

Methods: Mind the Gap
Webinar Series

When is the Stepped Wedge Study a Good Study Design Choice?



Presented by:

Karla Hemming, Ph.D.

University of Birmingham, UK



National Institutes of Health
Office of Disease Prevention



UNIVERSITY OF
BIRMINGHAM

WHEN IS A STEPPED-WEDGE CLUSTER RANDOMISED TRIAL A GOOD STUDY DESIGN CHOICE?

JAN 2022;

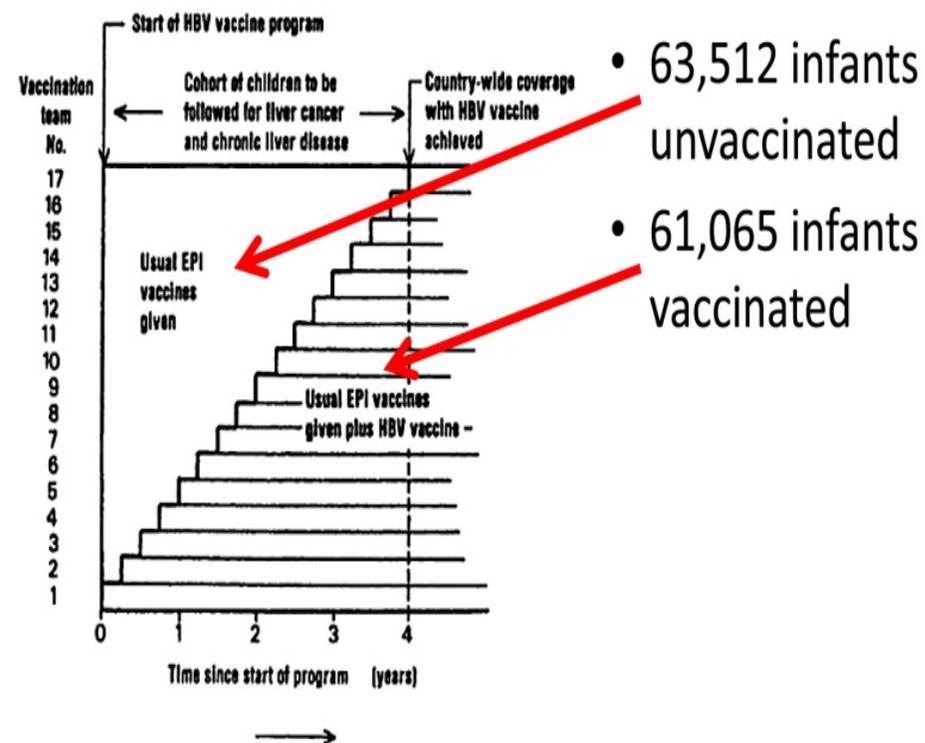
KARLA HEMMING, UNIVERSITY OF BIRMINGHAM
WITH ACKNOWLEDGEMENT TO MONICA TALJAARD,
OHRI

BRIEF HISTORY - THE FIRST SW-CRT

► Gambia Hepatitis Intervention Study (GHIS):

- Nation-wide trial of the HBV vaccine
- Vaccine programme rolled-out in a phased way (random) over 4 years (1986 to 1990)
- Outcome: Hepato Cellular Carcinoma (HCC) ascertained from a national cancer registry. 30 years follow-up
- *“This was because instantaneous universal vaccination in the country was impossible for logistic and financial reasons.”*

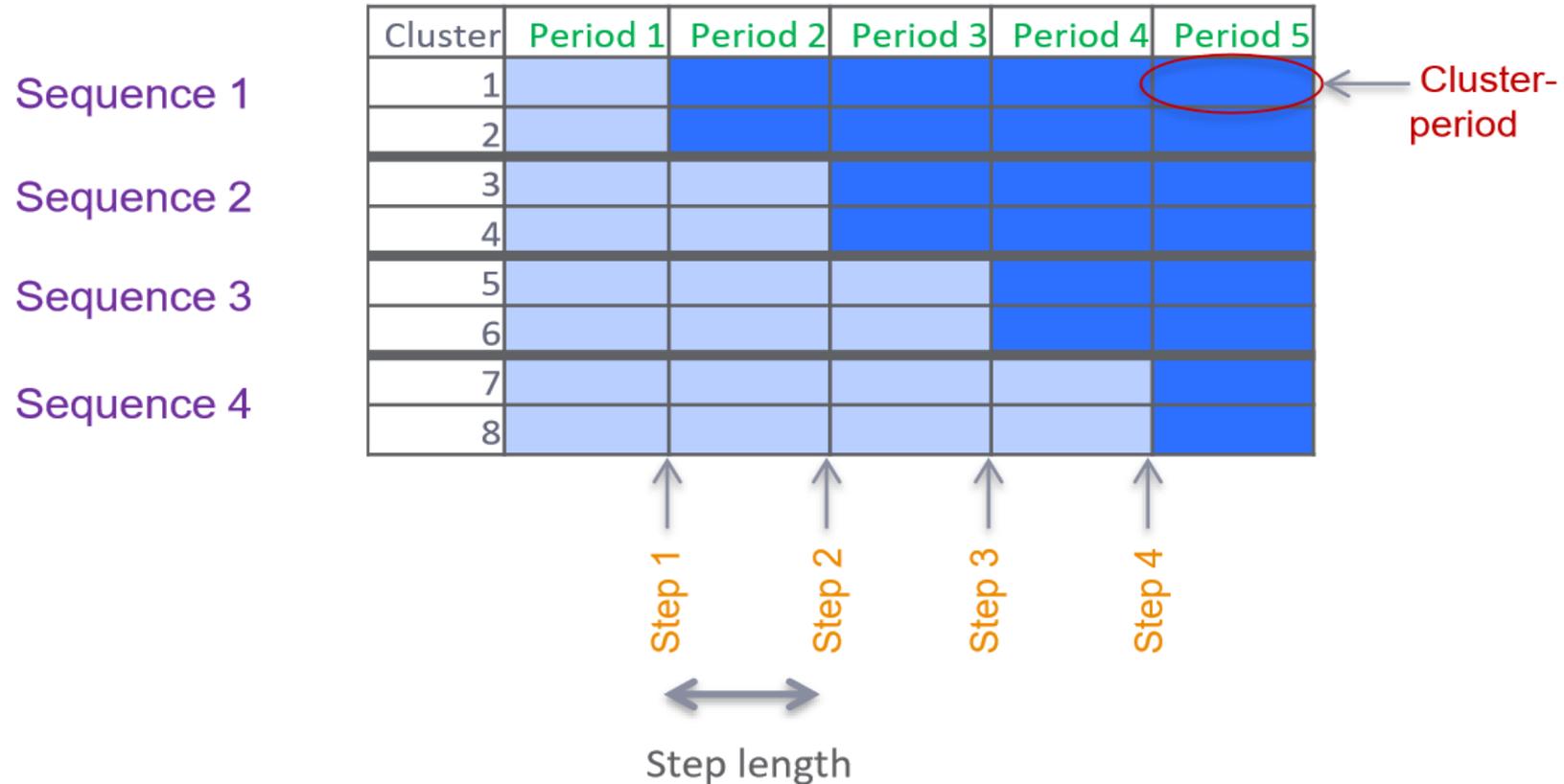
Stepped wedge design



Hall et al., 1987

Shimakawa Y, Lemoine M, Mendy M, Njai HF, D'Alessandro U, Hall A, Thursz M, Njie R. Population-based interventions to reduce the public health burden related with hepatitis B virus infection in the gambia, west Africa. Trop Med Health. 2014 Jun;42(2 Suppl):59-64; The Gambia Hepatitis Intervention Study. The Gambia Hepatitis Study Group. Cancer Res. 1987 Nov 1;47(21):5782-7

