Methods for the Analysis of Pretest-Posttest Binary Outcomes from Cluster Randomization Trials

(Spine title: Analysis of Pretest-Posttest Binary Outcomes from Cluster Randomization Trials)

(Thesis format: Monograph)

by

ASM <u>Borhan</u>

Graduate Program in Epidemiology & Biostatistics

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

School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© ASM Borhan, 2012

Abstract

In this thesis we investigate methods for assessing the intervention effect in completely randomized, cluster randomization trials where participants are measured prior to random assignment and again after completion of the intervention, i.e. a pretest-posttest design. Attention is further limited to binary outcomes. We compare the performance of six test statistics used to test the intervention effect. Test statistics are obtained from cluster-specific and population-averaged extensions of logistic regression. A simulation study is used to estimate type I error and power for the test statistics. In addition, we examine the effect on power of correctly assuming a common pretest-posttest association. Cluster-specific models yielded satisfactory 5% type I error rates while a longitudinal approach yielded the lowest power. Assumptions about the pretest-posttest association had little effect on power. Data from a school-based randomized trial are used to illustrate results.

KEYWORDS: Cohort design; Generalized estimating equations; Wald test.

Acknowledgments

My sincere appreciation goes to my supervisors, Dr. Gerarda Darlington, and Dr. Neil Klar for their incredible guidance, advice and encouragement. This thesis would have not been completed without their support.

I am indebted to Dr. Allan Donner for his help and encouragement since admission into this program. I am also sincerely grateful to Dr. John Koval, Dr. GY Zou, Dr. Yun-Hee Choi, and all other faculty members of the Department of Epidemiology and Biostatistics for their excellent lectures in Biostatistics and enthusiastic help. Dr. Stephanie Dixon also deserves many thanks.

I wish to thank my friends and many student colleagues for providing a stimulating environment to learn and grow.

My final words go to my family. I must thank my wife Syeda Munira, for her consistent love and support. I also must thank my parents.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Introduction

Randomized control trials may be designed to assess an intervention effect using the study outcome measured at baseline and again after the intervention is completed. Methodologies have been developed to evaluate the intervention effect using this pretest-posttest trial design when the allocation unit is an individual subject (Cronbach and Furby, 1970; Bonate, 2000; Senn, 1994). However, less attention has been given in the case of cluster randomized trials (CRT), especially when the outcome of interest is binary.

1.1.1 Cluster Randomization Trial

Randomized trials involving randomization of intact groups of subjects, instead of independent individuals, are commonly referred to as cluster randomized trials (Donner and Klar, 2000). The rate of adopting cluster randomization trials is increasing (Bland, 2004). Allocation units are diverse in such studies, and can include families or households, classrooms or schools, neighborhoods, and even entire communities.

For example, Flay *et al.* (1995) discuss a school-based, smoking prevention CRT, titled the Television, School, and Family Smoking Prevention and Cessation Project

(TVSFP). This trial examined the efficacy of a smoking prevention intervention in terms of tobacco and health knowledge, coping skills, and the prevalence of tobacco use. A total 7351 students drawn from 47 schools from Los Angeles and San Diego, California participated. Baseline or pretest data were collected from students prior to random assignment, while the posttest data were collected 1 and 2 years after the intervention.

Random allocation of clusters typically results in correlation among the outcomes of subjects from the same cluster. This degree of similarity is measured using an intracluster correlation coefficient (ICC), denoted by the Greek letter ρ. Interpretation of the ICC is analogous to the standard Pearson correlation coefficient between any two observations from subjects in the same cluster (Donner and Klar, 2000).

Correlation of responses among clustered subjects invalidates application of statistical techniques which assume independent observations. Standard statistical methodology needs to be adjusted for this clustering effect. Correlation in the responses inflates the variance of the estimated intervention effect. For continuous or binary data, this inflation can be quantified by the design effect, or variance inflation factor, given by $1 + (\bar{m} - 1)\rho$, where $\bar{m}$ is the average cluster size (Donner and Klar, 2000). Large inflation in the estimated variance can be found even with a small intracluster correlation coefficient when the average cluster size is large. For example, Hedeker et $al.$ (1994) estimated $\hat{\rho} = 0.02$ at the school-level using the data from the TVSFP (Flay et $al.$, 1995). Therefore, there is approximately a two-fold increase in the variance due to clustering of students within school since the average number of participants per school is $\bar{m} = 57$ so that the estimated variance inflation factor is $1 + (57 - 1)0.02 = 2.12$.

The clustering effect needs to be accounted for in the analysis in order to maintain valid statistical inferences. Specifically, failure to account for clustering when testing the intervention effect leads to an inflated type I error rate. In addition,

ignoring clustering when estimating sample size results in low power to detect the intervention effect even for test statistics which provide valid type I error rates by accounting for the effects of clustering (Donner and Klar, 2000).

Depending on the allocation of clusters, most cluster randomization trials can be classified as using one of three basic types of designs: (a) completely randomized, (b) matched-pair, or (c) stratified. Completely randomized designs omit pre-stratification and matching on baseline prognostic factors. This design is most suited for trials enrolling fairly large numbers of clusters (Donner and Klar, 2000). Random assignment of one of the two clusters in a stratum to each intervention group is termed a matched-pair design. The stratified design extends the matched-pair design where more than two clusters are randomly allocated to intervention groups within strata.

Random allocation creates groups that are identical at baseline on average (Altman and Dore, 1990). For any one trial there may be differences in baseline characteristics across intervention groups. Assessing the intervention effect considering only post intervention data is based on the assumption that groups are comparable at baseline. Unbalanced groups at baseline do not invalidate statistical inferences but they do reduce power to detect intervention effects. Incorporation of pretest data increases the power of the statistical test. In order to detect an intervention effect more precisely, both pretest and posttest data are considered in this thesis.

The primary purpose of this study is to examine the intervention effect using dichotomous data from the pretest-posttest setting of cluster randomized trials. This study focuses on the completely randomized design and is limited to one pretest and one posttest binary outcome measurement per subject. Moreover, attention is limited to trials having a single experimental and a control arm. In addition, small numbers of large clusters are considered in this thesis. Furthermore, we examine the impact of assumptions about the pretest-posttest associations at individual- and cluster-level in detecting the effect of intervention. This work extends that of Klar and Darlington

(2004) who limited attention to Gaussian data.

1.1.2   The Pretest-Posttest Design

1.1.2.1   Analysis of Pretest-Posttest Binary Data: Individually Randomized Trials

Methods for analysing pretest-posttest binary measurements may be based on (a) posttest observations only (b) posttest observations adjusted for baseline measurement, and (c) analysis of change from pretest to posttest measurements (Twisk, 2003; Bonate, 2000). The first two approaches may use logistic regression. For the second method, a logistic regression version of analysis of covariance can be used (Twisk and Proper, 2004). However, for the third method there is an ongoing debate regarding how to define change (Cronbach and Furby, 1970; Plewis, 1985; Twisk and Proper, 2004). Absolute change between pretest and posttest measurements is the most commonly used method of defining change in the context of quantitative data. For analysing change from pretest to posttest measurements we consider a longitudinal data analysis (LDA) approach (Liang and Zeger, 2000; Ukoumunne and Thompson, 2001).

1.1.2.2   Analysis of Pretest-Posttest Binary Data: Cluster Randomization Trials

The methodologies described in the previous section can be extended for analysing pretest-posttest dichotomous observations in the context of cluster randomization trials. Extensions take into account the anticipated clustering effect. Diggle *et al.* (2002) documented cluster-specific and population-averaged extensions of logistic regression for analysing correlated binary data. Both these extensions will be described further in Chapter 2.

1.2 Methodological developments: Correlated Binary Data: Pretest-Posttest Design

The literature for correlated binary data is vast. Therefore in this section attention is limited to analysis of pretest-posttest measures in cluster randomized trials with binary outcomes.

Ukoumunne and Thompson (2001) compared methods for repeated binary measurements in the context of cluster randomized trials. They focused on cross-sectional clustered data with fixed cluster size. These are cluster randomization trials where independent samples of subjects are selected per cluster at two or more time points. They evaluated the effect of intervention based on follow-up responses alone, follow-up responses adjusted for baseline responses, and change from baseline to follow-up responses.

Austin (2010) extended the work of Ukoumunne and Thompson (2001) and examined the empirical power of different methods for the analysis of cross-sectional repeated binary measurements. In testing the intervention effect it was found that the random-effects model for analysing change from baseline to follow-up responses yielded the lowest power. It was also noted that methods which accounted for baseline responses tended to have marginally greater power compared to methods which did not adjust for baseline responses in some scenarios.

We focus on the approach investigated by Ukoumunne and Thompson (2001) and Austin (2010) for assessing the effect of intervention from a cohort design where the same subjects are assessed over time.

Nixon and Thompson (2003) investigated the effect of intervention based on a method considering only follow-up outcome i.e. random effects logistic regression and method adjusted for baseline outcome (ANCOVA). Similar to Ukoumunne and

Thompson (2001) they focused on repeated cross-sectional binary data. They found both large cluster sizes and large between-cluster variances at baseline increased the precision of the intervention effect.

Sashegyi *et al.* (2000) discussed methods for analysing correlated binary data from cluster randomization trials using a cohort design allowing two or more repeated measurements per subject. To account the clustering that arises due to repeated observations from the same subjects from the same cluster over time and correlation among the observations from the same cluster they proposed a composite model, which combines empirical Bayes and generalized estimating equation (GEE) models. Coefficients from the composite model can be interpreted as cluster-specific log odds ratios. LDA can be considered as a special case of this composite model.

Localio *et al.* (2006) assessed the intervention effect based on a method which considers change from baseline to follow-up binary responses i.e. LDA. They focused on two designs: (1) the repeated cross-sectional design in which different subjects from the same cluster are measured at different occasions, and (2) the cohort design in which individuals from the same cluster are measured over time. The coefficient estimate corresponding to the interaction between time and treatment was the estimate of interest. They investigated the performance of several methods for estimating this interaction term. The authors found that Bayesian methods using Gibbs sampling yielded the best results in terms of bias and coverage compared to GEE, penalized qausi likelihood (PQL), and quadrature methods for a cohort design.

Donner and Zou (2007) discussed techniques for correlated binary data in the presence of a baseline measurement when there is dependency among the clusters. The data were collected from a design where subjects' mouths are divided into two segments (left or right). The authors derived a chi-square statistic which takes into account the correlation among observations in the same segment as well as the dependencies among observations in different segments of a subject's mouth. In the

presence of baseline measurements they considered an ANCOVA approach. Moreover, they considered a time effect for analysing change from baseline to follow-up. However, the authors only focused on the GEE extension of logistic regression approach for analysis. Moreover, this work is not strictly relevant for our work because we limit attention to independent clusters.

1.3   Objectives of the Study

The estimated effect of an intervention will be investigated using extensions of logistic regression adjusted for clustering. Models being compared will use (i) posttest measurement only, (ii) logistic ANCOVA of the posttest measurements adjusting for the pretest measurement, and (iii) a longitudinal approach (LDA). Both cluster-specific and population-averaged extensions of these models will be considered. For each model, simulation studies will be used to investigate aspects of the estimated intervention effect (i.e. odds ratio) including;

1. empirical type I error (H_0: true odds ratio=1);

2. power for those methods which have valid type I error;

3. bias of estimated *log* odds ratio; and

4. precision of estimated *log* odds ratio.

In addition, comparison among these models will be made using data from the TVSFP (Flay *et al.*, 1995). This study was introduced in Section 1.1.

1.4   Organization of the Thesis

This thesis has six chapters. Chapter 1 introduces cluster randomization describing the role of the completely randomized design and the pretest-posttest design. The objectives and rationale for this study are also discussed in this chapter. A detailed description of methods for analysing binary data from pretest-posttest cluster randomized trials are discussed in chapter 2. Design of the simulation study is described in chapter 3. Results from this simulation study are presented in chapter 4. Analysis of data from the TVSFP (Flay *et al.*, 1995) is then provided in chapter 5 using methods described in chapter 2. Analyses of data from the TVSFP are discussed and compared to those obtained from the simulation study in chapter 6. Chapter 6 also provides a summary of overall study findings and suggestions for further research.

Chapter 2

ASSESSING THE INTERVENTION EFFECT: APPLICATION TO PRETEST-POSTTEST CLUSTER RANDOMIZATION TRIALS

2.1 Introduction

In this chapter we describe methods for assessing the intervention effect using data from a completely randomized, cluster randomized trial with a pretest-posttest design. As noted in chapter 1, limited attention has been given to this analytical challenge when the outcome is binary. The methods discussed in this chapter will be investigated by a simulation study whose design appears in Chapter 3.

There are five sections in this chapter. Notation used throughout the thesis are defined in section 2.2. Extensions of logistic regression models for correlated binary data are described in section 2.3 and methods for testing the intervention effect are discussed in section 2.4. Finally, the chapter is summarized in section 2.5.

2.2 Notation

In this thesis, attention is limited to two-arm (experimental and control) trials within the context of a completely randomized design. The selected notation reflects this

feature of the study design. Let x_{ijs} and y_{ijs}, $i = 0, 1$, $j = 1, \ldots, J_i$, $s = 1, \ldots, n_{ij}$, denote the binary pretest and posttest observations for the sth subject from the jth cluster assigned to the ith intervention group. Here $i = 0$ denotes the control group while $i = 1$ denotes the treatment group. Moreover, n_{ij} denotes the cluster size of cluster j within intervention group i.

