) Robust analysis of stepped wedge trials using cluster-level summaries within periods. *Stat Med*. Jul 20;37(16):2487-2500.

SUPPLEMENTARY MATERIAL

SAS CODE: HUSSEY & HUGHES MODEL

- ▶ Cross-sectional SW-CRT design; continuous outcome

```
PROC MIXED DATA=indiv;  
CLASS cluster period;  
MODEL outcome = intervention period /SOLUTION DDFM=KR;  
RANDOM intercept/ SUBJECT=cluster VCORR;  
RUN;
```

- ▶ Note: Consider DDFM=BW to reduce computation time

SAS CODE: HOOPER MODEL

- ▶ Cross-sectional SW-CRT design; continuous outcome

```
PROC MIXED DATA=indiv;  
CLASS cluster period;  
MODEL outcome = intervention period /SOLUTION DDFM=KR;  
RANDOM intercept period / SUBJECT=cluster VCORR;  
RUN;
```

- ▶ Note: Consider DDFM=BW to reduce computation time

SAS CODE: KASZA & FORBES MODEL

- ▶ Cross-sectional SW-CRT design; continuous outcome

```
PROC MIXED DATA=indiv;  
CLASS cluster period;  
MODEL outcome = intervention period /SOLUTION DDFM=KR;  
RANDOM period / SUBJECT=cluster TYPE=AR(1) VCORR;  
RUN;
```

- ▶ Note: In case of computational issues,
 - Consider using PROC HPMIXED
 - Consider DDFM=BW to reduce computation time

TIMING OF ENROLMENT AND EXPOSURE

- ▶ Short exposure duration:
 - Individuals are recruited (or included) in continuous time as they become eligible (e.g., arrive at hospital emergency department) and are exposed for a short time
 - (Their outcomes may be assessed immediately or after a long follow-up)
 - Each individual experiences either the control or intervention condition
- ▶ Long exposure duration:
 - Most (or all) individuals may become exposed to the trial from the start and participate to the end
 - Most (or all) individuals experience both control and intervention conditions

Copas AJ e.a. (2015) Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. *Trials*; 16:352

TIMING AND NATURE OF OUTCOME MEASUREMENT

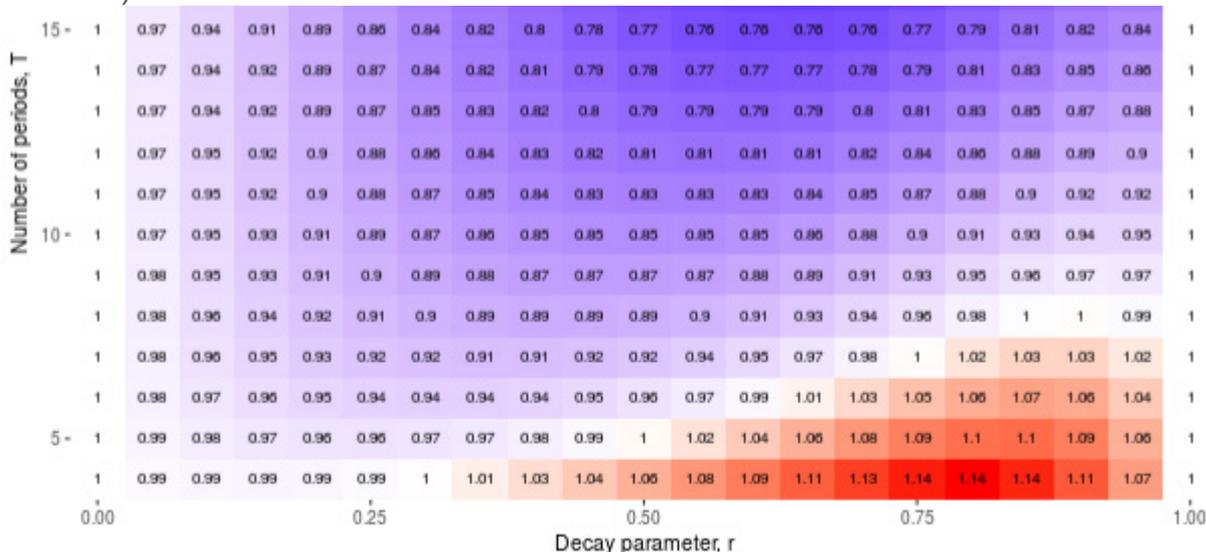
- ▶ Repeated measures on individuals at discrete calendar times linked to timing of the steps
- ▶ Repeated measures on individuals at times linked to the start of individual exposures (e.g., before and after exposure)
- ▶ Measurements on a small fraction of individuals within large clusters at discrete calendar times
- ▶ Single measurement from each individual at a fixed time after the start of their exposure (possibly after a long follow-up period such as 1 year)
- ▶ Time to event, where time begins at the start of exposure

Copas AJ e.a. (2015) Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. *Trials*; 16:352

IMPLICATIONS OF MISSPECIFIED CORRELATION?

