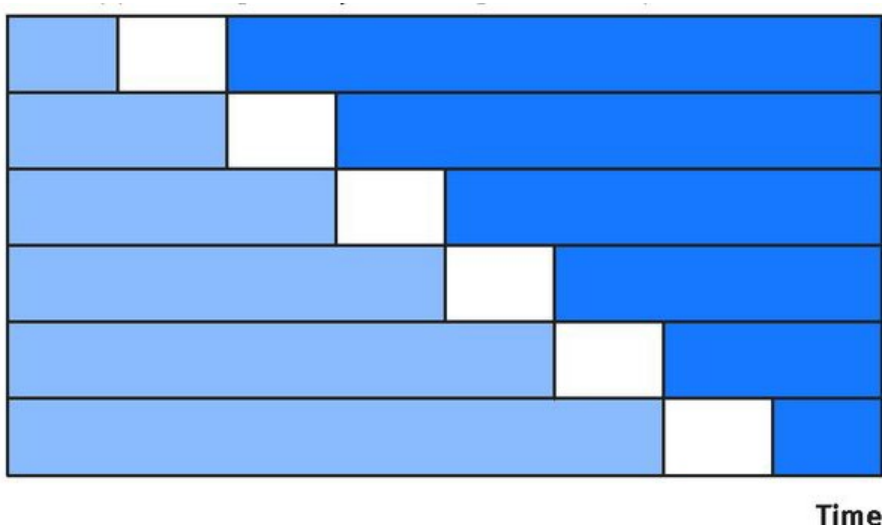
Stepped Wedge Designs in Practice

What, Why, When and How



Goldbag Research Seminar 2021-11-08

Thomas E. Love, Ph.D.

Professor of Medicine and of Population & Quantitative Health Sciences,
Case Western Reserve University School of Medicine

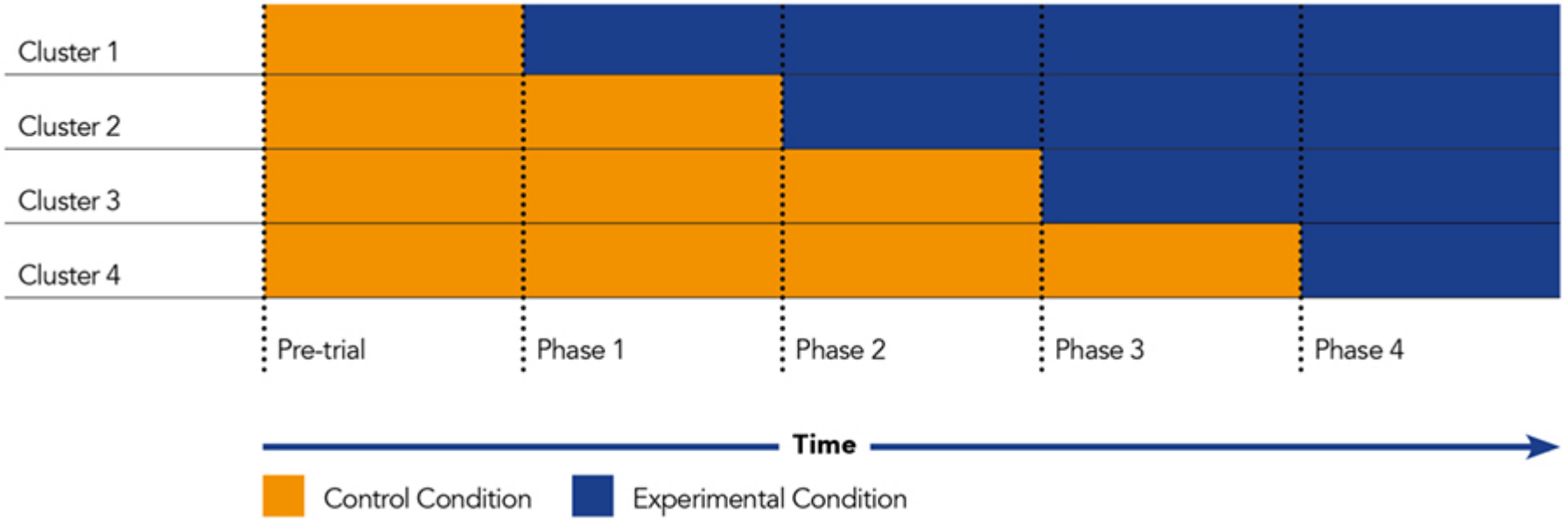
Director, Biostatistics & Data Sciences, Population Health Research Institute,
The MetroHealth System

Thomas dot Love at case dot edu

https://github.com/THOMASELOVE/stepped-wedge

A Brief Overview

- Stepped wedge designs are a form of cluster randomized trial that involves the sequential transition of clusters (hospitals / groups / communities) from control to intervention conditions in a randomized order. The crossover is in one direction.



Hemming 2020 IJE and Research Outreach

https://github.com/THOMASELOVE/stepped-wedge

Why Stepped Wedge vs. Parallel CRTs?

- Stepped wedge designs are at greater risk of bias compared with parallel cluster randomized trials.
  - In particular, there is a greater risk of bias due to misspecification of secular trends at the analysis stage.
- Cluster randomized trials are already at greater risk of bias compared with individual level randomized trials.
- Despite these facts, stepped wedge designs are increasingly popular.

