The Pennsylvania State University

The Graduate School

INTRA-CLUSTER CORRELATION CUT-OFF VALUES FOR SELECTING THE STEPPED-WEDGE CLUSTER RANDOMIZED TRIAL

A Thesis in

Clinical Research

by

Yining Ma

© 2019 Yining Ma

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2019

The thesis of Yining Ma was reviewed and approved* by the following:

Vernon Chinchilli

Distinguished Professor and Chair, Department of Public Health Sciences

Head of the Department of Public Health Sciences

Thesis Advisor

Lan Kong

Professor, Department of Public Health Sciences

Li Wang

Associate Professor, Department of Public Health Sciences

*Signatures are on file in the Graduate School

ABSTRACT

The stepped-wedge cluster randomized trial is an alternative to a parallel cluster randomized trial. Compared to a parallel cluster randomized trial, a stepped-wedge design refers to the randomization of one or more clusters transitioning from the control to the intervention at regularly-spaced time intervals. This paper explores the smallest value of the intra-cluster correlation for selecting a stepped-wedge cluster randomized trial instead of a parallel cluster randomized design. We generated data to attain 1,000 datasets for each simulation scenario. We designated nine time periods for the stepped-wedge stages. We adapted the same set of 1,000 simulated datasets for the parallel cluster randomized trials by randomizing four control and four intervention groups. We allowed the intra-cluster correlation to range from 0.01 to 0.20 to compare statistical power between the stepped-wedge cluster randomized design and the parallel cluster randomized design. Our results show that the stepped-wedge cluster randomized design can yield higher statistical power than the parallel cluster randomized design when the intra-cluster correlation coefficient exceeds 0.02.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first like to express my sincere gratitude to my mentor Dr. Chinchilli for the great support of my Master's study and research, for his motivation, enthusiasm, patience, and knowledge. When I ran into a trouble spot or had a question about my courses or research, he always encouraged me and improved my outlook. He is an effective mentor who guided me through my study and my life at Penn State. I really appreciate it.

Besides my mentor, I would like to thank Dr. Kong and Dr. Wang as the readers for my thesis. I am indebted to them for their very valuable contributions on this thesis.

Finally, I must express my very profound gratitude to my parents and to my family members for providing me with support and love. This is my first time to study abroad. Thanks so much for believing in me. It has become my power to pursue my dream.

Chapter 1

INTRODUCTION

A parallel cluster randomized trial is the most common type among cluster randomized trials for randomizing participants in intervention or control groups at the same time (Barker, McElduff, D'Este, & Campbell, 2016). The stepped-wedge cluster randomized trial is a novel study design and it is an alternative to a parallel cluster randomized trial (Hemming et al., 2015). A stepped-wedge design refers to the randomization of one or more clusters transitioning from control to intervention at regularly-spaced time intervals. This cross-over mechanism proceeds until all the control clusters are exposed to the intervention. For example, Baker et al implemented a web-based mindfulness training course to one of 30 care homes each week in Wales, in which 4 to 10 staff members' stress levels were measured in each home until all 30 homes received the intervention (Baker, Huxley, Dennis, Islam, & Russell, 2015). The stepped-wedge cluster randomized trial has been used to deliver an effective intervention at a cluster level (Brown & Lilford, 2006). The earliest and most widely known stepped-wedge study is the Gambia hepatitis intervention study in 1980s, in which the term "stepped wedge" was first coined. The effectiveness of hepatitis B vaccine had been tested among high risk populations, but the long-term protective effect of preventing chronic liver diseases and hepatocellular carcinoma (HCC) required evaluation. The Gambia Hepatitis Study Group considered adopting the stepped-wedge design: To implement the randomized vaccinations at the individual level was difficult and ethical issues did exist. To consider "stepped wedge," control groups were available and were compared during each step. Besides, Hepatitis B virus (HBV) vaccines can be widely applied at the end of this study as a nationwide delivery system's goal ("The Gambia Hepatitis Intervention Study. The Gambia Hepatitis Study Group," 1987).