Attention is further limited to fixed cluster sizes and a fixed number of clusters in each intervention group. That is $n_{ij} = n$ and $J_i = J$ for $i = 0, 1$.

The pretest and posttest binary variables x_{ijs} and y_{ijs} are defined as

$$
x_{ijs}, y_{ijs} = \begin{cases} 1, & \text{if event occurs} \\ 0, & \text{otherwise} \end{cases}
$$

Table 2.1 provides definitions of important quantities needed to discuss the methods that will be compared.

2.3 Models for correlated binary data

2.3.1 Extensions of logistic regression

The correlation among cluster members can be accounted for by incorporating two sources of random variations: between-cluster random variation and within-cluster random variation. Two extensions of logistic regression are considered in this thesis which allow for clustering. These extensions have been described as cluster-specific (CS) models or as population-averaged (PA) models (Donner and Klar, 2000; Diggle *et al.*, 2002). Cluster-specific and population-averaged extensions of logistic regression are described in sections 2.3.2 and 2.3.3, respectively. The relationship between the regression coefficients from cluster-specific and population-averaged models is described in section 2.3.4.

Table 2.1: Notation for several measurements among pretest and posttest observations

Measurements	Pretest	Posttest
Number of successes in (i,j)th cluster	$x_{ij} = \sum\limits_{s=1}^{n} x_{ijs}$	$y_{ij} = \sum\limits_{s=1}^{n} y_{ijs}$
Event rates in the (i,j)th cluster	$\hat{p}_{ij.} = \dfrac{x_{ij}}{n}$	$\hat{p}'_{ij.} = \dfrac{y_{ij}}{n}$
Event rates in ith intervention group	$\hat{p}_{i..} = \dfrac{\sum\limits_{j=1}^{J}\sum\limits_{s=1}^{n} x_{ijs}}{Jn}$	$\hat{p}'_{i..} = \dfrac{\sum\limits_{j=1}^{J}\sum\limits_{s=1}^{n} y_{ijs}}{Jn}$
Overall event rates	$\hat{p}_{...} = \dfrac{\sum\limits_{i=0}^{1}\sum\limits_{j=1}^{J}\sum\limits_{s=1}^{n} x_{ijs}}{2Jn}$	$\hat{p}'_{...} = \dfrac{\sum\limits_{i=0}^{1}\sum\limits_{j=1}^{J}\sum\limits_{s=1}^{n} y_{ijs}}{2Jn}$

2.3.2 Cluster-specific model

A cluster-specific extension of logistic regression may be written as:

$$logit(P_{ijs} = Pr(y_{ijs} = 1 \mid T_i, e_{ij})) = \alpha_1 + \beta_{CS}T_i + e_{ij} \qquad (2.1)$$

where

$e_{ij} \sim N(0, \sigma^2)$,

$logit(P_{ijs}) = log(P_{ijs}/(1 - P_{ijs}))$.

The intervention variable T_i is defined as

$$T_i = \begin{cases} 1, & \text{if treatment group (i=1)} \\ 0, & \text{if control group (i=0).} \end{cases}$$

The constant α_1 in model (2.1) denotes the baseline *log* odds of an event. Given e_{ij}, the y_{ijs}'s are assumed to be independent and follow a Bernoulli distribution with parameter P_{ijs}. The variable T_i is modeled as a fixed effect while e_{ij} denotes the between-cluster random source of variation. This is why models including both fixed and random effects may be described as mixed models (Fitzmaurice *et al.*, 2004; Heo and Leon, 2005; Neuhaus *et al.*, 1991).

2.3.3 Population-averaged model

There are several population-averaged extensions of logistic regression (Neuhaus *et al.*, 1991; Liang and Zeger, 1986; Williams, 1975; Prentice, 1986). In this thesis, we consider the generalized estimating equations (GEE) approach proposed by Liang and Zeger (1986). Both individual-level and cluster-level covariates can be incorporated in this approach. A GEE extension of logistic regression can be written as:

$$logit(Pr(y_{ijs} = 1 \mid T_i)) = \alpha_2 + \beta_{PA} T_i \tag{2.2}$$

This model estimates the effect of intervention averaged over all the clusters. The variance of the estimated regression coefficient from a population-averaged model can be obtained using either model-based or robust estimators (Donner and Klar, 2000). The model-based estimator is based on specification of a working correlation matrix. This unknown working correlation matrix describes the correlation between responses within clusters. The 'exchangeable' correlation matrix typically adopted for analysis of data from cluster randomized trials, assumes that responses among cluster members are equally correlated (Donner and Klar, 2000). By way of contrast

the between-cluster information is also used to calculate the robust variance estimator. The robust method has the advantage of providing a consistent estimate of the variance of the estimated regression coefficient even if the working correlation matrix is misspecified (Donner and Klar, 2000). Both model-based and robust variance estimators are asymptotically equivalent when the working correlation matrix is correctly specified (Donner and Klar, 2000).

It is possible to construct Wald-based tests and confidence intervals using the GEE approach. However, the validity of statistical inferences requires a large number of clusters to be included in the study especially for Wald-based tests and confidence intervals constructed using robust variance estimators (Klar and Donner, 2001). In spite of this limitation we limit attention to Wald-based tests following what is likely typical in practice.

2.3.4 Relationship between Cluster-specific and Population-averaged Regression Coefficients

The relationship between regression coefficients from cluster-specific and population-averaged models depends on the study outcome. For binary data, interpretation of parameters from both extensions of logistic regression models is different (Neuhaus and Jewell, 1993). Population-averaged models provide population mean log odds ratios while cluster-specific models provide cluster-specific log odds ratios.

According to Neuhaus *et al.* (1991), the relationship between population-averaged and cluster-specific regression parameters can be expressed as $\beta_{PA} \approx \beta_{CS}(1 - \rho(0))$. The term $\rho(0)$ is the ICC for the cluster-specific model when $\beta_{CS} = 0$. Since $0 < \rho < 1$, it is evident that the population-averaged effect is smaller than the cluster-specific effect. Also in the absence of an intervention effect, when $\beta_{CS} = 0$ then $\beta_{PA} = 0$.

2.4   Effect of intervention

2.4.1   Methods of Interest

In this section we summarize the methods that can be used for statistical inferences on the intervention effect using both pretest and posttest data. Methods include logistic analysis of covariance (ANCOVA), and longitudinal data analysis (LDA) approach. Logistic ANCOVA is discussed in section 2.4.1.1 while the LDA approach is discussed in section 2.4.1.2. Both cluster-specific and population-averaged extensions of these methods are presented as each will be evaluated by simulation.

2.4.1.1   Logistic ANCOVA

In ANCOVA, the posttest measurement is used as the outcome variables while the pretest measurement is used as a covariate. Klar and Darlington (2004) used a mixed effects extension of ANCOVA to investigate the intervention effect using Gaussian data from pretest-posttest cluster randomization trials. Statistical inferences on the intervention effect can be obtained by fitting cluster-specific and population-averaged models such as

Cluster-specific logistic ANCOVA

$$logit(Pr(y_{ijs} = 1 \mid T_i, e_{ij})) = \alpha_3 + \beta_{31}T_i + \beta_{32}x_{ijs} + e_{ij} \tag{2.3}$$

Population-averaged logistic ANCOVA

$$logit(Pr(y_{ijs} = 1 \mid T_i)) = \alpha_4 + \beta_{41}T_i + \beta_{42}x_{ijs} \tag{2.4}$$

2.4.1.2   Longitudinal Data Analysis (LDA) Approach

Liang and Zeger (2000) and Diggle *et al.* (2002) describe analysing pretest and posttest data using a longitudinal approach. They consider a time variable t, $t = 0, 1$ to represent the pretest and posttest data where $t = 0$ is assigned to pretest data while $t = 1$ is assigned to posttest data. They build the model by considering both time and time by treatment interaction terms. Localio *et al.* (2006) discussed several methods for estimating this interaction term. Cluster-specific and population-averaged extensions of this model are considered for testing the intervention effect such that

Cluster-specific LDA approach

$$logit(Pr(z_{ijst} = 1|T_i, t, e_{ij})) = \alpha_5 + \beta_{51}T_i + \beta_{52}t + \beta_{53}T_i \times t + e_{ij} \qquad (2.5)$$

Population-averaged LDA approach

$$logit(Pr(z_{ijst} = 1 \mid T_i, t)) = \alpha_6 + \beta_{61}T_i + \beta_{62}t + \beta_{63}T_i \times t \qquad (2.6)$$

where

$$z_{ijst} = \begin{cases} x_{ijs}, & \text{for (t=0)} \\ y_{ijs}, & \text{for (t=1)}. \end{cases}$$

The regression coefficient of the time by treatment interaction term is the parameter of interest in each of these models.

2.5   Summary

To allow for clustering we considered extensions of logistic regression. These models are further extended for statistical inferences on an intervention effect using both

pretest and posttest binary measurements. These methods are first evaluated using a simulation study and then demonstrated using the data from the TVSFP (Flay *et al.*, 1995).

Chapter 3

SIMULATION STUDY - DESIGN

3.1 Introduction

The design of a simulation study used for evaluating the six methods, described
in Chapter 2, is provided in this chapter. These methods can be distinguished by
whether a cluster-specific or a population-averaged model is used. The methods are
the cluster-specific and population-averaged extensions of logistic regression, logistic
ANCOVA, and longitudinal data analysis (LDA). As previously noted, our focus is on
testing the intervention effect using both posttest and pretest binary outcomes from
completely randomized cluster randomized trials.

This chapter consists of five sections. Values of parameters used for the simu-
lation study are discussed in section 3.2. Methods used to generate the pretest and
posttest data are described in section 3.3. The methods of evaluation are presented
in section 3.4. The chapter concludes with a brief summary in section 3.5. Design of
the simulation study follows recommendations provided by Burton *et al.* (2006).

3.2 Parameters

Several parameters were used for this simulation study. The parameters considered are
the degree of intracluster correlation, the number of clusters, number of subjects per

cluster, baseline *log* odds (α), subject-level association (γ_I), cluster-level association (γ_C), and the intervention effect (β). We limit attention to only one value of α in this study. Parameter values are presented in Table 3.1.

The simulation study used a factorial design for the six parameters that varied. The possible values of the parameters intracluster correlation coefficient, β, number of clusters, cluster size, γ_I , and γ_C lead to 192 parameter combinations. One thousand simulations were conducted for each of these combination. The simulation study was conducted using SAS/STAT software, Version 9.3 of the SAS system for Unix. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA. The reason for choosing this number of iterations was that the approximate 95% confidence interval for a 5% rejection rate is (0.036, 0.064). Thus statistical tests with type I error rates less than 0.036 are deemed overly conservative, and tests with type I error rates greater than 0.064 are overly liberal.

3.2.1   Choice of intracluster correlation coefficient

We consider the measure of ICC used by Rodríguez and Elo (2003), for binary outcomes based on a latent variable formulation of a mixed-effects logistic regression model. Heo and Leon (2005) also considered this measurement of ICC in their study. This ICC is given by

$$\rho_{logit} = \frac{\sigma_u^2}{\sigma_u^2 + \pi^2/3}.$$

where σ_u^2 is the between-cluster variance component, $\pi^2/3$ is the within-cluster variance component, and $\pi = 3.14159$.

Klar and Darlington (2004) selected the value 0.03 as the ICC. This value is approximately representative of school-based trials (Donner and Klar, 2000; Flay

et al., 1995; Hedeker *et al.*, 1994). Hedeker *et al.* (1994) investigated data from the TVSFP study (Flay *et al.*, 1995) and reported that the ICC at the school-level was approximately 0.02. Therefore, for the simulation study, the values of ICC used are $\rho = 0.01$, and 0.02. These ICC's values correspond to between cluster variances $\sigma^2 = 0.033$ and 0.07 respectively.

3.2.2   Number of clusters, cluster size

In this study we focus on a small number of large clusters as is typical of school-based or community randomized trials. The number of clusters and cluster sizes were based on the paper of Klar and Darlington (2004). Therefore, the number of clusters was chosen to vary from 15 to 50, while clusters of sizes 15 to 100 were used in the simulation. Attention was further limited to clusters of fixed size.

3.2.3   Model Fitting and Test statistic

There are several methods available for fitting cluster-specific models (Yasui *et al.*, 2004). In this thesis we concentrate on adaptive Gaussian quadrature (Hedeker and Gibbons, 1994; Pinheiro and Bates, 1995).

For testing the intervention effect we consider a Wald type test statistic (Agresti, 2002) given by

$$\chi^2 = \left\{ \frac{\hat{\beta} - 0}{\widehat{SE}(\hat{\beta})} \right\}^2$$

where $\hat{\beta}$ is the estimated intervention effect from one of the six models described in Chapter 2, and $\widehat{SE}(\hat{\beta})$ is its estimated standard error. This test statistic is, asymptotically, a chi-square random variable with one degree of freedom when the null hypothesis is true (Agresti, 2002). Moreover, we focus on model-based and robust

estimates of $SE(\hat{\beta})$ for cluster-specific and population-averaged models, respectively. Furthermore, we limit attention to an exchangeable working correlation matrix for population-averaged models.

SAS procedures, PROC GLIMMIX (SAS Institute Inc., 2011) and PROC GEN-MOD (SAS Institute Inc., 2011) were used to fit cluster-specific and population-averaged models respectively.

