

©Copyright 2016

Lingyu Li

# Stepped Wedge Design for Multiple Interventions

Lingyu Li

A thesis  
submitted in partial fulfillment of the  
requirements for the degree of

Master of Science

University of Washington

2016

Committee:

James P. Hughes

Susanne May

Program Authorized to Offer Degree:  
Biostatistics

University of Washington

**Abstract**

Stepped Wedge Design  
for Multiple Interventions

Lingyu Li

Chair of the Supervisory Committee:  
Professor James P. Hughes  
Department of Biostatistics

This thesis includes a brief summary of statistical models used in stepped wedge design (Chapter 1) and an introduction to variants for the stepped wedge design with multiple interventions (Chapter 2), including Concurrent, Replacement, Supplementation, and Factorial designs. Under the basic model with random cluster effect, nonparametric fixed time effects, and fixed treatment effect, estimates of the treatment effects and variances of the estimates are presented and compared. In some specific settings, the Supplementation design is found to be preferable to the Replacement design, and provides more precise estimates. In Chapter 3, the model with random time effects is discussed. Performances of different methods, including generalized estimating equation and linear mixed models, are presented in terms of the mean and standard deviation of the estimates of the treatment effect, coverage of the 95% confidence intervals for the estimates, and mean of the estimates of the slope. Some limitations of this work are discussed in Chapter 4.

# TABLE OF CONTENTS

	Page
List of Figures . . . . .	iii
Chapter 1: Standard Stepped Wedge Design . . . . .	1
1.1 Generalized Linear Model . . . . .	2
1.2 Statistical Models for Stepped Wedge Designs . . . . .	2
1.3 Numerical Estimates and Scientific Intuitions . . . . .	8
Chapter 2: Stepped Wedge Design with Multiple Interventions . . . . .	12
2.1 Statistical Models for Multiple Interventions . . . . .	12
2.2 Different Designs for Multiple interventions . . . . .	13
2.3 Analyses for Concurrent Design . . . . .	15
2.4 Analyses for Replacement Design . . . . .	17
2.5 Analyses for Supplementation Design . . . . .	23
2.6 Analyses for Factorial Design . . . . .	26
2.7 Across Designs Comparison . . . . .	26
Chapter 3: Random Time Effects Model . . . . .	28
3.1 Data Simulation . . . . .	28
3.2 Simulation Results . . . . .	29
Chapter 4: Discussion and Limitation . . . . .	35
Glossary . . . . .	36
Bibliography . . . . .	37
Appendix A: Covariance Matrix for Random Treatment Effects Model with Multiple Interventions . . . . .	38

A.1	Concurrent design with 2 interventions . . . . .	38
A.2	Replacement design with 2 interventions . . . . .	39
A.3	Supplementaion design with 2 interventions . . . . .	39
A.4	Factorial design with 2 interventions . . . . .	39

## LIST OF FIGURES

Figure Number	Page
2.1 Variances of treatment effects vs. <i>ICC</i> for the concurrent design (Table 2.1) and designs with additional time periods using nonparametric time effect model.	16
2.2 Variances of treatment effects vs. <i>ICC</i> for the replacement design (Table 2.2) under the nonparametric time effect model. . . . .	19
2.3 Variances of treatment effects vs. <i>ICC</i> for the original replacement design (Table 2.2) and designs with additional time periods under the nonparametric time effect model. . . . .	20
2.4 Variances of treatment effects vs. <i>ICC</i> for replacement designs for 5 clusters and with different numbers of time intervals for intervention 1, under the nonparametric time effect model. . . . .	21
2.5 Variances of treatment effects vs. <i>ICC</i> for replacement designs for 5 clusters and with different numbers of time intervals for intervention 1, under the linear time effect model. . . . .	22
2.6 Variances of treatment effects vs. <i>ICC</i> for replacement designs for 5 clusters and with different numbers of time intervals for intervention 1, under the no time effect model. . . . .	23
2.7 Variances of treatment effects vs. <i>ICC</i> for the supplementation design (Table 2.3) under the nonparametric time effect model. . . . .	25
2.8 Variances of treatment effects vs. <i>ICC</i> for the replacement design (Table 2.2) and the supplementation design (Table 2.3) under the nonparametric time effect model. . . . .	27
3.1 Mean of $\hat{\theta}$ vs. S.D. of the slope. . . . .	29
3.2 S.D. of $\hat{\theta}$ vs. S.D. of the slope. . . . .	29
3.3 Coverages of the 95% confidence intervals for the treatment effect using non-inflated standard errors versus standard deviation of random slope for the time effect. . . . .	30

3.4	Mean of the non-inflated standard errors versus the standard deviations of the estimates of treatment effect. The dash line is a line with intercept equal to 0, and slope equal to 1. Circles, triangles (point up), plus, cross, and diamond symbols correspond to the cases when $s = 0, 0.003, 0.006, 0.009,$ and $0.012,$ respectively. . . . .	31
3.5	Mean of the inflated standard errors versus the standard deviations of the estimates of treatment effect. The dash line is a line with intercept equal to 0, and slope equal to 1. Circles, triangles (point up), plus, cross, and diamond symbols correspond to the cases when $s = 0, 0.003, 0.006, 0.009,$ and $0.012,$ respectively. . . . .	32
3.6	Coverages of the 95% confidence intervals for the treatment effect using inflated standard errors versus standard deviation of random slope for the time effect. . . . .	33
3.7	Mean of the estimates of slope versus standard deviation of random slope for the time effect. . . . .	34

## ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my thesis advisor, Professor James Hughes, for his excellent guidance, caring, patience, and providing me with an excellent atmosphere for doing research. I would also like to thank Professor Susanne May for providing helpful comments and helping me figure out the results.

I would like to thank Gitana Garofalo, who is our graduate program advisor, for helping and supporting me through my graduate study.

Finally, I would like to thank my parents. They were always supporting me and encouraging me with their best wishes.

## **DEDICATION**

to my parents, Tongjun and Guijuan

## Chapter 1

**STANDARD STEPPED WEDGE DESIGN**

A stepped wedge design is a unidirectional crossover design for cluster randomized trials (CRT). At baseline in a typical stepped wedge design, none of the clusters receive the intervention. The intervention is implemented sequentially to the trial participants (either as individuals or clusters of individuals) over a number of time periods. The order in which the different individuals or clusters receive the intervention is determined at random[1]. Once introduced, the intervention is never removed and all clusters receive the intervention of interest at the end of the study. Table 1.1 shows a stepped wedge design for 3 clusters.

	Time Interval 1	Time Interval 2	Time Interval 3	Time Interval 4
Cluster 1	0	1	1	1
Cluster 2	0	0	1	1
Cluster 3	0	0	0	1

Table 1.1: Standard Stepped Wedge Design with 3 Clusters

Compared to a parallel design or a classic crossover design, a standard stepped wedge design typically involves more time periods which extends the length of the study. However, there are some advantages of using a stepped wedge design. First, instead of implementing the intervention to half of the clusters at the same time, a stepped wedge design allows the researchers to implement the intervention in steps, which may be easier in practice. In addition, a stepped wedge design is more appropriate than a crossover design when the intervention cannot be removed after implementation due to practical reasons or ethical concerns, for example, knowledge or training cannot be removed after being given.

In this chapter, I discussed the standard stepped wedge design with one intervention, including statistical models for this design, statistical estimates of treatment effect, and a comparison of variances of the treatment effect in different designs.

### **1.1 Generalized Linear Model**

Before explaining the statistical models for a stepped wedge design, let me briefly introduce the generalized linear model (GLM)[2] with identity link function and normally distributed errors. It can be written as Equation 1.1,

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (1.1)$$

where  $\mathbf{Y}$  is a vector of outcomes,  $\mathbf{Z}$  is a design matrix,  $\boldsymbol{\beta}$  is a vector of parameters, and  $\boldsymbol{\epsilon}$  is a vector of the errors (not necessarily independent). Let  $\mathbf{V} = \text{Var}(\boldsymbol{\epsilon})$ . Then the estimates of the parameters are

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Y}) \quad (1.2)$$

and the covariance matrix of  $\hat{\boldsymbol{\beta}}$  is

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1} \quad (1.3)$$

### **1.2 Statistical Models for Stepped Wedge Designs**

In a stepped wedge design, the treatment effect is partially confounded with time by design. To estimate the treatment effect, it is critical to properly account for time effects in the model. Time effects can be modeled in many different ways, such as fixed nonparametric time effects, linear time effects, or random time effects (i.e. time effects that vary by cluster), etc. Chapters 1 and 2 focus on the models with fixed time effects, and the model with random time effects will be discussed in chapter 3.

### 1.2.1 Nonparametric Time Effects Model

Hussey and Hughes describe a model with nonparametric time effects and fixed treatment effect for the standard stepped wedge design[3]. Their model also includes random cluster effects to model the correlation between observations within the same cluster.

For a design with  $I$  clusters,  $T$  time intervals, and  $N$  individuals sampled per cluster per time interval, let  $Y_{ijk}$  be the response corresponding to individual  $k$  in time period  $j$  from cluster  $i$  ( $i$  in  $1, \dots, I$ ;  $j$  in  $1, \dots, J$ ;  $k$  in  $1, \dots, N$ ), and let  $\mu_{ij}$  be the mean response for cluster  $i$  in time interval  $j$ . The model for  $\mu_{ij}$  can be written as

$$\mu_{ij} = \mu + \alpha_i + \beta_j + X_{ij}\theta \quad (1.4)$$

where  $\alpha_i$  is a random effect for cluster  $i$  ( $i$  in  $1, \dots, I$ ) such that  $\alpha_i \sim N(0, \tau^2)$ ;  $\beta_j$  is a fixed effect corresponding to time interval  $j$  ( $j$  in  $1, \dots, T - 1$ , and  $\beta_T = 0$  for identifiability);  $X_{ij}$  is an indicator of the treatment mode in cluster  $i$  in time interval  $j$  (1=intervention; 0=control) and  $\theta$  is the treatment effect.

Then, the individual level responses and response means for clusters may be modeled as

$$Y_{ijk} = \mu_{ij} + e_{ijk} \quad (1.5)$$

$$Y_{ij} = \mu_{ij} + e_{ij} \quad (1.6)$$

respectively, where  $e_{ijk} \stackrel{iid}{\sim} N(0, \sigma_e^2)$ ;  $e_{ij} = \sum_k e_{ijk}/N \stackrel{iid}{\sim} N(0, \sigma^2)$  and  $\sigma^2 = \sigma_e^2/N$ . The intraclass correlation (ICC) characterizes the correlation between individuals from the same cluster, and is defined as

$$\rho = \frac{\tau^2}{\tau^2 + \sigma_e^2} \quad (1.7)$$

In this thesis, I assume that  $N$  is always equal and only consider the situation where  $\tau^2$  and  $\sigma^2$  are known. Therefore, the estimates as well as the variance of the fixed effects may be

obtained using a GLM analysis on the cluster means,  $Y_{ij}$ . Specifically, let  $\mathbf{V}$  be an  $IT \times IT$  block diagonal matrix, where each  $T \times T$  block within  $\mathbf{V}$  describes the correlation structure between the cluster response means for different time intervals, and has the structure

$$\begin{pmatrix} \sigma^2 + \tau^2 & \tau^2 & \dots & \tau^2 \\ \tau^2 & \sigma^2 + \tau^2 & \dots & \tau^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tau^2 & \tau^2 & \dots & \sigma^2 + \tau^2 \end{pmatrix} \quad (1.8)$$

Let  $\mathbf{Z}$  be the design matrix with  $IT$  rows and  $(T + 1)$  columns, where the first column corresponds to the intercept; the second to the  $T^{th}$  columns correspond to the nonparametric time effects, and the  $(T + 1)^{th}$  column corresponds to the treatment effect. For a simple stepped wedge design with 2 clusters (shown in Table 1.2), the design matrix is presented as Equation 1.9.