TERMINOLOGY

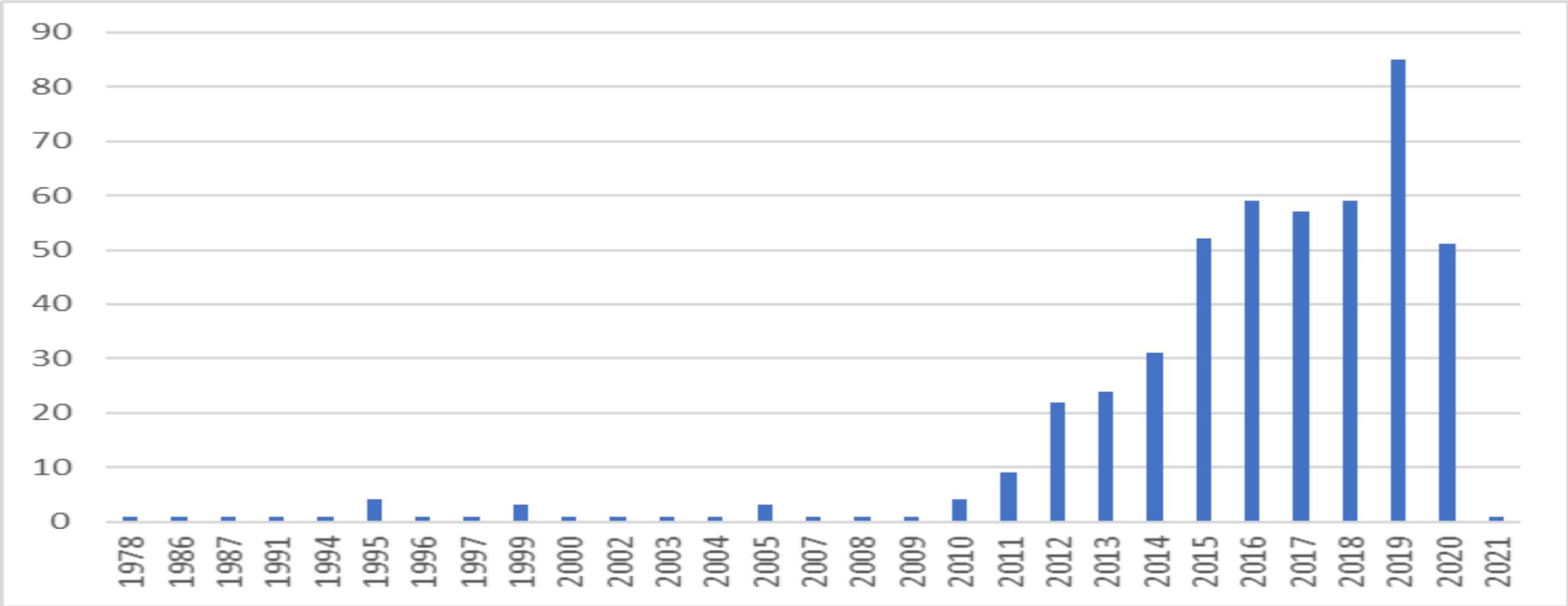


Reporting of Stepped-Wedge Cluster Randomised Trials: Extension of the CONSORT 2010 statement with explanation and elaboration
 K Hemming, M Taljaard, JE McKenzie, R Hooper, A Copas, JA Thompson, M Dixon-Woods, A Aldcroft, Doussau, M Grayling, C Kristunas, CE Goldstein, MK Campbell, A Girling, S Eldridge, MJ Campbell, RJ Lilford, C Weijer, A Forbes, JM Grimshaw BMJ 2019

OBJECTIVES

- ▶ Motivate the need for careful justification for design choice
 - Outline main *Risks of Bias (RoB)* for *parallel-CRTs*
 - Consider additional *Risks of Bias (RoB)* associated with a *stepped-wedge* trial
 - Consequently hypothesize the SW-CRT is at greater risk of bias than the parallel-CRT
- ▶ Propose a set of *justifications* for when the SW-CRT might represent a preferable design compared to the parallel-CRT
- ▶ Outline how these differ from the *common misconceptions* of appropriate justifications

THE POPULARITY OF THE SW-CRT IS CLEARLY GROWING



Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC Med Res Methodol. 2006;6:54
Citations of “step* wedge” classified as “randomised controlled trial” by PubMed Feb 2021 (total 433)

REPORTED REASONS FOR ADOPTING A SW-CRT

The most commonly cited reasons for choosing a SW-CRT are its *perceived logistical, social and ethical benefits*

Reported Rationale	Number (%)
None	12 (38%)
<i>Reported reasons</i>	
All sites get intervention	7 (22%)
Resource constraints (staggering)	12 (38%)
Ethical	8 (25%)
Methodological (power)	3 (9%)

Taljaard M, Hemming K, Shah L, Giraudeau B, Grimshaw JM, Weijer C. Inadequacy of ethical conduct and reporting of stepped wedge cluster randomized trials: Results from a systematic review. Clin Trials. 2017 Aug;14(4):333-341.

WHY IS CAREFUL JUSTIFICATION NEEDED?

- ▶ Council for International Organizations of Medical Sciences (*CIOMS*) says:
 - The ethical justification for undertaking health-related research involving humans is its *scientific and social value...*
 - Scientific value refers to the ability of a study to produce *reliable, valid information...*
 - Therefore, researchers, sponsors, research ethics committees, and health authorities, must ensure that proposed studies are scientifically sound...
- ▶ To produce reliable information the study should be at *low risk of bias*

ASSESSING RISK OF BIAS IN RANDOMISED TRIALS

- ▶ Risks of bias in randomised trials have been carefully described in the Cochrane systematic review *Risk of Bias tool (RoB2)* [Higgins 2016]
- ▶ An adaptation of the main guidance has been made for parallel cluster trials

Higgins JPT, Sterne JAC, Savović J, Page MJ, Hróbjartsson A, Boutron I, Reeves B, Eldridge S. A revised tool for assessing risk of bias in randomized trials In: Chandler J, McKenzie J, Boutron I, Welch V (editors). Cochrane Methods. Cochrane Database of Systematic Reviews 2016, Issue 10 (Suppl 1).

