ELSEVIER

Journal of Clinical Epidemiology 66 (2013) 752−758

Journal of Clinical Epidemiology

Stepped wedge designs could reduce the required sample size in cluster randomized trials

Willem Woertman[a,1], Esther de Hoop[a,*,1], Mirjam Moerbeek[b], Sytse U. Zuidema[c], Debby L. Gerritsen[d], Steven Teerenstra[a]

[a]*Department for Health Evidence, Radboud University Medical Centre, 133, PO Box 9101, 6500 HB, Nijmegen, The Netherlands*
[b]*Department of Methodology and Statistics, Utrecht University, PO Box 80089, 3508 TB, Utrecht, The Netherlands*
[c]*Department of General Practice, University of Groningen, University Medical Centre Groningen, PO Box 72, 9700 AB, Groningen, The Netherlands*
[d]*Department of Primary and Community Care, Centre for Family Medicine, Geriatric Care and Public Health, Radboud University Medical Centre, 117, PO Box 9101, 6500 HB, Nijmegen, The Netherlands*

Accepted 8 January 2013; Published online 22 March 2013

Abstract

Objective: The stepped wedge design is increasingly being used in cluster randomized trials (CRTs). However, there is not much information available about the design and analysis strategies for these kinds of trials. Approaches to sample size and power calculations have been provided, but a simple sample size formula is lacking. Therefore, our aim is to provide a sample size formula for cluster randomized stepped wedge designs.

Study Design and Setting: We derived a design effect (sample size correction factor) that can be used to estimate the required sample size for stepped wedge designs. Furthermore, we compared the required sample size for the stepped wedge design with a parallel group and analysis of covariance (ANCOVA) design.

Results: Our formula corrects for clustering as well as for the design. Apart from the cluster size and intracluster correlation, the design effect depends on choices of the number of steps, the number of baseline measurements, and the number of measurements between steps. The stepped wedge design requires a substantial smaller sample size than a parallel group and ANCOVA design.

Conclusion: For CRTs, the stepped wedge design is far more efficient than the parallel group and ANCOVA design in terms of sample size.

Keywords: Cluster randomized trial; Stepped wedge design; Parallel group design; ANCOVA; Sample size; Design effect

1. Introduction

Randomized controlled trials are considered the gold standard in evaluating health care interventions [1]. However, cluster randomized trials (CRTs) are increasingly being used in the health care setting [2]. In these trials, complete social units, or groups of individuals (such as families, nursing homes [NHs], or general practices), are randomized to different treatments. They are mostly used to prevent contamination and in situations where individual randomization is not possible or not desirable for logistic, financial, or ethical reasons [3].

The most commonly used trial design is the parallel group design in which each cluster is randomized to either an intervention or control condition [4]. Within this design, each cluster receives only one kind of treatment during the study, and usually all clusters start simultaneously. An extension of the parallel group design is the analysis of covariance (ANCOVA) design where a baseline measurement is added to the design and included as a covariate in the analysis [5]. In contrast, in the crossover design, every cluster will receive both the intervention and the control treatment. Yet, the order of the interventions is randomized for each cluster [3,4,6]. However, it is not always possible to conduct a crossover design because it assumes that the carryover effects are absent [3,4,6]. This means that the estimated treatment effects should be independent of the order in which the treatments were assigned. So, the effects of the first treatment should have disappeared by the time the second treatment is started, which may be unrealistic if, for example, the first treatment is the reinforcement of

Conflict of interest: Neither the article nor any parts of it have been published or submitted before. No external funding has been received and no conflict of interest is present.

[1] W.W and E.de.H. are the joint contributions.

* Corresponding author. Tel.: +31-24-3667262; fax: +31-24-3613505.
E-mail address: E.deHoop@ebh.umcn.nl (E. de Hoop).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2013.01.009

What is new?

- Approaches to power calculations for cluster randomized stepped wedge designs have been provided, but a simple sample size formula is lacking. Therefore, we present a sample size formula for these kinds of trials.
- We derived a formula in which, besides the cluster size and intracluster correlation, the number of steps and measurements can be varied.
- The stepped wedge design requires a substantial smaller sample size than a parallel group or analysis of covariance design.

a hygiene protocol and the second implies falling back to usual care in hospital wards.