3.2.4 Convergence

We checked convergence for each of these methods considered in this thesis. For any iteration, if a method fails to converge, this iteration was discarded and a new iteration was conducted until there were 1000.

3.2.5 Type I Error, Power, Bias, and Standard Error Comparisons

Firstly, we computed the type I error rates for all the methods. Type I error was estimated based on null hypothesis of no intervention effect i.e. $H_0 : \beta = 0$.

Power comparison was limited to those methods which had valid type I error rates. Moreover, power was estimated for detecting an alternative hypothesis of *log* odds ratio of β at nominal level $\alpha = 0.05$ (two-sided).

Klar and Darlington (2004) selected two values of the intervention effect, 0.3 and 0.35 for power comparison. Heo and Leon (2005) used the values of 0.3 and 0.5 for power comparison in their study on binary outcome data in the context of cluster randomization trials. Therefore, we considered the value of the *log* odds ratio of intervention effect 0.3 for power comparison.

Table 3.1: Values of parameters used for simulation

Parameters	Values
ICC, ρ	0.01, 0.02
Number of clusters, J	15, 30, 50
Cluster size, n	15, 30, 50, 100
Subject-level association, γ_I	0, 0.5
Cluster-level association, γ_C	0, 0.5
Log-odds ratio for intervention effect, β	0, 0.30

3.3 Generation of Data

Klar and Darlington (2004) used mixed-effects linear regression for generating pretest and posttest quantitative observations. For generating binary data, Heo and Leon (2005) used mixed-effects logistic regression. We considered the following mixed-effects logistic regression approaches for generating pretest and posttest dichotomous observations. In these approaches, we considered the same cluster-effect random term as the same individuals from the same clusters are measured over time.

Pretest score

Pretest data (x_{ijs}) were generated using the following random-effects logistic regression model:

$$logit(p_{ijs}) = \alpha + v_{ij} \tag{3.1}$$

where

$$v_{ij} \sim N(0, \sigma^2),$$

$logit(p_{ijs}) = log(p_{ijs}/(1 - p_{ijs}))$, and

$p_{ijs} = Pr(x_{ijs} \mid v_{ij})$.

Posttest score

Posttest data (y_{ijs}) were generated based on the model as follows:

$$logit(p'_{ijs}) = \alpha + \beta T_i + \gamma_I(x_{ijs} - \hat{P}_{ij.}) + \gamma_C(\hat{P}_{ij.} - \hat{P}_{...}) + v_{ij} \qquad (3.2)$$

where $v_{ij} \sim N(0, \sigma^2)$,

$logit(p'_{ijs}) = log(p'_{ijs}/(1 - p'_{ijs}))$, and

$p'_{ijs} = Pr(y_{ijs}|T_i, v_{ij})$.

The regression coefficients γ_I and γ_C, respectively, measure association at the individual-level and cluster-level respectively. The choice of the values of γ_I and γ_C were based on Klar and Darlington (2004). Therefore, data were generated using the values $\gamma_I = 0, 0.5$ and $\gamma_C = 0, 0.5$. Intervention groups are represented by two values, $T_i = 0$ represents the control group while $T_i = 1$ represents the treatment group.

Data were generated using the following steps:

1. Let $\alpha = 0$;

2. Generate between-cluster random variables v_{ij} from $N(0, \sigma^2)$;

3. Calculate $p_{ijs} = Pr(x_{ijs}|v_{ij})$;

4. Generate the pretest score x_{ijs} for each observation randomly from a Bernoulli distribution with success probability p_{ijs};

5. Calculate the proportion of successes in the (i, j)th cluster $(\hat{p}_{ij.})$ and the proportion of successes among the pretest dataset $(\hat{p}_{...})$;

6. Calculate $p'_{ijs} = Pr(y_{ijs}|T_i, v_{ij})$;

7. Generate the posttest score y_{ijs} for each observation randomly from a Bernoulli distribution with success probability p'_{ijs};

8. Apply to each of the six methods with SAS procedure to obtain the parameter estimates and the test statistics.

3.4 Evaluation measures

Comparisons among the statistical methods are based on type I error rate, statistical power, bias, standard error, and standard deviation of estimated regression coefficient. They are computed as follows:

1. For type I error rate we consider the null hypothesis of no intervention effect that is, $H_0 : \beta = 0$. The proportion of p-values less than 0.05 under this null hypothesis is measured to calculate the type I error rate.

2. The statistical power is computed based on the alternative hypothesis $\beta = 0.3$. The proportion of p-values less than 0.05 under this alternative hypothesis is computed to calculate the statistical power.

3. The bias is computed by taking the difference between the average of 1000 estimates of β and the true value. Positive bias represents an overestimation of the intervention effect while negative bias represents an underestimation.

4. The standard error is computed as the average of 1000 standard errors of the estimates of β.

5. The standard deviation is computed using the 1000 estimates of β.

3.5 Summary

A simulation study designed for generating pretest and posttest binary outcomes and evaluating the six methods is presented here. Several parameters and their values are chosen to compare these methods in terms of type I error rates, power, bias, standard error, and standard deviation.

Chapter 4

SIMULATION STUDY - RESULTS

4.1   Introduction

Results obtained from the simulation study described in Chapter 3 are summarized in this chapter. Issues of convergence are summarized in section 4.2. The results related to each of the four study objectives are discussed in sections 4.3 through 4.6, respectively. Validity of the statistical tests are summarized in section 4.3. Power of the methods with valid type I error rates are discussed in section 4.4. The bias of the estimated regression coefficient and its precision are summarized in section 4.5 and 4.6, respectively. Finally section 4.7 contains an overall summary of this chapter.

4.2   Convergence

Iterative methods are required to obtain the maximum likelihood estimates and to fit the models. For any method, convergence occurs when the difference in estimates of parameters from one iteration to the next is less than some maximum value. The rate of convergence was 100% for all the models. The absence of issues related to convergence may have been due to, in part, omitting both very rare or very common events.

4.3 Test validity: type I error rates

The results of type I error rates are presented in tables 4.1 through 4.6. These six tables are distinguished according to the number of clusters ($J = 15, 30, 50$) and the degree of intracluster correlations ($\rho = 0.01, 0.02$). The overly liberal and overly conservative significance levels are highlighted. Type I error rates for both values of $\rho = 0.01, 0.02$ were very similar.

Population-averaged extensions of logistic ANCOVA, the longitudinal data analysis (LDA) approach, and logistic regression with posttest measurements only (PO) methods tended to produce overly liberal type I error rates when there was a small numbers of clusters i.e. $J = 15$. In contrast to these models, the cluster-specific extensions of PO and ANCOVA approaches yielded approximately nominal type I error rates for all parameter combinations investigated. These methods yielded marginally overly liberal and overly conservative type I error rates for a few cases which can be ignored. All the methods yielded type I error rates close to nominal level as the number of clusters and cluster sizes increase.

4.4 Power of valid tests

Empirical power for the methods described in Chapter 2 is discussed in this section. Power results for population-averaged extension of PO, ANCOVA, and LDA are omitted for parameter combinations having overly liberal type I error rates; i.e. when there were few clusters per intervention group ($J = 15$). Moreover, methods which showed overly liberal and overly conservative type I error rates at any parameter combination were also excluded.

Empirical power values for $\beta = 0.30$ are presented in tables 4.7 through 4.12. These six tables are again distinguished by the number of clusters and the degree of

intracluster correlation. The cluster-specific extensions of PO and ANCOVA are valid across all parameter combinations. The empirical power for these two methods was comparable in many scenarios. However, cluster-specific ANCOVA yielded marginally greater power compared to cluster-specific PO especially when $J = 15$ and $n = 30$ and when $J = 30$ and $n = 15$. Cluster-specific LDA and population-averaged extension of PO, ANCOVA and LDA are not valid in all parameter combinations. The population-averaged extension of PO and ANCOVA yielded power close to the cluster-specific extension when it is valid except for a few cases. In general, LDA yielded the lowest power. For different combinations of γ_I and γ_C, cluster-specific ANCOVA yielded power similar in magnitude or marginally lower compared to when we generated the data considering common γ_I and γ_C. On the other hand, power yielded by cluster-specific ANCOVA varied little compared to cluster-specific PO for different combinations of γ_I and γ_C. For example, when $\gamma_I = 0.5$ power values for cluster-specific PO were more variable compared to cluster-specific ANCOVA. In most of the cases, the empirical power values for all the methods remain the same or increase slightly as the value of ICC increased from 0.01 to 0.02.

Empirical power values for all of these methods was close to 100% when there were 50 clusters per intervention group and 100 subjects per cluster. These tables were not included.

4.5 Absolute bias of the estimated regression coefficients

Bias results of the estimated regression coefficient for the intervention effect are presented in tables 4.13 through 4.24. These twelve tables are divided according to number of clusters ($J = 15, 30, 50$), value of the *log* odds ratio of the intervention effect ($\beta = 0, 0.3$), and the degree of ICC ($\rho = 0.01, 0.02$). The bias was calculated as $\bar{\hat{\beta}} - \beta$. When $\beta = 0$, the bias of all the methods was close to 0 except when the cluster

size was small ($n = 15$). The LDA approach yielded more biased results compared to PO and ANCOVA when $J = 30$ and 50 and there were 15 subjects per cluster.

Similar to $\beta = 0$, the LDA approach yielded the most biased results when $\beta = 0.30$. There was little difference in bias between PO and ANCOVA. Moreover, cluster-specific and population-averaged extensions yielded almost the same bias for PO, ANCOVA, and LDA. There was little increase in bias in the case of the PO and LDA approaches when $\beta = 0.3$ is compared to $\beta = 0$.

In brief, the LDA approach yielded greater bias compared to ANCOVA and PO. In general, bias was similar or slightly higher when $\gamma_I \neq \gamma_C$. Again bias remained almost the same for both values of $\rho = 0.01, 0.02$.

4.6   Precision of the estimated regression coefficients

4.6.1   Standard error of the estimated regression coefficient

Precision of the estimated regression coefficient for the intervention effect was evaluated using the standard error (SE). Tables 4.25 through 4.36 contain the results for standard errors. As earlier, these twelve tables are distinguished according to number of clusters ($J = 15, 30, 50$), value of the log odds ratio of the intervention effect ($\beta = 0, 0.3$), and the value of ICC ($\rho = 0.01, 0.02$). Population-averaged extensions of PO, ANCOVA, and LDA approaches yielded smaller SEs compared to the cluster-specific extensions of these approaches. SEs obtained from the LDA approach were the largest. SEs of the PO and ANCOVA methods were very similar. For each method, SEs obtained for different combinations of individual and cluster-level association for all methods did not vary. Also SEs obtained from each method remained almost the same for both values of $\beta = 0, 0.30$.

Overall, the SE of each method was smaller for larger cluster size given the num-

ber of clusters. Furthermore, the SE decreased as the number of clusters increased. In addition, SEs remain almost same for the values of the ICC ($\rho = 0.01, 0.02$).

4.6.2 Standard deviation of the regression coefficient estimates

The observed standard deviations of the regression coefficient estimates are provided in Tables 4.37 through 4.42. These six tables are distinguished according to number of clusters ($J = 15, 30, 50$) and true value of $\beta = 0, 0.3$.

In most of the cases the averages of the standard errors of the regression coefficient estimates and the standard deviations of the regression coefficient estimates were similar in magnitude for all the methods. However, the difference was greater when the number of clusters and cluster sizes were small. The difference between the standard error and the standard deviation tended to decrease as the number of clusters and cluster size increased. The pattern of standard deviations is almost same for two values of ICC ($\rho = 0.01, 0.02$). Results for standard deviations for $\rho = 0.01$ are not showed here.