When the intra-cluster correlation coefficient (ICC) is small, such as less than 0.10, the literature indicates that stepped-wedge cluster design yields an inefficient performance compared to the simple parallel cluster trial (Hemming & Girling, 2013; Hemming, Haines, Chilton, Girling, & Lilford, 2015). However, when the ICC reaches or exceeds 0.10, the literature indicates that a stepped-wedge cluster trial has a greater statistical power and becomes preferable (Hemming et al., 2015). These authors state that a larger ICC can deliver a higher statistical power. Hence, selecting a stepped-wedge cluster design instead of a parallel cluster randomized design is appropriate in this situation. Although a stepped-wedge cluster design has been used for a few decades (Barker et al., 2016), it still is unclear which level of ICC should be considered for using it because the aforementioned articles in the literature did not actually report the results of any simulation studies confirming a cut-off value of 0.10 for the ICC. The aim of this study is to explore the smallest value of the ICC for selecting a stepped-wedge cluster randomized trial instead of a parallel cluster randomized design. It will provide evidence for stakeholders and researchers to choose suitable cluster designs, stepped-wedge or parallel cluster trials, according to the ICC.

Figure **1**: Stepped-wedge cluster randomized controlled study design with 8 clusters and 8 intervention phases (yellow cells refer to the control and blue cells indicate intervention groups)

Clusters	Phase 0	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8
1									
2									
3									
4									
5									
6									
7									
8									

Chapter 2

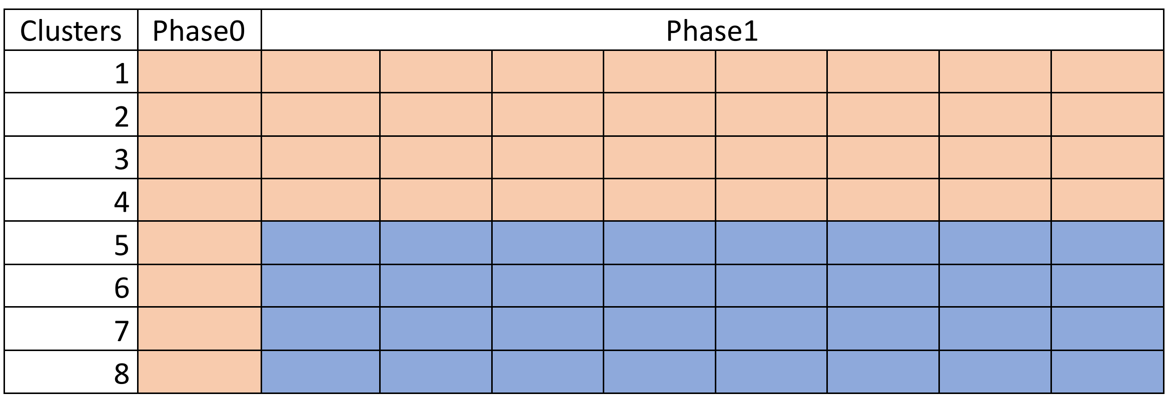
METHODS

We conducted a computer simulation study by generating data in SAS 9.4. We generated 1000 datasets for each scenario. We designated nine periods for the stepped-wedge stages. At period 0, all clusters were in the control stage to get baseline data. For each ensuing period (one through eight) one of the clusters randomly received the intervention. The mean for the control group was 0. For the intervention group, the mean was 0.5. The mean difference was 0.5. Data were randomly generated by following a normal distribution. Next, we simulated the data for four participants within each cluster. We decided on four participants because it yields reasonable levels of statistical power that do not approach the floor of 0.0 and do not approach the ceiling of 1.0, relative to the variance 1.0 or 2.0 we used. In each dataset, we generated 288 sample observations. For parallel cluster randomized trials, we randomized four clusters to control and four clusters to intervention by using the same simulation data that we had generated for the stepped-wedge designs, as indicated in Figure 2. For the parallel design, this yielded 144 samples were in the control group and 144 in the intervention group. We controlled for variance by setting it equal to 1.0 for one scenario, and setting it equal to 2.0 for another scenario. We also generated scenarios in which the ICC ranges from 0.01 to 0.20.

The linear mixed-effects model we used is as follows: Y_{ij} (ith observation in jth group) = mean (depends on group) + α_j (cluster variable) + ε_{ij} (i = 1, 2, 3,4, j = 1, 2, 3, 4, 5, 6, 7, 8).