RISKS OF BIAS PARALLEL-CRTS

- ▶ **Domain 1a:** bias arising from the randomisation process
- ▶ **Domain 1b:** bias arising from the timing of identification and recruitment of participants in relation to the timing of the randomisation
- ▶ **Domain 2:** bias due to deviations from the intended intervention
- ▶ **Domain 3:** bias due to missing outcome data
- ▶ **Domain 4:** bias due to the measurement of the outcome
- ▶ **Domain 5:** bias due to the selection of the reported result

RISKS OF BIAS PARALLEL-CRTS

- ▶ **Domain 1a:** bias arising from the randomisation process
- ▶ **Domain 1b:** bias arising from the timing of identification and recruitment of participants in relation to the timing of the randomisation
- ▶ **Domain 2:** bias due to deviations from the intended intervention
- ▶ **Domain 3:** bias due to missing outcome data
- ▶ **Domain 4:** bias due to the measurement of the outcome
- ▶ **Domain 5:** bias due to the selection of the reported result

ADDITIONAL DOMAIN*

▶ Domain 6: Analytical biases

*Not included in RoB 2.0 but might be of importance with a small number of clusters.

DOMAIN 1A: BIAS ARISING FROM THE RANDOMISATION PROCESS

- ▶ In a parallel-CRT, randomisation refers to the process of allocating clusters to arms
- ▶ Biases can arise if this allocation is not random or is not adhered to at the level of the cluster
- ▶ Two essential ingredients:
 1. Truly *random*
 2. *Concealed* at the time of recruitment of clusters (perhaps under appreciated)

HOW RANDOMISATION BIAS CAN ARISE

“I’ve been told that the next clusters really needs to get this new intervention. Can they go in the treatment arm?”



“The next cluster on the allocation list will get the control. So hold off randomising them for a few days”



Subversion of the allocation
Usually unintentional but non the less consequential

HOW TO PREVENT RANDOMISATION BIAS

- ▶ Concealed randomisation implementation
 - Independent person undertakes the randomisation (e.g., “Trials unit” or “Data coordinating centre”)
 - Randomisation concealed from the cluster until the cluster participation is ratified (i.e. all cluster level approvals obtained)

So, apply the same level of rigor to the randomisation process as
in a patient randomised trial!

DOMAIN 1B: BIAS ARISING FROM IDENTIFICATION OR RECRUITMENT OF PARTICIPANTS WITHIN CLUSTERS

- ▶ When identification and recruitment of participants occurs with *knowledge of the treatment allocation* this can lead to differential recruitment and identification between treatment conditions
- ▶ This is about internal validity not external validity
- ▶ Sometimes referred to as “selection bias” (but this is ambiguous)

IDENTIFICATION AND RECRUITMENT BIASES EXPLAINED

- ▶ Arises when identification or recruitment take place with knowledge of the treatment condition to which the cluster has been randomized:
 - Concern in trials with *post randomisation recruitment*
 - Also of concern in trials with *post randomisation identification* of participants
 - Probably most important source of bias in a cluster trial

VERY HARD TO DETECT FROM BASELINE TABLES

Table 1 Demographic data of the study groups

	AMP (n = 142)	PEN (n = 141)
Gestational age (weeks)		
Median (quartiles)	31 (27–34)	31 (28–35)
>36 weeks – n (%)	29 (20)	28 (20)
<28 weeks – n (%)	41 (29)	34 (24)
<26 weeks – n (%)	24 (17)	21 (15)
Birthweight (g)		
Median (quartiles)	1467 (920–2553)	1500 (960–2343)
<1501 g – n (%)	73 (51.4)	72 (51.1)
<1001 g – n (%)	36 (25.4)	39 (27.7)
<751 g – n (%)	19 (13)	15 (11)
M/F sex – n	78/64	85/56
Mean Apgar score at 5 min	6.6 ± 1.5	6.3 ± 1.6
Ventilated – n (%)	99 (70)	116 (82)
Surfactant – n (%)	81 (56)	88 (62)
Caesarean section – n (%)	77 (54)	80 (57)
Multiple pregnancies – n (%)	34 (24)	23 (17)
Chorionamnionitis – n (%)	21 (15)	30 (21)
PROM >16 h – n (%)	25 (18)	28 (20)
Prenatal glucocorticoids	86 (61)	71 (50)
Maternal antibiotic therapy – n (%)		
During pregnancy	38 (27)	27 (19)
During delivery	51 (36)	46 (33)

PROM, premature rupture of membranes.

Big red flag is differences in numbers recruited across arms

Difficult to identify if imbalances chance or a reflection of bias

Metsvaht T, Ilmoja ML, Parm Ü, Maipuu L, Merila M, Lutsar I. Comparison of ampicillin plus gentamicin vs. penicillin plus gentamicin in empiric treatment of neonates at risk of early onset sepsis. *Acta Paediatr.* 2010 May;99(5):665-672.

EMPIRICAL EVIDENCE IN CRTS (LOTS!)

Baseline Imbalance

- ▶ On average over many CRTs there is a *significant difference in the average age* of participants between treatment and control arm
- ▶ This does not happen in individually randomised trials (on average)

Process evidence

- ▶ Many CRTs *recruit participants after randomisation*
- ▶ Most do not clearly report if this recruitment was blind to allocation
- ▶ Thus, many CRTs do not appear to conceal the allocation.

RECRUITMENT BIAS

“My cluster has been allocated to the control arm. I don’t think I will ask any very sick patients to participate.”



“You should try to recruit everyone who is eligible. Better still – lets hand over the recruitment to someone who doesn’t know the allocation”



PREVENTION OF IDENTIFICATION AND RECRUITMENT BIAS

▶ Prevention strategies:

1. Recruit all participants before randomisation (not always possible)
2. Blind treatment conditions (not always possible)

▶ Mitigation strategies:

1. Minimizing the number of eligibility criteria. Including only *objective eligibility criteria* (minimize identification bias)
2. *Recruitment* by someone independent to the trial who is *blind* to cluster status (minimize recruitment bias)