Table 4.1: Type I error rate (%) for testing the intervention effect; $\rho = \mathbf{0.01}$ and $\mathbf{J = 15}$ clusters per intervention group using extensions of logistic regression (overly liberal and overly conservative type I error rates are in bold font)

| | | | Methods | | | | | |
| | Association Level | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	4.8	**7.4**	4.7	**7.4**	5.4	**8.9**
		0.5	5.6	**8.5**	5.3	**8.9**	4.2	**8.0**
	0.5	0	**6.6**	**8.3**	**6.7**	**8.8**	6.2	**9.3**
		0.5	**6.9**	**8.8**	**6.7**	**8.9**	4.9	**9.7**
30	0	0	4.6	**7.6**	4.7	**7.5**	5.0	**8.2**
		0.5	4.5	**7.5**	4.5	**7.6**	**3.4**	**7.2**
	0.5	0	4.8	**8.9**	5.0	**8.4**	4.5	**7.7**
		0.5	4.9	**8.5**	4.8	**8.1**	**2.8**	**7.9**
50	0	0	4.4	**7.5**	4.3	**7.5**	4.7	**7.5**
		0.5	4.6	**7.1**	4.2	**7.0**	**3.4**	**7.3**
	0.5	0	6.1	**7.7**	5.9	**7.6**	4.1	**9.1**
		0.5	6.1	**8.2**	6.3	**8.3**	3.6	**8.8**
100	0	0	5.5	**8.7**	5.4	**8.9**	5.9	**8.1**
		0.5	5.9	**9.1**	5.8	**9.1**	3.9	**7.6**
	0.5	0	5.2	**7.6**	5.2	**7.9**	3.6	6.0
		0.5	5.4	**8.2**	5.3	**8.4**	**2.8**	**6.5**

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.2: Type I error rate (%) for testing the intervention effect; $\rho = \mathbf{0.01}$ and $\mathbf{J = 30}$ clusters per intervention group using extensions of logistic regression (overly liberal and overly conservative type I error rates are in bold font)

| | | | Methods | | | | | |
| | Association Level | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	5.0	5.6	5.2	5.7	5.6	**7.3**
		0.5	**6.5**	**7.6**	**6.7**	**7.8**	3.7	**6.5**
	0.5	0	5.3	6.3	5.0	6.0	5.0	**6.7**
		0.5	4.6	**6.5**	4.9	**6.5**	4.2	**7.4**
30	0	0	5.9	**7.7**	5.9	**7.9**	4.8	6.0
		0.5	6.2	**7.2**	6.3	**7.3**	**3.3**	5.6
	0.5	0	3.6	5.6	3.5	5.8	**3.0**	4.8
		0.5	4.5	5.7	4.0	5.3	**2.3**	5.4
50	0	0	3.8	5.8	3.8	5.9	4.6	6.3
		0.5	5.4	**7.4**	5.3	**7.4**	4.9	**7.1**
	0.5	0	4.9	**6.9**	5.6	**6.8**	5.5	6.4
		0.5	5.8	6.4	5.9	**6.5**	4.3	**6.8**
100	0	0	5.4	6.2	5.3	6.0	4.8	6.2
		0.5	5.8	6.0	5.8	5.9	3.7	5.7
	0.5	0	4.8	6.1	4.9	6.3	4.6	6.3
		0.5	5.0	6.1	4.7	**6.6**	**3.1**	5.7

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.3: Type I error rate (%) for testing the intervention effect; $\rho = \mathbf{0.01}$ and $\mathbf{J = 50}$ clusters per intervention group using extensions of logistic regression (overly liberal and overly conservative type I error rates are in bold font)

| | | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	3.9	5.0	3.9	5.1	5.1	6.2
		0.5	4.3	5.1	4.4	4.9	4.7	**6.8**
	0.5	0	4.0	4.6	4.1	4.9	4.9	5.7
		0.5	4.4	4.8	4.4	4.8	3.5	5.3
30	0	0	4.7	6.0	4.7	6.1	5.5	5.9
		0.5	5.8	6.0	5.4	6.0	4.7	**6.9**
	0.5	0	3.9	4.8	4.6	4.7	6.1	6.2
		0.5	4.7	5.1	4.4	5.3	4.0	5.6
50	0	0	5.2	5.5	5.3	5.5	5.3	5.4
		0.5	**6.6**	**7.2**	**6.6**	**7.2**	4.2	6.0
	0.5	0	4.0	4.9	4.6	5.0	4.3	5.1
		0.5	4.3	5.5	4.3	5.8	**3.4**	5.2
100	0	0	4.7	6.2	4.9	6.2	4.4	6.0
		0.5	5.5	5.8	5.5	5.8	4.8	**7.1**
	0.5	0	5.2	5.5	4.9	5.5	5.0	5.4
		0.5	5.1	5.5	5.2	5.3	**3.4**	5.4

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.4: Type I error rate (%) for testing the intervention effect; $\rho = \mathbf{0.02}$ and $\mathbf{J = 15}$ clusters per intervention group using extensions of logistic regression (overly liberal and overly conservative type I error rates are in bold font)

			Methods					
	Association Level		PO[a]		ANCOVA		LDA[b]	
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	5.4	**7.2**	5.2	**7.0**	4.7	**8.0**
		0.5	5.6	**8.6**	5.3	**8.0**	4.6	**8.8**
	0.5	0	6.4	**8.9**	**6.5**	**9.0**	6.1	**9.2**
		0.5	**6.6**	**9.2**	5.8	**9.9**	4.4	**9.7**
30	0	0	5.6	**7.9**	5.6	**8.0**	5.2	**7.9**
		0.5	5.8	**7.4**	5.8	**7.5**	4.1	**7.0**
	0.5	0	4.5	**8.4**	4.7	**7.9**	4.3	**7.7**
		0.5	4.9	**8.0**	4.9	**8.2**	**3.2**	**8.0**
50	0	0	4.3	**8.0**	4.1	**8.1**	4.8	**8.0**
		0.5	3.9	**7.3**	3.9	**7.2**	**3.4**	**7.7**
	0.5	0	**6.7**	**9.0**	**6.7**	**9.1**	4.9	**7.8**
		0.5	6.3	**8.7**	5.7	**8.8**	**3.4**	**7.8**
100	0	0	6.0	**8.4**	6.1	**8.8**	6.3	**8.0**
		0.5	5.9	**8.8**	6.1	**8.9**	5.8	**8.9**
	0.5	0	4.9	**7.9**	4.9	**7.6**	4.0	5.9
		0.5	5.4	**8.6**	5.0	**7.6**	**2.2**	**6.6**

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.5: Type I error rate (%) for testing the intervention effect; $\rho = \mathbf{0.02}$ and $\mathbf{J} = \mathbf{30}$ clusters per intervention group using extensions of logistic regression (overly liberal and overly conservative type I error rates are in bold font)

			Methods					
	Association Level		PO[a]		ANCOVA		LDA[b]	
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	5.1	6.2	5.4	6.1	5.4	**6.6**
		0.5	6.2	**7.5**	6.4	**7.3**	4.6	**6.8**
	0.5	0	4.8	5.8	5.2	5.5	5.1	**6.7**
		0.5	4.9	**6.7**	5.4	**6.5**	4.3	**7.3**
30	0	0	6.1	**7.8**	5.9	**7.9**	5.3	6.7
		0.5	6.6	**8.0**	6.5	**7.9**	4.0	6.2
	0.5	0	**3.3**	5.3	3.8	5.2	3.6	4.8
		0.5	3.9	5.6	**3.5**	5.6	**2.5**	5.1
50	0	0	4.3	5.9	4.3	5.9	5.0	**6.6**
		0.5	4.7	**6.8**	4.7	**6.7**	4.8	**7.5**
	0.5	0	4.9	**6.7**	4.8	**6.6**	5.8	**6.7**
		0.5	5.8	**6.7**	4.9	**6.8**	4.1	**7.4**
100	0	0	4.3	5.9	4.3	5.8	5.3	5.7
		0.5	6.4	**6.8**	6.2	**7.0**	4.4	6.4
	0.5	0	4.6	5.9	4.5	6.2	4.8	5.4
		0.5	5.3	6.1	5.0	6.2	3.8	5.4

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.6: Type I error rate (%) for testing the intervention effect; $\rho = \mathbf{0.02}$ and $\mathbf{J = 50}$ clusters per intervention group using extensions of logistic regression (overly liberal and overly conservative type I error rates are in bold font)

| | | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	4.5	5.5	4.5	5.5	4.8	5.6
		0.5	4.9	4.9	4.8	4.9	5.0	**6.6**
	0.5	0	4.6	5.8	4.8	5.8	5.5	6.1
		0.5	3.9	4.9	4.3	4.9	3.7	5.9
30	0	0	4.7	6.4	4.7	6.4	6.1	**6.8**
		0.5	4.6	5.1	4.5	5.3	4.3	**7.0**
	0.5	0	4.2	5.3	4.9	5.2	5.8	6.0
		0.5	4.7	5.5	4.8	5.6	4.7	**6.5**
50	0	0	5.4	6.2	5.6	6.2	5.3	5.2
		0.5	5.5	6.0	5.6	6.0	3.8	4.8
	0.5	0	4.4	5.3	4.4	5.4	4.4	4.6
		0.5	4.2	5.4	4.8	5.6	3.3	5.5
100	0	0	4.8	6.4	4.6	6.3	4.6	5.3
		0.5	5.8	**6.5**	5.8	**6.6**	3.7	5.8
	0.5	0	5.6	6.0	5.5	6.4	5.1	4.9
		0.5	5.2	6.2	5.4	6.0	3.7	5.4

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.7: Empirical power (%) for testing the intervention effect; $\beta = 0.3$, $\rho = \mathbf{0.01}$ and $\mathbf{J = 15}$ clusters per intervention group using extensions of logistic regression (methods with overly liberal and overly conservative type I error rates were omitted)

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	34.5	-	34.7	-	20.8	-
		0.5	36.0	-	35.5	-	18.2	-
	0.5	0	33.3	-	34.2	-	21.4	-
		0.5	35.0	-	35.5	-	19.0	-
30	0	0	58.1	-	58.4	-	35.3	-
		0.5	57.6	-	57.8	-	-	-
	0.5	0	58.0	-	57.8	-	33.7	-
		0.5	58.4	-	59.7	-	-	-
50	0	0	84.5	-	84.1	-	56.3	-
		0.5	84.9	-	84.5	-	-	-
	0.5	0	82.6	-	82.6	-	51.7	-
		0.5	83.4	-	83.5	-	51.2	-
100	0	0	98.0	-	98.0	-	83.8	-
		0.5	97.8	-	97.7	-	84.8	-
	0.5	0	98.4	-	98.6	-	82.7	83.9
		0.5	98.3	-	98.3	-	-	87.9

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.8: Empirical power (%) for testing the intervention effect; $\beta = 0.3$, $\rho = \mathbf{0.01}$ and $\mathbf{J = 30}$ clusters per intervention group using extensions of logistic regression (methods with overly liberal and overly conservative type I error rates were omitted)

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	60.6	62.5	60.9	62.3	32.9	-
		0.5	61.3	-	61.5	-	31.4	37.8
	0.5	0	61.1	63.4	60.7	62.2	31.9	34.5
		0.5	59.3	60.3	60.3	60.9	30.9	-
30	0	0	87.8	-	87.8	-	60.5	61.4
		0.5	87.5	-	87.7	-	-	68.5
	0.5	0	87.8	88.5	88.7	87.9	-	61.5
		0.5	86.1	87.1	87.3	87.6	-	65.6
50	0	0	98.6	98.5	98.6	98.4	83.4	83.0
		0.5	98.5	-	98.5	-	83.4	-
	0.5	0	98.5	98.2	98.5	98.0	81.3	83.3
		0.5	98.7	98.2	98.7	98.5	81.3	85.1
100	0	0	100	100	100	100	98.5	97.9
		0.5	100	100	100	100	99.1	98.8
	0.5	0	100	100	100	100	97.9	97.6
		0.5	100	100	100	100	-	98.8

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.9: Empirical power (%) for testing the intervention effect; $\beta = 0.3$, $\rho = \mathbf{0.01}$ and $\mathbf{J = 50}$ clusters per intervention group using extensions of logistic regression (methods with overly liberal and overly conservative type I error rates were omitted)

			Methods					
	Association Level		PO[a]		ANCOVA		LDA[b]	
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	86.0	85.4	85.2	85.3	54.6	55.0
		0.5	84.1	82.9	83.8	82.8	56.7	61.2
	0.5	0	82.6	83.0	82.4	82.7	53.8	54.5
		0.5	83.0	81.9	83.0	83.1	54.6	59.0
30	0	0	98.1	98.2	98.1	98.1	81.5	82.3
		0.5	98.2	98.1	98.2	98.2	84.2	87.4
	0.5	0	98.3	98.1	98.1	98.3	81.3	82.0
		0.5	98.7	98.7	99.0	98.8	83.2	87.2
50	0	0	99.9	99.8	99.9	99.8	96.1	96.0
		0.5	100	-	100	-	97.4	98.0
	0.5	0	99.9	100	99.9	100	95.3	95.3
		0.5	99.8	100	99.9	100	-	97.5

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.10: Empirical power (%) for testing the intervention effect; $\beta = 0.3$, $\rho = \mathbf{0.02}$ and $\mathbf{J = 15}$ clusters per intervention group using extensions of logistic regression (methods with overly liberal and overly conservative type I error rates were omitted)

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	35.7	-	35.6	-	20.4	-
		0.5	35.7	-	35.0	-	19.0	-
	0.5	0	35.5	-	35.4	-	21.0	-
		0.5	36.1	-	35.7	-	19.6	-
30	0	0	58.4	-	58.6	-	36.4	-
		0.5	57.2	-	57.3	-	32.8	-
	0.5	0	59.6	-	60.3	-	34.7	-
		0.5	60.5	-	61.5	-	-	-
50	0	0	85.5	-	85.0	-	55.6	-
		0.5	85.9	-	85.1	-	-	-
	0.5	0	82.3	-	82.0	-	50.8	-
		0.5	81.8	-	82.1	-	-	-
100	0	0	98.2	-	98.2	-	83.7	-
		0.5	98.0	-	97.9	-	85.1	-
	0.5	0	98.3	-	98.1	-	81.9	82.5
		0.5	98.2	-	98.3	-	-	85.7

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.11: Empirical power (%) for testing the intervention effect; $\beta = 0.3$, $\rho = \mathbf{0.02}$ and $\mathbf{J = 30}$ clusters per intervention group using extensions of logistic regression (methods with overly liberal and overly conservative type I error rates were omitted)

