

# When is a stepped-wedge cluster randomised trial a good design choice?

*Prof Karla Hemming and Prof Monica Taljaard examine a new type of study design, the stepped-wedge cluster randomised trial, that offers the chance to improve the evidence base on which policy decisions are made. They emphasise the importance of justifying the use of a stepped-wedge cluster randomised trial and explore a number of situations that reveal potential conditions where a stepped-wedge cluster randomised trial is preferable to other study designs. The team also explain why care should be taken when utilising this type of study.*

Randomised trials are often used to reduce bias when testing new medical treatments. Participants are randomly allocated to either the treatment group, where they receive the new treatment or experimental drug, or the control group, where they receive the current standard of care or a placebo. Randomisation means that any differences in patient outcomes, such as how long they survive after a cancer diagnosis, can be attributed to the experimental drug and all other possible explanations can be ruled out. Without randomisation, for instance, if doctors were to decide who should be treated with the experimental drug, any apparent differences in outcomes could be attributed to other disparities among the patients, rather than whether or not they were treated with the experimental drug.

Over the past decades, the use of clinical trials has undoubtedly led to an

increase in life expectancies. Medical care, however, is made up of much more than just drug interventions. Precisely how this care is organised and delivered is also significant. For instance, it has been hypothesised that limiting the number of hours junior doctors are permitted to work each week could decrease the number of errors made. Nevertheless, policy decisions such as this are rarely informed by randomised trials, as the logistics and feasibility of conducting such studies is often perceived to be insurmountable. For example, while recruiting 1,000 participants for a patient randomised trial is not an easy task; translating this into recruiting 1,000 hospitals that would allow researchers to randomly decide whether their junior doctors' rotas are to be limited is infeasible.

Yet, Prof Karla Hemming (University of Birmingham) and Prof Monica Taljaard (Ottawa Hospital Research Institute) examine a new type of study design that brings hope and the chance to improve the evidence base on which these policy decisions are made. This study design is the stepped-wedge cluster randomised trial (SW-CRT). Through an exploration of SW-CRT, they ascertain when this method is preferable to other study designs and why care should be taken when using it.

## CLUSTER RANDOMISED TRIAL

The cluster randomised trial is an established study design for pragmatic research, where the study can take place in real-world or typical practice settings. This is particularly useful for the pragmatic evaluations of health policy interventions, including changes to the delivery of services

and educational or public health type interventions.

A cluster randomised controlled trial involves the randomisation of groups, or clusters of subjects, such as hospitals, public health units or communities, rather than individual participants. In a parallel cluster randomised trial (parallel-CRT), half of the clusters are randomly assigned to the intervention condition, while the other half are assigned to the control condition.

## THE STEPPED-WEDGE CLUSTER RANDOMISED TRIAL

In a SW-CRT, the clusters move sequentially from control to intervention conditions in a randomised order. This fits with what happens in a conventional (unevaluated) roll-out where usually, some logistical constraint will prevent clusters receiving treatment simultaneously, so the clusters tend to receive the treatment in steps or waves. A stepped wedge describes the shape that is produced from a schematic illustration of the design (see figure 1). The crossover is in one direction, usually from control to intervention, and once it has been implemented, the intervention is not removed.

In a SW-CRT, the new policy under evaluation is gradually and randomly rolled out to all hospitals, or clusters, until the new policy is universal. Often, this design is viewed as ethically advantageous, since it allows all of the clusters to receive the



novel intervention eventually. Novel interventions, however, don't always work and they can lead to increased harm, which is why we want to evaluate them in the first place. There has been a surge in the use of the stepped-wedge design for both service delivery

## There has been a surge in the use of the stepped-wedge design for both service delivery and policy evaluations.

and policy evaluations due to this perceived ethical advantage. However, this increase in uptake hasn't always been appropriate – and it is starting to be used when the alternative parallel cluster trial might be a better fit.

## GREATER RISK OF BIAS

When compared with the conventional parallel-CRT, the SW-CRT is at

greater risks of bias because of the staggered nature of the roll-out, in that the observations under the control condition are collected earlier than those under the intervention condition. Because there is a natural tendency for practices to gradually improve or worsen over time, it becomes difficult to separate the effect of this "secular trend" from the effect of the intervention. Any other randomised design seeks to minimise confounders, that is, alternative factors that could explain the observed results other than the treatments being studied. Conversely, the SW-CRT induces a confounder by design. Moreover, the SW-CRT may be at greater risks of other biases than the conventional parallel-CRT. These include bias that occurs when data collected under

