**HOW TO CONDUCT A STEP-WEDGED DESIGN: AN INTRODUCTION**

**by Simon Moss**

|  |
| --- |
| **Introduction** |

If you want to assess the effects of some intervention or treatment—such as a revolutionary drug or teaching method—most researchers recommend a randomized control trial. However, for various reasons, a step-wedged design might be preferable. In particular, this design is more appropriate if

* you cannot as readily apply the intervention to all participants at the same time
* you can more readily examine a modest number of participants over time rather than a large number of participants once

|  |
| --- |
| **Description of the step-wedge design** |

To illustrate this design, suppose you wanted to explore whether some supplement, such as turmeric, improves memory. The following figure represents how the step-wedge design could be applied to answer this question.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | March | April | May | Jun | July |
| Image result for cluster of people silhouette |  | Tumeric | Tumeric | Tumeric | Tumeric |
| Image result for cluster of people silhouette |  |  | Tumeric | Tumeric | Tumeric |
| Image result for cluster of people silhouette |  |  |  | Tumeric | Tumeric |
| Image result for cluster of people silhouette |  |  |  |  | Tumeric |
| Memory test Memory test Memory test Memory test Memory test | | | | | |

As this figure shows

* researchers assess the participants—in this instance, their memory—at several times
* after the first assessment, a cluster of participants receive the intervention—in this instance, the turmeric.
* after the next assessment, another cluster of participants receive the intervention
* after the next assessment, another cluster of participants receive the intervention, and so forth
* once introduced, the intervention is maintained until the study ends.

Several variations have been introduced, demonstrating the step-wedged design is quite flexible. For example

* in some variations, not all participants are assessed at every time; instead, participants can begin late or withdraw early, called continuous recruitment
* in some variations, all participants are assessed at least twice before anyone receives the intervention—to establish a baseline, called a control period
* in some variations, each cluster comprises a different number of participants; some clusters might comprise only one participant; other clusters might comprise many participants

|  |
| --- |
| **Benefits** |

The step-wedge design offers several vital benefits. For example

* this design is ethical because, typically, all participants receive the intervention at least during some of the time
* the design is practical because you do not need to organize all participants to complete the intervention at the same time
* the design enables researchers to explore how the time of year or time since the intervention commenced affect the outcomes
* you are more likely to generate significant effects with a modest number of participants—because you can compare the same people over time and different people at the same time (Woertman et al., 2013)

|  |
| --- |
| **Drawbacks** |

Nevertheless, you should be aware of some limitations or complications of the step-wedge design. In particular

* because all participants eventually receive the intervention, you cannot compare whether people who receive the intervention benefit more in the future.
* the data are not entirely accurate unless the design comprises a sufficient number of clusters—although the precise minimum is contentious; perhaps include more than 5 clusters.
* usually, the participants and researchers will be aware of when individuals receive the intervention—so these studies are seldom double blind.
* in practice, more people receive the intervention later in the study; consequently, the benefits of this intervention could partly be ascribed to changes in the measure—such as the measure of memory performance—over time.
* to offset this problem, but also to accommodate the complexities in this design, researchers need to conduct sophisticated techniques to analyse the data,

|  |
| --- |
| **Statistical tests** |

The most challenging feature revolves around statistical analyses. The following table outlines the statistical tests that are most often utilized to analyse the data that are collected in this design. All these techniques are designed to accommodate the observation that individuals within each cluster will generate more similar findings than individuals from distinct clusters. In contrast, simple techniques, such as t-tests, will tend to disregard this observation—and, hence, the p values will tend to be inaccurate.

|  |  |  |
| --- | --- | --- |
| Test | Circumstances in which  this test is useful when | Helpful information |
| Linear Mixed Models | * each cluster comprises the same number of participants * the measure is normally distributed within each group   If these circumstances are fulfilled, this option is the most likely to generate significant results |  |
| Generalized Linear Mixed Models | * each cluster does not comprise the same number of participants or the measure is not normally distributed within each group |  |
| Generalized Estimating Equations | * each cluster does not comprise the same number of participants or the measure is not normally distributed within each group * you are primarily interested in the overall effect of this intervention instead of the effect for specific clusters   This technique is also most suitable whenever the outcome—such as memory performance—is binary, such as “good” versus “bad”. However, this technique can also be used when the outcome is numerical. | In SPSS, roughly speaking   * In the data file, create a new row each time a person is measured. So, the same person might appear in several rows. * Choose “Analyze”, “Generalized linear model”, and then “Generalized estimating equation”.   **Repeated tab**   * In the subjects variable, specify the ID column—a distinct ID number for each participant * In the within-subject variables, specify a variable that changes over time, such as time * For Working Covariance Matrix, retain the default unless you want to indicate how the scores are correlated over time.   **Test of model tab**   * Indicate whether the measure is interval or linear, ordinal, counts, and so forth   **Response tab**   * In the Dependent variable box, specify the measure, such as memory performance   **Predictors tab**   * In Factors, specify the columns that represents the cluster in which participants were assigned and whether the participants had received the intervention or not while this measure was assessed.   For more information, see the document on Generalized estimating equations on this CDU webpage about research methods |

For more information, Google the relevant statistics and your preferred statistical package. Hopefully, in the future, I will extend this document to offer more insight about the necessary statistical tests.

|  |
| --- |
| **References** |

Baio, G., Copas, A., Ambler, G., Hargreaves, J., Beard, E., & Omar, R. Z. (2015). Sample size calculation for a stepped wedge trial. Trials, 16(1), 354. doi:10.1186/s13063-015-0840-9.

Hargreaves, J. R., Copas, A. J., Beard, E., Osrin, D., Lewis, J. J., Davey, C., Thompson, J. A., Baio, G., Fielding, K. L, & Prost, A. (2015). Five questions to consider before conducting a stepped wedge trial. Trials. 16(1), 350. doi:10.1186/s13063-015-0841-8.

Hussey, M. A., & Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. Contemporary Clinical Trials, 28(2), 182–91. doi:10.1016/j.cct.2006.05.007.

Hemming, K., Lilford, R., & Girling, A. J. (2015). Stepped-wedge cluster randomised controlled trials: A generic framework including parallel and multiple-level designs. Statistics in Medicine, 34(2), 181-196. doi:10.1002/sim.6325.

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. doi:10.1136/bmj.h391.

Keriel-Gascou, M., Buchet-Poyau, K., Rabilloud, M., Duclos, A., & Colin, C. (2014). A stepped wedge cluster randomized trial is preferable for assessing complex health interventions. Journal of Clinical Epidemiology, 67(7), 831–833. doi:10.1016/j.jclinepi.2014.02.016. PMID 24774471.

Woertman, W., de Hoop, E., Moerbeek, M., Zuidema, S. U., Gerritsen, D. L., Teerenstra, S. (2013). Stepped wedge designs could reduce the required sample size in cluster randomized trials. Journal of Clinical Epidemiology, 66(7), 752–758. doi:10.1016/j.jclinepi.2013.01.009. PMID 23523551.