- ▶ Implications of not allowing for a correlation decay?
(see look-up table at <https://monash-biostat.shinyapps.io/MisspecCorrStruct>)

Implications of assuming Model 2 when Model 3 model holds ($p=0.03$, $m=100$)



- ▶ Ratio < 1: Variance of treatment effect estimate underestimated (p-values too small; CI too narrow)
- ▶ Ratio > 1: Variance of treatment effect estimate over-estimated (p-values too large; CI too wide)

MULTIPLE PERIODS BEFORE AND AFTER

- ▶ Designs with multiple periods before and after the rollout of the intervention are not recommended (Hargreaves et al. Trials 2015)

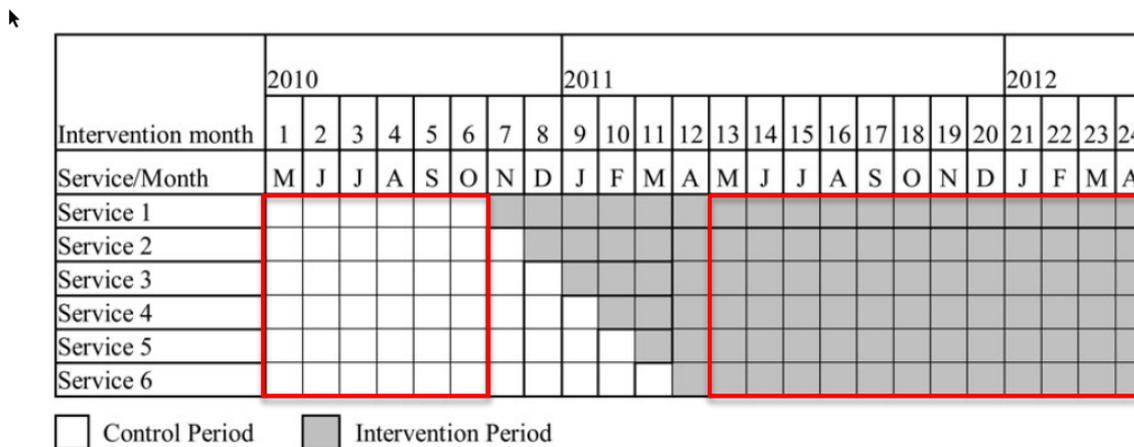


Figure 1. Timeline of antimicrobial stewardship stepped-wedge rollout to the clinical services. The “stepped” introduction of the intervention occurred over months 7–12; the white “wedge” reflects the control period, and the gray “wedge” the intervention period.

Clinical Infectious Diseases 2014;59(6):867–74

INCOMPLETE DESIGNS

- ▶ Incomplete designs can be used to reduce the data collection burden

Incomplete SW design

	Time		
Cluster	1	2	3
1	Intervention		
2	Intervention		
3	Control	Intervention	
4	Control	Intervention	
5		Control	Intervention
6		Control	Intervention
7			Control
8			Control