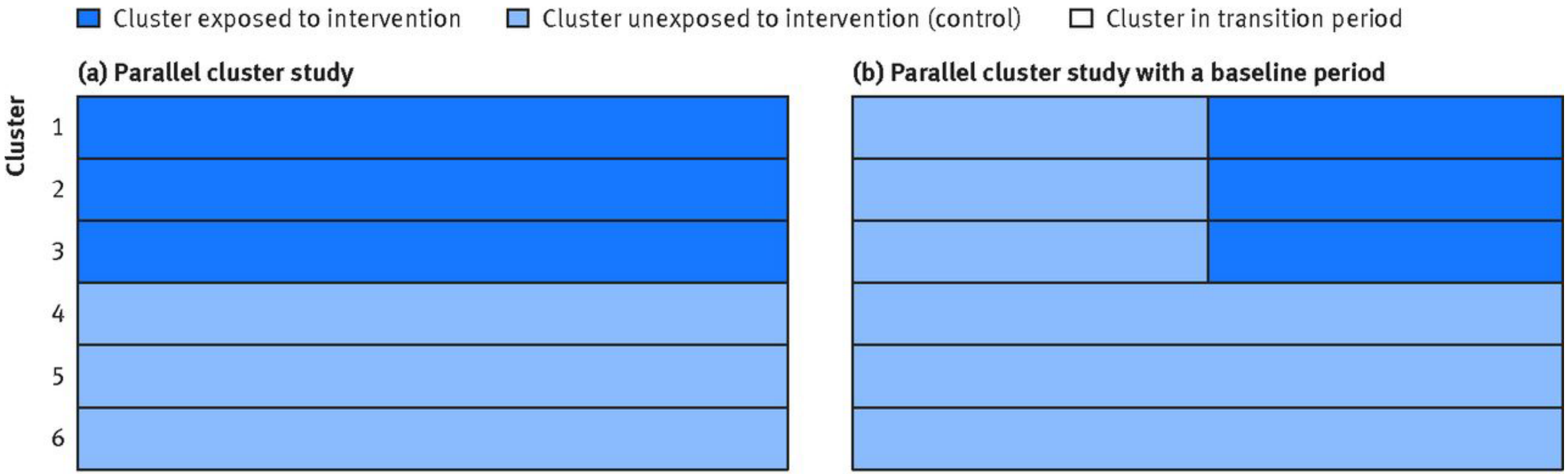
When might a stepped wedge be an appropriate choice?

- The stepped wedge allows us to do a randomized evaluation which otherwise would not be possible due to policy or other constraints
- The stepped wedge facilitates recruitment of clusters as it enhances the acceptability of a randomized evaluation
- The stepped wedge is the only feasible design because the intervention requires the roll-out of a scarce resource
- The stepped wedge has increased statistical power over other plausible study designs (especially true if we have few clusters.)
- Parallel CRTs are still, in general, the better choice when feasible.

Comparing RCTs and CRTs

- **Randomized Controlled Trial**
  - unit that is randomized is the unit of analysis
  - for large n, important attributes are likely to be distributed similarly across groups
- **CRT (Cluster-Randomized Trial)**
  - interventions focus on systems, prevention, behavior
  - useful when "contamination" is likely to be a problem and/or blinding is impossible
  - unit randomized is not the sole unit of interest
  - "clustering" within units is the critical issue

Cluster Randomized Trial Schematics



Hemming, Haines et al. BMJ 2015; 350:h391 https://doi.org/10.1136/bmj.h391

https://github.com/THOMASELOVE/stepped-wedge

Patients within Providers within Practices within Study Groups: A 4-Level Parallel CRT (Intervention applied to Study Group I)