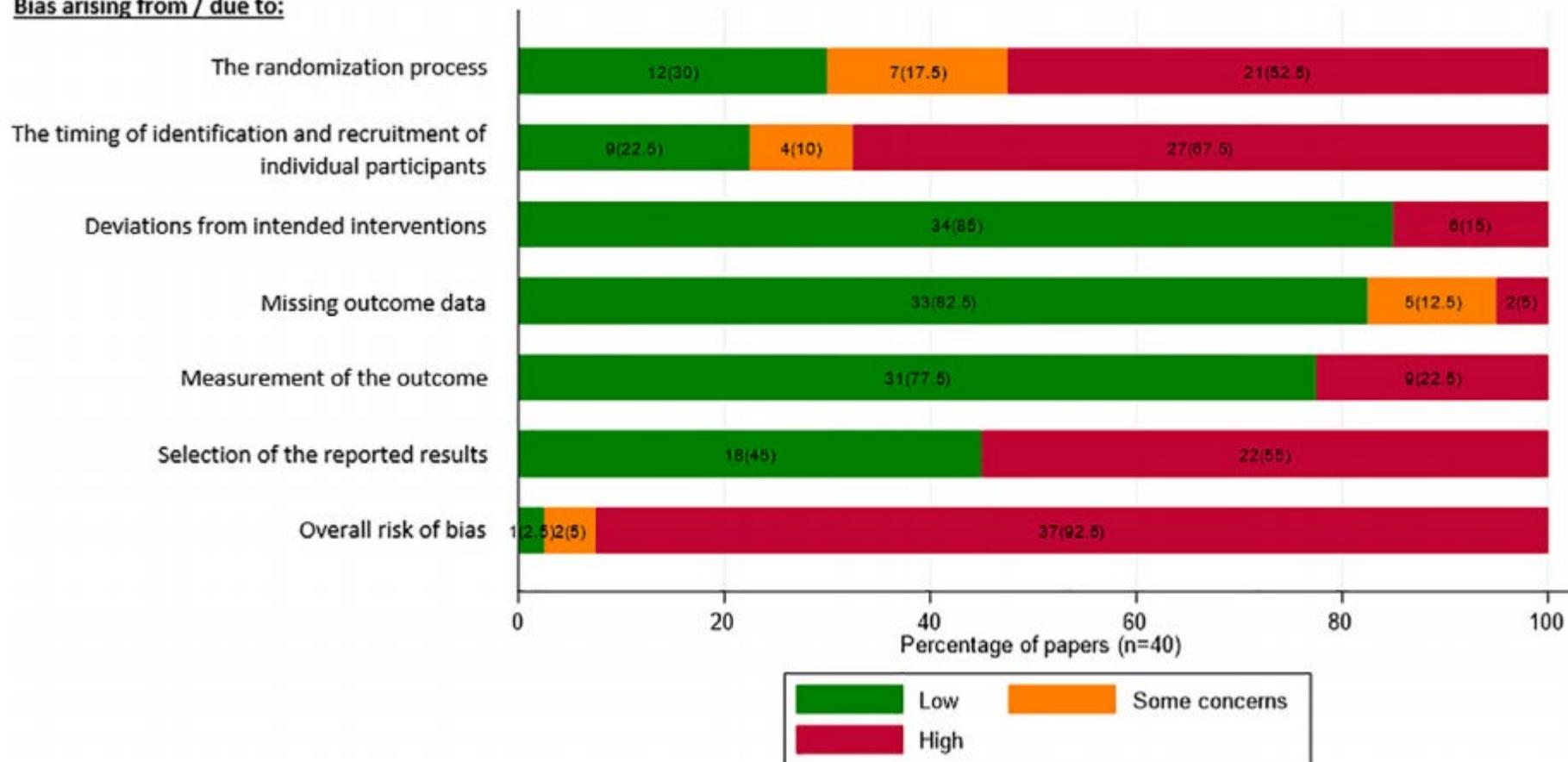
DOMAIN 6: ANALYTICAL BIASES

- ▶ Analytical biases refer to bias in the estimation of *treatment effects* and *confidence intervals* because of misspecification of the statistical model.
- ▶ Analysis must allow for
 - Clustered nature of the design
 - Small numbers of clusters (t-test on $K-2$ DoF or “small sample corrections”)
- ▶ This sort of bias might be considered “a fixable fault” as it does not necessarily require pre-planning (although it is data greedy)

Leyrat C, Morgan KE, Leurent B, Kahan BC. Cluster randomized trials with a small number of clusters: which analyses should be used? Int J Epidemiol. 2018 Feb 1;47(1):321-331; Kahan BC, Forbes G, Ali Y, Jairath V, Bremner S, Harhay MO, Hooper R, Wright N, Eldridge SM, Leyrat C. Increased risk of type I errors in cluster randomised trials with small or medium numbers of clusters

SUMMARY: RISK OF BIAS IN CRTS

Bias arising from / due to:



Easter C, Thompson JA, Eldridge S, Taljaard M, Hemming K. Cluster randomized trials of individual-level interventions were at high risk of bias. *J Clin Epidemiol.* 2021 Jun 29;138:49-59.

SW -CRTS

- ▶ No formal adoption of the RoB2.0 tool for SW-CRTS
- ▶ Here we outline some important extra considerations

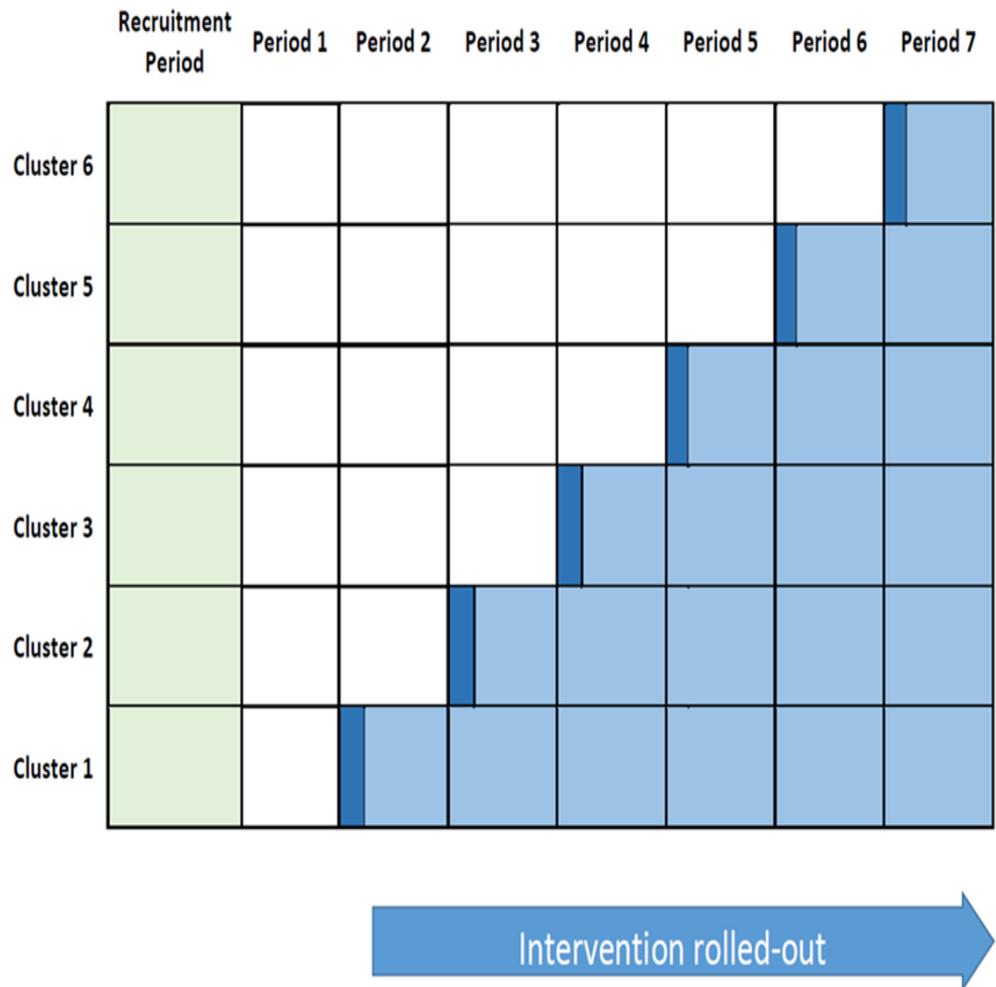
RECAP: STEPPED WEDGE CLUSTER RANDOMISED DESIGN