	Time Interval 1	Time Interval 2	Time Interval 3
Cluster 1	0	1	1
Cluster 2	0	0	1

Table 1.2: Standard Stepped Wedge Design with 2 Clusters

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \quad (1.9)$$

Then, the Model 1.6 can be written as,

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{e} \quad (1.10)$$

where  $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V})$ .

### 1.2.2 No Time Effect Model

If there is no time effect, the statistical model can be simplified as

$$\mu_{ij} = \mu + \alpha_i + X_{ij}\theta \quad (1.11)$$

The estimates and variances of the fixed effects can be calculated by using GLM. In this case,  $\mathbf{V}$  remains the same as the nonparametric time effects model, while the design matrix here has dimension  $IT \times 2$ . Under the no time effect model, the design matrix for the simple 2-cluster stepped wedge design (Table 1.2) is

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \quad (1.12)$$

### 1.2.3 Linear Time Effects Model

In this model, I assume a linear relationship between time and  $\mu_{ij}$ . The statistical model can be written as

$$\mu_{ij} = \mu + \alpha_i + t_j\beta + X_{ij}\theta \quad (1.13)$$

where  $t_j$  = beginning time point of time interval  $j$ , and  $t_1 = 0$ . Thus,  $\beta$  is the slope for the time effects.  $\mathbf{V}$  remains the same as the nonparametric time effects model, but the design

matrix here has dimension  $IT \times 3$ . Under the linear time effects model, the design matrix for the 2-cluster stepped wedge design (Table 1.2) is

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & t_2 & 1 \\ 1 & t_3 & 1 \\ 1 & 0 & 0 \\ 1 & t_2 & 0 \\ 1 & t_3 & 1 \end{pmatrix} \quad (1.14)$$

#### 1.2.4 Random Treatment Effect Model

All the above models assume that the intervention effect,  $\theta$ , is constant across clusters. However, the effect of the intervention may vary from cluster to cluster due to variation in the quality of implementation or other factors. A statistical model with random treatment effect is

$$\mu_{ij} = \mu + \alpha_i + \beta_j + X_{ij}\theta + X_{ij}\gamma_i \quad (1.15)$$

where  $\gamma_i \sim N(0, \eta^2)$  is a random treatment effect for cluster  $i$  [4]. In addition, there could be a correlation between the cluster random effect and the treatment random effect, which can be written as  $Cov(\alpha_i, \gamma_i) = \nu$ . The fixed effect estimates and variances can be calculated using GLM. The design matrix,  $\mathbf{Z}$ , remains the same as the nonparametric time effects model. Covariance matrix,  $\mathbf{V}$ , remains an  $IT \times IT$  block diagonal matrix, but the  $T \times T$  blocks have different structures now. Let me take the simple stepped wedge design with 3 clusters (Table 1.1) as an example. The covariance matrix  $\mathbf{V}$  is

$$\begin{pmatrix} \mathbf{V}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_3 \end{pmatrix} \quad (1.16)$$

where  $\mathbf{V}_1$  is

$$\begin{pmatrix} \sigma^2 + \tau^2 & \tau^2 & \tau^2 & \tau^2 \\ \tau^2 & \sigma^2 + \tau^2 + \eta^2 + 2\nu & \tau^2 + \eta^2 & \tau^2 + \eta^2 \\ \tau^2 & \tau^2 + \eta^2 & \sigma^2 + \tau^2 + \eta^2 + 2\nu & \tau^2 + \eta^2 \\ \tau^2 & \tau^2 + \eta^2 & \tau^2 + \eta^2 & \sigma^2 + \tau^2 + \eta^2 + 2\nu \end{pmatrix} \quad (1.17)$$

$\mathbf{V}_2$  is

$$\begin{pmatrix} \sigma^2 + \tau^2 & \tau^2 & \tau^2 & \tau^2 \\ \tau^2 & \sigma^2 + \tau^2 & \tau^2 & \tau^2 \\ \tau^2 & \tau^2 & \sigma^2 + \tau^2 + \eta^2 + 2\nu & \tau^2 + \eta^2 \\ \tau^2 & \tau^2 & \tau^2 + \eta^2 & \sigma^2 + \tau^2 + \eta^2 + 2\nu \end{pmatrix} \quad (1.18)$$

$\mathbf{V}_3$  is

$$\begin{pmatrix} \sigma^2 + \tau^2 & \tau^2 & \tau^2 & \tau^2 \\ \tau^2 & \sigma^2 + \tau^2 & \tau^2 & \tau^2 \\ \tau^2 & \tau^2 & \sigma^2 + \tau^2 & \tau^2 \\ \tau^2 & \tau^2 & \tau^2 & \sigma^2 + \tau^2 + \eta^2 + 2\nu \end{pmatrix} \quad (1.19)$$

### 1.2.5 Random Time Effects Model

As an extension to the nonparametric time effects model, the random effects of time can be taken into consideration. The statistical model is

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \zeta_{ij} + X_{ij}\theta \quad (1.20)$$

where  $\zeta_{ij} \sim N(0, \kappa_j^2)$  is a random time effect for cluster  $i$  in time interval  $j$ . When using GLM to get the estimates, the design matrix has dimension  $IT \times 3$  and the covariance matrix is a  $IT \times IT$  matrix. If assuming that the random time effect and the random cluster effect are independent, each  $T \times T$  block in the covariance matrix has the following structure:

$$\begin{pmatrix} \sigma^2 + \tau^2 + \kappa_1^2 & \tau^2 & \cdots & \tau^2 \\ \tau^2 & \sigma^2 + \tau^2 + \kappa_2^2 & \cdots & \tau^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tau^2 & \tau^2 & \cdots & \sigma^2 + \tau^2 + \kappa_T^2 \end{pmatrix} \quad (1.21)$$

### 1.3 Numerical Estimates and Scientific Intuitions

The nonparametric time effects model (Model 1.4) is the most commonly used model for stepped wedge design. To understand it better, I used Mathematica[5] to calculate the estimate of the intervention effect, as well as the variance of the estimate, in specific designs.

#### 1.3.1 Stepped Wedge Design with 2 Clusters and 3 Time Intervals

First, I investigated the estimates in the simplest stepped wedge design with 2 clusters and 3 time intervals (as shown in Table 1.2). The estimate of the intervention effect can be written as

$$\begin{aligned} \hat{\theta} &= \frac{1}{\sigma^2 + 2\tau^2} \left[ \sigma^2(Y_{12} - Y_{22}) + \tau^2(-Y_{11} + 2Y_{12} - Y_{13} + Y_{21} - 2Y_{22} + Y_{23}) \right] \\ &= \frac{1}{1 + \delta} \left[ (Y_{12} - Y_{22}) + \delta \left( -\frac{1}{2}Y_{11} + Y_{12} - \frac{1}{2}Y_{13} \right) + \delta \left( \frac{1}{2}Y_{21} - Y_{22} + \frac{1}{2}Y_{23} \right) \right] \end{aligned} \quad (1.22)$$

which is equal to

$$\begin{aligned} \hat{\theta} &= \frac{1}{1 + \delta} (Y_{12} - Y_{22}) \\ &+ \frac{\delta}{1 + \delta} \left[ \frac{1}{4} [Y_{12} - Y_{11} - (Y_{22} - Y_{21})] + \frac{1}{4} [Y_{13} - Y_{11} - (Y_{13} - Y_{12} + Y_{22} - Y_{21})] \right. \\ &\left. + \frac{1}{4} [Y_{23} - Y_{21} - (Y_{13} - Y_{12} + Y_{22} - Y_{21})] + \frac{1}{4} [Y_{23} - Y_{22} - (Y_{13} - Y_{12})] \right] \end{aligned} \quad (1.23)$$

where  $\delta = 2\tau^2/\sigma^2$ . Note that when  $ICC = 0$  (i.e. independent case),  $\delta = 0$  and  $\delta$  increases with the increase of  $ICC$  (Equation 1.7). Therefore, under independent cases, the estimate

of the intervention effect is only based on the between-cluster comparison ( $Y_{12} - Y_{22}$ ) in time interval 2. When  $ICC > 0$ , the estimate of the intervention effect can be written as a weighted sum of this between-cluster comparison and within-cluster comparisons (e.g.  $Y_{12} - Y_{11} - (Y_{22} - Y_{21})$  is the comparison between  $Y_{12}$  and  $Y_{11}$  minus an estimate of the temporal change between time 1 and time 2,  $Y_{22} - Y_{21}$ ). The weight for the within-cluster comparison will increase with the increase of  $ICC$ , and when  $ICC \rightarrow 1$ ,  $\delta \rightarrow \infty$  and  $\frac{\delta}{1 + \delta} \rightarrow 1$ , so  $\hat{\theta}$  can be derived from the within-cluster comparisons, and each of the treatment-baseline comparison is equally weighted.

For the design in Table 1.2, the estimate of the variance of the intervention effect can be written as

$$\hat{Var}(\hat{\theta}) = \frac{2\sigma^2(\sigma^2 + 3\tau^2)}{\sigma^2 + 2\tau^2} \quad (1.24)$$

which is equal to

$$\hat{Var}(\hat{\theta}) = \sigma^2\left(2 + \frac{\delta}{1 + \delta}\right) \quad (1.25)$$

Therefore, when  $ICC = 0$ , the estimate of variance is  $2\sigma^2$  and as  $ICC \rightarrow 1$ , the variance  $\rightarrow 3\sigma^2$ .

### 1.3.2 Stepped Wedge Design with 2 Clusters and 4 Time Intervals

To investigate the influence of the additional time intervals at the end of the study, I calculate the estimates in the stepped wedge design with 2 clusters and 4 time intervals (as shown in Table 1.3). Assume that  $N$ , which is the number of observations per cluster per time interval, is always equal.

The estimate of the intervention effect is

$$\hat{\theta}' = \frac{1}{\sigma^2 + 3\tau^2} \left[ \sigma^2(Y_{12} - Y_{22}) + \tau^2(-Y_{11} + 3Y_{12} - Y_{13} - Y_{14} + Y_{21} - 3Y_{22} + Y_{23} + Y_{24}) \right] \quad (1.26)$$

	Time Interval 1	Time Interval 2	Time Interval 3	Time Interval 4
Cluster 1	0	1	1	1
Cluster 2	0	0	1	1

Table 1.3: Stepped wedge design for 2 clusters with 4 time intervals.