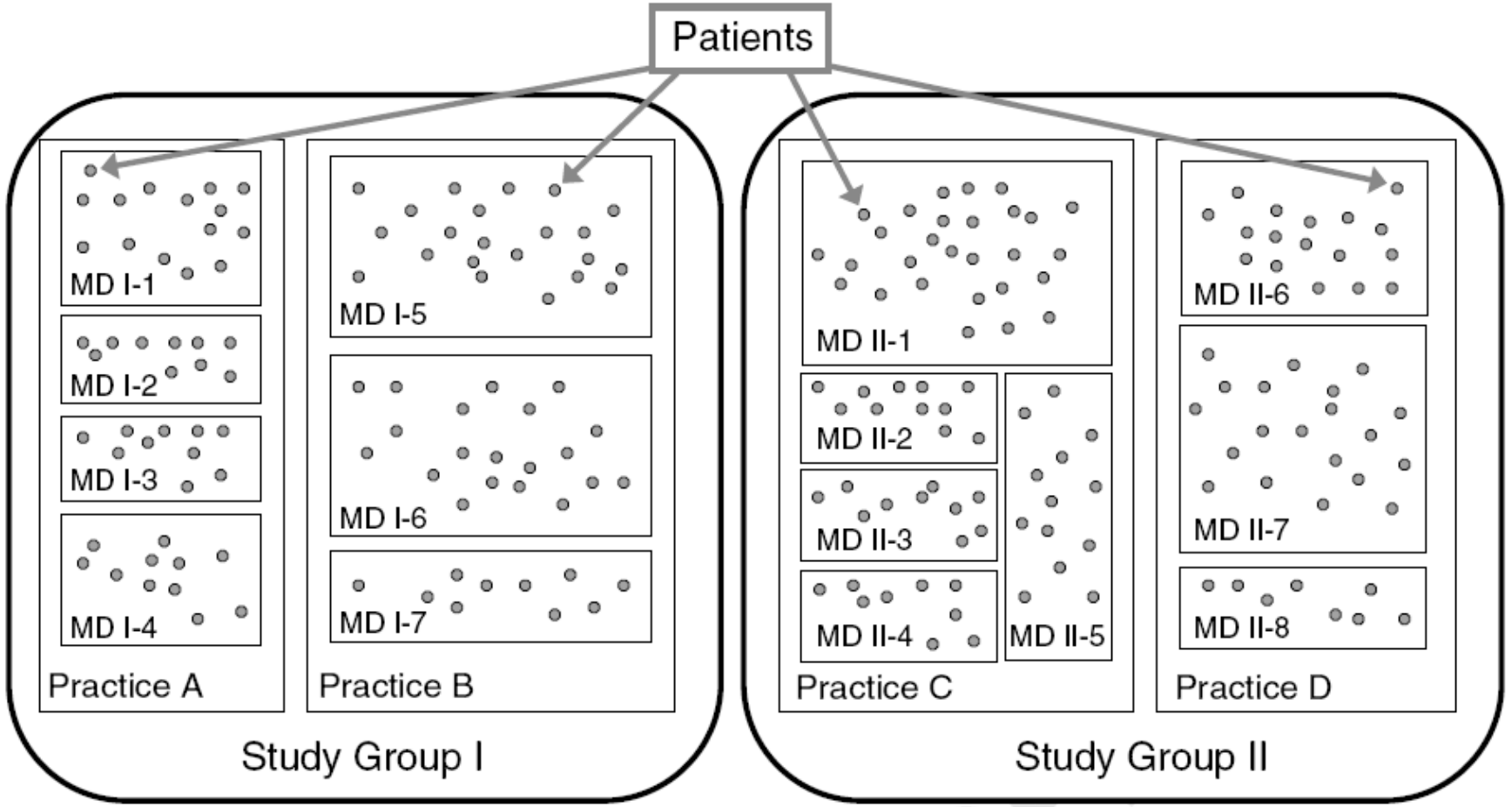


Figure 36.1. A simple four-level cluster-randomised trial.

Cebul RD Dawson NV Love TE (2007) *Textbook of Clinical Trials*, 2nd Ed., Wiley.

https://github.com/THOMASELOVE/stepped-wedge

Key Concerns when doing Cluster-Randomized Trials (CRTs)

- Unit of randomization (assignment) is usually different than the unit of analysis
- Clustering has design (sample size) and analytic implications
  - Need larger samples than an individual RCT
    - Standard RCT sample size formulas lead to underpowered CRTs
  - Need better pre-trial data for balancing
  - Need more sophisticated statistical methods for analyses
    - Standard RCT approaches will bias p values downward

Implementing Cluster-Randomized Trials

- Some implementation challenges lead to “observational” studies
  - Political issues
  - Temporal trends
- Challenges regarding
  - Randomization
  - Susceptibility
  - Performance
  - Detection
  - Transfer

https://github.com/THOMASELOVE/stepped-wedge

Implementing Cluster-Randomized Trials

- Some implementation challenges lead to "observational" studies
  - Political issues
  - Temporal trends
- Challenges regarding
  - **Randomization**
  - Susceptibility
  - Performance
  - Detection
  - Transfer

Not all sites will allow randomization
- We "need" to be in one group or the other

Sites may need to be randomized together
- functionally linked (staff moves regularly between 2 sites)

https://github.com/THOMASELOVE/stepped-wedge

Implementing Cluster-Randomized Trials

- Some implementation challenges lead to "observational" studies
  - Political issues
  - Temporal trends
- Challenges regarding
  - Randomization
  - **Susceptibility**
  - Performance
  - Detection
  - Transfer

Susceptibility to outcome at baseline may differ due to (for instance) trends over time
- May threaten "comparability" across intervention/non-intervention clusters

https://github.com/THOMASELOVE/stepped-wedge

Implementing Cluster-Randomized Trials

- Some implementation challenges lead to "observational" studies
  - Political issues
  - Temporal trends
- Challenges regarding
  - Randomization
  - Susceptibility
  - **Performance**
  - Detection
  - Transfer

Something new influences the outcome, but only at certain sites at certain times
- Intervention fidelity
- Co-interventions

https://github.com/THOMASELOVE/stepped-wedge

Implementing Cluster-Randomized Trials

- Some implementation challenges lead to “observational” studies
  - Political issues
  - Temporal trends
- Challenges regarding
  - Randomization
  - Susceptibility
  - Performance
  - **Detection**
  - Transfer

Detection of the outcome
- Software upgrade adds automatic recording of useful data elements
- Gold standard test is “easier” to order at some sites compared to others

https://github.com/THOMASELOVE/stepped-wedge

Implementing Cluster-Randomized Trials

- Some implementation challenges lead to "observational" studies
  - Political issues
  - Temporal trends
- Challenges regarding
  - Randomization
  - Susceptibility
  - Performance
  - Detection
  - **Transfer**

Transfer issues
- Drop-outs and crossovers
- Comparisons that don't consider clusters appropriately

https://github.com/THOMASELOVE/stepped-wedge

The Impact of Clustering

- If there are important cross-group differences in important factors at baseline, this affects:
  - Study Power [Effective Sample Size]
  - Study Design:
    - Pre-randomization cluster “balancing” will improve balance of important prognostic factors
  - Study Analysis:
    - Need analytic techniques that account for clustering: e.g. GEE, hierarchical models, etc.