For a variance of 1 or 2, we changed the intra-cluster correlation from 0.01 to 0.20 to compare statistical power between the stepped-wedge cluster randomized design and the parallel cluster randomized design. The statistical power for every ICC scenario for these two designs was the probability of rejecting the null hypothesis (estimated by the number of rejections among 1000 data sets). We then compared the statistical power of the two designs. We anticipated that we will identify an appreciable amount of ICC when the stepped-wedge cluster randomized design yields a higher statistical power than the parallel cluster design.

Figure **2**: Parallel cluster randomized controlled study design with 8 clusters to randomized as control and intervention groups (yellow cells refer to the control and blue cells indicate intervention groups)



Chapter 3

RESULTS

We compared the statistical power in the stepped-wedge cluster randomized design and the statistical power in the parallel cluster randomized design using the number of rejections among 1000 data sets. The intra-cluster correlation (ICC) had an optimal boundary for the stepped-wedge cluster randomized design to have a higher statistical power than the parallel cluster randomized design. When the variance is 1 or 2, we allowed the ICC to range from 0.01 to

0.20 to compare statistical power between the stepped-wedge cluster randomized (SWCR) design and the parallel cluster randomized (CR) design. In Figure 3, there clearly exists higher statistical power for the stepped-wedge cluster randomized design than the parallel cluster randomized design when the ICC exceeded 0.02 when the variance equals 1. When the variance changed from 1 to 2, as the Figure 4 shows, it was apparent that the statistical power followed the same pattern, i.e., the stepped-wedge cluster randomized design has higher power when the ICC exceeded 0.02, which is illustrated in Table 1. Overall, our results indicate that the stepped-wedge cluster randomized design can yield higher statistical power than the parallel cluster randomized design when the intra-cluster correlation exceeds 0.02.

Figure **3**. Comparing statistical power between the stepped-wedge cluster randomized design and the parallel cluster randomized design for variance = 1.0 (orange lines refer to stepped-wedge cluster randomized design and grey lines represent the parallel cluster randomized design)

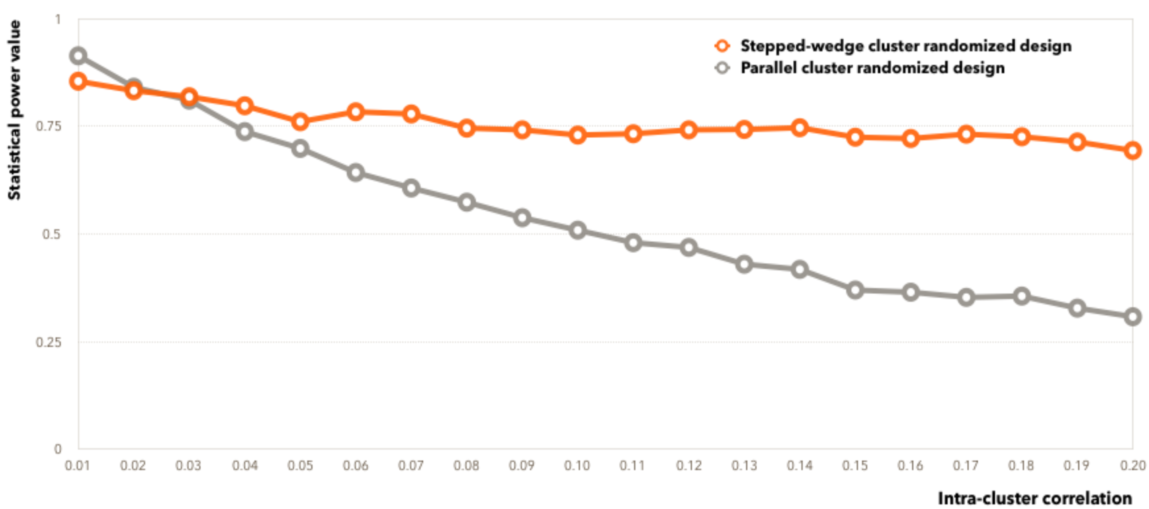


Figure **4**. Comparing statistical power between the stepped-wedge cluster randomized design and the parallel cluster randomized design for variance = 2.0 (orange lines refer to stepped-wedge cluster randomized design and grey lines represent the parallel cluster randomized design)