I was not able to get a meaningful form for this equation, because here are different ways to estimate the change in the mean over time. For example, for the difference between time effects for time interval 2 and time interval 4, it could be estimated directly as  $Y_{14} - Y_{12}$  or could be estimated by adding the difference between time 4 and time 3, and the difference between time 3 and time 2. Things become too complicated to be easily interpreted.

The estimate of the variance of intervention effect is

$$\hat{Var}(\hat{\theta}') = \frac{2\sigma^2(\sigma^2 + 4\tau^2)}{\sigma^2 + 3\tau^2} \quad (1.27)$$

which is equal to

$$\hat{Var}(\hat{\theta}') = \sigma^2 \left( \frac{4}{3} + \frac{2/3}{1 + \delta'} \right) \quad (1.28)$$

where  $\delta' = \frac{3\tau^2}{\sigma^2}$ . When  $ICC = 0$ , i.e.  $\delta' = 0$ , the variance of the estimate is  $2\sigma^2$  as before. Therefore, additional time intervals at the end of the study will not change the variance of the estimate of intervention effect under the independent case. However, the ratio of  $\hat{Var}(\hat{\theta})$  and  $\hat{Var}(\hat{\theta}')$  is

$$\begin{aligned} \frac{\hat{Var}(\hat{\theta})}{\hat{Var}(\hat{\theta}')} &= \frac{(\sigma^2 + 3\tau^2)^2}{(\sigma^2 + 2\tau^2)(\sigma^2 + 4\tau^2)} \\ &= \frac{\sigma^4 + 6\sigma^2\tau^2 + 9\tau^4}{\sigma^4 + 6\sigma^2\tau^2 + 8\tau^4} \\ &\geq 1 \end{aligned} \quad (1.29)$$

The ratio is strictly greater than 1 when  $\tau^2 > 0$ . Therefore, for  $ICC > 0$ , the variance of estimate of intervention effect decreases with additional time intervals at the end of study.

Hussey and Hughes[3] gave a general form of the variance of  $\hat{\theta}$ , which is

$$Var(\hat{\theta}) = \frac{I\sigma^2(\sigma^2 + T\tau^2)}{(IU - W)\sigma^2 + (U^2 + ITU - TW - IV)\tau^2} \quad (1.30)$$

where  $U = \sum_{ij} X_{ij}$ ,  $W = \sum_j (\sum_i X_{ij})^2$ , and  $V = \sum_i (\sum_j X_{ij})^2$ .

Therefore, the variances of the intervention effects can be calculated and compared for any of the standard stepped wedge designs.

## Chapter 2

# STEPPED WEDGE DESIGN WITH MULTIPLE INTERVENTIONS

In this chapter, I will introduce the statistical models used for stepped wedge designs with multiple interventions, as well as different variations of the designs. I will also investigate the variances of the estimates of different interventions under different designs. The assumption that there is no interaction between different interventions is made for all of the models. I also assumed 100 subjects in each cluster and the baseline prevalence equal to 0.05 when calculating variances of treatment effects under specific designs, which means  $\sigma_e^2 = 0.05 \times (1 - 0.05)$ .

### 2.1 Statistical Models for Multiple Interventions

For stepped wedge design with multiple interventions, the model for the mean of cluster  $i$  in time interval  $j$ ,  $\mu_{ij}$ , under the nonparametric time effects model can be written as

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \sum_{m=1}^M X_{mij} \theta_m \quad (2.1)$$

where  $X_{mij}$  is the indicator of the treatment mode in cluster  $i$  in interval  $j$  for intervention  $m$ , and  $\theta_m$  is the treatment effect for intervention  $m$ .

The estimates and variances of intervention effects can be calculated using GLM, as I did for the standard stepped wedge designs. Notice that for designs with  $M$  interventions, the design matrix has  $IT$  rows and  $(T + M)$  columns, and the  $(T + m)^{th}$  column corresponds to the indicators of the  $m^{th}$  treatment, where  $m = 1, \dots, M$ .

For the random treatment effects model with multiple treatments, the model for  $\mu_{ij}$  can be written as

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \sum_{m=1}^M X_{mij}\theta_m + \sum_{m=1}^M X_{mij}\gamma_{mi} \quad (2.2)$$

where  $\gamma_{mi} \sim N(0, \eta_m^2)$  is the random treatment effect for treatment  $m$  and cluster  $i$ . Assume that  $Cov(\alpha_i, \gamma_{mi}) = \nu_m$  and  $Cov(\gamma_{mi}, \gamma_{ni}) = \xi_{mn}$  for  $m \neq n$ .

## 2.2 Different Designs for Multiple interventions

In this section, I will introduce 4 kinds of designs for 2 interventions: concurrent design, replacement design, supplementation design and factorial design. For the first three designs, they can be easily extended to multiple interventions, while our discussion of the factorial design is limited to the case of 2 interventions.

### 2.2.1 Concurrent Design

Table 2.1 presents the concurrent design for 2 interventions and 6 clusters, with 4 time intervals.

	Time Interval 1	Time Interval 2	Time Interval 3	Time Interval 4
Cluster 1	0	1	1	1
Cluster 2	0	0	1	1
Cluster 3	0	0	0	1
Cluster 4	0	2	2	2
Cluster 5	0	0	2	2
Cluster 6	0	0	0	2

Table 2.1: Concurrent stepped wedge design for 2 interventions and 6 clusters with 4 time intervals.

Basically, the concurrent design is a design where the interventions are implemented simultaneously and separately as in the standard stepped wedge design. For every cluster,

only one intervention is implemented, and at the end of the study, clusters received different interventions.

### 2.2.2 Replacement Design

Table 2.2 provides an example of replacement design for 2 interventions and 3 clusters with 6 time intervals.

	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Cluster 1	0	1	1	2	2	2
Cluster 2	0	0	1	1	2	2
Cluster 3	0	0	0	1	1	2

Table 2.2: Replacement stepped wedge design for 2 interventions and 3 clusters with 6 time intervals.

In replacement designs, intervention 1 is replaced by intervention 2 (which is replaced by intervention 3, etc). Every cluster receives all of the interventions, but there is only one intervention implemented per time interval. At the end of the study, all of the clusters receive the same intervention, i.e. the last intervention.

### 2.2.3 Supplementation Design

Table 2.3 shows an example of a supplementation design for 2 interventions and 3 clusters with 6 time intervals.

In supplementation designs, interventions are added to the previous interventions. Similar to the replacement designs, every cluster receives all of the interventions, and at the end of the study, all of the clusters receive all of the interventions. The replacement design and the supplementation design can be very similar to each other, and therefore, comparable. I will compare the treatment effects in the two designs later in Section 2.7.

	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Cluster 1	0	1	1	1+2	1+2	1+2
Cluster 2	0	0	1	1	1+2	1+2
Cluster 3	0	0	0	1	1	1+2

Table 2.3: Supplementation stepped wedge design for 2 interventions and 3 clusters with 6 time intervals.

#### 2.2.4 Factorial Design

Table 2.4 presents an example of factorial design for 2 interventions and 4 clusters with 5 time intervals.

	Time 1	Time 2	Time 3	Time 4	Time 5
Cluster 1	0	1	1	1+2	1+2
Cluster 2	0	0	1	1	1+2
Cluster 3	0	0	2	2	1+2
Cluster 4	0	2	2	1+2	1+2

Table 2.4: Factorial stepped wedge design for 2 interventions and 4 clusters with 5 time intervals.

The factorial design is a mixture of concurrent and supplementation designs. The advantage of factorial design is that it allows researchers to use fewer clusters (comparing to concurrent design), and fewer time intervals (comparing to replacement design and supplementation design) to evaluate two interventions at the same time.

### 2.3 Analyses for Concurrent Design

Since the concurrent design is symmetric for both of the interventions, under the fixed treatment effects models, the variances of the two intervention effects are equal, i.e. it has

the same power to evaluate the two fixed treatment effects. Adding more time intervals of treatment at the end of the design will decrease the variances for both of the treatment effects. Figure 2.1 shows the variance-ICC relationship for interventions in the original concurrent design (Table 2.1), and designs with additional time periods under the nonparametric time effects model (Equation 2.1).

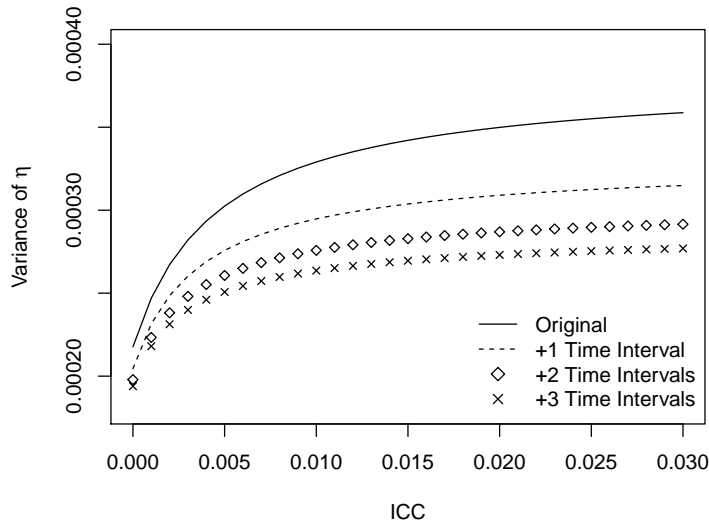


Figure 2.1: Variances of treatment effects vs.  $ICC$  for the concurrent design (Table 2.1) and designs with additional time periods using nonparametric time effect model.

Examples of blocks in covariance matrix of the concurrent design (Table 2.1), under the random treatment effects model (Equation 2.2), are shown in Appendix A. For simplicity, I assumed that  $Cov(\alpha_i, \gamma_{mi}) = \nu_m = 0$  (i.e. the random cluster effect and the random treatment effect for intervention  $m$  are independent) and  $Cov(\gamma_{mi}, \gamma_{ni}) = \xi_{mn} = 0$  for  $m \neq n$  (i.e. the random effects for two different interventions,  $m$  and  $n$ , are independent), and found that if  $\eta_1^2 = \eta_2^2$ , the variances of the fixed treatment effects for interventions 1 and 2 are equal. If  $\eta_1^2$  and  $\eta_2^2$  are not equal, the variances of the fixed treatment effects are still equal to each other when  $\tau^2 = 0$ , but when  $\tau^2 > 0$ , the intervention with larger variance for the random treatment effect has larger variance for the fixed treatment effect, too.

## 2.4 Analyses for Replacement Design

To get a basic understanding of the replacement design, I used Mathematica to calculate the numerical results for the estimates of treatment effects in the simplest replacement design with 2 clusters and 4 time intervals (Table 2.5), under the nonparametric time effect model.

	Time Interval 1	Time Interval 2	Time Interval 3	Time Interval 4
Cluster 1	0	1	2	2
Cluster 2	0	0	1	2

Table 2.5: The simplest replacement stepped wedge design for 2 interventions with 2 clusters and 4 time intervals.