Additional Reporting Requirements for CRTs (CONSORT Statement, 2004)

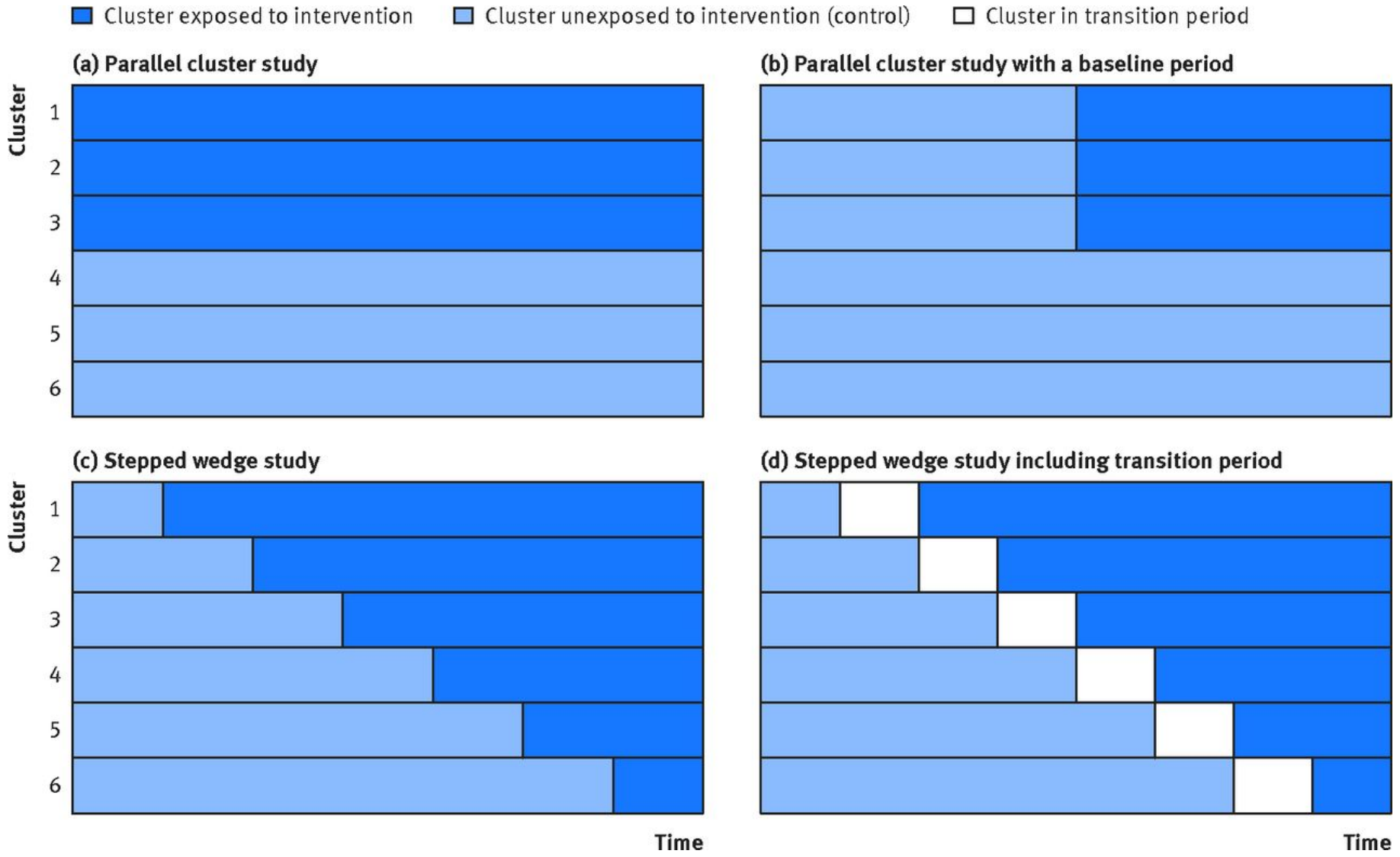
1. Rationale for adopting a cluster design
2. How clustering was incorporated into the sample size calculations (include ICCs)
3. How clustering effects were incorporated into the analysis
4. The flow of both clusters and individuals through the trial, from assignment to analysis.

Campbell MK et al. for the CONSORT Group (2004)

The Stepped Wedge Design

- Random and sequential crossover of clusters from control to intervention until all clusters are exposed.
  - Building on ideas like “waiting list designs” or “phased” implementations of interventions
- Particularly well suited to studies that don’t require the recruitment of individual patients, so they’re becoming more common (especially) in assessing service delivery interventions.
- More clusters are exposed to the intervention near the end of the study, as compared to the early stages of the study.
- Like all cluster-randomized trials, there are increased risks of selection biases in a stepped wedge design…

Schematic of Stepped Wedge vs. Parallel CRT studies



Hemming, Haines et al. BMJ 2015; 350:h391 https://doi.org/10.1136/bmj.h391

https://github.com/THOMASELOVE/stepped-wedge

Gambia Hepatitis Intervention Study

- Investigating the effectiveness of a Hepatitis B vaccine in preventing liver disease.
  - Substantial preliminary evidence of efficacy against hepatitis B
  - No randomized evidence on protection against liver disease.
  - Phased but random implementation of the vaccination
  - Sequential rollout: geographical areas of the country were randomly allocated to incorporate the vaccination into the existing childhood vaccination schedule in 10-12 week steps, so that complete national coverage happened over 4 years.
  - Follow-up of the cohort for liver disease outcomes is ongoing.

