
Article:
Campbell, M.J. orcid.org/0000-0003-3529-2739, Hemming, K. and Taljaard, M. (2019) The stepped wedge cluster randomised trial: what it is and when it should be used. Medical Journal of Australia, 210 (6). 253-254.e1. ISSN 0025-729X

https://doi.org/10.5694/mja2.50018

Introduction

The basic premise of a stepped-wedge cluster randomised trial (SW-CRT) is that all clusters start in the control condition, and they switch to the intervention condition in an order determined by randomisation. SW-CRTs differ from cluster cross-over trials in that the switch is only in one direction, from control to intervention condition.

An illustration of the design of one particular trial is shown in the Figure[1]. This design includes 10 clusters which are randomised to five sequences (two clusters per sequence) which determine the order in which the clusters will receive the intervention. The steps are defined as the points when the intervention is delivered in each sequence (five steps in this case). Outcomes are measured in different periods (at six discrete time points in this case). The design can be generalised so that different numbers of clusters are randomised to each sequence and the periods between the steps do not have to be all the same length. There can be periods of time between measurement occasions to protect against carry-over of the intervention effect, as in a cross-over trial, or to allow the intervention to be delivered and exert its expected effects. SW-CRTs differ from a wait-list-design in which clusters allocated to the control eventually receive the intervention but the outcome following the intervention does not contribute to the evaluation. Data obtained from the first and last measurement are sometimes omitted to reduce the duration of study without loss of power, as long as the cluster sizes are increased in compensation[2]. If a two-sequence design is restricted to the first two measurement occasions for sequences one and two, the design reduces to a cluster trial with baseline measurements.[3]

SW-CRTs are typically regarded as falling into one of two broad types: cohort designs, where the same subjects are measured in each period; and cross-sectional designs where different subjects are measured in each period. Cohort trials can be closed, when there are no new recruits after the study is started, or open, when subjects may join the study after it has started [4] Outcomes can be measured at discrete points in time (for example using cross-sectional community surveys) or continuously in time (for example, when patients present to a hospital emergency department continuously over time and have their outcomes measured using routinely collected hospital data).[5]

Rationale

There are several reasons why a SW-CRT may be preferred over a parallel arm cluster trial.[6,7] The SW-CRT may be more powerful than a parallel arm design; that is because within cluster comparisons can reduce the variance of the treatment effects.[8] Another appealing reason is that policy makers may wish to roll out the intervention under a strong belief that it will be beneficial; in this case the adoption of a SW-CRT offers an opportunity for a rigorous evaluation of the intervention during routine implementation[9]. Some justify the use of the design citing ethical reasons for ensuring all of a population receive the intervention at some time[10]; although others have argued against this justification, since this corresponds to an a priori belief that the intervention is beneficial, which questions the need for a randomised evaluation[7]. However, in many SW trials there remains the possibility that the intervention may be either ineffective in a particular setting or that it may lead to harm, irrespective of an a priori belief in its benefits.[7] There are other advantages to the use of a stepped wedge design, including enhanced ability to recruit. A justification that is sometimes used is that it may not be feasible to introduce the intervention to all clusters

simultaneously; however, this should not be the sole justification as the parallel arm cluster trial likewise can be conducted with a staggered roll-out.[6]

Limitations

Before adopting a SW-CRT it is important to weigh up the advantages and disadvantages of the design compared to a parallel arm cluster randomised trial (CRT). SW-CRTs can be logistically more complicated and there are more decisions to make such as the timing of the intervention and the length of the steps. Crucially, since clusters are randomised as to the time of implementing the intervention, all cluster approvals must be in place before the study can start. The design is also vulnerable to drop-out or under recruitment at either the cluster or patient level, as the addition of non-randomized clusters once the trial has started is questionable and extending the duration of the study to recuperate any loss in recruitment might have negligible benefits in terms of power. While the SW-CRT may require fewer clusters than a parallel arm CRT (depending on the cluster size, intra-cluster correlation coefficient (ICC) and number of measurement periods), it may require a larger number of subjects and/or measurements, and impose a higher burden on patients, which in cohort studies[8] may lead to patients dropping out and reducing the power of the study. A SW-CRT may take longer to complete, particularly if the treatment effect takes a long time to exert its effect. Perhaps the most important limitation of the design is the inducement of confounding by time: the SW-CRT must always adjust for time because of this confounding.

Examples

The Devon Active Villages (DAVE) trial is an example of a cross-sectional SW-CRT[11]. Here 128 rural villages (clusters) were randomised to one of four sequences. The intervention was tailored to each village and provided 12 weeks of physical activity opportunities, including at least three different types of activities per village. Support was provided for a further 12 months. Each measurement occasion was separated by several months to avoid the peak holiday periods. A random sample of households within each village was selected to receive a postal survey at baseline (in the month prior to commencement of the first intervention period) and within a week of the end of each of the four intervention periods. The primary outcome of interest was the proportion of adults reporting sufficient physical activity to meet internationally recognised guidelines. Thus each household is only likely to appear once in the trial. Because of the staggered nature of the intervention, the trial was spread over nearly two years, whereas a conventional parallel group trial might have been conducted over a shorter duration.

