Journal of Clinical Epidemiology 72 (2016) 75–83

Journal of Clinical Epidemiology

ORIGINAL ARTICLES

Different methods to analyze stepped wedge trial designs revealed different aspects of intervention effects

J.W.R. Twisk[a,b,*], E.O. Hoogendijk[b,c,d], S.A. Zwijsen[b,c], M.R. de Boer[b,e]

[a]Department of Epidemiology and Biostatistics, VU University Medical Centre, Amsterdam, The Netherlands
[b]EMGO+ Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands
[c]Department of General Practice & Elderly Care Medicine, VU University Medical Center, Amsterdam, The Netherlands
[d]Gérontopôle, Toulouse University Hospital, Toulouse, France
[e]Section Methodology and Applied Biostatistics, Department of Health Sciences, VU University, Amsterdam, The Netherlands

Accepted 3 November 2015; Published online 14 November 2015

Abstract

Objectives: Within epidemiology, a stepped wedge trial design (i.e., a one-way crossover trial in which several arms start the intervention at different time points) is increasingly popular as an alternative to a classical cluster randomized controlled trial. Despite this increasing popularity, there is a huge variation in the methods used to analyze data from a stepped wedge trial design.

Study Design and Setting: Four linear mixed models were used to analyze data from a stepped wedge trial design on two example data sets. The four methods were chosen because they have been (frequently) used in practice. Method 1 compares all the intervention measurements with the control measurements. Method 2 treats the intervention variable as a time-independent categorical variable comparing the different arms with each other. In method 3, the intervention variable is a time-dependent categorical variable comparing groups with different number of intervention measurements, whereas in method 4, the changes in the outcome variable between subsequent measurements are analyzed.

Results: Regarding the results in the first example data set, methods 1 and 3 showed a strong positive intervention effect, which disappeared after adjusting for time. Method 2 showed an inverse intervention effect, whereas method 4 did not show a significant effect at all. In the second example data set, the results were the opposite. Both methods 2 and 4 showed significant intervention effects, whereas the other two methods did not. For method 4, the intervention effect attenuated after adjustment for time.

Conclusion: Different methods to analyze data from a stepped wedge trial design reveal different aspects of a possible intervention effect. The choice of a method partly depends on the type of the intervention and the possible time-dependent effect of the intervention. Furthermore, it is advised to combine the results of the different methods to obtain an interpretable overall result.

Keywords: Baseline adjustment; Intervention; Longitudinal studies; Stepped wedge trial design; Statistical methods; Time adjustment

1. Introduction

The stepped wedge trial design is a one-way crossover trial in which several arms start with the intervention at different time points (see Fig. 1). The starting point of the intervention is randomized, and although this randomization theoretically can be on an individual level, it is mostly on a cluster level, such as hospitals, departments, or neighborhoods. In the literature, there is some debate about the usefulness of a stepped wedge trial design. This is not only related to practical and logistical reasons, but also to statistical power issues [1–6]. However, within epidemiology, the stepped wedge trial design is increasingly popular as an alternative to the classical cluster randomized controlled trial (RCT).

Besides the discussion about the usefulness of a stepped wedge trial design (a discussion which will not be covered in this article), there is also much confusion about the way data from a stepped wedged trial design should be analyzed. In a systematic review, Brown and Lilford [7] mentioned that ''no two studies use the same method in analyzing data,'' whereas Mdege et al. [8] concluded that there was a huge variation in statistical methods used, varying from simple cross-sectional statistical methods, such as

* Corresponding author. Tel.: 020 4444495.
E-mail address: jwr.twisk@vumc.nl (J.W.R. Twisk).

http://dx.doi.org/10.1016/j.jclinepi.2015.11.004
0895-4356/

What is new?

Key findings

- There is huge variation in methods used to analyse data from a stepped wedge trial.
- The necessity of the adjustment of time and the necessity of the adjustment for the baseline value of the outcome depends on the method used.
- Different methods reveal different aspects of the intervention effect.
- The choice of a method partly depends on the type of intervention and the possible time-dependent effect of the intervention.

What this adds to what was known?

- The paper provides an overview of different methods that can be used to analyse data from a stepped wedge trial design.
- The paper evaluates different methods that can be used to analyse data from a stepped wedge trial design.
- The paper provides a table with pros and cons of the different methods that can be used to analyse data from a stepped wedge trial design.

What is the implication and what should change now?

- Researchers can use this paper as a reference to choose a suitable method to analyse data from a stepped wedge trial design.
- Researchers should combine the results of the different methods to obtain an interpretable overall result.

t-tests or Mann−Whitney U tests to more complicated methods, such as mixed models.

Most stepped wedge trial designs are longitudinal in nature. This means that the same group of subjects is followed over time and the different clusters receive the intervention at different points in time. There are also stepped wedge trial designs that are cross-sectional regarding the subjects. In those stepped wedge trial designs at each interval, new subjects are included and depending on the timing and the cluster in which they are randomized they receive either the intervention or the control condition. It is also possible that the stepped wedge trial design is a combination of both. The focus of the present article is on stepped wedge trial designs that are (partly) longitudinal in nature.

The most important issue to be considered in the analysis of data from a longitudinal stepped wedge trial design is the one-way crossover nature of the design. Because of that, the effect of the intervention can be measured partly within the subject (each subject moves at a certain point in time from the control to the intervention condition) and partly between the subjects (at a certain point in time, the intervention group can be compared with the control group). Ideally, these two aspects of the effect should be combined in the analysis. To do that it is necessary that data from a stepped wedge trial design are analyzed with a method that is capable to combine these effects; that is, a mixed model analysis [9]. Therefore, in the present article, only variations of mixed models will be considered as appropriate ways to analyze data from stepped wedge trial designs.

	Time					
Arm(s)	1	2	3	4	5	6
1	0	X	X	X	X	X
2	0	0	X	X	X	X
3	0	0	0	X	X	X
4	0	0	0	0	X	X
5	0	0	0	0	0	X

0 = control condition; X = intervention condition

Fig. 1. Schematic illustration of a cluster-stepped wedge design with five arms and six repeated measurements.

Besides the combination of the within and between-subject effects, in the analysis of data from a stepped wedge trial design, also the time variable can play an important role. In a classical RCT, the time variable is of no interest because the control and the intervention group are measured at the same time points; that is, the intervention variable is time independent [10], and therefore, adjustment for time cannot influence the estimated intervention effect. In a stepped wedge trial design, this is different because the intervention variable is time dependent and can influence the estimated intervention effect. Finally, it should be evaluated whether an adjustment for baseline differences in the outcome variable should be made. Although in classical RCTs, there is a debate going on whether an adjustment for the baseline value of the outcome variable is necessary to get a valid estimate of the intervention effect [11−13], most researchers argue that this adjustment should always be done [14]. In the present article, we will also take this issue into consideration regarding a stepped wedge trial design.

To contribute to the debate regarding the analyses of stepped wedge trial data, the purpose of this article is to compare several statistical methods that can be used to analyze data from a stepped wedge trial design on two example data sets.

2. Methods

2.1. Data sets

2.1.1. The ACT trial

The frail older adults: care in transition (ACT) trial was based on a geriatric care model and ran over a 24-month