The Gambia Hepatitis Study Group. The Gambia hepatitis intervention study. Cancer Res 1987; 47: 5782-5787.

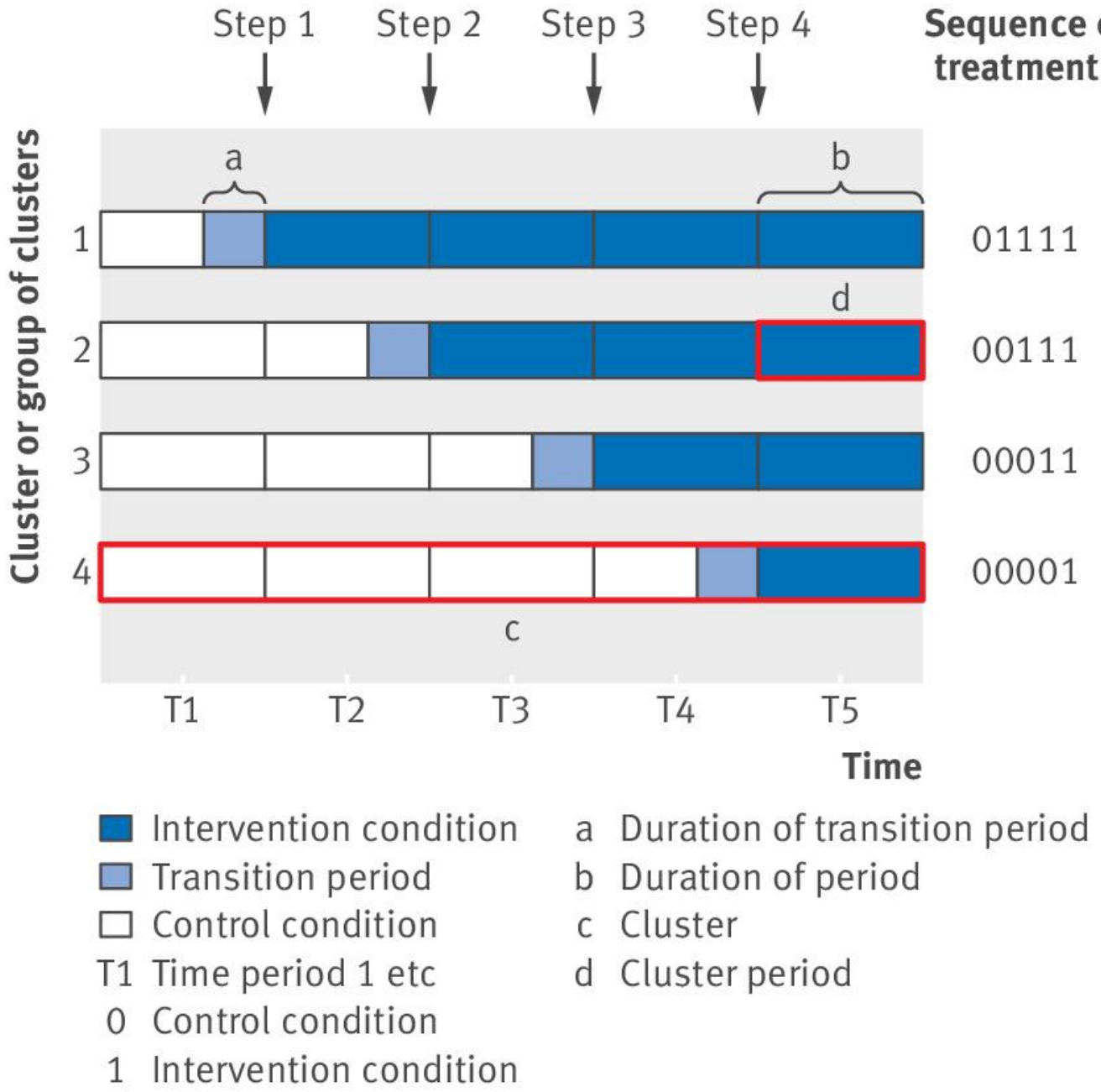
Multi-structured depression management in nursing homes

- 17 nursing homes in the Netherlands were randomly assigned to one of five dates to introduce an intervention designed to promote the diagnosis and management of depression between 2009-2011.
- 3-4 homes randomized approximately every four months.
- Residents and staff were blinded to the allocation (informed consent was elicited from residents at the start of the study,)
- Primary outcome (prevalence of depression) was analyzed using a linear mixed model, adjusted for time and a random nursing home effect, and allowing for repeated measures – estimated 7.3% reduction (95% CI 0.9, 13.7) after the intervention was introduced.