The estimate of the treatment effect for intervention 1 is:

$$\begin{aligned}\hat{\theta}_1 &= \frac{1}{\sigma^2 + 2\tau^2} \left[ \sigma^2(Y_{12} - Y_{22}) + \tau^2(-Y_{11} + 2Y_{12} - Y_{14} + Y_{21} - 2Y_{22} + Y_{24}) \right] \\ &= \frac{1}{1 + \delta} \left[ (Y_{12} - Y_{22}) + \delta \left( -\frac{1}{2}Y_{11} + Y_{12} - \frac{1}{2}Y_{14} \right) + \delta \left( \frac{1}{2}Y_{21} - Y_{22} + \frac{1}{2}Y_{24} \right) \right]\end{aligned}\quad (2.3)$$

where  $\delta = \frac{2\tau^2}{\sigma^2}$ . It is equal to

$$\begin{aligned}\hat{\theta}_1 &= \frac{1}{1 + \delta}(Y_{12} - Y_{22}) \\ &+ \frac{\delta}{1 + \delta} \left[ \frac{1}{4}[Y_{12} - Y_{11} - (Y_{22} - Y_{21})] + \frac{1}{4}[Y_{14} - Y_{11} - (Y_{14} - Y_{12} + Y_{22} - Y_{21})] \right] \\ &+ \frac{1}{4}[Y_{24} - Y_{21} - (Y_{14} - Y_{12} + Y_{22} - Y_{21})] + \frac{1}{4}[Y_{24} - Y_{22} - (Y_{14} - Y_{12})]\end{aligned}\quad (2.4)$$

The estimate can be divided into two parts: between-cluster comparison and within-cluster comparison. When  $ICC = 0$ , i.e.  $\delta = 0$ , the estimate is derived from the between-cluster comparison,  $Y_{12} - Y_{22}$ . When  $ICC > 0$ , i.e.  $\delta > 0$ , the estimate is the sum of weighted between-cluster comparison and within-cluster comparison. In this design, the treatment effect for intervention 1 can be estimated using within-cluster comparisons in four ways. The first is the direct treatment-control comparison,  $Y_{12} - Y_{11}$ , minus the estimate of the

temporal change between time 2 and time 1, which is  $Y_{22} - Y_{21}$ . The other three parts are a little bit tricky. The second is  $Y_{14} - Y_{11} - (Y_{14} - Y_{12} + Y_{22} - Y_{21})$ , where  $Y_{14} - Y_{11}$  is the estimate of the treatment effect for intervention 2 plus the estimate of the temporal change between time 4 and time 1,  $Y_{14} - Y_{12}$  is the estimate of the difference between the treatment effects for intervention 2 and intervention 1 plus the estimate of the temporal change between time 4 and time 2, and  $Y_{22} - Y_{21}$  is the estimate of the temporal change between time 2 and time 1. Therefore,  $Y_{14} - Y_{11} - (Y_{14} - Y_{12} + Y_{22} - Y_{21})$  is an estimate of the treatment effect for intervention 1 by using within-cluster comparison. The third and the fourth terms can be explained similarly.

The estimate of treatment effect for intervention 2 is

$$\begin{aligned}\hat{\theta}_2 &= \frac{1}{\sigma^2 + 2\tau^2} \left[ \sigma^2(Y_{13} - Y_{23} + Y_{12} - Y_{22}) + 2\tau^2(-Y_{11} + Y_{12} + Y_{13} - Y_{14} + Y_{21} - Y_{22} - Y_{23} + Y_{24}) \right] \\ &= \frac{1}{1 + \delta} \left[ (Y_{13} - Y_{23} + Y_{12} - Y_{22}) + \delta(-Y_{11} + Y_{12} + Y_{13} - Y_{14}) + \delta(Y_{21} - Y_{22} - Y_{23} + Y_{24}) \right]\end{aligned}\tag{2.5}$$

which is equal to

$$\begin{aligned}\hat{\theta}_2 &= \frac{1}{1 + \delta} (Y_{13} - Y_{23} + Y_{12} - Y_{22}) \\ &\quad + \frac{\delta}{1 + \delta} \left[ (Y_{24} - Y_{22}) - (Y_{14} - Y_{13}) - (Y_{23} - Y_{22}) + (Y_{12} - Y_{11}) - (Y_{22} - Y_{21}) \right]\end{aligned}\tag{2.6}$$

Similarly to the estimate for intervention 1, the  $\theta_2$  can be divided into two parts: between-cluster comparison and within-cluster comparison.  $(Y_{13} - Y_{23} + Y_{12} - Y_{22})$  is the between-cluster comparison part. For the within-cluster comparison,  $(Y_{24} - Y_{22})$  is the estimate of treatment effect for intervention 2 plus the estimate of temporal change between time 4 and time 2;  $(Y_{14} - Y_{13})$  is the estimate of temporal change between time 4 and time 3;  $(Y_{23} - Y_{22})$  is the estimate of temporal change between time 3 and time 2 plus the treatment effect for intervention 1; and  $(Y_{12} - Y_{11}) - (Y_{22} - Y_{21})$  is the estimate of treatment effect for intervention 1.

When  $ICC = 0$ ,  $\hat{\theta}_1 = Y_{12} - Y_{22}$ , and  $\hat{\theta}_2 = Y_{13} - Y_{23} + Y_{12} - Y_{22}$ , that is, to estimate treatment effect for intervention 1, there is a direct treatment-control comparison, while to estimate treatment effect for intervention 2, it can only use indirect comparisons. Therefore, it is reasonable to guess that there is less power to evaluate the treatment effect for intervention 2, i.e. the variance of the treatment effect for intervention 2 is larger.

The variance- $ICC$  relationships for both interventions in the replacement design (Table 2.2) are shown in Figure 2.2. The variance of the treatment effect for intervention 2 is larger as I expected.

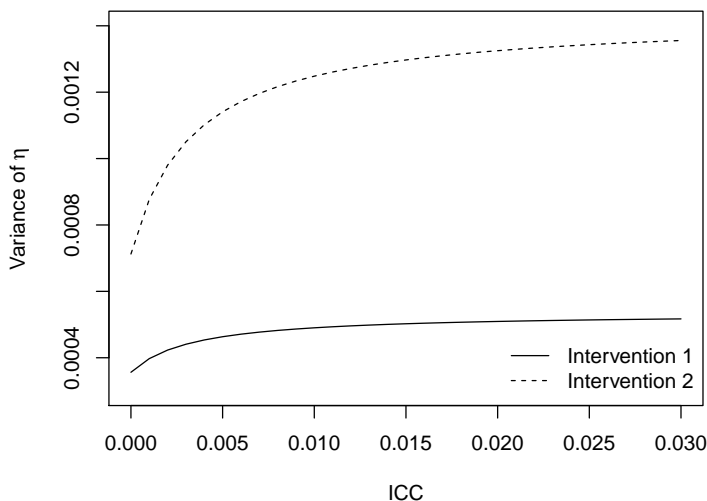


Figure 2.2: Variances of treatment effects vs.  $ICC$  for the replacement design (Table 2.2) under the nonparametric time effect model.

Adding additional time periods of intervention 2 in the end of the design will decrease both of the variances of the intervention effects when  $ICC > 0$ , as shown in Figure 2.3.

I want to investigate the influence of the number of time intervals of intervention 1 on the variances of both treatment effects, too. Figure 2.4 shows the variance- $ICC$  relationships for both interventions in the replacement design with 5 clusters and different numbers of time intervals of intervention 1 under the nonparametric time effect model (see Table 2.6

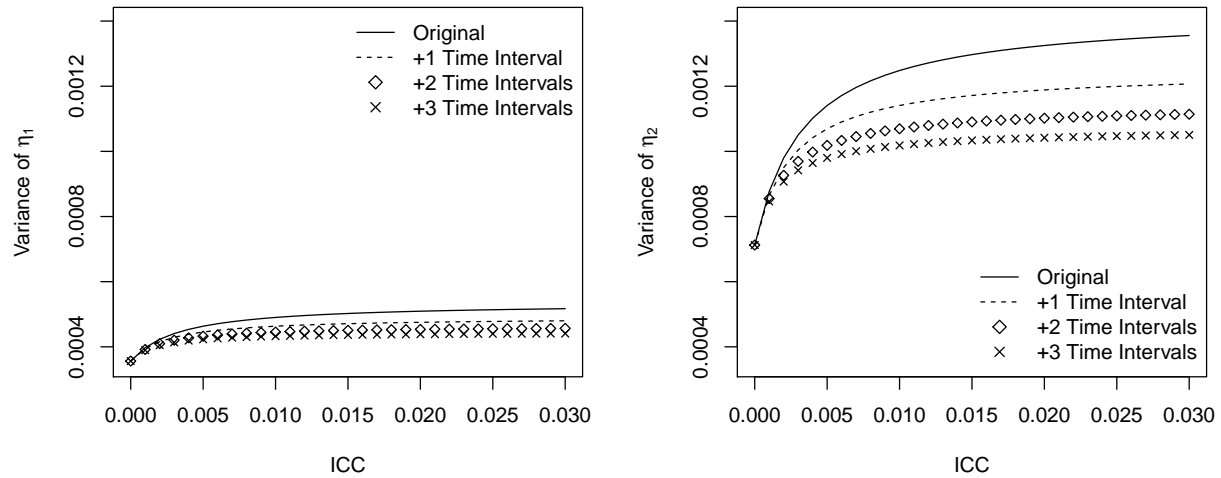


Figure 2.3: Variances of treatment effects vs.  $ICC$  for the original replacement design (Table 2.2) and designs with additional time periods under the nonparametric time effect model.

for an example with 2 time intervals of intervention 1). I used the designs with from 1 to 4 time intervals of intervention 1, because when there are more than 4 time intervals for intervention 1, they are just additional time periods where all clusters receive intervention 1. The additional time intervals will not change the variances in the independent case, and will decrease the variances of treatment effects when  $ICC > 0$ .

	1	2	3	4	5	6	7	8
Cluster 1	0	1	1	2	2	2	2	2
Cluster 2	0	0	1	1	2	2	2	2
Cluster 3	0	0	0	1	1	2	2	2
Cluster 4	0	0	0	0	1	1	2	2
Cluster 5	0	0	0	0	0	1	1	2

Table 2.6: The replacement stepped wedge design for 2 interventions with 5 clusters and 2 time intervals for intervention 1. Every column corresponds to one time interval.

Interestingly, the variances of the treatment effects show no consistent pattern (Figure 2.4). One potential explanation is that when adding more time intervals for intervention 1, it will gain power to evaluate the treatment effect for intervention 1 by having more direct treatment-control comparisons and more data points, but it will also lose some power to estimate more time effects; for estimating the treatment effect for intervention 2, with the increase of the number of time intervals for intervention 1, it will lose some power because of the decrease of the number of direct comparisons between intervention 2 and control and the need to estimate more time effects, but a little power will be gained by having more indirect comparisons.