	Intervention
	Control

Incomplete SW design

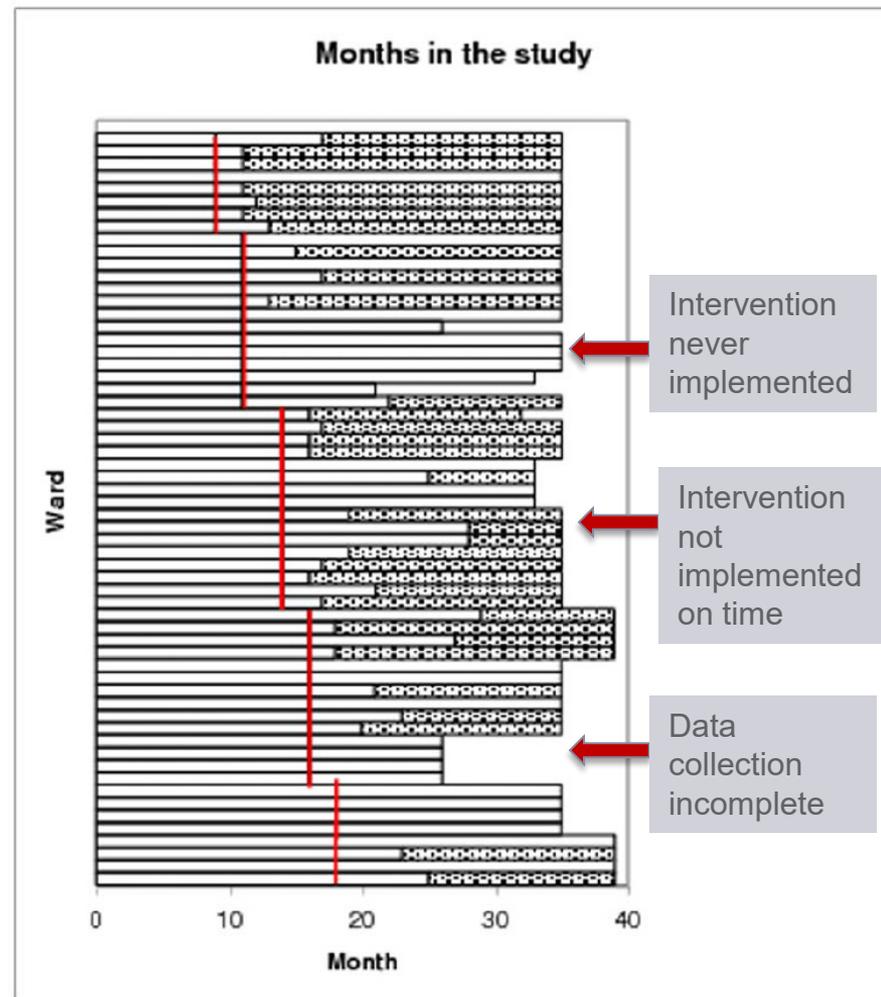
	Time		
Cluster	1	2	3
1	Intervention		Intervention
2	Intervention		Intervention
3	Control	Intervention	
4	Control	Intervention	
5		Control	Intervention
6		Control	Intervention
7	Control		Control
8	Control		Control

	Intervention
	Control

Hooper R, Bourke L. The dog-leg: an alternative to a cross-over design for pragmatic clinical trials in relatively stable populations. *Int J Epidemiol.* 2014; 43(3): 930-6

WHEN THINGS GO WRONG

- ▶ The Feedback Intervention Trial – Improving Hand Hygiene Compliance in UK Healthcare Workers (Fuller et al, 2012)



WAIT LIST DESIGN VS STEPPED WEDGE

Standard wait list design

Groups	1
1	Control
2	Control
3	Control
4	Control

Trial ends

Intervention



Stepped wedge trial

Groups	Time					
	1	2	3	4	5	6
1	Control	Control	Control	Control	Control	Intervention
2	Control	Control	Control	Control	Intervention	Intervention
3	Control	Control	Control	Intervention	Intervention	Intervention
4	Control	Control	Intervention	Intervention	Intervention	Intervention
5	Control	Intervention	Intervention	Intervention	Intervention	Intervention



- Standard wait-list design: Offer control clusters the intervention at the end (but may not contribute data to evaluation)

WHAT QUALIFIES AS A SW-CRT?

- Must have a minimum of 3 sequences
- May have 2 sequences and 3 periods
- May not have all clusters starting in control and ending in intervention
- May not have complete data

✓

	1	2	3
1	Light Blue	Light Blue	Dark Blue
2	Light Blue	Dark Blue	Dark Blue

Diagram illustrating a valid SW-CRT design with 2 sequences and 3 periods. A red checkmark is above the table. A mouse cursor is pointing at the top right of the table.

✗

	1	2
1	Light Blue	Light Blue
2	Light Blue	Dark Blue

Diagram illustrating an invalid SW-CRT design because it does not have a minimum of 3 sequences. A red X is above the table.

✓

	1	2
1	Light Blue	Light Blue
2	Light Blue	Dark Blue
3	Dark Blue	Dark Blue

Diagram illustrating a valid SW-CRT design with 3 sequences and 2 periods. A red checkmark is above the table.

✓

	1	2	3
1	Dark Blue	Light Blue	Dark Blue
2	Light Blue	Dark Blue	Light Blue
3	Light Blue	Light Blue	Dark Blue
4	Light Blue	Light Blue	Light Blue

Diagram illustrating a valid SW-CRT design with 4 sequences and 3 periods. A red checkmark is above the table.

CHOICE OF NUMBER OF STEPS

- ▶ Depends on total trial duration and sample size requirement
- ▶ Greater power is achieved with more steps (maximum with one cluster per step)
 - But may not be possible to implement intervention in more than one cluster per time
 - Or it may be too expensive to implement intervention in only one cluster at a time (e.g., training)
 - More steps also means more data collection points which may be expensive
- ▶ Step length may need to be chosen with consideration of the lag time