- ▶ The SW-CRT design is becoming increasingly a common design choice
- ▶ It shares many characteristics in common with CRTs, especially other repeated measure CRT designs
- ▶ However, it brings out some significant complexities:
 1. Temporal confounding
 2. Within-cluster contamination
 3. Deviations from randomisation
- ▶ In this section we will look to understand how these additional complexities can lead to bias

1. ANALYTICAL (TEMPORAL) BIASES

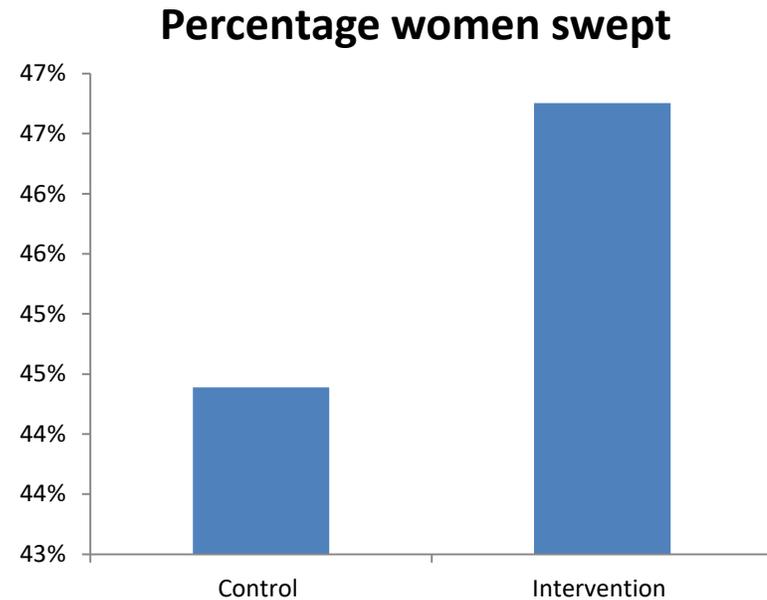
- ▶ Chosen analytical approach to ensure unbiased estimates of the intervention effect and its standard error
- ▶ Bias:
 - Analysis complicated by the fact that underlying changes over time – called secular trends – may be confounded with the intervention effect
 - *Thus, an apparent effect due to the intervention may in fact be due to natural changes over time*
- ▶ Standard errors:
 - Within-cluster correlations might need to take a more complicated form than a simple exchangeable

WHY TIME IS A CONFOUNDER



- ▶ A variable is a confounder if it is associated with both:
 - Treatment
 - Outcome
- ▶ In SW-CRT, time is a confounder by design!
 - Likelihood of *treatment* increases with time
 - And if *outcome changes over time* in absence of any intervention (secular trend)

EXAMPLE: RESULTS FROM A SW-CRT WHERE THERE IS AN UNDERLYING SECULAR TREND

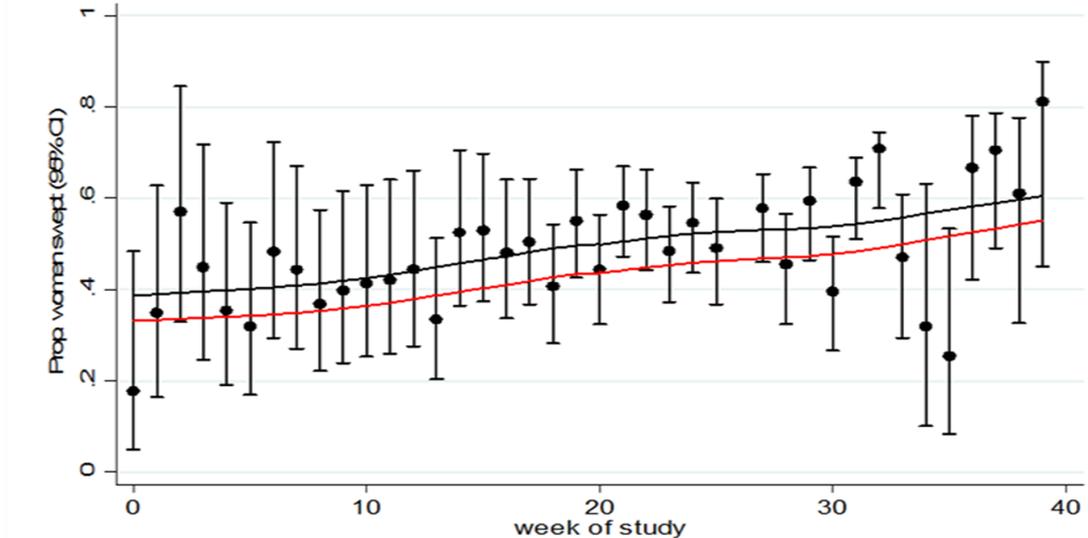


Unadjusted RR = 1.06 95% CI (0.97, 1.16)