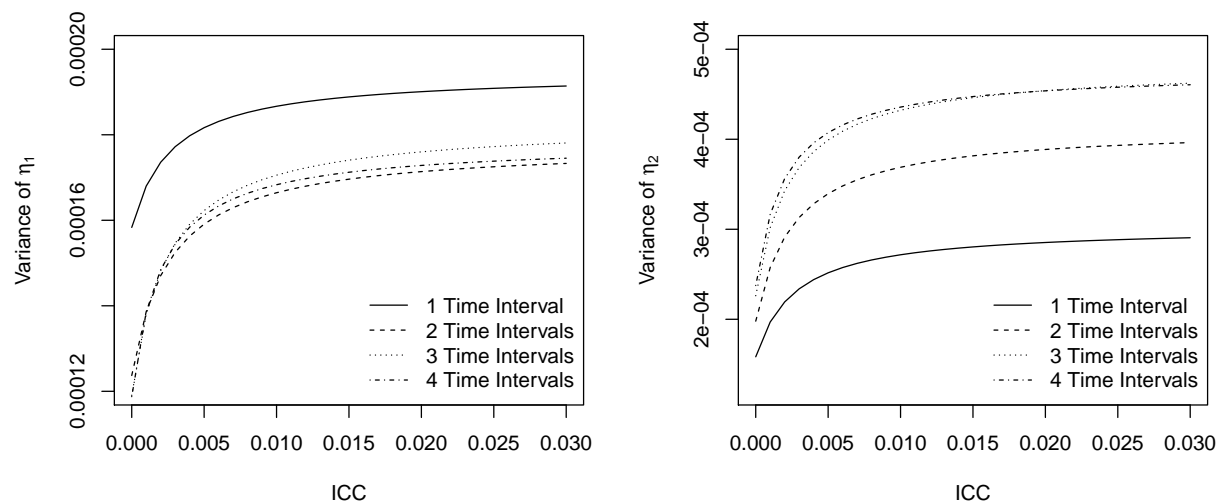


Figure 2.4: Variances of treatment effects vs.  $ICC$  for replacement designs for 5 clusters and with different numbers of time intervals for intervention 1, under the nonparametric time effect model.

To better understand these interesting results, I investigated the variance- $ICC$  relationships for the same replacement designs under the linear time effect model (Figure 2.5) and the no time effect model (Figure 2.6), respectively.

Under the linear time effect model, no matter how many time intervals are in the design, there is only one parameter for time effect. Therefore, with more time intervals for

intervention 1, the variance of the treatment effect for intervention 1 always decreases, and interestingly, the variance of treatment effect for intervention 2 always increases with more time intervals for intervention 1.

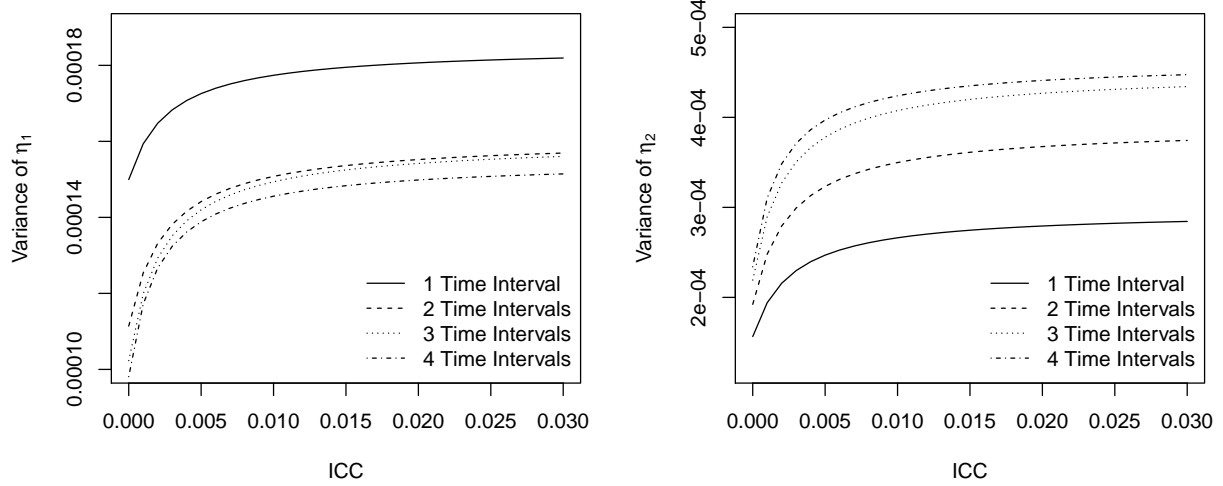


Figure 2.5: Variances of treatment effects vs.  $ICC$  for replacement designs for 5 clusters and with different numbers of time intervals for intervention 1, under the linear time effect model.

Under the no time effect model, it can compare intervals in different time periods to get the between-cluster comparison. Therefore, the variances of treatment effects for both interventions will decrease with the increase of the number of time intervals for intervention 1.

Under the random time effects, an example of a block in the covariance matrix is shown in Appendix A. For simplicity, I assumed that  $Cov(\alpha_i, \gamma_{mi}) = \nu_m = 0$  (i.e. the random cluster effect and the random treatment effect for intervention  $m$  are independent) and  $Cov(\gamma_{mi}, \gamma_{ni}) = \xi_{mn} = 0$  for  $m \neq n$  (i.e. the random effects for two different interventions,  $m$  and  $n$ , are independent). Under this assumption, if  $\eta_1^2 \leq \eta_2^2$ , the variance of the fixed treatment effect for interventions 1 is smaller than the variance for intervention 2. If  $\eta_1^2 \geq \eta_2^2$ , there might be a crossing point in the variances as  $\tau^2$  increasing. The variance of the fixed

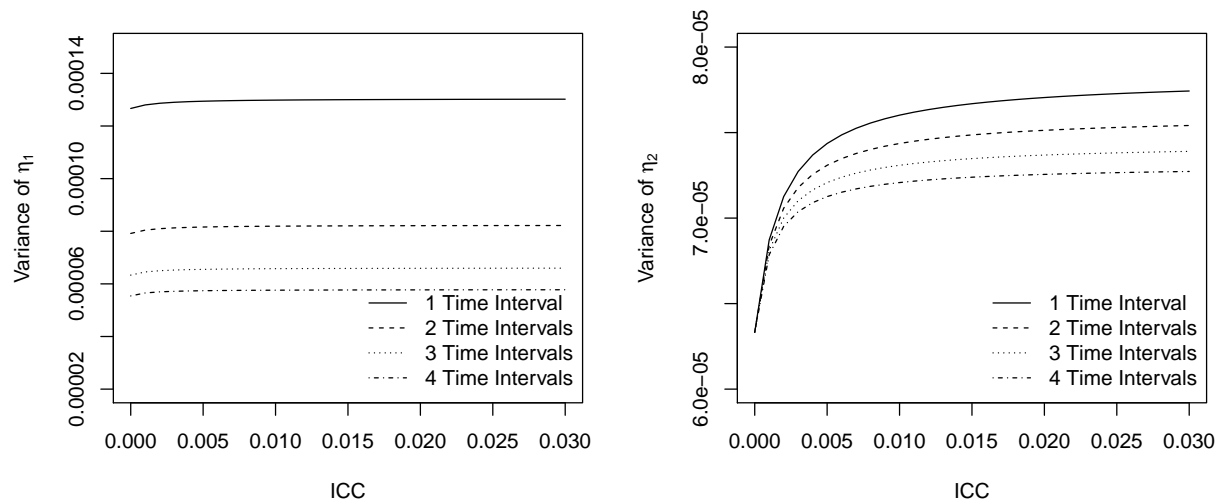


Figure 2.6: Variances of treatment effects vs.  $ICC$  for replacement designs for 5 clusters and with different numbers of time intervals for intervention 1, under the no time effect model.

treatment effect for intervention 1 might be larger than the variance for intervention 2 when  $ICC$  is large enough.

## 2.5 Analyses for Supplementation Design

Similar to the analyses for the replacement design, to get a basic understanding of supplementation design, I used Mathematica to calculate the numerical results for the estimates of treatment effects in the simplest supplementation design with 2 clusters and 4 time intervals (Table 2.7), under the nonparametric time effect model.

	Time Interval 1	Time Interval 2	Time Interval 3	Time Interval 4
Cluster 1	0	1	1+2	1+2
Cluster 2	0	0	1	1+2

Table 2.7: The simplest supplementation stepped wedge design for 2 interventions with 2 clusters and 4 time intervals.

The estimate of the treatment effect for intervention 1 is:

$$\begin{aligned}\hat{\theta}_1 &= \frac{1}{\sigma^2 + 2\tau^2} \left[ \sigma^2(Y_{12} - Y_{22}) + \tau^2(-Y_{11} + 2Y_{12} - Y_{14} + Y_{21} - 2Y_{22} + Y_{24}) \right] \\ &= \frac{1}{1 + \delta} \left[ (Y_{12} - Y_{22}) + \delta \left( -\frac{1}{2}Y_{11} + Y_{12} - \frac{1}{2}Y_{14} \right) + \delta \left( \frac{1}{2}Y_{21} - Y_{22} + \frac{1}{2}Y_{24} \right) \right]\end{aligned}\quad (2.7)$$

where  $\delta = \frac{2\tau^2}{\sigma^2}$ . Rewriting gives

$$\begin{aligned}\hat{\theta}_1 &= \frac{1}{1 + \delta}(Y_{12} - Y_{22}) \\ &+ \frac{\delta}{1 + \delta} \left[ \frac{1}{4}[Y_{12} - Y_{11} - (Y_{22} - Y_{21})] + \frac{1}{4}[Y_{14} - Y_{11} - (Y_{14} - Y_{12} + Y_{22} - Y_{21})] \right] \\ &+ \frac{1}{4}[Y_{24} - Y_{21} - (Y_{14} - Y_{12} + Y_{22} - Y_{21})] + \frac{1}{4}[Y_{24} - Y_{22} - (Y_{14} - Y_{12})]\end{aligned}\quad (2.8)$$

$\hat{\theta}_1$  in the supplementation design is the same as the result in the replacement design, and it can be explained similarly.

The estimate of treatment effect for intervention 2 is

$$\begin{aligned}\hat{\theta}_2 &= \frac{1}{\sigma^2 + 2\tau^2} \left[ \sigma^2(Y_{13} - Y_{23}) + \tau^2(-Y_{11} + 2Y_{13} - Y_{14} + Y_{21} - 2Y_{23} + Y_{24}) \right] \\ &= \frac{1}{1 + \delta} \left[ (Y_{13} - Y_{23}) + \delta \left( -\frac{1}{2}Y_{11} + Y_{13} - \frac{1}{2}Y_{14} \right) + \delta \left( \frac{1}{2}Y_{21} - Y_{23} + \frac{1}{2}Y_{24} \right) \right]\end{aligned}\quad (2.9)$$

which is equal to

$$\begin{aligned}\hat{\theta}_2 &= \frac{1}{1 + \delta}(Y_{13} - Y_{23}) \\ &+ \frac{\delta}{1 + \delta} \left[ \frac{1}{4}[Y_{13} - Y_{11} - (Y_{23} - Y_{21})] + \frac{1}{4}[Y_{14} - Y_{11} - (Y_{14} - Y_{13} + Y_{23} - Y_{21})] \right] \\ &+ \frac{1}{4}[Y_{24} - Y_{21} - (Y_{14} - Y_{13} + Y_{23} - Y_{21})] + \frac{1}{4}[Y_{24} - Y_{23} - (Y_{14} - Y_{13})]\end{aligned}\quad (2.10)$$

Unlike the results in replacement design,  $\hat{\theta}_2$  in supplementation design has the same form as  $\hat{\theta}_1$  and I would expect that the variances of treatment effects for both interventions would be the same under fixed intervention effect models.