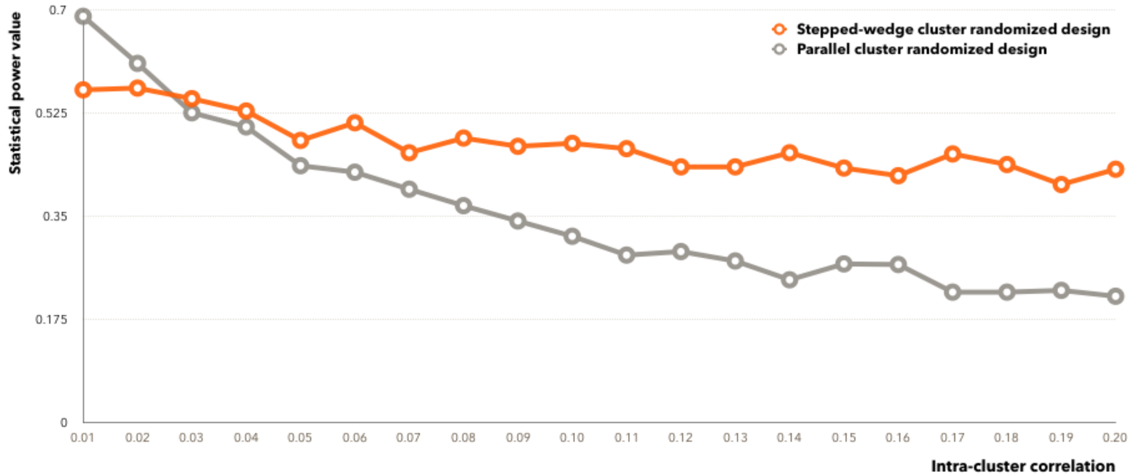


Table **1**. Changes in statistical power values for the stepped-wedge cluster randomized trial and the parallel cluster randomized trial

ICC	STATPOWER_SWCR Variance = 1.0	STATPOWER_CR Variance = 1.0	STATPOWER_SWCR Variance = 2.0	STATPOWER_CR Variance = 2.0
0.01	0.855	0.914	0.565	0.690
0.02	0.833	0.841	0.568	0.610
0.03	0.819	0.811	0.550	0.526
0.04	0.798	0.738	0.529	0.502
0.05	0.761	0.699	0.479	0.436
0.06	0.784	0.643	0.509	0.425
0.07	0.779	0.607	0.458	0.396
0.08	0.746	0.574	0.483	0.368
0.09	0.742	0.538	0.469	0.342
0.1	0.730	0.509	0.474	0.316
0.11	0.733	0.480	0.465	0.284
0.12	0.742	0.469	0.434	0.290
0.13	0.743	0.430	0.434	0.274
0.14	0.747	0.418	0.458	0.242
0.15	0.725	0.370	0.432	0.269
0.16	0.722	0.365	0.419	0.268
0.17	0.732	0.353	0.456	0.221
0.18	0.726	0.356	0.438	0.221
0.19	0.714	0.328	0.404	0.224
0.2	0.694	0.308	0.430	0.214

Chapter 4

DISCUSSION

This thesis clarified a selection criterion between the stepped-wedge cluster randomized design and the parallel cluster randomized design, according to the intra-cluster correlation coefficient (ICC). The cluster randomized designs have increased and been widely used in health-related research in the past few years (Barker et al., 2016; Campbell, Donner, & Klar, 2007). The reason for choosing such a design in a pragmatic research study is that the cluster randomization may decrease the risk of contamination. It also is easier to apply the intervention in clusters than an individual level (Zhan et al., 2014). To undertake the stepped-wedge cluster randomized design, ethical and logistical issues always have been considered as the top priority items (Beard et al., 2015; Hargreaves et al., 2015; Prost et al., 2015).