Apparent effect positive

Looks like intervention increasing proportion of women swept

Adjustment for time effects reverses apparent direction of effect



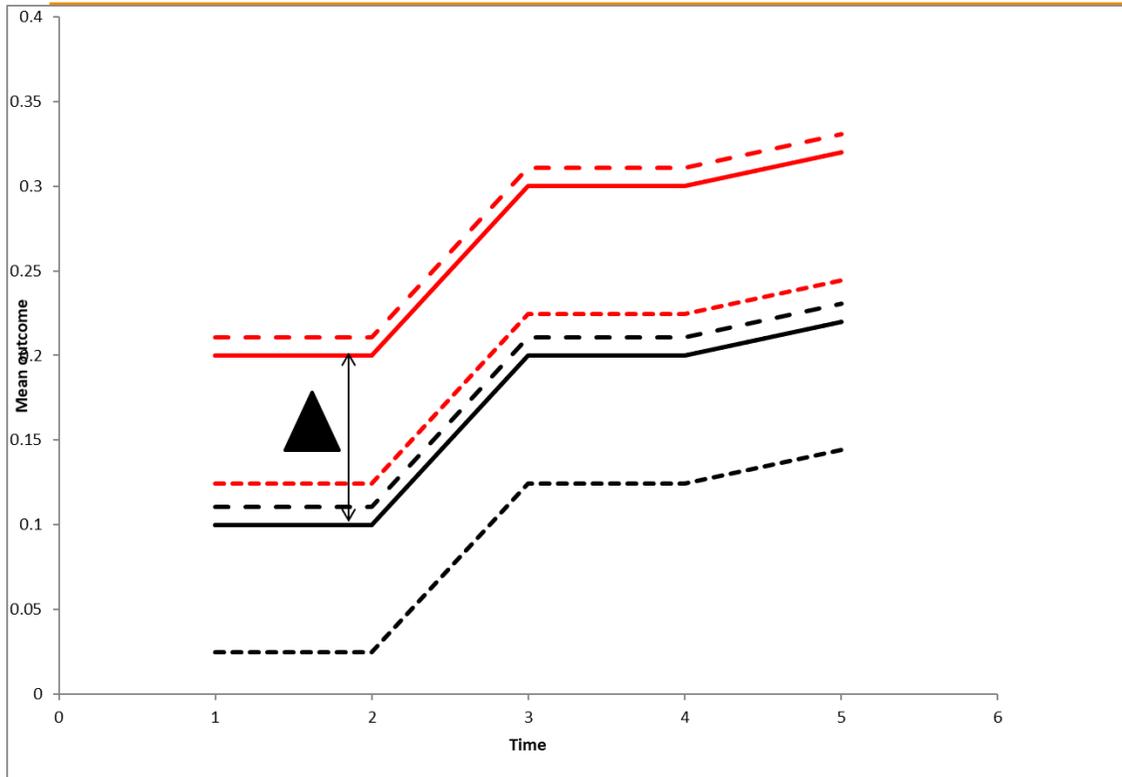
Black: secular trend (control clusters)

Red: adjusted for time (intervention clusters)

Adjusted RR = 0.88 95% CI (0.69, 1.12)

Kenyon S et al . Evaluation of a bespoke training to increase uptake by midwifery teams of NICE Guidance for membrane sweeping to reduce induction of labour: a stepped wedge cluster randomised design. *Trials*. 2017 Jul 27;18(1):357.

MODEL BASED ASSUMPTIONS



Cluster 1: -----
Cluster 2: - - -
Cluster 3:

- ▶ All clusters follow the same secular trend
- ▶ Illustrated on graph by varying dashed lines
- ▶ Common intervention effect
 - Red treatment condition
 - Black control condition

WHAT DOES THIS MEAN IN PRACTICE?

- ▶ Analysis complicated by the fact that adjustment for time important
- ▶ But, most model based analysis make the assumption of a common secular trend
 - Likely *untenable* in many situations
 - *Not testable*
- ▶ Worth noting these types of biases are unlikely to affect parallel CRTs

Thompson JA, Fielding KL, Davey C, Aiken AM, Hargreaves JR, Hayes RJ. Bias and inference from misspecified mixed-effect models in stepped wedge trial analysis. *Stat Med*. 2017 Oct 15;36(23):3670-3682.

Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials*. 2007;28(2):182–191

HOW TO PREVENT ANALYTICAL BIASES

- ▶ At the design stage:
 - Randomising a large number of clusters
 - More homogenous clusters (to meet assumption of a common secular trend)
- ▶ At the analysis stage:
 - Always adjust for secular trends and allow for non-exchangeable correlations
 - Permutation tests

“Time is a strong candidate for having a confounding effect in almost every SW-CRT and this logical reasoning should dictate its criteria for inclusion rather than a reliance on statistical significance testing or lack of, which is not useful for determination of confounding”

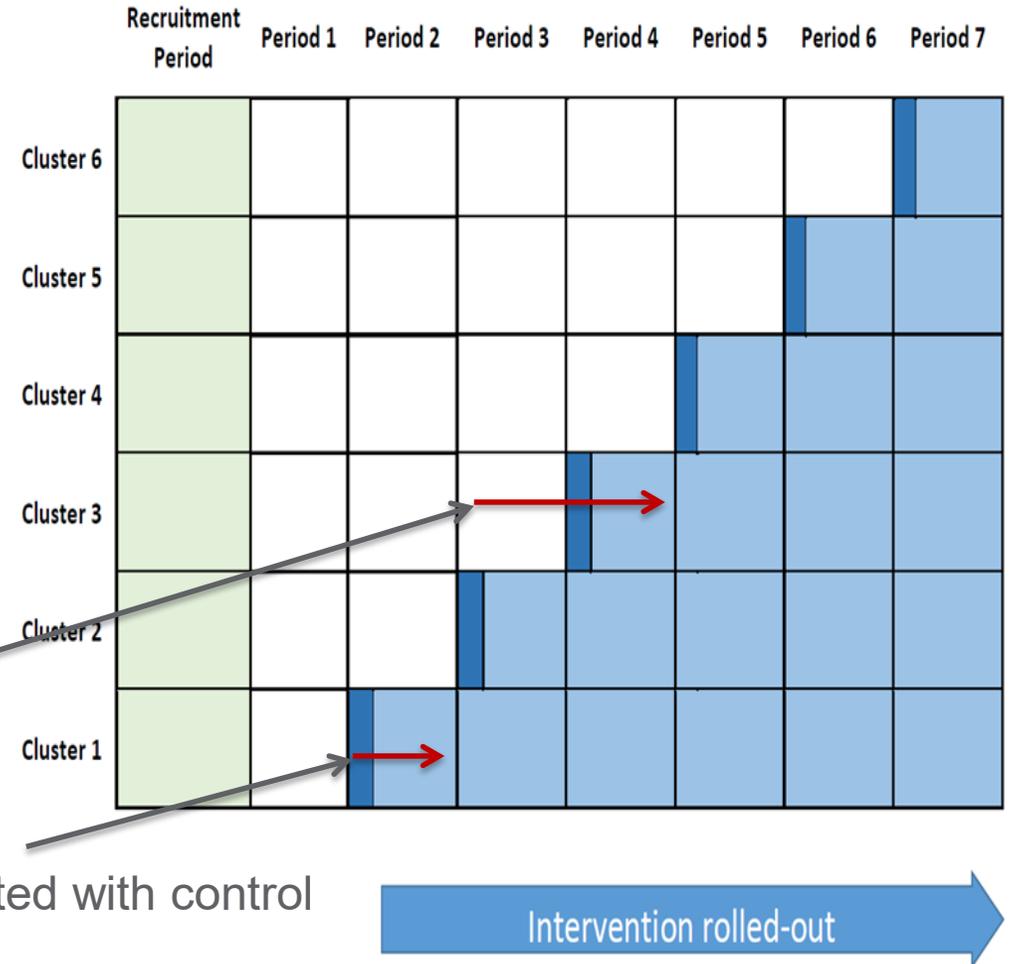
Hemming K, Taljaard M, Forbes A. Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples. *Trials*. 2017 Mar 4;18(1):101. Thompson J, Davey C, Hayes R, Hargreaves J, Fielding K. swpermute: Permutation tests for Stepped-Wedge Cluster-Randomised Trials. *Stata J*. 2019;19(4):803-819