The variance-*ICC* relationships for both interventions in the supplementation design (Table 2.3) under the nonparametric time effect model are shown in Figure 2.7. The two variances of the treatment effects are equal to each other as I expected.

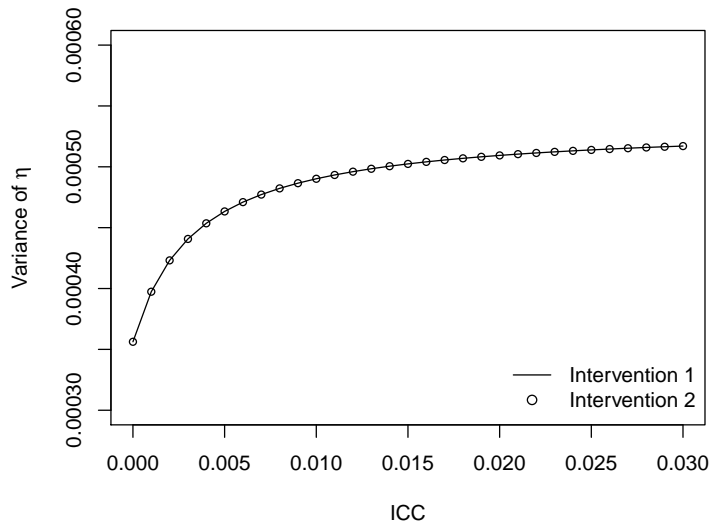


Figure 2.7: Variances of treatment effects vs. *ICC* for the supplementation design (Table 2.3) under the nonparametric time effect model.

Similar to the results for replacement designs, adding additional time periods of treatment in the end of the design will decrease both of the variances of the intervention effects when  $ICC > 0$ . And there is no consistent pattern for designs with different numbers of time intervals of only intervention 1.

An example of a block in the covariance matrix for the supplementation design under random treatment effects model is presented in the Appendix A. If I assume that  $Cov(\alpha_i, \gamma_{mi}) = \nu_m = 0$  (i.e. the random cluster effect and the random treatment effect for intervention  $m$  are independent) and  $Cov(\gamma_{mi}, \gamma_{ni}) = \xi_{mn} = 0$  for  $m \neq n$  (i.e. the random effects for two different interventions,  $m$  and  $n$ , are independent), the variance of the fixed treatment effect for intervention 1 will be smaller than the variance for intervention 2 when  $\eta_1^2 \leq \eta_2^2$  and  $ICC > 0$ . If  $\eta_1^2 > \eta_2^2$ , there will be a crossing point in the variance-*ICC* plots

and the variance of the fixed treatment effect for intervention 1 will be larger when *ICC* is large enough.

## 2.6 Analyses for Factorial Design

The factorial design is symmetric and is designed to introduce both interventions simultaneously. Therefore, the variances of the treatment effects for both interventions are equal under fixed treatment effect models. Additional time periods of treatment at the end of the study will decrease the variances of both treatment effects.

Under the random treatment effect model, examples of blocks in the covariance matrix are shown in Appendix A. The variances of fixed treatment effects for both interventions will be equal if  $\eta_1^2 = \eta_2^2$ . When  $\eta^2$ s are unequal, the intervention with larger  $\eta^2$  will have larger variance for the fixed treatment effect with all nonzero *ICC*.

## 2.7 Across Designs Comparison

Designs with the different numbers of clusters or numbers of time intervals are not comparable to each other. Based on the characteristics of each design, only the replacement and supplementation designs are comparable. Therefore, I compared the variances of the treatment effects in these two designs (Table 2.2 and Table 2.3), which have the same number of clusters and time intervals. Plots of variance-*ICC* relationships are shown in Figure 2.8.

Variances of treatment effects in the supplementation design and the variance of treatment effect for intervention 1 in the corresponding replacement design are equal to each other under the nonparametric time effect model, while the variance of treatment effect for intervention 2 in the replacement design is larger. That is, it would have more power to evaluate the treatment effect for intervention 2 using the supplementation design. Therefore, if possible, I would prefer supplementation design to replacement design just on statistical ground. There might be other considerations that make one or the other design preferable or infeasible.

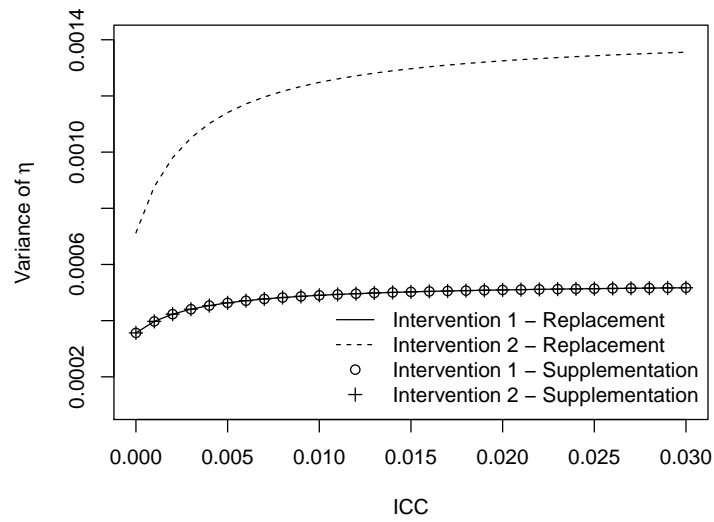


Figure 2.8: Variances of treatment effects vs.  $ICC$  for the replacement design (Table 2.2) and the supplementation design (Table 2.3) under the nonparametric time effect model.

## Chapter 3

### RANDOM TIME EFFECTS MODEL

In the previous chapters, I only considered models with fixed time effects. However, in real cases, there might be random time effects for different clusters. In this chapter, I will use simulations to compare the performances of linear mixed models (LMM) with and without the random effect for time, and generalized estimating equation (GEE).

#### 3.1 Data Simulation

There are a number of ways to simulate data under the random time effects model. For simplicity, I used the random linear time effects model, which is the linear time effects model (1.13) with random slopes. It can be written as

$$Y_{ij} = \mu_{ij} + e_{ij}$$

and

$$\mu_{ij} = \mu + \alpha_i + t_j\beta_i + X_{ij}\theta \tag{3.1}$$

where  $t_j =$  beginning time in interval  $j$ , and  $\beta_i \sim N(\beta, s^2)$ . In the simulations, I assumed  $t_j = j$  and  $\mu = 0.3$ .

I considered three cases: the true treatment effect ( $\theta$ ) equal to -0.2, 0, and 0.2. For each case, data were simulated with the standard deviation of the random slope ( $s$ ) equal to 0 (i.e. the fixed linear time effect model), 0.003, 0.006, 0.009, and 0.012, and with the mean of the slope ( $\beta$ ) equal to -0.01, 0, and 0.01. For every situation with specific  $\theta$ ,  $s$  and  $\beta$ , 1000 datasets were simulated under the assumptions that  $N$  (number of observations per time interval per cluster) = 100,  $\tau$  (standard deviation of the random cluster effect) = 0.05,

and  $\sigma_e$  (standard deviation of an individual-level error) = 0.5 ( $\sigma$  (standard deviation of a cluster-level error) = 0.05). I used the standard stepped wedge design with 10 clusters and 11 time periods, which means  $j = 0, \dots, 10$ .

For each simulation, I analyzed the data using LMM with random intercept, LMM with random intercept and random slope, and GEE with exchangeable correlation. Because of the small number of clusters, I also looked at the inflated standard errors for both of the LMMs and the GEE model.

$$\begin{aligned} \text{Inflated S.E.} &= \sqrt{\frac{\text{Number of Clusters}}{\text{Number of Clusters} - \text{Number of Fixed Effects}}} \times \text{S.E.} \\ &= \sqrt{\frac{10}{7}} \times \text{S.E.} \end{aligned} \quad (3.2)$$

### 3.2 Simulation Results

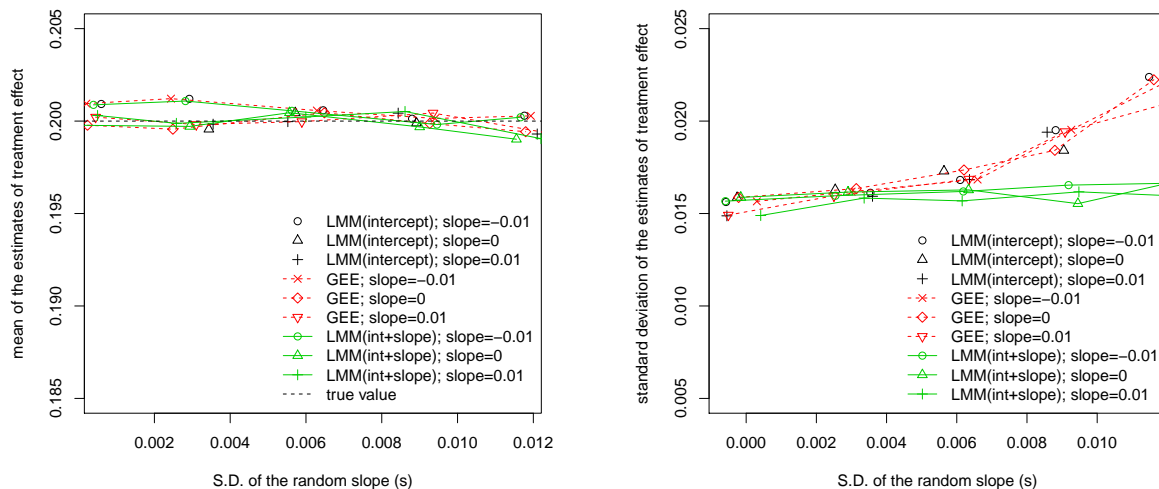


Figure 3.1: Mean of  $\hat{\theta}$  vs. S.D. of the slope. Figure 3.2: S.D. of  $\hat{\theta}$  vs. S.D. of the slope.

Figures 3.1 and 3.2 show the mean and standard deviation of the estimates of treatment effect when the true treatment effect( $\theta$ ) = 0.2. Legend is in the form: model; mean of

the slope. According to Figure 3.1, the mean of the estimates of treatment effect fluctuate around the true value. Compared to the standard deviation of the estimates, the deviations from the true value are not large enough to suggest a bias. Figure 3.2 shows that generally, the standard deviation of the estimates of treatment effect increase with the increase of the standard deviation of the random slope. The standard deviation of the random slope from the true model (i.e. LMM with random intercept and random slope) is the smallest among all the three models under the same assumption.