Leontjevas R et al. Lancet 2013; 381: 2255-2264, as cited in Hemming 2015.

https://github.com/THOMASELOVE/stepped-wedge

A Stepped Wedge Design with Four Clusters and Five Measurement Periods



Hemming K Taljaard M et al (2018) https://www.bmj.com/content/363/bmj.k1614

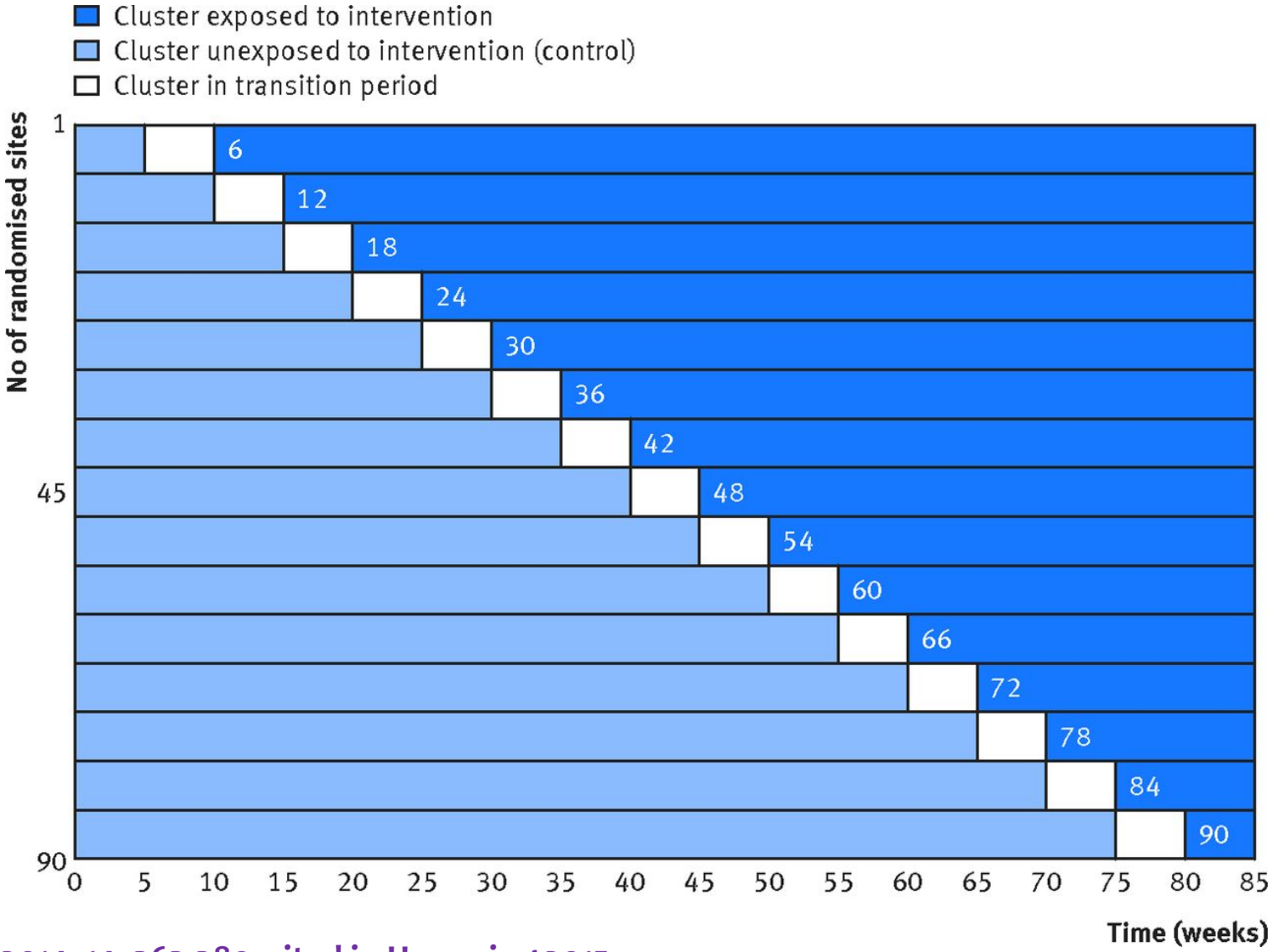
The EPOCH Trial

- Cross-sectional stepped wedge cluster randomised trial of a service delivery intervention to improve the care of patients undergoing emergency laparotomy
- Complex intervention, including quality improvement and an integrated care pathway.
- Rolled out sequentially to 90 hospitals, with six clusters of 15 geographically close hospitals (that is, clustering within clustering) switching from control to intervention every 5 weeks at 15 different time points, with a 5 week transition period in each cluster (see figure in next slide).

Hemming K, Girling A. Stata J 2014; 14: 363-380, cited in Hemming 2015.

https://github.com/THOMASELOVE/stepped-wedge

The EPOCH Trial Schematic



Hemming K, Girling A. Stata J 2014; 14: 363-380, cited in Hemming 2015.

https://github.com/THOMASELOVE/stepped-wedge

The EPOCH Trial

- The primary outcome is 90-day mortality, and no individual patient recruitment is needed.
- Approximately 18 patients will meet the inclusion criteria (>40 years old and undergoing emergency laparotomy) per hospital per 5-week step, so the total sample size is about 27,500 patients.
- The trial is powered to detect a change in 90-day mortality from 25% to 22% at 90% power and 5% significance.