2. DEVIATIONS FROM INTENDED INTERVENTION

- ▶ ***Within-cluster contamination*** (carry-over effects) refers to biases due to collected data under the control condition becoming contaminated by the intervention condition (or visa-versa)
- ▶ This can occur at two levels:
 - Individual-level
 - Cluster-level

Contaminated with intervention

Contaminated with control



WITHIN-CLUSTER CONTAMINATION AT LEVEL OF CLUSTER

- ▶ Observation collected under the control condition is *contaminated by the intervention* condition
 - e.g., when a provider in a site that is still in the control condition deliberately or inadvertently implements the intervention before the allocated time
- ▶ Alternatively, an observation collected under the intervention condition might become *contaminated by the control* condition
 - e.g., when a provider in a site that has already crossed to the intervention condition deliberately or inadvertently persists in applying the control intervention

WITHIN-CLUSTER CONTAMINATION AT LEVEL OF INDIVIDUAL

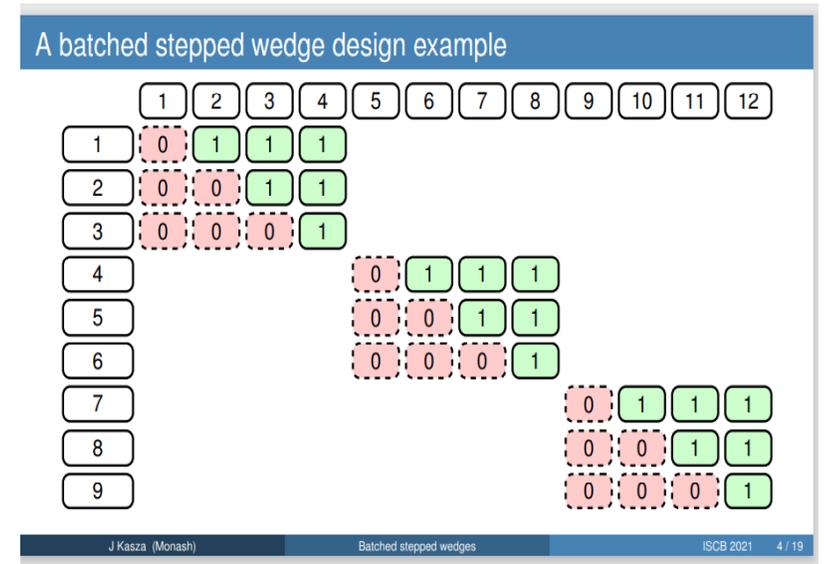
- ▶ In a setting where participants have a *long exposure to the intervention*:
 - e.g., patients in intensive care units:
 - where some patients have a prolonged stay and may still be in the intensive care unit at the time of crossing over
 - AND, they have a prolonged exposure to the intervention, perhaps the intervention is a change to the way care is delivered
 - It is possible that observations from individuals included in the control condition become contaminated with the intervention condition

HOW TO PREVENT WITHIN CLUSTER CONTAMINATION

- ▶ Solution:
 - *Transition periods* / careful design
 - In trials where participants have a *short exposure* to the intervention, e.g., in the so-called continuous recruitment short exposure design, this type of contamination is unlikely to arise.
- ▶ Worth noting this sort of contamination is unlikely to happen in a parallel CRT as clusters do not transition between treatment conditions

3. RANDOMIZATION BIASES

- ▶ *Non-adherence* to allocation schedule
 - Some clusters “going first” as they are “more needy”
 - Clusters join / randomized when ready to participate in trial
- ▶ Prevention strategies:
 - Revealing allocation as late as possible?
 - Revealing allocation as early as possible?
 - *Batch-design*



SUMMARY SO FAR

- ▶ SW-CRTs are subject to several risks of bias that might challenge the strength of the evidence generated from this design
- ▶ Some of these risks of bias may affect other types of cluster randomised designs too, but many appear to be greater under the SW-CRT
- ▶ The use of the SW-CRT must thus be carefully considered

JUSTIFICATIONS FOR THE SW-CRT

- ▶ Outline four justifications
 - None are hard and fast rules
 - Often these apparent justifications do not stand up to scrutiny
- ▶ The more that are satisfied the more likely the SW-CRT is the right choice
- ▶ Outline common misconceptions

WHAT NEEDS TO BE JUSTIFIED

- In the SW-CRT design justification includes:
 1. The use of cluster randomisation
 2. The need to roll out the intervention to all clusters
 3. The need for staggered roll-out of the intervention

WHY JUSTIFICATION IS NEEDED

- ▶ The SW-CRT requires a model based analysis:
 - Assumptions underpinning these models might not be appropriate
- ▶ The SW-CRT is more complicated in its design:
 - Risk that study might fail to deliver on its objectives
- ▶ The SW-CRT might increase number of individuals and clusters exposed to an intervention of unknown effectiveness:
 - Risk of increasing the number of participants exposed to a potentially harmful intervention

JUSTIFICATION 1: THE SW-CRT PROVIDES A MEANS TO CONDUCT A RANDOMISED EVALUATION

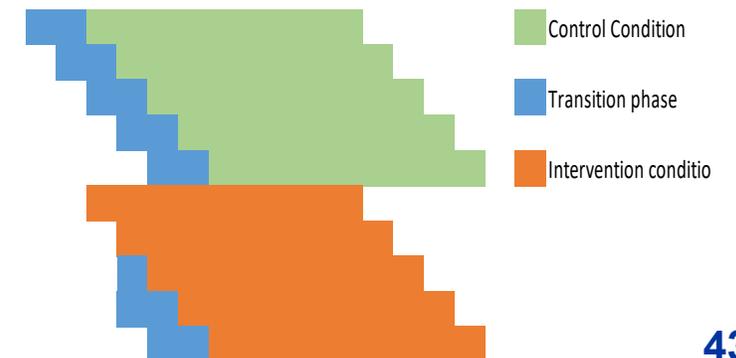
- ▶ Interventions are frequently rolled out without any robust randomised evaluation.
 - Sometimes the roll-out might be sequential because of a limited resource or capacity to roll-out to the entire health system simultaneously or
 - Because a gradual implementation allows the possibility to learn from earlier implementation in such a way that the intervention is adapted as more is learnt.
- ▶ If stakeholders can be *convinced to randomise* the order of the roll-out:
 - Using the SW-CRT might provide a means to both obtain a robust evaluation and allow staggering of the roll-out.
 - However, if stakeholders can be convinced of the benefits of randomising the order of the roll-out to align with a SW-CRT, then it might also be possible to *convince the stakeholders of the benefits of a parallel-CRT*