Herein, we will focus on the stepped wedge design, which is a type of crossover design in which (different) clusters switch treatments in only one direction at different time points (steps) [7−10]. Typically, all clusters start in the control condition. Then, the clusters switch to the intervention at consecutive time points, where the time of the switch is randomized for every cluster. Eventually, all clusters will have switched from one condition to the other (see Fig. 1).

The stepped wedge design is especially useful when the intervention is thought to do more good than harm (i.e., when there is no equipoise) [8−10]. In that situation, it is unethical to withhold or withdraw the intervention from a proportion of the subjects as would occur in a parallel group or crossover design, respectively. Besides, it may be impossible to implement the intervention in half of all clusters simultaneously because of practical, logistical, or financial reasons [8−10]. Then, the stepwise treatment implementation of the stepped wedge design offers a solution.

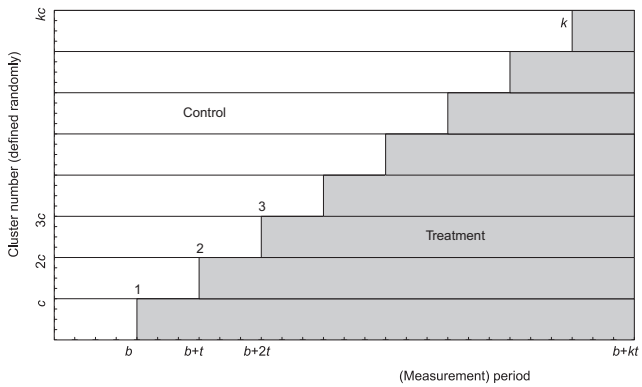

Fig. 1. Illustration of the stepped wedge design, where different (groups of) clusters switch from control to treatment at different time points.

In addition, there are other advantages of the stepped wedge design. First, the clusters act as their own controls because they receive both the control and treatment conditions. Therefore, the intervention effect can be estimated from both between- and within-cluster comparisons. This results in more statistical power and smaller required sample sizes than in a parallel group design [8]. Furthermore, it is possible to control for time with the stepped wedge design [9]. By modeling the effects of time, it is possible to study whether the time spent in the intervention condition influences the effectiveness of the treatment. Finally, recruitment of clusters and/or subjects may be easier within this design because everyone will receive the treatment during the trial.

In this article, we present a relatively simple sample size formula for stepped wedge CRTs. A recent review showed that the stepped wedge design is increasingly being used over the last couple of years [10]. Yet, it was noted that the reporting of stepped wedge CRTs needs to be improved, especially the reporting of sample size and power calculations. Hussey and Hughes [8] provide approaches to sample size and power calculations. However, their approach does not provide a sample size formula. Therefore, we propose a simpler sample size approach using a design effect (sample size correction factor).

In Section 2, we describe a trial in which the stepped wedge design is being used. Throughout the article, we will use this trial as an example. In Section 3, a sample size formula will be presented, and a comparison with the parallel group and ANCOVA design will be made in Section 4. We will conclude with a summary and discussion in Section 5.

2. Example—the Act in Case of Depression study

Depression is a common health problem in NH residents. However, it is often undetected and undertreated. Therefore, the Nijmegen University Network of Nursing Homes developed the Act in Case of Depression (AiD) program [11]. This is a multidisciplinary care program to identify and treat depression, and to monitor the treatment effects. Because the AiD program involves the training and cooperation of nursing staff, physicians, psychologists, and recreational therapists in the NHs, this program is naturally implemented at the unit level (ward) of the NHs.

The AiD study is a CRT using a stepped wedge design that aims to assess the efficacy of the AiD program in NH units. There are two main reasons why a stepped wedge design was chosen. First, the number of available units was small. Therefore, a parallel group design would not have sufficient power to detect a relevant treatment effect (see Section 4). Second, it was impractical to implement the program in half of the participating units simultaneously because of the substantial training effort that was required. Hence, stepwise implementation of the program was preferred. Obviously, a crossover design was impossible for