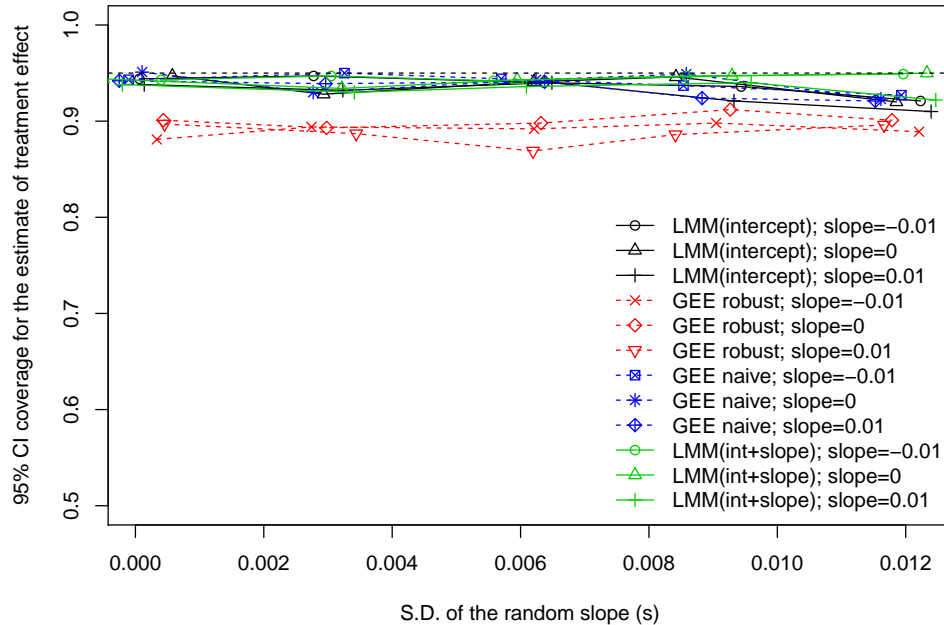


Figure 3.3: Coverages of the 95% confidence intervals for the treatment effect using non-inflated standard errors versus standard deviation of random slope for the time effect.

Coverages of the 95% confidence intervals for the treatment effect when  $\theta = 0.2$  are shown in Figure 3.3. Overall, the true model, LMM with random intercept and random slope, has the best performance, as the coverage is around 95% with varying  $s$ . The LMM with random intercept and the model using the naive standard error from GEE provide good coverage when  $s$  is small but decrease as  $s$  increases. The coverage from the model using

the robust standard error from GEE is around 90%, but there is not a clear decrease in the coverage with the increase of  $s$ .

Since the number of clusters in the dataset is small (10 clusters), it might be helpful to use inflated standard errors as defined above in Equation 3.2. The mean of the standard errors from simulations should be close to the standard deviation of  $\hat{\theta}$ . Figures 3.4 and 3.5 show the correlations between the mean of the non-inflated standard errors and the standard deviations of the estimates of treatment effect, and the correlation between the inflated mean standard errors and the standard deviations, respectively, for the case when  $\theta = 0.2$ . According to the figures, the mean of the original standard errors are generally smaller than the standard deviations and the mean non-inflated standard errors from the true model (LMM with random intercept and random slope) are the closest to the corresponding standard deviations, while the mean of the inflated standard errors are generally higher than the standard deviations and the mean inflated robust standard errors from the GEE are the closest to the corresponding standard deviations. Based on these results, I would expect to have better coverage for 95% confidence intervals for treatment effects with inflated standard errors.

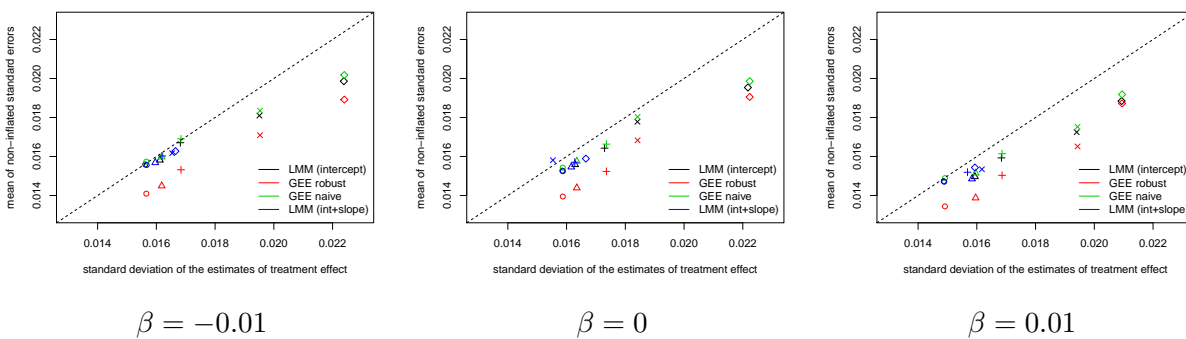


Figure 3.4: Mean of the non-inflated standard errors versus the standard deviations of the estimates of treatment effect. The dash line is a line with intercept equal to 0, and slope equal to 1. Circles, triangles (point up), plus, cross, and diamond symbols correspond to the cases when  $s = 0, 0.003, 0.006, 0.009$ , and  $0.012$ , respectively.

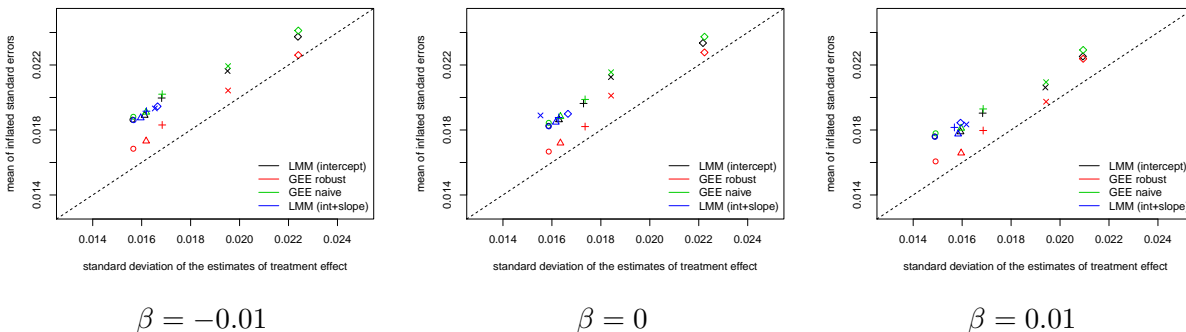


Figure 3.5: Mean of the inflated standard errors versus the standard deviations of the estimates of treatment effect. The dash line is a line with intercept equal to 0, and slope equal to 1. Circles, triangles (point up), plus, cross, and diamond symbols correspond to the cases when  $s = 0, 0.003, 0.006, 0.009,$  and  $0.012,$  respectively.

Coverages of the 95% confidence intervals for the treatment effect using the inflated standard errors are shown in Figure 3.6. The coverage increases for all of the models with the correction. Typically, the coverage from the model using inflated robust standard errors from GEE increases and gets closer to 0.95. However, note that all the three models, the LMM with random intercept, GEE with naive standard errors, and the LMM with random intercept and random slope, provide good coverage without the inflation.

Finally, I am interested to see how these three models perform to estimate the slope (coefficient of time). Figure 3.7 shows the mean of the estimates of slope for the three models when  $\theta = 0.2$ . The mean of the estimates of slope are close to each other under all the three models, and there are very small deviations from the true value when  $s$  greater than 0 in some cases. For all of the above results, I found similar patterns with varying  $\theta$ .

In conclusion, the LMM with random intercept and random slope has the best performance, which is not surprising since it is the true model. The estimates of treatment effect from the LMM with only random intercept and the GEE model with exchangeable correlation fluctuate around the true value and close to each other and the standard errors from the LMM with only random intercept are close to the naive standard errors from the GEE

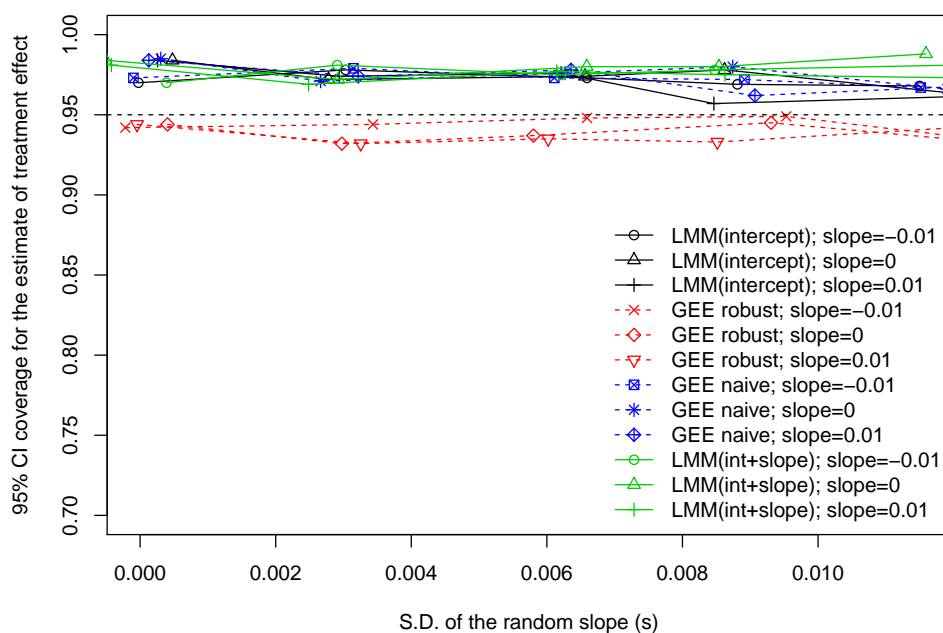


Figure 3.6: Coverages of the 95% confidence intervals for the treatment effect using inflated standard errors versus standard deviation of random slope for the time effect.

model, too. The robust standard error from the GEE model is the smallest among the four kinds of standard errors under the same assumption. It has the lowest coverage of the 95% confidence interval for treatment effect. However, the coverage would not decrease much with the increase of the standard deviation of the random slope and will be acceptable after simple correction.