JUSTIFICATION 2: THE SW-CRT FACILITATES CLUSTER RECRUITMENT

- ▶ Sometimes, stakeholders might be reluctant to participate in a trial unless they are *guaranteed to receive the intervention*
 - Desires and a-priori beliefs might mean that stakeholders are more likely to participate in the trial when designed as a SW-CRT.
- ▶ Researchers could attempt to demonstrate that clusters are indeed more likely to participate in a SW-CRT trial
 - Should be *fully informed* about alternatives such as a parallel-CRT and waitlist designs
 - And, the benefits of randomisation to a parallel design

JUSTIFICATION 3: THE SW-CRT CREATES A LOGISTICALLY FEASIBLE DESIGN

- ▶ Sometimes there can be *limited resources* or capacity to roll-out to the entire health system *simultaneously*
 - *For example, can not feasibly roll out intervention to 10 clusters at the same time*
- ▶ Sequential roll-out in a SW-CRT can bring about its own logistical issues for example:
 - Organizing research ethics approvals for all centres in advance of the trial and ensuring that all centres are ready to implement the intervention according to the allocated schedule can be challenging
 - A parallel-CRT can also be conducted in a staggered way



JUSTIFICATION 4: THE SW-CRT HAS INCREASED STATISTICAL POWER

- ▶ The number of available clusters may be restricted based either on availability, willingness to participate, or limited trial budgets
 - In these circumstances, the SW-CRTs may achieve the desired power with fewer clusters than a parallel-CRT
 - In fact, with a small number of clusters, 80% or 90% power might not even be achievable in a parallel-CRT, whereas a SW-CRT can achieve 80% or 90% power with the same number of clusters
- ▶ Yet, in the setting of a **small number of clusters**:
 - SW-CRTs are the more powerful design they are also at greater risk of **analytical bias**
 - Small sample corrections must be used

OTHER CONSIDERATIONS

- ▶ Study duration:
 - Imperative to provide an evaluation of the intervention's effectiveness in a shorter amount of time
 - Whether parallel CRT or SW-CRT depends on specific circumstances
- ▶ Time to realise effect of intervention:
 - Intervention needs time to start working and affect outcomes.
 - This might take time for complex interventions

COMMON MISCONCEPTION

“The SW-CRT is ethically appropriate when the intervention is expected to do more good than harm.”

We're thinking stepped wedge. We feel there is a strong ethical argument for giving everyone the intervention.



Then you should give everyone the intervention.



freshspectrum.com &
steppedwedgehog.blog

Binik A. Delaying and withholding interventions: ethics and the stepped wedge trial. J Med Ethics. 2019 Oct;45(10):662-667.

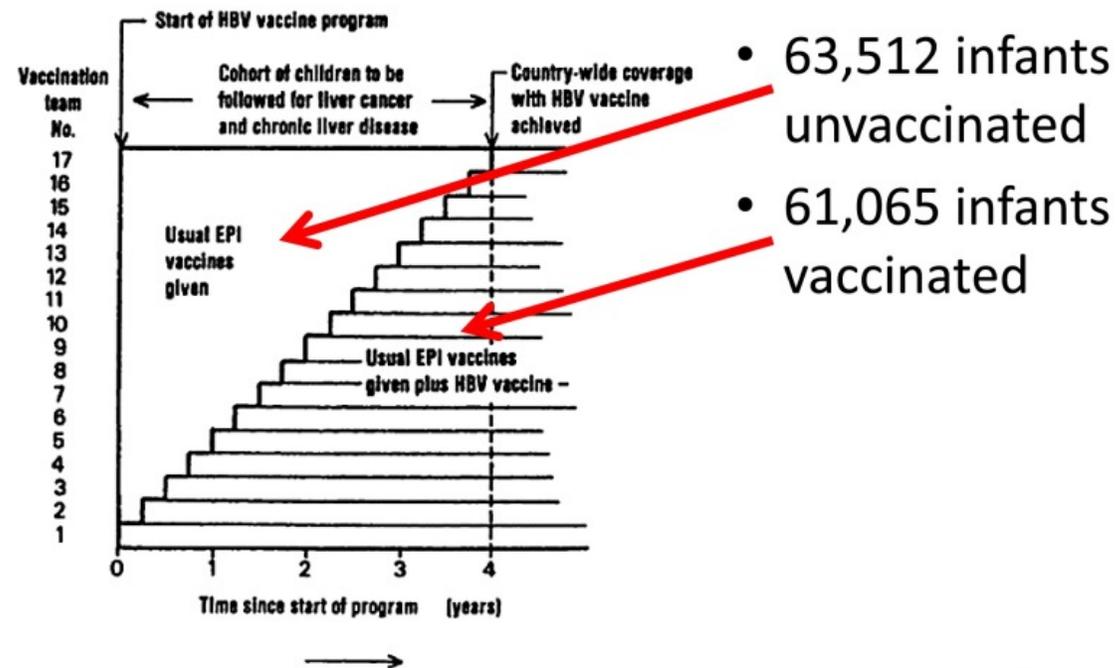
Graphic by Richard Hooper

WAS THE GAMBIA STUDY WELL JUSTIFIED?

► Gambia Hepatitis Intervention Study:

- Nation-wide trial of the HBV vaccine
- Vaccine programme rolled-out in a phased way (random) over 4 years (1986 to 1990)
- Outcome: Hepato Cellular Carcinoma (HCC) ascertained from a national cancer registry. 30 years follow-up
- *“This was because instantaneous universal vaccination in the country was impossible for logistic and financial reasons.”*

Stepped wedge design



Hall et al., 1987

Shimakawa Y, Lemoine M, Mendy M, Njai HF, D'Alessandro U, Hall A, Thursz M, Njie R. Population-based interventions to reduce the public health burden related with hepatitis B virus infection in the gambia, west Africa. Trop Med Health. 2014 Jun;42(2 Suppl):59-64. The Gambia Hepatitis Intervention Study. The Gambia Hepatitis Study Group. Cancer Res. 1987 Nov 1;47(21):5782-7

JUSTIFICATION IN THE GAMBIA STUDY

- ▶ Cluster randomisation necessary:
 - Individual-level intervention
 - Interest in direct and indirect effects
- ▶ SW-CRT well justified:
 - There was very possible evidence of effectiveness in other settings and so the fact that every cluster would ultimately receive the intervention was likely appealing
 - More importantly, the SW design allowed a randomised evaluation in what likely would have alternatively been an role out without any randomisation.

SUMMARY

- ▶ The use of the SW-CRT is growing
 - Yet the SW-CRT is likely to be at greater *risks of bias* compared to the conventional parallel cluster randomised trial (parallel-CRT)
 - For this reason, the CONSORT extension for SW-CRTs requires that investigators provide a clear *justification* for the choice of study design

We argue that the mere popularity and novelty of the SW-CRT should not be a factor in its adoption. In situations when a conventional parallel-CRT is feasible it is likely to be the preferred design.

AND FINALLY

<https://www.youtube.com/watch?v=yf7RNsElmpo>