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	60.5	62.9	60.6	63.0	33.9	33.8
		0.5	60.2	-	60.5	-	31.6	37.7
	0.5	0	59.7	61.7	60.3	61.1	31.3	34.0
		0.5	58.4	60.6	60.4	61.6	31.6	-
30	0	0	88.1	-	88.0	-	60.5	61.1
		0.5	88.7	-	88.5	-	60.6	67.3
	0.5	0	89.7	89.7	90.0	89.4	59.6	61.6
		0.5	88.1	88.8	88.7	89.7	-	66.1
50	0	0	98.6	98.4	98.6	98.4	82.6	82.1
		0.5	98.0	98.7	98.0	98.7	83.1	-
	0.5	0	98.6	98.1	98.3	97.7	80.8	79.7
		0.5	98.5	98.2	98.5	98.3	81.7	-
100	0	0	100	100	100	100	98.5	97.6
		0.5	100	100	100	100	99.0	98.8
	0.5	0	100	100	100	100	97.8	97.5
		0.5	100	100	100	100	98.5	98.4

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.12: Empirical power (%) for testing the intervention effect; $\beta = 0.3$, $\rho = \mathbf{0.02}$ and $\mathbf{J = 50}$ clusters per intervention group using extensions of logistic regression (methods with overly liberal and overly conservative type I error rates were omitted)

| | | | Methods | | | | | |
| | Association Level | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	85.5	84.8	85.3	85.0	56.1	55.8
		0.5	84.9	83.8	84.7	83.9	55.4	60.7
	0.5	0	82.5	83.4	81.9	82.8	51.9	54.0
		0.5	81.2	81.6	81.7	82.0	53.2	59.1
30	0	0	98.1	98.1	98.1	98.1	80.9	81.1
		0.5	97.4	97.5	97.5	97.5	82.9	85.2
	0.5	0	98.3	98.3	98.2	98.1	81.5	81.8
		0.5	98.6	98.8	99.0	99.0	81.9	86.1
50	0	0	99.9	99.8	99.9	99.8	95.9	95.7
		0.5	100	100	100	100	97.8	97.9
	0.5	0	100	100	100	100	95.2	95.2
		0.5	99.8	100	99.9	100	-	97.2

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.13: Bias of the estimated regression coefficient; $\beta = \mathbf{0}$, $\rho = \mathbf{0.01}$ and $\mathbf{J} = \mathbf{15}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	-0.007	-0.007	-0.007	-0.007	-0.006	-0.006
		0.5	0.001	0.001	0.001	0.001	0.002	0.002
	0.5	0	0.019	0.018	0.019	0.019	0.019	0.019
		0.5	0.017	0.017	0.018	0.018	0.018	0.018
30	0	0	0.001	0.001	0.001	0.001	0.003	0.003
		0.5	0.001	0.001	0.000	0.000	0.003	0.003
	0.5	0	0.001	0.001	0.002	0.002	0.004	0.004
		0.5	0.001	0.001	0.001	0.001	0.003	0.003
50	0	0	0.003	0.003	0.003	0.003	0.002	0.002
		0.5	-0.003	-0.003	-0.003	-0.003	-0.004	-0.004
	0.5	0	0.002	0.002	0.002	0.002	0.001	0.001
		0.5	0.002	0.002	0.002	0.002	0.000	0.000
100	0	0	-0.005	-0.005	-0.005	-0.005	-0.002	-0.002
		0.5	-0.004	-0.004	-0.004	-0.004	-0.001	-0.001
	0.5	0	0.000	0.000	0.001	0.001	0.003	0.003
		0.5	0.001	0.001	0.001	0.001	0.003	0.003

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.14: Bias of the estimated regression coefficient; $\beta = \mathbf{0}$, $\rho = \mathbf{0.01}$ and $\mathbf{J} = \mathbf{30}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	-0.003	-0.003	-0.003	-0.003	-0.009	-0.009
		0.5	-0.008	-0.008	-0.007	-0.007	-0.014	-0.014
	0.5	0	0.003	0.003	0.002	0.002	-0.003	-0.003
		0.5	0.003	0.003	0.002	0.002	-0.003	-0.003
30	0	0	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
		0.5	0.003	0.003	0.003	0.003	0.003	0.003
	0.5	0	0.004	0.004	0.004	0.004	0.004	0.004
		0.5	0.004	0.004	0.004	0.004	0.004	0.004
50	0	0	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
		0.5	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	0.5	0	-0.001	-0.001	-0.001	-0.001	-0.000	-0.000
		0.5	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
100	0	0	-0.002	-0.002	-0.002	-0.002	-0.001	-0.001
		0.5	-0.000	-0.000	-0.000	-0.000	0.001	0.001
	0.5	0	-0.000	-0.000	-0.000	-0.000	0.001	0.001
		0.5	-0.001	-0.001	-0.001	-0.001	0.001	0.001

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.15: Bias of the estimated regression coefficient; $\beta = 0$, $\rho = 0.01$ and $\mathbf{J} = 50$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.005	0.005	0.005	0.005	0.011	0.011
		0.5	-0.000	-0.000	-0.000	-0.000	0.005	0.005
	0.5	0	0.001	0.001	0.001	0.001	0.006	0.006
		0.5	0.000	0.000	0.001	0.001	0.006	0.006
30	0	0	0.000	0.000	0.000	0.000	0.002	0.002
		0.5	0.000	0.000	0.000	0.000	0.002	0.002
	0.5	0	0.001	0.001	0.001	0.001	0.002	0.002
		0.5	0.000	0.000	0.000	0.000	0.002	0.002
50	0	0	-0.002	-0.002	-0.002	-0.002	-0.000	-0.000
		0.5	-0.002	-0.002	-0.002	-0.002	0.001	0.001
	0.5	0	0.001	0.001	0.002	0.002	0.003	0.003
		0.5	0.001	0.001	0.001	0.001	0.003	0.003
100	0	0	-0.001	-0.001	-0.001	-0.001	-0.000	-0.000
		0.5	0.001	0.001	0.001	0.001	0.001	0.001
	0.5	0	-0.000	-0.000	-0.000	-0.000	0.001	0.001
		0.5	-0.000	-0.000	-0.000	-0.000	0.001	0.001

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.16: Bias of the estimated regression coefficient; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.01}$ and $\mathbf{J = 15}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.002	0.001	0.003	0.002	0.002	0.002
		0.5	-0.007	-0.008	-0.006	-0.007	-0.006	-0.007
	0.5	0	0.001	-0.001	0.007	0.005	0.001	0.000
		0.5	0.001	-0.001	0.007	0.006	0.001	-0.000
30	0	0	-0.004	-0.005	-0.004	-0.005	-0.002	-0.002
		0.5	-0.006	-0.006	-0.005	-0.006	-0.003	-0.003
	0.5	0	-0.009	-0.009	-0.004	-0.004	-0.006	-0.006
		0.5	-0.005	-0.006	-0.000	-0.001	-0.003	-0.003
50	0	0	0.011	0.011	0.012	0.011	0.009	0.009
		0.5	0.007	0.007	0.008	0.007	0.006	0.005
	0.5	0	0.001	0.000	0.005	0.005	-0.001	-0.001
		0.5	0.002	0.002	0.007	0.007	0.001	0.000
100	0	0	-0.002	-0.002	-0.002	-0.002	0.000	0.000
		0.5	-0.003	-0.003	-0.003	-0.003	-0.000	-0.001
	0.5	0	-0.004	-0.004	0.001	0.001	-0.002	-0.002
		0.5	-0.005	-0.005	0.000	0.000	-0.002	-0.002

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.17: Bias of the estimated regression coefficient; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.01}$ and $\mathbf{J} = \mathbf{30}$ clusters per intervention group

| | Association Level | | PO[a] | | ANCOVA | | LDA[b] | |
| | | | | | Methods | | | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	-0.001	-0.002	-0.000	-0.001	-0.008	-0.008
		0.5	-0.002	-0.003	-0.001	-0.002	-0.008	-0.009
	0.5	0	-0.004	-0.005	0.000	-0.001	-0.011	-0.011
		0.5	-0.004	-0.005	0.001	-0.000	-0.010	-0.011
30	0	0	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
		0.5	0.000	-0.000	0.001	0.001	0.001	0.000
	0.5	0	-0.007	-0.007	-0.002	-0.003	-0.007	-0.007
		0.5	-0.007	-0.008	-0.003	-0.003	-0.007	-0.007
50	0	0	0.001	0.001	0.002	0.001	0.001	0.001
		0.5	0.001	0.001	0.001	0.001	0.001	0.001
	0.5	0	-0.004	-0.004	0.000	0.000	-0.004	-0.004
		0.5	-0.005	-0.005	0.000	-0.000	-0.005	-0.005
100	0	0	-0.002	-0.003	-0.002	-0.002	-0.001	-0.001
		0.5	-0.001	-0.001	-0.001	-0.001	0.000	0.000
	0.5	0	-0.006	-0.006	-0.001	-0.001	-0.004	-0.004
		0.5	-0.005	-0.005	-0.001	-0.001	-0.004	-0.004

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.18: Bias of the estimated regression coefficient; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.01}$ and $\mathbf{J} = \mathbf{50}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.005	0.005	0.006	0.005	0.010	0.009
		0.5	0.004	0.004	0.005	0.004	0.009	0.009
	0.5	0	-0.000	-0.001	0.005	0.004	0.005	0.004
		0.5	-0.003	-0.003	0.003	0.002	0.002	0.002
30	0	0	0.001	0.001	0.001	0.001	0.003	0.003
		0.5	0.004	0.004	0.004	0.004	0.006	0.005
	0.5	0	-0.000	-0.001	0.004	0.004	0.001	0.001
		0.5	0.001	0.000	0.006	0.005	0.002	0.002
50	0	0	-0.003	-0.003	-0.003	-0.003	-0.001	-0.001
		0.5	-0.003	-0.003	-0.002	-0.003	-0.000	-0.001
	0.5	0	-0.007	-0.007	-0.002	-0.002	-0.005	-0.005
		0.5	-0.007	-0.007	-0.002	-0.002	-0.004	-0.005
100	0	0	-0.001	-0.001	-0.001	-0.001	-0.000	-0.001
		0.5	-0.001	-0.001	-0.001	-0.001	0.000	0.000
	0.5	0	-0.006	-0.006	-0.001	-0.002	-0.005	-0.006
		0.5	-0.006	-0.006	-0.002	-0.002	-0.006	-0.006

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.19: Bias of the estimated regression coefficient; $\beta = \mathbf{0}$, $\rho = \mathbf{0.02}$ and $\mathbf{J} = \mathbf{15}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	-0.007	-0.007	-0.007	-0.007	-0.005	-0.005
		0.5	-0.003	-0.003	-0.003	-0.003	-0.001	-0.001
	0.5	0	0.016	0.016	0.017	0.017	0.018	0.018
		0.5	0.016	0.016	0.016	0.016	0.017	0.017
30	0	0	0.001	0.001	0.001	0.001	0.004	0.004
		0.5	-0.003	-0.002	-0.003	-0.003	0.000	0.000
	0.5	0	-0.000	-0.000	0.000	0.000	0.002	0.002
		0.5	0.000	0.000	0.001	0.001	0.003	0.003
50	0	0	0.002	0.002	0.002	0.002	-0.000	-0.000
		0.5	-0.000	-0.000	-0.000	-0.000	-0.003	-0.003
	0.5	0	0.002	0.002	0.001	0.001	-0.001	-0.001
		0.5	0.002	0.002	0.001	0.001	-0.001	-0.001
100	0	0	-0.005	-0.005	-0.005	-0.005	-0.003	-0.003
		0.5	-0.005	-0.005	-0.005	-0.005	-0.003	-0.003
	0.5	0	0.001	0.001	0.001	0.001	0.003	0.003
		0.5	0.001	0.001	0.001	0.001	0.003	0.003

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.20: Bias of the estimated regression coefficient; $\beta = \mathbf{0}$, $\rho = \mathbf{0.02}$ and $\mathbf{J = 30}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	-0.003	-0.003	-0.003	-0.003	-0.009	-0.009
		0.5	-0.011	-0.011	-0.011	-0.011	-0.017	-0.017
	0.5	0	0.003	0.003	0.003	0.003	-0.004	-0.004
		0.5	0.003	0.003	0.002	0.002	-0.004	-0.004
30	0	0	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
		0.5	0.004	0.004	0.004	0.004	0.005	0.005
	0.5	0	0.005	0.005	0.005	0.005	0.005	0.005
		0.5	0.004	0.004	0.004	0.004	0.004	0.004
50	0	0	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
		0.5	0.000	0.000	0.000	0.000	-0.000	-0.000
	0.5	0	0.000	0.000	0.000	0.000	-0.000	-0.000
		0.5	0.000	0.000	0.000	0.000	-0.000	-0.000
100	0	0	-0.002	-0.002	-0.002	-0.002	0.000	0.000
		0.5	-0.001	-0.001	-0.001	-0.001	0.001	0.001
	0.5	0	-0.001	-0.001	-0.001	-0.001	0.001	0.001
		0.5	-0.001	-0.001	-0.001	-0.001	0.001	0.001