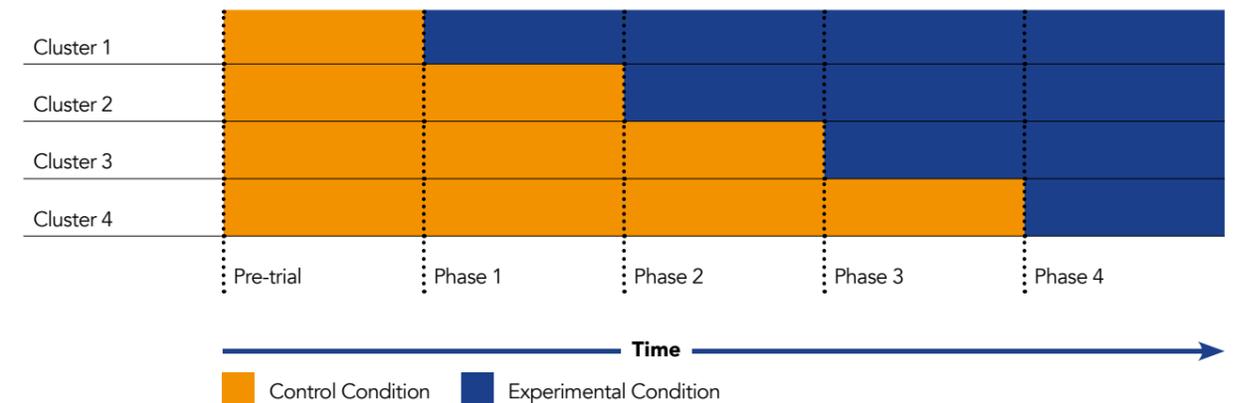


Figure 1. A pictorial representation of a stepped wedge trial: the clusters move sequentially from control (orange) to experimental conditions (blue) in a randomised order. The intervention will gradually be rolled out in all clusters.

the control condition becomes contaminated by the intervention condition, or vice versa. This is known as within-cluster contamination. Within-cluster contamination is more likely to occur in a SW-CRT, since every cluster is exposed to both control and intervention conditions. Such bias underpins the development of the CONSORT (Consolidated Standards of Reporting Trials) extension for stepped-wedge cluster randomised trials. This reporting guideline highlights the additional complexities of the design and requires that investigators provide a clear justification for using this design.

#### WHEN IS A STEPPED-WEDGE CLUSTER RANDOMISED TRIAL APPROPRIATE?

Methodologists are increasingly recognising the importance of justifying the use of a SW-CRT. Prof Hemming and Prof Taljaard have explored a number of situations and reveal potential conditions where a SW-CRT might be an appropriate choice. These justifications often overlap.

Often, interventions are rolled out without any robust randomised evaluation. Limited resources or capacity can mean that the roll-out is staggered. If the stakeholders, such as nurses, GPs or hospital management, can be persuaded to randomise the roll-out, a SW-CRT can be carried out and provide a means to conduct an evaluation which otherwise would not be possible.

Permission to carry out cluster randomised trials is often required from gatekeepers, such as general practice managers, ward matrons and lead consultants. They can be reluctant to participate in a trial unless they are assured that they will have the opportunity to receive the intervention which might be expected to offer some benefits and the expectation that the intervention is better than no intervention. Here, the SW-CRT enables cluster recruitment as it makes randomised evaluation more acceptable to cluster gatekeepers and other stakeholders.

Due to pragmatic and logistical constraints, such as the roll-out of a

scarce resource, the SW-CRT may be the only feasible design. While a parallel-CRT can also be conducted in a staggered way, it becomes infeasible if the roll-out of the intervention is constrained to only a couple of clusters at a time.

The SW-CRT can have increased statistical power over other study designs, especially when the number of available clusters is restricted, due to availability, willingness to participate or limited trial budgets. Thus, an SW-CRT may achieve the desired statistical power with fewer clusters than a parallel-CRT.

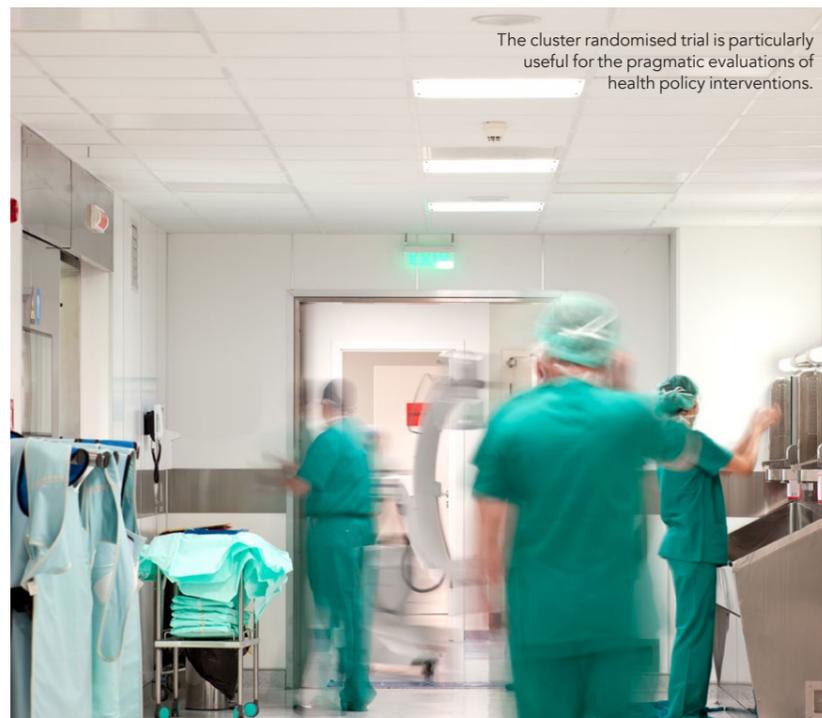
## The stepped-wedge design is at risk of bias because of the staggered nature of the roll-out, particularly when there are only a small number of clusters.