Hemming K, Girling A. Stata J 2014; 14: 363-380, cited in Hemming 2015.

https://github.com/THOMASELOVE/stepped-wedge

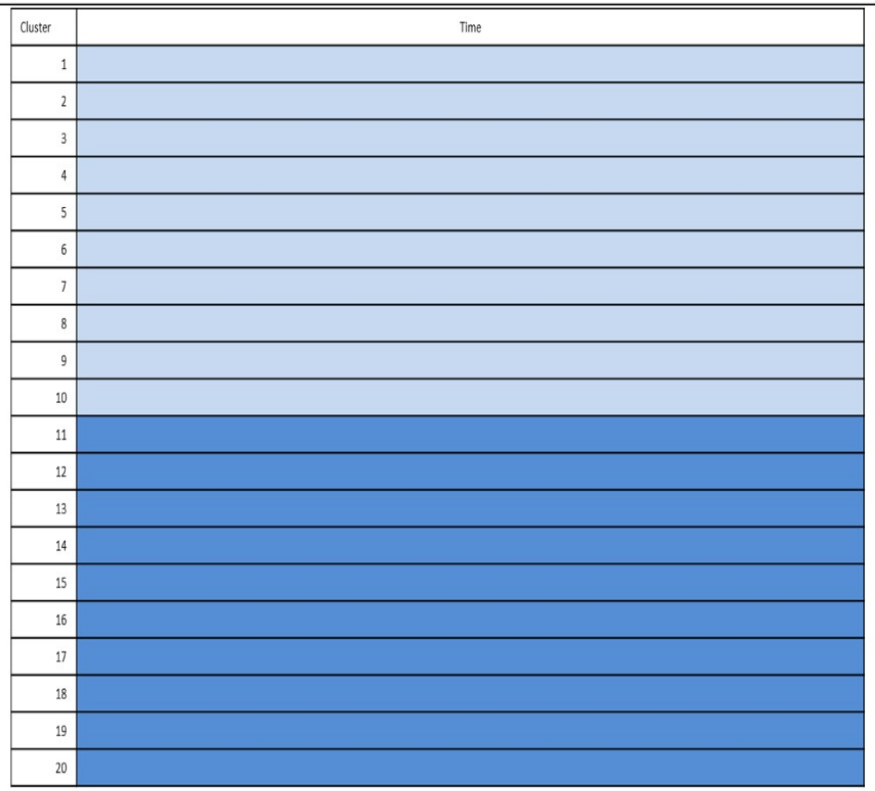
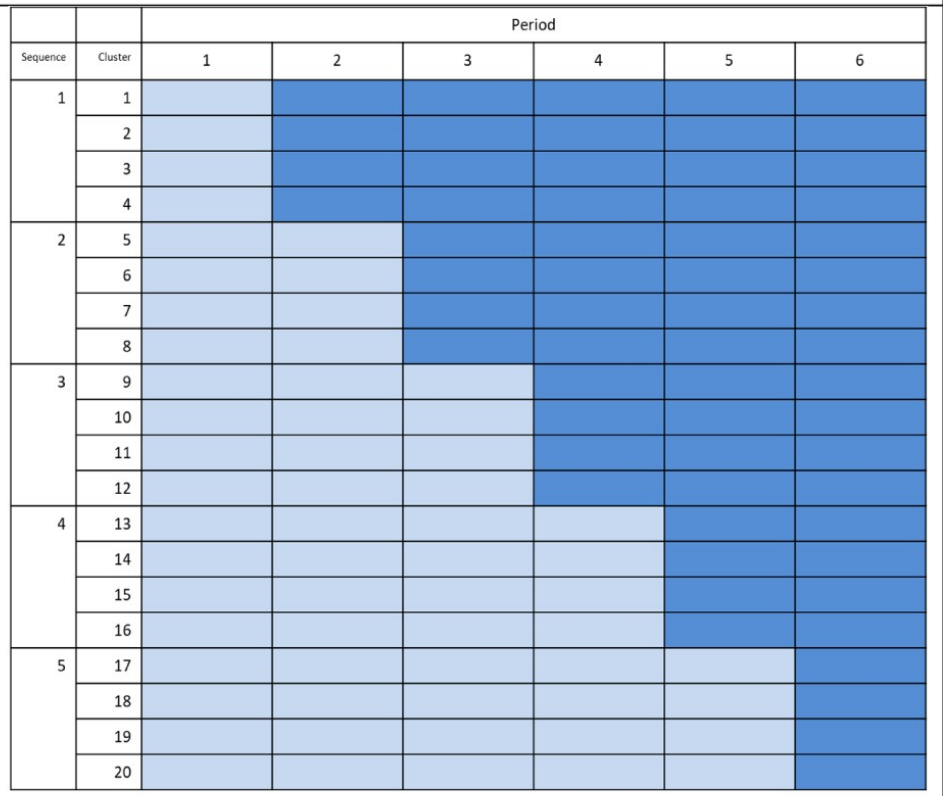
Stepped Wedge Design Features

- The clusters switching at each step are usually independent of one another; however, they might be related in some way, in which case a multilevel element is introduced into the design (as in EPOCH.)
- Usually designed so that approximately equal numbers of clusters switch at each step.
- Some designs (like EPOCH) make special allowance for the length of time it takes to embed the intervention into a cluster. During such transition periods the cluster cannot be considered as either exposed or not exposed