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.21: Bias of the estimated regression coefficient; $\beta = 0$, $\rho = 0.02$ and $\mathbf{J} = 50$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.004	0.004	0.004	0.004	0.009	0.009
		0.5	0.005	0.005	0.005	0.005	0.010	0.010
	0.5	0	0.001	0.001	0.002	0.002	0.006	0.006
		0.5	-0.001	-0.001	0.000	0.000	0.005	0.005
30	0	0	0.000	0.000	0.000	0.000	0.002	0.002
		0.5	-0.001	-0.001	-0.001	-0.001	0.001	0.001
	0.5	0	0.000	0.000	0.000	0.000	0.001	0.001
		0.5	-0.000	-0.000	-0.000	-0.000	0.001	0.001
50	0	0	-0.001	-0.001	-0.001	-0.001	0.000	0.000
		0.5	-0.002	-0.002	-0.002	-0.002	-0.001	-0.001
	0.5	0	0.001	0.001	0.002	0.002	0.003	0.003
		0.5	0.001	0.001	0.001	0.001	0.003	0.003
100	0	0	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
		0.5	-0.000	-0.000	-0.000	-0.000	0.000	0.000
	0.5	0	0.000	0.000	0.000	0.000	0.001	0.001
		0.5	0.000	0.000	0.000	0.000	0.001	0.001

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.22: Bias of the estimated regression coefficient; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.02}$ and $\mathbf{J = 15}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.001	0.000	0.002	0.001	0.002	0.002
		0.5	-0.003	-0.004	0.001	-0.003	-0.002	-0.003
	0.5	0	0.001	-0.001	0.007	0.006	0.002	0.001
		0.5	0.000	-0.001	0.007	0.005	0.002	0.001
30	0	0	-0.004	-0.004	-0.004	-0.004	-0.001	-0.002
		0.5	-0.008	-0.008	-0.008	-0.009	-0.005	-0.006
	0.5	0	-0.005	-0.005	0.000	-0.001	-0.003	-0.003
		0.5	-0.004	-0.005	0.001	0.001	-0.002	-0.002
50	0	0	0.012	0.011	0.012	0.012	0.009	0.009
		0.5	0.009	0.009	0.009	0.009	0.007	0.006
	0.5	0	0.001	0.000	0.005	0.005	-0.002	-0.002
		0.5	0.000	-0.000	0.005	0.004	-0.003	-0.003
100	0	0	-0.002	-0.002	-0.002	-0.002	0.000	-0.000
		0.5	-0.001	-0.001	-0.001	-0.001	0.001	0.000
	0.5	0	-0.004	-0.004	0.001	0.001	-0.002	-0.002
		0.5	-0.005	-0.005	-0.000	-0.001	-0.003	-0.004

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.23: Bias of the estimated regression coefficient; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.02}$ and $\mathbf{J} = \mathbf{30}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	-0.000	-0.001	0.000	-0.001	-0.008	-0.008
		0.5	-0.004	-0.005	-0.003	-0.004	-0.011	-0.011
	0.5	0	-0.004	-0.005	0.000	-0.001	-0.011	-0.011
		0.5	-0.003	-0.004	0.002	0.001	-0.009	-0.010
30	0	0	0.000	-0.000	0.000	-0.000	0.000	0.000
		0.5	0.003	0.003	0.003	0.003	0.003	0.003
	0.5	0	-0.005	-0.005	-0.000	-0.001	-0.005	-0.005
		0.5	-0.005	-0.005	0.000	-0.000	-0.004	-0.005
50	0	0	0.001	0.001	0.002	0.001	0.001	0.000
		0.5	0.002	0.002	0.002	0.002	0.001	0.001
	0.5	0	-0.004	-0.004	0.001	0.001	-0.004	-0.004
		0.5	-0.004	-0.005	0.000	0.000	-0.005	-0.005
100	0	0	-0.002	-0.003	-0.002	-0.002	-0.001	-0.001
		0.5	-0.002	-0.002	-0.002	-0.002	0.000	-0.000
	0.5	0	-0.006	-0.006	-0.001	-0.001	-0.004	-0.004
		0.5	-0.006	-0.006	-0.001	-0.001	-0.004	-0.004

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.24: Bias of the estimated regression coefficient; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.02}$ and $\mathbf{J} = \mathbf{50}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.004	0.004	0.005	0.004	0.009	0.009
		0.5	0.007	0.006	0.007	0.006	0.012	0.011
	0.5	0	-0.002	-0.003	0.003	0.002	0.003	0.002
		0.5	-0.004	-0.005	0.001	0.001	0.001	0.000
30	0	0	0.001	0.000	0.001	0.001	0.002	0.001
		0.5	0.001	0.001	0.001	0.001	0.002	0.002
	0.5	0	-0.000	-0.001	0.004	0.004	0.001	0.001
		0.5	-0.000	-0.001	0.005	0.004	0.001	0.001
50	0	0	-0.003	-0.003	-0.003	-0.003	-0.001	-0.001
		0.5	-0.003	-0.003	-0.003	-0.003	-0.001	-0.001
	0.5	0	-0.007	-0.007	-0.002	-0.002	-0.005	-0.005
		0.5	-0.007	-0.007	-0.002	-0.002	-0.005	-0.005
100	0	0	-0.001	-0.001	-0.001	-0.001	-0.000	-0.001
		0.5	-0.001	-0.001	-0.001	-0.001	-0.000	-0.001
	0.5	0	-0.006	-0.006	-0.001	-0.002	-0.006	-0.006
		0.5	-0.006	-0.006	-0.001	-0.001	-0.005	-0.006

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.25: Standard error of the estimated regression coefficient; $\beta = \mathbf{0}$, $\rho = \mathbf{0.01}$ and $\mathbf{J} = \mathbf{15}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.189	0.179	0.189	0.180	0.268	0.255
		0.5	0.189	0.181	0.189	0.181	0.268	0.238
	0.5	0	0.189	0.180	0.191	0.184	0.268	0.254
		0.5	0.189	0.181	0.191	0.183	0.268	0.239
30	0	0	0.134	0.127	0.134	0.127	0.189	0.179
		0.5	0.134	0.127	0.134	0.127	0.189	0.169
	0.5	0	0.134	0.126	0.135	0.129	0.189	0.178
		0.5	0.134	0.127	0.135	0.128	0.189	0.167
50	0	0	0.103	0.098	0.103	0.098	0.146	0.138
		0.5	0.103	0.099	0.103	0.099	0.146	0.129
	0.5	0	0.103	0.097	0.104	0.099	0.146	0.138
		0.5	0.103	0.098	0.104	0.099	0.146	0.129
100	0	0	0.073	0.069	0.073	0.069	0.103	0.098
		0.5	0.073	0.069	0.073	0.069	0.103	0.092
	0.5	0	0.073	0.069	0.074	0.071	0.103	0.099
		0.5	0.073	0.069	0.074	0.070	0.103	0.093

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.26: Standard error of the estimated regression coefficient; $\beta = \mathbf{0}$, $\rho = \mathbf{0.01}$ and $\mathbf{J = 30}$ clusters per intervention group

	Association Level		PO[a]		ANCOVA		LDA[b]	
					Methods			
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.134	0.131	0.134	0.131	0.189	0.186
		0.5	0.134	0.131	0.134	0.131	0.189	0.173
	0.5	0	0.134	0.129	0.135	0.133	0.189	0.184
		0.5	0.134	0.131	0.135	0.132	0.189	0.173
30	0	0	0.094	0.092	0.094	0.092	0.133	0.131
		0.5	0.094	0.093	0.094	0.092	0.134	0.122
	0.5	0	0.094	0.091	0.095	0.093	0.133	0.129
		0.5	0.094	0.091	0.095	0.092	0.133	0.122
50	0	0	0.073	0.071	0.073	0.071	0.103	0.101
		0.5	0.073	0.072	0.073	0.072	0.103	0.095
	0.5	0	0.073	0.071	0.074	0.073	0.103	0.101
		0.5	0.073	0.072	0.074	0.072	0.103	0.095
100	0	0	0.052	0.050	0.052	0.050	0.073	0.071
		0.5	0.052	0.051	0.052	0.051	0.073	0.067
	0.5	0	0.052	0.049	0.052	0.051	0.073	0.071
		0.5	0.052	0.050	0.052	0.051	0.073	0.067

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.27: Standard error of the estimated regression coefficient; $\beta = \mathbf{0}$, $\rho = \mathbf{0.01}$ and $\mathbf{J} = \mathbf{50}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.103	0.101	0.103	0.102	0.146	0.143
		0.5	0.103	0.103	0.103	0.103	0.146	0.135
	0.5	0	0.103	0.101	0.104	0.103	0.146	0.143
		0.5	0.103	0.102	0.104	0.103	0.146	0.135
30	0	0	0.073	0.072	0.073	0.072	0.103	0.102
		0.5	0.073	0.073	0.073	0.073	0.103	0.096
	0.5	0	0.073	0.072	0.074	0.073	0.103	0.102
		0.5	0.073	0.072	0.074	0.073	0.103	0.095
50	0	0	0.057	0.056	0.057	0.056	0.080	0.079
		0.5	0.057	0.056	0.057	0.056	0.080	0.074
	0.5	0	0.057	0.055	0.057	0.056	0.080	0.079
		0.5	0.057	0.056	0.057	0.056	0.080	0.074
100	0	0	0.040	0.039	0.040	0.039	0.057	0.056
		0.5	0.040	0.039	0.040	0.039	0.057	0.053
	0.5	0	0.040	0.039	0.040	0.040	0.057	0.056
		0.5	0.040	0.039	0.040	0.039	0.057	0.052

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.28: Standard error of the estimated regression coefficient; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.01}$ and $\mathbf{J = 15}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.190	0.180	0.191	0.181	0.268	0.256
		0.5	0.190	0.183	0.191	0.183	0.268	0.239
	0.5	0	0.190	0.180	0.192	0.185	0.268	0.253
		0.5	0.190	0.181	0.193	0.184	0.268	0.239
30	0	0	0.134	0.127	0.134	0.127	0.189	0.179
		0.5	0.134	0.127	0.135	0.127	0.189	0.168
	0.5	0	0.134	0.126	0.136	0.129	0.189	0.179
		0.5	0.134	0.127	0.136	0.128	0.189	0.168
50	0	0	0.104	0.098	0.104	0.098	0.146	0.138
		0.5	0.104	0.099	0.104	0.099	0.147	0.130
	0.5	0	0.104	0.098	0.105	0.100	0.147	0.139
		0.5	0.104	0.098	0.105	0.099	0.147	0.130
100	0	0	0.073	0.069	0.074	0.069	0.104	0.099
		0.5	0.073	0.074	0.074	0.070	0.104	0.093
	0.5	0	0.073	0.069	0.074	0.072	0.104	0.099
		0.5	0.073	0.070	0.074	0.071	0.104	0.094

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.29: Standard error of the estimated regression coefficient; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.01}$ and $\mathbf{J} = \mathbf{30}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.134	0.131	0.135	0.131	0.189	0.186
		0.5	0.134	0.132	0.135	0.132	0.189	0.175
	0.5	0	0.134	0.130	0.136	0.133	0.189	0.185
		0.5	0.134	0.131	0.136	0.132	0.189	0.174
30	0	0	0.095	0.092	0.095	0.092	0.134	0.130
		0.5	0.095	0.093	0.095	0.093	0.134	0.122
	0.5	0	0.095	0.092	0.096	0.094	0.134	0.130
		0.5	0.095	0.092	0.096	0.093	0.134	0.122
50	0	0	0.074	0.072	0.074	0.072	0.104	0.102
		0.5	0.074	0.073	0.073	0.073	0.104	0.096
	0.5	0	0.074	0.072	0.074	0.073	0.104	0.101
		0.5	0.074	0.072	0.074	0.073	0.104	0.095
100	0	0	0.052	0.051	0.052	0.051	0.073	0.072
		0.5	0.052	0.051	0.052	0.051	0.073	0.068
	0.5	0	0.052	0.051	0.052	0.052	0.073	0.072
		0.5	0.052	0.051	0.052	0.052	0.073	0.067

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.30: Standard error of the estimated regression coefficient; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.01}$ and $\mathbf{J} = \mathbf{50}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.104	0.102	0.104	0.103	0.147	0.144
		0.5	0.104	0.103	0.104	0.103	0.147	0.136
	0.5	0	0.104	0.102	0.105	0.104	0.147	0.145
		0.5	0.104	0.103	0.105	0.104	0.147	0.136
30	0	0	0.074	0.072	0.074	0.072	0.103	0.103
		0.5	0.074	0.073	0.074	0.073	0.104	0.096
	0.5	0	0.074	0.072	0.074	0.073	0.104	0.102
		0.5	0.073	0.072	0.074	0.073	0.104	0.096
50	0	0	0.057	0.056	0.057	0.056	0.080	0.079
		0.5	0.057	0.056	0.057	0.056	0.080	0.075
	0.5	0	0.057	0.055	0.057	0.057	0.080	0.079
		0.5	0.057	0.056	0.057	0.056	0.080	0.074
100	0	0	0.040	0.039	0.040	0.039	0.057	0.056
		0.5	0.040	0.039	0.040	0.039	0.057	0.053
	0.5	0	0.040	0.039	0.041	0.040	0.057	0.056
		0.5	0.040	0.039	0.041	0.039	0.057	0.052