#### OTHER CONSIDERATIONS

There may be an imperative to provide an evaluation of the intervention's effectiveness in a short amount of time, in which case the overall study duration can dictate the choice of trial method. Depending on the trial's circumstances, the SW-CRT may take more time than parallel-CRT, making the latter the preferred choice.

Study designs may need to allow time to realise the effect of the intervention. This is usually relatively straightforward in the evaluation of non-complex interventions, such as giving a drug to a patient, so the patient is thus exposed. When evaluating complex interventions, however, it might take considerable time for an intervention to become fully embedded and influence outcomes. While transition periods can be incorporated to allow for this delay, they might need to be quite long. This can increase the duration of the SW-CRT when compared with a parallel-CRT.

Prof Hemming and Prof Taljaard explain that as the number of arguments in favour of an SW-CRT increases, it is likely that the benefits of using the SW-CRT will outweigh its risks. They argue, however, that popularity and novelty should not be a factor in adopting the SW-CRT, and where a conventional parallel-CRT is feasible, it is likely to be the preferred design.



# Behind the Research

## Prof Karla Hemming & Prof Monica Taljaard

E: [k.hemming@bham.ac.uk](mailto:k.hemming@bham.ac.uk) T: +44 73874398841  
Resources for stepped wedge and cluster trials: W: <https://clusterrcts.shinyapps.io/rshinyapp/>  
W: <https://clusterrandomisedtrials.qmul.ac.uk/>

### Research Objectives

Prof Hemming and Prof Taljaard examine the stepped-wedge cluster randomised trial and determine situations in which this type of study is preferred to other study designs.

### Detail

#### Address

Institute of Applied Health Research  
University of Birmingham, UK

#### Bio

**Karla Hemming** is Professor of Biostatistics, Institute of Applied Health Research, University of Birmingham. Prof Hemming's research interests include how to design cluster and stepped-wedge trials in order to maximise their statistical efficiency and minimise their risk of bias.

**Monica Taljaard** is Senior Scientist, Clinical Epidemiology Program Ottawa Hospital Research Institute. As a biostatistician and member of the Ottawa Methods Center, Dr Taljaard works with researchers and clinicians from a variety of backgrounds in the design and analysis of cluster randomised trials, standard clinical trials, and observational studies.

#### Funding

- This research was partly funded by the UK NIHR Collaborations for Leadership in Applied Health Research and Care West Midlands initiative.
- Karla Hemming is funded by a NIHR Senior Research Fellowship SRF-2017-10-002.

#### Collaborators

The Ottawa Hospital Research Institute Canada

### References

- Hemming K., Taljaard M. (2020). Reflection on modern methods: when is a stepped-wedge cluster randomized trial a good study design choice? *International Journal of Epidemiology*, 49(3), 1043–1052. Available at: <https://doi.org/10.1093/ije/dyaa077>
- Hemming K., Haines T.P., Chilton P.J., Girling A.J., Lilford R.J. (2015). The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*, 350, h391.
- Hemming K., Taljaard M., McKenzie J.E. et al. (2018). Reporting of stepped wedge cluster randomized trials: extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ*, 363, k1614.
- Huffman M.D., Mohanan P.P., Devarajan R. et al., Acute Coronary Syndrome Quality Improvement in Kerala (ACS QUIK) Investigators. (2018). Effect of a Quality Improvement Intervention on Clinical Outcomes in Patients in India With Acute Myocardial Infarction: The ACS QUIK Randomized Clinical Trial. *JAMA*, 319(6), 567–578.

### Personal Response

#### What future extensions of the SW-CRT approach might further enhance its applicability?

Policy decisions in health care need to be informed by high-quality evidence. The stepped-wedge study, by randomising the order of any planned policy roll-out across participating units, is appealing but is likely to face several important risks of bias. This is especially true when the roll-out is across only a small number of units (e.g. hospitals). When the policy is implemented once-off at national level or rolled out (perhaps randomly) across a small number of participating units, observing what happens using an alternative design such as an interrupted time series is likely to be as, if not more, robust.