Our simulation study used data to represent a population with randomized clusters that might be encountered in practice. We generated data for our simulation study from pre-specified values for the means, variances, and the ICC. However, for the cluster randomized design and the stepped-wedge cluster randomized design in real data situations, the ICC reflects the level of correlation between any pair of individuals within a cluster (Eldridge, Ukoumunne, & Carlin, 2009).

We did not consider a carryover effect in our simulation study. We assumed that the individuals within each cluster proceed into the next stage on an individual basis. This carryover effect may impact the statistical power of the study (Hargreaves et al., 2015).

Our goal was to determine the minimum amount of ICC that identifies the conditions in which the stepped-wedge cluster randomized trial is more efficient than the parallel cluster randomized trial. We assumed that both trials had the same sample size. Other studies showed stepped-wedge cluster randomized trial requires fewer clusters than a parallel cluster randomized design (Woertman et al., 2013). If the sample size is different, smaller or larger, further simulation study investigations are needed. Also, more extensive simulation studies are needed to determine if the means and variances have any impact on efficiency. Finally, we only investigated a linear mixed-effects model with a normally distributed outcome variable. Additional studies with a nonlinear mixed-effects model for binary and ordinal outcome variables are needed.

Chapter 5

CONCLUSIONS

The evidence from this study suggests how to proceed when considering whether to invoke a stepped-wedge cluster randomized trial or a parallel cluster randomized trial. The boundary value of 0.02 for the ICC, if the researcher has an educated guess for its value in advance of the study, can be applied to select the stepped-wedge cluster randomized design or parallel cluster randomized design.

REFERENCES

Baker, C., Huxley, P., Dennis, M., Islam, S., & Russell, I. (2015). Alleviating staff stress in care homes for people with dementia: protocol for stepped-wedge cluster randomised trial to evaluate a web-based Mindfulness- Stress Reduction course. *BMC Psychiatry*, *15*, 317. doi:10.1186/s12888-015-0703-7

Barker, D., McElduff, P., D'Este, C., & Campbell, M. J. (2016). Stepped wedge cluster randomised trials: a review of the statistical methodology used and available. *BMC Med Res Methodol*, *16*, 69. doi:10.1186/s12874-016-0176-5

Beard, E., Lewis, J. J., Copas, A., Davey, C., Osrin, D., Baio, G., . . . Prost, A. (2015). Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014. *Trials, 16*, 353. doi:10.1186/s13063-015-0839-2

Brown, C. A., & Lilford, R. J. (2006). The stepped wedge trial design: a systematic review. *BMC medical research methodology, 6*(1), 54.

Campbell, M. J., Donner, A., & Klar, N. (2007). Developments in cluster randomized trials and Statistics in Medicine. *Stat Med, 26*(1), 2-19. doi:10.1002/sim.2731

Eldridge, S. M., Ukoumunne, O. C., & Carlin, J. B. (2009). The intra-cluster correlation coefficient in cluster randomized trials: a review of definitions. *International Statistical Review, 77*(3), 378-394.

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

Hemming, K., & Girling, A. (2013). The efficiency of stepped wedge vs. cluster randomized trials: stepped wedge studies do not always require a smaller sample size. *J Clin Epidemiol, 66*(12), 1427-1428. doi:10.1016/j.jclinepi.2013.07.007

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

Prost, A., Binik, A., Abubakar, I., Roy, A., De Allegri, M., Mouchoux, C., . . . Osrin, D. (2015). Logistic, ethical, and political dimensions of stepped wedge trials: critical review and case studies. *Trials, 16*, 351. doi:10.1186/s13063-015-0837-4

The Gambia Hepatitis Intervention Study. The Gambia Hepatitis Study Group. (1987). *Cancer Res, 47*(21), 5782-5787.

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. *J Clin Epidemiol, 66*(7), 752-758. doi:10.1016/j.jclinepi.2013.01.009

Zhan, Z., van den Heuvel, E. R., Doornbos, P. M., Burger, H., Verberne, C. J., Wiggers, T., & de Bock, G. H. (2014). Strengths and weaknesses of a stepped wedge cluster randomized design: its application in a colorectal cancer follow-up study. *J Clin Epidemiol, 67*(4), 454-461. doi:10.1016/j.jclinepi.2013.10.018