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.31: Standard error of the estimated regression coefficient; $\beta = \mathbf{0}$, $\rho = \mathbf{0.02}$ and $\mathbf{J} = \mathbf{15}$ clusters per intervention group

| | Association Level | | PO[a] | | ANCOVA | | LDA[b] | |
| | | | | | Methods | | | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.189	0.179	0.189	0.180	0.268	0.256
		0.5	0.189	0.180	0.189	0.180	0.268	0.238
	0.5	0	0.189	0.179	0.191	0.183	0.268	0.255
		0.5	0.189	0.181	0.191	0.183	0.268	0.239
30	0	0	0.134	0.127	0.134	0.128	0.189	0.181
		0.5	0.134	0.127	0.134	0.127	0.189	0.169
	0.5	0	0.134	0.126	0.135	0.129	0.189	0.179
		0.5	0.134	0.127	0.135	0.128	0.189	0.168
50	0	0	0.104	0.098	0.104	0.098	0.146	0.139
		0.5	0.103	0.099	0.104	0.099	0.146	0.131
	0.5	0	0.103	0.098	0.104	0.099	0.146	0.139
		0.5	0.103	0.098	0.104	0.099	0.146	0.131
100	0	0	0.073	0.069	0.073	0.069	0.103	0.101
		0.5	0.073	0.069	0.073	0.069	0.103	0.094
	0.5	0	0.073	0.069	0.074	0.071	0.103	0.101
		0.5	0.073	0.069	0.074	0.070	0.103	0.095

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.32: Standard error of the estimated regression coefficient; $\beta = \mathbf{0}$, $\rho = \mathbf{0.02}$ and $\mathbf{J = 30}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.134	0.131	0.134	0.131	0.189	0.187
		0.5	0.134	0.131	0.134	0.131	0.189	0.174
	0.5	0	0.134	0.129	0.135	0.133	0.189	0.185
		0.5	0.134	0.131	0.135	0.132	0.189	0.173
30	0	0	0.094	0.092	0.095	0.092	0.134	0.131
		0.5	0.094	0.093	0.095	0.093	0.134	0.123
	0.5	0	0.094	0.091	0.095	0.093	0.133	0.131
		0.5	0.094	0.092	0.095	0.092	0.134	0.123
50	0	0	0.073	0.072	0.073	0.072	0.103	0.102
		0.5	0.073	0.072	0.073	0.072	0.103	0.096
	0.5	0	0.073	0.071	0.074	0.073	0.103	0.102
		0.5	0.073	0.071	0.074	0.072	0.103	0.096
100	0	0	0.052	0.050	0.052	0.050	0.073	0.073
		0.5	0.052	0.051	0.052	0.051	0.073	0.068
	0.5	0	0.052	0.050	0.052	0.051	0.073	0.073
		0.5	0.052	0.050	0.052	0.051	0.073	0.069

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.33: Standard error of the estimated regression coefficient; $\beta = \mathbf{0}$, $\rho = \mathbf{0.02}$ and $\mathbf{J = 50}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.103	0.102	0.104	0.102	0.146	0.144
		0.5	0.103	0.102	0.104	0.102	0.146	0.135
	0.5	0	0.103	0.101	0.104	0.103	0.146	0.144
		0.5	0.104	0.102	0.104	0.103	0.146	0.136
30	0	0	0.073	0.072	0.073	0.072	0.103	0.103
		0.5	0.073	0.073	0.073	0.073	0.103	0.096
	0.5	0	0.073	0.071	0.074	0.073	0.103	0.102
		0.5	0.073	0.072	0.074	0.073	0.103	0.096
50	0	0	0.057	0.056	0.057	0.056	0.081	0.080
		0.5	0.056	0.056	0.056	0.056	0.080	0.075
	0.5	0	0.056	0.055	0.057	0.056	0.080	0.079
		0.5	0.057	0.056	0.057	0.056	0.080	0.075
100	0	0	0.040	0.039	0.040	0.039	0.056	0.057
		0.5	0.040	0.039	0.040	0.039	0.057	0.054
	0.5	0	0.040	0.039	0.040	0.040	0.057	0.057
		0.5	0.040	0.039	0.040	0.040	0.057	0.054

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.34: Standard error of the estimated regression coefficient; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.02}$ and $\mathbf{J} = \mathbf{15}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.191	0.181	0.191	0.181	0.268	0.256
		0.5	0.191	0.183	0.191	0.183	0.268	0.240
	0.5	0	0.191	0.180	0.192	0.184	0.268	0.254
		0.5	0.191	0.181	0.193	0.183	0.268	0.238
30	0	0	0.134	0.127	0.135	0.127	0.189	0.179
		0.5	0.134	0.127	0.135	0.127	0.189	0.169
	0.5	0	0.134	0.127	0.135	0.130	0.189	0.181
		0.5	0.134	0.128	0.136	0.129	0.189	0.170
50	0	0	0.104	0.098	0.104	0.098	0.147	0.140
		0.5	0.104	0.099	0.104	0.099	0.147	0.132
	0.5	0	0.104	0.098	0.105	0.100	0.147	0.141
		0.5	0.104	0.098	0.105	0.099	0.147	0.132
100	0	0	0.073	0.069	0.074	0.069	0.104	0.102
		0.5	0.073	0.070	0.074	0.070	0.104	0.095
	0.5	0	0.073	0.069	0.074	0.072	0.104	0.102
		0.5	0.074	0.071	0.074	0.071	0.104	0.095

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.35: Standard error of the estimated regression coefficient; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.02}$ and $\mathbf{J = 30}$ clusters per intervention group

| | Association Level | | PO[a] | | ANCOVA | | LDA[b] | |
| | | | Methods | | | | | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.135	0.131	0.135	0.131	0.189	0.186
		0.5	0.135	0.132	0.135	0.132	0.189	0.175
	0.5	0	0.135	0.130	0.136	0.133	0.189	0.185
		0.5	0.135	0.131	0.136	0.133	0.189	0.174
30	0	0	0.095	0.092	0.095	0.092	0.134	0.131
		0.5	0.095	0.093	0.095	0.092	0.134	0.123
	0.5	0	0.095	0.092	0.096	0.094	0.134	0.131
		0.5	0.095	0.092	0.096	0.093	0.134	0.123
50	0	0	0.074	0.072	0.074	0.072	0.104	0.103
		0.5	0.074	0.072	0.074	0.072	0.104	0.097
	0.5	0	0.074	0.072	0.074	0.073	0.104	0.103
		0.5	0.074	0.072	0.074	0.073	0.104	0.097
100	0	0	0.052	0.051	0.052	0.051	0.073	0.073
		0.5	0.052	0.051	0.052	0.051	0.073	0.069
	0.5	0	0.052	0.051	0.052	0.052	0.073	0.073
		0.5	0.052	0.051	0.052	0.052	0.073	0.069

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.36: Standard error of the estimated regression coefficient; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.02}$ and $\mathbf{J} = \mathbf{50}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.104	0.102	0.104	0.103	0.147	0.145
		0.5	0.104	0.103	0.104	0.103	0.147	0.136
	0.5	0	0.104	0.102	0.105	0.104	0.147	0.145
		0.5	0.104	0.103	0.105	0.104	0.147	0.136
30	0	0	0.074	0.072	0.074	0.072	0.104	0.103
		0.5	0.074	0.073	0.074	0.073	0.104	0.097
	0.5	0	0.074	0.072	0.074	0.073	0.104	0.102
		0.5	0.074	0.072	0.074	0.073	0.104	0.096
50	0	0	0.057	0.056	0.057	0.056	0.080	0.080
		0.5	0.057	0.056	0.057	0.056	0.080	0.075
	0.5	0	0.057	0.056	0.057	0.057	0.080	0.080
		0.5	0.057	0.056	0.057	0.057	0.080	0.075
100	0	0	0.040	0.040	0.040	0.040	0.057	0.057
		0.5	0.040	0.040	0.040	0.040	0.057	0.054
	0.5	0	0.040	0.039	0.040	0.040	0.057	0.057
		0.5	0.040	0.039	0.040	0.040	0.057	0.054

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.37: Observed standard deviation of the regression coefficient estimates; $\beta = \mathbf{0}$, $\rho = \mathbf{0.02}$ and $\mathbf{J} = \mathbf{15}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.193	0.192	0.193	0.192	0.276	0.276
		0.5	0.194	0.193	0.194	0.194	0.261	0.261
	0.5	0	0.195	0.194	0.200	0.199	0.279	0.278
		0.5	0.197	0.196	0.199	0.199	0.262	0.261
30	0	0	0.136	0.136	0.137	0.137	0.193	0.193
		0.5	0.136	0.135	0.136	0.136	0.177	0.177
	0.5	0	0.135	0.134	0.137	0.137	0.186	0.186
		0.5	0.136	0.136	0.137	0.136	0.175	0.175
50	0	0	0.102	0.101	0.102	0.102	0.145	0.145
		0.5	0.101	0.101	0.101	0.101	0.135	0.135
	0.5	0	0.108	0.108	0.111	0.110	0.148	0.147
		0.5	0.108	0.108	0.109	0.109	0.138	0.138
100	0	0	0.075	0.075	0.075	0.075	0.106	0.106
		0.5	0.076	0.076	0.077	0.077	0.101	0.101
	0.5	0	0.073	0.073	0.074	0.074	0.103	0.103
		0.5	0.074	0.074	0.074	0.074	0.097	0.097

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.38: Observed standard deviation of the regression coefficient estimates; $\beta = \mathbf{0}$, $\rho = \mathbf{0.02}$ and $\mathbf{J = 30}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.135	0.134	0.135	0.135	0.192	0.192
		0.5	0.137	0.137	0.138	0.137	0.180	0.179
	0.5	0	0.131	0.131	0.135	0.135	0.191	0.191
		0.5	0.131	0.131	0.133	0.133	0.179	0.179
30	0	0	0.098	0.098	0.098	0.098	0.134	0.133
		0.5	0.097	0.098	0.098	0.098	0.125	0.125
	0.5	0	0.089	0.089	0.091	0.091	0.126	0.126
		0.5	0.091	0.090	0.091	0.091	0.118	0.118
50	0	0	0.072	0.072	0.072	0.072	0.104	0.104
		0.5	0.074	0.074	0.074	0.074	0.100	0.100
	0.5	0	0.074	0.074	0.075	0.075	0.107	0.107
		0.5	0.074	0.074	0.075	0.075	0.100	0.100
100	0	0	0.052	0.052	0.052	0.052	0.074	0.074
		0.5	0.053	0.053	0.053	0.053	0.070	0.070
	0.5	0	0.051	0.051	0.052	0.052	0.074	0.074
		0.5	0.052	0.052	0.052	0.052	0.069	0.069

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.39: Observed standard deviation of the regression coefficient estimates; $\beta = \mathbf{0}$, $\rho = \mathbf{0.02}$ and $\mathbf{J} = \mathbf{50}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.099	0.099	0.099	0.099	0.145	0.145
		0.5	0.103	0.103	0.103	0.103	0.139	0.139
	0.5	0	0.101	0.101	0.104	0.104	0.151	0.151
		0.5	0.101	0.101	0.102	0.102	0.141	0.141
30	0	0	0.074	0.074	0.074	0.074	0.108	0.107
		0.5	0.072	0.072	0.072	0.072	0.100	0.100
	0.5	0	0.070	0.070	0.073	0.073	0.108	0.108
		0.5	0.071	0.071	0.072	0.072	0.101	0.101
50	0	0	0.058	0.058	0.058	0.058	0.082	0.082
		0.5	0.059	0.059	0.059	0.059	0.077	0.077
	0.5	0	0.055	0.055	0.056	0.056	0.078	0.078
		0.5	0.056	0.056	0.056	0.056	0.074	0.074
100	0	0	0.041	0.041	0.041	0.041	0.058	0.058
		0.5	0.041	0.041	0.041	0.041	0.054	0.054
	0.5	0	0.040	0.040	0.041	0.041	0.057	0.057
		0.5	0.040	0.040	0.041	0.040	0.053	0.053

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.40: Observed standard deviation of the regression coefficient estimates; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.02}$ and $\mathbf{J} = \mathbf{15}$ clusters per intervention group

| | Association Level | | PO[a] | | ANCOVA | | LDA[b] | |
| | | | Methods | | | | | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.199	0.198	0.199	0.199	0.281	0.280
		0.5	0.193	0.192	0.193	0.192	0.257	0.256
	0.5	0	0.189	0.189	0.195	0.194	0.269	0.269
		0.5	0.190	0.189	0.193	0.192	0.254	0.253
30	0	0	0.136	0.136	0.137	0.136	0.202	0.201
		0.5	0.135	0.135	0.135	0.135	0.186	0.186
	0.5	0	0.130	0.130	0.134	0.133	0.189	0.189
		0.5	0.134	0.134	0.135	0.135	0.179	0.179
50	0	0	0.103	0.103	0.103	0.103	0.143	0.143
		0.5	0.104	0.104	0.104	0.104	0.135	0.135
	0.5	0	0.104	0.104	0.106	0.106	0.144	0.144
		0.5	0.106	0.106	0.106	0.106	0.133	0.133
100	0	0	0.075	0.075	0.075	0.075	0.107	0.107
		0.5	0.076	0.076	0.076	0.076	0.101	0.101
	0.5	0	0.074	0.074	0.075	0.075	0.105	0.105
		0.5	0.073	0.073	0.074	0.074	0.098	0.098