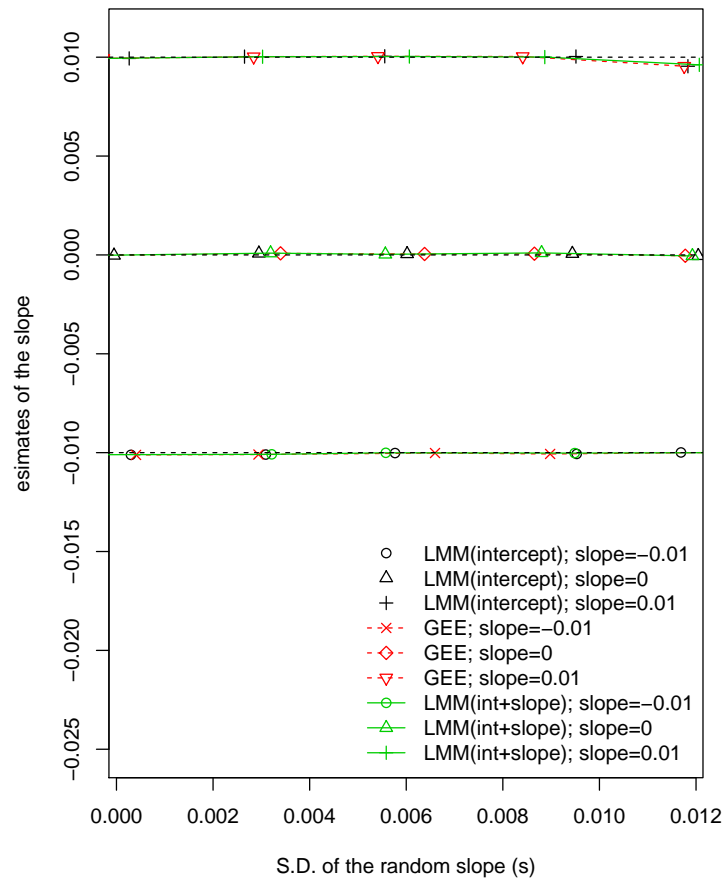


Figure 3.7: Mean of the estimates of slope versus standard deviation of random slope for the time effect.

## Chapter 4

### DISCUSSION AND LIMITATION

Classic stepped wedge designs usually include only one intervention. This thesis is aimed to extend the cases to multiple interventions, and compare the performances of some basic variants. However, there are some assumptions made in the analyses for simplicity, which might not be correct in real applications.

First, I assumed the numbers of individuals per time interval per cluster ( $N$ ) are always equal. The relationships between the variances of the estimates of the two treatment effects in the Concurrent, Replacement, Supplementation, and Factorial designs would be different if there are different numbers of individuals for different time intervals or clusters.

Second, for analyses under the random treatment effects model, I assumed that the random cluster effect and the random treatment effect for every intervention are independent, as well as the random effects for two different interventions are independent. These assumptions would be inappropriate if, for example, the treatment effect is larger when the cluster-level mean is higher. However, the situation will be complicated if all these correlations were taken into consideration. I just listed some examples of the blocks in the covariance matrices under the random treatment effects model in the Appendix.

Third, for the analyses under the random time effects model in Chapter 3, I made a strong assumption that the time trend for every cluster is linear. This assumption simplifies the analyses, but it also limits the generalizability of our conclusions. One of many other possible ways to simulate data for the random time effects model is to use the nonparametric time effects model with random effect for every time effect. Note that the correlation between consecutive time intervals might need to be considered in that model.

## GLOSSARY

CRT: stands for cluster randomized trials, which randomize groups, rather than individual subjects, to receive different interventions[6].

GLM: stands for generalized linear model, which is a statistical regression model.

LMM: stands for linear mixed models[7]. It is a statistical model, which includes fixed effects and random effects.

GEE: stands for generalized estimating equation[8], which is an extension of generalized linear models for the analysis of longitudinal data.

## BIBLIOGRAPHY

- [1] Celia A Brown and Richard J Lilford. The stepped wedge trial design: a systematic review. *BMC medical research methodology*, 6(1):1, 2006.
- [2] Peter McCullagh and John A Nelder. *Generalized linear models*, volume 37. CRC press, 1989.
- [3] Hussey M A, Hughes J P. Design and analysis of stepped wedge cluster randomized trials. *Contemporary clinical trials*, 28(2):182–191, 2007.
- [4] Hughes J P, Granston T S, Heagerty P J. Current issues in the design and analysis of stepped wedge trials. *Contemporary Clinical Trials*, 45:55–60, 2015.
- [5] Wolfram Research, Inc. Mathematica.
- [6] Todd A Florin and Samir S Shah. Cluster randomized trials. *AAP Grand Rounds*, 26(2):15–15, 2011.
- [7] Charles E McCulloch and John M Neuhaus. *Generalized linear mixed models*. Wiley Online Library, 2001.
- [8] Scott L Zeger, Kung-Yee Liang, and Paul S Albert. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, pages 1049–1060, 1988.

## Appendix A

### COVARIANCE MATRIX FOR RANDOM TREATMENT EFFECTS MODEL WITH MULTIPLE INTERVENTIONS

For a design with  $I$  clusters and  $T$  time intervals, the covariance matrix under random treatment effects model (Equation 2.2) is a block diagonal matrix with dimension  $IT \times IT$ . Examples of  $T \times T$  block matrix for different designs are presented below. For cases with 2 interventions, there is only one  $\xi$  term. Therefore, I assume that  $\xi_{12} = \xi$ .

#### A.1 Concurrent design with 2 interventions

For the design shown in Table 2.1, the first block in the covariance matrix, which corresponds to the first row in the table, is

$$\begin{pmatrix} \sigma^2 + \tau^2 & \tau^2 + \nu_1 & \tau^2 + \nu_1 & \tau^2 + \nu_1 \\ \tau^2 + \nu_1 & \sigma^2 + \tau^2 + \eta_1^2 + 2\nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 \\ \tau^2 + \nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 & \sigma^2 + \tau^2 + \eta_1^2 + 2\nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 \\ \tau^2 + \nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 & \sigma^2 + \tau^2 + \eta_1^2 + 2\nu_1 \end{pmatrix}$$

The fifth block in the covariance matrix, which corresponds to the fifth row in Table 2.1, is

$$\begin{pmatrix} \sigma^2 + \tau^2 & \tau^2 & \tau^2 + \nu_2 & \tau^2 + \nu_2 \\ \tau^2 & \sigma^2 + \tau^2 & \tau^2 + \nu_2 & \tau^2 + \nu_2 \\ \tau^2 + \nu_2 & \tau^2 + \nu_2 & \sigma^2 + \tau^2 + \eta_2^2 + 2\nu_2 & \tau^2 + \eta_2^2 + 2\nu_2 \\ \tau^2 + \nu_2 & \tau^2 + \nu_2 & \tau^2 + \eta_2^2 + 2\nu_2 & \sigma^2 + \tau^2 + \eta_2^2 + 2\nu_2 \end{pmatrix}$$

### A.2 Replacement design with 2 interventions

For the design shown in Table 2.2, the first block in the covariance matrix, which corresponds to the first row in the table, is

$$\begin{pmatrix} \sigma^2 + \tau^2 & \tau^2 + \nu_1 & \tau^2 + \nu_1 & \tau^2 + \nu_2 & \tau^2 + \nu_2 & \tau^2 + \nu_2 \\ \tau^2 + \nu_1 & \sigma^2 + \tau^2 + \eta_1^2 + 2\nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 & \tau^2 + \nu_1 + \nu_2 + \xi & \tau^2 + \nu_1 + \nu_2 + \xi & \tau^2 + \nu_1 + \nu_2 + \xi \\ \tau^2 + \nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 & \sigma^2 + \tau^2 + \eta_1^2 + 2\nu_1 & \tau^2 + \nu_1 + \nu_2 + \xi & \tau^2 + \nu_1 + \nu_2 + \xi & \tau^2 + \nu_1 + \nu_2 + \xi \\ \tau^2 + \nu_2 & \tau^2 + \nu_1 + \nu_2 + \xi & \tau^2 + \nu_1 + \nu_2 + \xi & \sigma^2 + \tau^2 + \eta_2^2 + 2\nu_2 & \tau^2 + \eta_2^2 + 2\nu_2 & \tau^2 + \eta_2^2 + 2\nu_2 \\ \tau^2 + \nu_2 & \tau^2 + \nu_1 + \nu_2 + \xi & \tau^2 + \nu_1 + \nu_2 + \xi & \tau^2 + \eta_2^2 + 2\nu_2 & \sigma^2 + \tau^2 + \eta_2^2 + 2\nu_2 & \tau^2 + \eta_2^2 + 2\nu_2 \\ \tau^2 + \nu_2 & \tau^2 + \nu_1 + \nu_2 + \xi & \tau^2 + \nu_1 + \nu_2 + \xi & \tau^2 + \eta_2^2 + 2\nu_2 & \tau^2 + \eta_2^2 + 2\nu_2 & \sigma^2 + \tau^2 + \eta_2^2 + 2\nu_2 \end{pmatrix}$$

### A.3 Supplementaion design with 2 interventions

For the design shown in Table 2.3, the third block in the covariance matrix, which corresponds to the third row in the table, is

$$\begin{pmatrix} \sigma^2 + \tau^2 & \tau^2 & \tau^2 & \tau^2 + \nu_1 & \tau^2 + \nu_1 & \tau^2 + \nu_1 + \nu_2 \\ \tau^2 & \sigma^2 + \tau^2 & \tau^2 & \tau^2 + \nu_1 & \tau^2 + \nu_1 & \tau^2 + \nu_1 + \nu_2 \\ \tau^2 & \tau^2 & \sigma^2 + \tau^2 & \tau^2 + \nu_1 & \tau^2 + \nu_1 & \tau^2 + \nu_1 + \nu_2 \\ \tau^2 + \nu_1 & \tau^2 + \nu_1 & \tau^2 + \nu_1 & \sigma^2 + \tau^2 + \eta_1^2 + 2\nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 + \nu_2 + \xi \\ \tau^2 + \nu_1 & \tau^2 + \nu_1 & \tau^2 + \nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 & \sigma^2 + \tau^2 + \eta_1^2 + 2\nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 + \nu_2 + \xi \\ \tau^2 + \nu_1 + \nu_2 & \tau^2 + \nu_1 + \nu_2 & \tau^2 + \nu_1 + \nu_2 & \tau^2 + \eta_1^2 + 2\nu_1 + \nu_2 + \xi & \tau^2 + \eta_1^2 + 2\nu_1 + \nu_2 + \xi & \sigma^2 + \tau^2 + \eta_1^2 + \eta_2^2 + 2(\nu_1 + \nu_2 + \xi) \end{pmatrix}$$

### A.4 Factorial design with 2 interventions

For the design shown in Table 2.4, the second block in the covariance matrix, which corresponds to the second row in the table, is

$$\begin{pmatrix} \sigma^2 + \tau^2 & \tau^2 & \tau^2 + \nu_1 & \tau^2 + \nu_1 & \tau^2 + \nu_1 + \nu_2 \\ \tau^2 & \sigma^2 + \tau^2 & \tau^2 + \nu_1 & \tau^2 + \nu_1 & \tau^2 + \nu_1 + \nu_2 \\ \tau^2 + \nu_1 & \tau^2 + \nu_1 & \sigma^2 + \tau^2 + \eta_1^2 + 2\nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 + \nu_2 + \xi \\ \tau^2 + \nu_1 & \tau^2 + \nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 & \sigma^2 + \tau^2 + \eta_1^2 + 2\nu_1 & \tau^2 + \eta_1^2 + 2\nu_1 + \nu_2 + \xi \\ \tau^2 + \nu_1 + \nu_2 & \tau^2 + \nu_1 + \nu_2 & \tau^2 + \eta_1^2 + 2\nu_1 + \nu_2 + \xi & \tau^2 + \eta_1^2 + 2\nu_1 + \nu_2 + \xi & \sigma^2 + \tau^2 + \eta_1^2 + \eta_2^2 + 2(\nu_1 + \nu_2 + \xi) \end{pmatrix}$$