An example of a cohort design is a trial of a resistance training programme on physical function in patients receiving dialysis.[12] A total of 171 patients from 15 community satellite haemodialysis clinics participated in the evaluation of a progressive resistance training programme which used resistance elastic bands in patients in a seated position during the first hour of haemodialysis treatment. The stepped-wedge design had three sequences, each containing five randomly allocated clinics which switched to the intervention at 12, 24 or 36 weeks. The primary outcome was an objective physical function measurement, taken at the end of each 12 week period up to 48 weeks. Including the baseline measurement, 113 patients were measured five times. Essentially this is a closed cohort design, since no new patients were recruited after study start. However there was considerable drop-out during the trial and a careful statistical analysis was required to account for this.

Design and Analysis

The standard method of analysis uses a mixed-effects model, frequently referred to as the Hussey and Hughes model, which includes random effects for the clusters and fixed effects for the time periods[13]. However, methodology for the SW-CRT design has developed substantially in recent years with new analytical models being proposed to allow for more flexible correlation structures.[14] Additional methodological complexities in SW-CRTs include the possibility of within cluster contamination over time and time varying treatment effects. SW-CRTs require more complex approaches to both sample size determination and analysis[14]. For example, additional random effects, for cluster-periods or for repeated measures on the same participants may be required to allow correlations to depend on time of measurement.

There are now computer programs (for example in Stata) to aid in the sample size calculation for a SW-CRT[8]. To adjust for clustering, sample size calculations require not only an estimate for the intra-cluster correlation coefficient, as for a conventional cluster trial, but also an estimate for the correlation between measurements over time, in addition to the usual parameters such as the effect size. Cohort designs require additional correlation coefficients. Often advance estimates for these correlations are difficult to obtain and sample size calculations therefore should incorporate sensitivity analyses to examine implications of a range of values for the correlation coefficients.

Unsurprisingly, given their complexities, reviews have shown that SW-CRTs are often poorly designed and analysed.[1,15.] There is now a CONSORT extension to SW-CRTs which is designed to improve the reporting of these trials.[2] Investigators planning a SW-CRT are advised to study the CONSORT extension to SW-CRTs, to ensure that all relevant aspects required in the reporting of the completed trial are considered in advance and not just during the write-up when it would be impossible to change the design and the relevant information may not be available. Many of the issues discussed here are covered in more detail in this elaboration paper.

Word count 1499

References

1. Barker D, McElduff P, D'Este C, Campbell MJ. Stepped wedge cluster randomised trials: a review of the statistical methodology used and available. BMC Medical Research Methodology 2016; 16(1):69

2. Thompson JA, Fielding K, Hargreaves J, Copas A. The optimal design of stepped wedge trials with equal allocation to sequences and a comparison to other trial designs. Clinical Trials 2017; (6):639-47.

3. Hemming K, Taljaard M, McKenzie JE, et al. Reporting of Stepped-Wedge Cluster Randomised Trials: Extension of the CONSORT 2010 statement with explanation and elaboration. To appear British Medical Journal 2018.

4. Copas AJ, Lewis JJ, Thompson JA, et al. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. Trials 2015;16(1):352.

5 Huffman MD, Mohanan PP, Devarajan R,et al. Effect of a quality improvement intervention on clinical outcomes in patients in India with acute myocardial infarction: the ACS QUIK randomized clinical trial. JAMA. 2018;319(6):567-78.

6. Hemming K, Haines TP, Chilton PJ, et al. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. British Medical Journal 2015;350:h391.

7. Prost A, Binik A, Abubakar I, et al. Logistic, ethical, and political dimensions of stepped wedge trials: critical review and case studies. Trials 2015; 16:351

8. Hemming K, Taljaard M. Sample size calculations for stepped wedge and cluster randomised trials: a unified approach. Journal of clinical epidemiology. 2016 Jan 1;69:137-46.

9. Mdege ND, Man MS, Taylor CA, Torgerson DJ. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. Journal of Clinical Epidemiology 2011;64(9):936-48.

10 Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC Medical Research Methodology 2006; 6:54.

11. Solomon E, Rees T, Ukoumunne OC, et al. The Devon Active Villages Evaluation (DAVE) trial of a community-level physical activity intervention in rural south-west England: a stepped wedge cluster randomised controlled trial. International Journal of Behavioral Nutrition and Physical Activity 2014; 11(1):94.

12. Bennett PN, Fraser S, Barnard R, et al. Effects of an intradialytic resistance training programme on physical function: a prospective stepped-wedge randomized controlled trial. Nephrology Dialysis Transplantation 2015; 31(8) :1302-9.

13 Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemporary Clinical Trials. 2007; 28(2):182-91.

14 Kasza J, Hemming K, Hooper R, et al. Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. Statistical Methods in Medical Research. 2017: https://doi.org/10.1177/0962280217734981.

15. Martin J, Taljaard M, Girling A, Hemming K. Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials. BMJ Open. 2016;6(2):e010166.

Figure: The Stepped wedge design The shaded area is when the intervention is applied: Figure from Barker, McElduff, D'Este, Campbell[1] (with permission)