Rationale for the Stepped Wedge

- Pragmatic study design to account for various constraints
  - A key stakeholder thinks there's already sufficient evidence of effectiveness, but the researcher is less certain.
  - Logistical constraints with complex interventions
    - A "fair" way to determine rollout order under logistical constraints?
- Very popular among people applying for grant funding
- Some better reporting tools are now available (CONSORT specs.)
- Some potential for improving statistical performance
- Some improvement over a non-experimental design

https://github.com/THOMASELOVE/stepped-wedge

	(a): Parallel cluster randomised trial (parallel-CRT)		(b): Stepped-wedge cluster randomised trial (SW-CRT)
Control clusters / Intervetnion clusters		Clusters allocated to sequences	
	Example includes 20 clusters, 10 allocated to the intervention condition and 10 to the control condition.		Example includes 20 clusters, with 4 clusters allocated to each of the 5 sequences and measurements taken over 6 time periods.
	Participants might all be recruited before randomistion; or recruited continously throughtout the trial (as intended in this depiction).		Participants might be all recruited before randomisation and repeatedly measured; cross-sectional samples might be taken once in each measurement period (on different or the same participants); or participants might be recruited continuously throughout the trial (and either measured once or repeatedly).

Hemming 2020 IJE

Stepped Wedge vs. Parallel Cluster Studies

Choose Parallel CRT if …

- Clusters are relatively homogeneous (ICC is small)
- Confounding effect over time is expected to be large
- Clusters are relatively small, and you have a large number of clusters available.
- You have the option to do it, pragmatically.

Consider Stepped Wedge if …

- You anticipate big differences between clusters (large ICC)
- You expect minimal underlying temporal trends
- Clusters are relatively large, or if you have a relatively small number of clusters
- You intend a baseline period with all sites “unexposed”.

https://github.com/THOMASELOVE/stepped-wedge

The ICC
(Intra-Cluster Correlation Coefficient)

$$\textbf{ICC} \text{ for an outcome} = \frac{\text{Variance in the outcome attributable to differences BETWEEN CLUSTERS}}{\text{TOTAL variance in the outcome}}$$

- ICC = 0.01 means 1% of the variance in the outcome is attributable to differences between clusters
- ICC is interpreted as if it were "R^2"

Reporting of Stepped Wedge Trials

- Several systematic reviews have demonstrated poor reporting of key features of stepped wedge designs
- CONSORT additional materials specified in Hemming 2018
- For example, under sample size calculation
  - Level of significance, Power, Target difference
  - Variation of outcome (standard deviation or control proportion)
  - Number of clusters, Average cluster size
  - Assumed correlation structure (ICC)
  - Assumed within-person correlation if measures are repeated

Misconception #1

- We should use a stepped wedge when there is an ethical imperative for all clusters to receive the intervention.
  - If the intervention is **known** to be effective, there is no ethical justification for withholding it from clusters or individuals. If it is argued that there is an ethical imperative for all clusters to receive the intervention, then there should be an ethical imperative for all individuals to receive the intervention without delay.
  - Only when equipoise holds does it become reasonable to expose some clusters to the intervention but not others, so that robust evidence can be generated.
- There is no ethical benefit of using the stepped wedge in the case of a known-to-be-effective intervention, though its use can be ethical in light of clinical equipoise.

Misconception #2

- The stepped wedge is commonly perceived to increase the number of individuals or clusters exposed to an intervention of unknown effectiveness or perhaps harm.
- In a stepped wedge design, all clusters do eventually receive the intervention.
- However, this doesn't mean that a parallel CRT wouldn't actually expose more clusters (it's not unusual for a parallel CRT to require more than twice the number of clusters in a stepped wedge design)

When should we use a stepped wedge?

- There is evidence in support of the intervention (for example, known to be effective at individual level but uncertain about policy level), or there is resistance to a parallel design in which only half of the clusters receive the intervention.
- The intervention is a service delivery or policy change that can be implemented without the need for individual participant consent.
- The outcome, or at least some important outcomes, may be available from routinely collected data (no patient questionnaires).
- The intra-cluster correlation is anticipated to be high or cluster sizes large so that a cross-sectional stepped wedge design is likely to be more efficient than the simple parallel cluster design.

Hemming, Haines et al. BMJ 2015; 350: h391 https://doi.org/10.1136/bmj.h391

Reasons to be cautious about a stepped-wedge cluster randomized trial

- When the intra-cluster correlation is low (or the cluster size small) the stepped wedge cross-sectional study can be an inefficient design compared with a simple parallel cluster design.
- The study has a cohort or open cohort design, for which there are currently no methods developed to determine power available.
- When the outcome requires individual participant data collection (without blinding), lack of concealment of allocation is likely to mean a risk of differential selection of participants between arms.
- It is unlikely that clusters will be able to follow the randomization schedule.

Hemming, Haines et al. BMJ 2015; 350:h391 https://doi.org/10.1136/bmj.h391

https://github.com/THOMASELOVE/stepped-wedge

Key Sources for this Presentation

- The work of Karla Hemming, Monica Taljaard and colleagues
  - Hemming K Haines TP et al BMJ 2015; 350;h391
  - Hemming K Taljaard M et al BMJ 2018; 363:k1614
  - Hemming K Taljaard M & Grimshaw J Trials 2019; 20: 68
  - Hemming K & Taljaard M IJE 2020: 49(3) 1043-1052
  - Hemming & Taljaard video: When is a stepped-wedge cluster randomized trial a good design choice? Research Outreach
  - Hemming, Taljaard et al. BMJ 2020; 371:m3800
- Cebul RD Dawson NV Love TE various short courses (SMDM, ICHPS) on Cluster Randomized Trials, 2005-2010.
- Hussey & Hughes 2007 Contemp Clin Trials 28(2); 182-191