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.41: Observed standard deviation of the regression coefficient estimates; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.02}$ and $\mathbf{J} = \mathbf{30}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.133	0.133	0.134	0.133	0.191	0.191
		0.5	0.137	0.137	0.137	0.137	0.179	0.179
	0.5	0	0.133	0.133	0.137	0.136	0.192	0.191
		0.5	0.132	0.132	0.134	0.133	0.178	0.178
30	0	0	0.098	0.097	0.098	0.098	0.135	0.135
		0.5	0.098	0.096	0.098	0.098	0.127	0.127
	0.5	0	0.093	0.092	0.095	0.095	0.131	0.131
		0.5	0.093	0.093	0.094	0.094	0.123	0.122
50	0	0	0.074	0.074	0.074	0.074	0.105	0.105
		0.5	0.075	0.075	0.075	0.075	0.102	0.102
	0.5	0	0.073	0.073	0.074	0.074	0.105	0.105
		0.5	0.074	0.074	0.075	0.075	0.099	0.099
100	0	0	0.052	0.052	0.052	0.052	0.075	0.075
		0.5	0.053	0.053	0.053	0.053	0.070	0.070
	0.5	0	0.052	0.052	0.053	0.053	0.074	0.074
		0.5	0.053	0.053	0.053	0.053	0.069	0.069

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

Table 4.42: Observed standard deviation of the regression coefficient estimates; $\beta = \mathbf{0.3}$, $\rho = \mathbf{0.02}$ and $\mathbf{J = 50}$ clusters per intervention group

| | Association Level | | Methods | | | | | |
| | | | PO[a] | | ANCOVA | | LDA[b] | |
Cluster Size	Individual	Cluster	CS[c]	PA[d]	CS	PA	CS	PA
15	0	0	0.098	0.098	0.098	0.098	0.145	0.145
		0.5	0.102	0.102	0.102	0.102	0.137	0.136
	0.5	0	0.101	0.101	0.103	0.103	0.148	0.147
		0.5	0.102	0.102	0.103	0.103	0.141	0.141
30	0	0	0.075	0.075	0.075	0.075	0.109	0.109
		0.5	0.074	0.074	0.074	0.074	0.102	0.102
	0.5	0	0.073	0.073	0.075	0.075	0.110	0.110
		0.5	0.072	0.072	0.073	0.073	0.102	0.101
50	0	0	0.057	0.057	0.057	0.057	0.081	0.081
		0.5	0.057	0.057	0.057	0.057	0.075	0.075
	0.5	0	0.056	0.056	0.057	0.057	0.079	0.079
		0.5	0.057	0.057	0.057	0.057	0.075	0.075
100	0	0	0.041	0.041	0.041	0.041	0.059	0.059
		0.5	0.040	0.040	0.040	0.040	0.054	0.054
	0.5	0	0.040	0.040	0.041	0.041	0.056	0.056
		0.5	0.041	0.041	0.041	0.041	0.053	0.053

[a]PO: Logistic regression with posttest measurements only

[b]LDA: Both pretest and posttest measurements used as the outcome. Regression coefficient corresponding to $group \times time$ is the parameter of interest

[c]CS: Cluster-Specific

[d]PA: Population-Averaged

4.7 Summary

Results of different evaluation measures such as, type I error rates, empirical power, absolute bias, standard errors obtained using the methods described in Chapter 2 are discussed in this chapter. These results were obtained according to the simulation design described in Chapter 3.

Chapter 5

EXAMPLE: THE TELEVISION, SCHOOL, AND FAMILY SMOKING PREVENTION AND CESSATION PROJECT

5.1 Introduction

In this chapter, methods described in Chapter 2 are applied to assess the effect of intervention using data from the TVSFP i.e. Television, School, and Family Smoking Prevention and Cessation Project (Flay *et al.*, 1995) which was previously described in Chapter 1. The TVSFP is a cluster randomization trial which enrolled 47 schools and 340 classrooms from Los Angeles and San Diego and took place between 1986 and 1988.

A social influences program may help delay the onset of smoking among adolescents by helping students build skills to resist social influence to smoke (Flay, 1985; Glynn, 1990). Mass media anti-smoking campaigns may also have an effect on preventing smoking (Flay, 1987). There is growing interest in school-based trials to prevent smoking (Flay *et al.*, 1995; Cameron *et al.*, 1999; Hollingworth *et al.*, 2012). Many such smoking prevention trials were based on cluster randomization. For example, as part of the TVSFP a school-based social resistance curriculum was investigated to prevent smoking by increasing tobacco and health knowledge of students.

This chapter is divided into five sections. Design of the TVSFP is discussed

in section 5.2. Results obtained from this study are presented in section 5.3. A discussion of these results is provided in section 5.4. Finally this chapter ends with a summary in section 5.5.

5.2 Design of the TVSFP Study

Data for this trial were obtained from 7351 seventh grade students who participated in the pretest assessment. Students were pretested in January 1986 and completed a post intervention questionnaire in April, 1986. The second survey took place after one year of follow-up in April 1987. Again data were collected after two years of follow-up in April 1988. For our analysis, we limit attention to data from 1600 students from 28 Los Angeles schools. These data were previously examined by Hedeker *et al.* (1994) and Klar and Darlington (2004), and are available at *http* : *//tigger.uic.edu/~hedeker/mix.html*.

We limit attention to one intervention, social-resistance classroom curriculum (CC), from this factorial design with four study conditions. Furthermore, we concentrate on one of the primary study outcome variables, the tobacco and health knowledge scale (THKS) score. The THKS varied from zero to seven. A score of zero indicates that none of the knowledge questions were answered correctly while a score of seven indicates that all knowledge questions were answered correctly. For the purpose of analysis we dichotomize the THKS score as 0-2 and 3-7 following Hedeker (1999).

5.3 Results

5.3.1 Descriptive Analysis

Characteristics of the 1600 students are presented in table 5.1 by intervention group. There are the same number of schools in each of the classroom curriculum (CC) and control groups. However, the number of participating students per school varies by intervention group. The average number of students per school is approximately 55 in the CC group (min=23, max=114) while for the control group it is 60 (min=18, max=137).

Table 5.1: The TVSFP: measures among students identified during baseline survey

Measures	Classroom Curriculum (CC)	Control
Number of Students	763	837
Number of Schools	14	14
Students per School		
Average	55	60
Range	(23-114)	(18-137)

Denoting the pretest and posttest scores as preTHKS and postTHKS, respectively, the number of students and the percentage classified according to the dichotomized preTHKS and postTHKS scores for both the control and intervention groups are given in Table 5.2. The number of events (student with THKS score 3-7 defined as event) and the percentage of events by intervention groups are provided in Table 5.3. The event rate is calculated as the percentage of students whose THKS score was between three and seven. It increased approximately from 32% to 62% between pretest and posttest observations in the classroom curriculum intervention group (CC) and from 36% to 45% in the control group. Analyses were performed to

investigate whether the observed increases in knowledge scores of CC vs control were statistically significantly different.

Table 5.2: Number (%) of students classified according to dichotomized preTHKS and postTHKS score by intervention groups in the TVSFP

	postTHKS score			
	CC (n=763)		Control (n=837)	
preTHKS Score	0	1	0	1
0^a	226 (29.6)	291 (38.1)	338 (40.4)	195 (23.3)
1^b	66 (8.7)	180 (23.6)	123 (14.7)	181 (21.6)

^{a}0 denotes a THKS score from 0-2

^{b}1 denotes a THKS score from 3-7

Table 5.3: Number of event (%) among preTHKS and postTHKS by intervention groups in the TVSFP

	Classroom Curriculum (n=763)	Control (n=837)
preTHKS	246 (32%)	304 (36%)
postTHKS	471 (62%)	376 (45%)

5.3.2   Effect of social-resistance classroom curriculum (CC) on THKS score

Effect of the CC vs control on the THKS score is summarized in Table 5.3. Each of the methods described previously are applied. For all methods, the estimated odds ratio of increasing THKS score approximately equals 2 comparing students in the classroom curriculum group (CC) to students in the control group. The 95% confidence intervals for all methods do not contain the odds ratio of one. Similarly, tests of the effect of intervention of CC on THKS score from all the methods were statistically significant ($p < 0.01$).

Table 5.4: Estimated effect of CC vs Control on student THKS using data from the TVSFP

| | Extensions of Logistic Regression | | | | | |
| | ANCOVA | | LDA | | Posttest Only | |
Effect of CC	CS [a]	PA [b]	CS	PA	CS	PA
Odds Ratio	2.20	2.15	2.43	2.37	2.09	2.02
95%CI[c]	1.54- 3.15	1.54- 3.00	1.81-3.26	1.74- 3.22	1.42-3.07	1.41- 2.89
P-value	<0.0001	<0.0001	<0.0001	<0.0001	0.0002	0.0001

[a]CS: Cluster-Specific

[b]PA: Population-Averaged

[c]CI: Confidence Interval

5.4 Discussion

Klar and Darlington (2004) also studied the effect of CC on THKS. The notable difference with our approach is that these authors focused on methods where THKS was modeled as a continuous outcome. Based on the results provided by Klar and Darlington (2004) there was also a statistically significant effect of CC ($p < 0.01$) on increasing THKS.

The odds ratio estimates of cluster-specific and population-averaged models are very similar likely due to the small ICC ($\hat{\rho} = 0.02$) obtained using data from the TVSFP. As noted by Neuhaus *et al.* (1991) the odds ratio from cluster-specific models is larger than that from population-averaged models.

The TVSFP was designed using a 2×2 factorial design which includes television (TV) in addition to the classroom curriculum. The effect of CC can be investigated adjusting for other factors (such as TV, site) considered in this project.

5.5   Summary

The methods described in Chapter 2 are illustrated using data from the TVSFP. The effect of social-resistance classroom curriculum (CC) on the study outcome dichotomized variable tobacco and health knowledge scale (THKS) score is examined using these methods.

Chapter 6

DISCUSSION

6.1 Introduction

In this chapter, we summarize study findings. The primary focus of this thesis was to compare models for assessing the intervention effect using pretest-posttest binary data from cluster randomization trials. This chapter comprises four sections. Key thesis findings are summarized in section 6.2. Study limitations and scope for further research are discussed in section 6.3. Finally, section 6.4 contains a summary of this chapter.

6.2 Key findings

There were six Wald test statistics being compared. Three of the test statistics were based on cluster-specific extensions of logistic regression while the remaining three test statistics were obtained using population-averaged extensions of logistic regression.

A simulation study and data from the TVSFP (Flay *et al.*, 1995) were used to illustrate the use of the test statistics. The simulation study is described in Chapter 3. In the simulation study, the number of clusters varied from 15 to 50, and the cluster sizes ranged from 15 to 100 subjects per cluster. It is evident from the simulation study that the cluster-specific extension of logistic regression yielded satisfactory type I error

rates at the 5% nominal level. On the other hand, the population-averaged extension of logistic regression yielded overly liberal type I error rates when the number of clusters was small. For the number of clusters $J = 15$, the type I error rates were at least 7%. These results agree with results obtained by other researchers and this approach requires a large number of clusters (Pan and Wall, 2002). Poor performance of GEE in the case of a small number of clusters also was reported by other authors (Bellamy *et al.*, 2000).

The cluster-specific extension of logistic ANCOVA and logistic regression based on posttest measurements only provided type I error rates close to the nominal level (5%) when we generated the data considering different values for pretest and posttest association. That is, type I error rates were not affected with different pretest-posttest associations at the cluster- and individual-level. Empirical power of the methods investigated in this study was marginally affected when we generated the data using different values at the individual-level and the cluster-level. These results are consistent with the results presented by Klar and Darlington (2004). However, they showed power can be regained by fitting a model incorporating the terms which represent the individual and cluster level association.

The cluster-specific ANCOVA and PO are valid for all parameter combinations. Among these two methods, in some scenarios, the cluster-specific ANCOVA yielded marginally greater power compared to logistic regression based on posttest only. The LDA approach where both pretest and posttest measurements were considered as the outcomes yielded the lowest power compared with the other competing methods. These results mirror the findings of Austin (2010). We used adaptive Gaussian quadrature and GEE for estimating the coefficients corresponding to the interaction term in the case of LDA. However, Localio *et al.* (2006) reported that a Bayesian approach yielded better estimates compared to GEE and adaptive quadrature. Several advantages of using the methods that accounted for baseline measurements were

reported by Austin (2010). Vickers and Altman (2001) concluded that methods accounting for baseline measurements usually have higher statistical power compared to analysing change from baseline. Moreover, cluster-specific ANCOVA provided comparable or slightly more precise estimates compared to competing methods. Nixon and Thompson (2003) concluded that improved precision can be achieved by adjusting for baseline.

Overall, empirical power for each method increased with cluster size. Similarly, for a given cluster size, power increased as the number of clusters increases.

The results of this study lead to the recommendation that cluster-specific logistic ANCOVA is appropriate for testing the effect of intervention in case of pretest-posttest binary outcome from completely randomized cluster randomization trials. Also population-averaged models are not appropriate when the number of clusters are small.

6.3 Study limitations and possible future research

Several methods are available to generate correlated binary data. We considered a cluster-specific model to generate the pretest and posttest binary data. However, Neuhaus and Jewell (1993) commented that interpretation of the estimated covariate effect obtained from cluster-specific model may be difficult when the covariates are investigated at the cluster level. The authors also noted that cluster-specific models would be more appropriate for testing covariate effects that vary within clusters rather than the intervention effect where every subject in a cluster is assigned to the same intervention. They preferred population-averaged models such as the GEE approach for testing the effect of intervention. In this thesis, we considered population-averaged extensions of logistic regression for testing the intervention effect but not for generating the data.

In practice, the number of subjects per cluster varies from cluster to cluster. In the simulation study, we limited our attention to an equal number of subjects per cluster. Furthermore, we concentrated on only two values of intracluster correlation coefficient. Also we considered only a few values for subject- and cluster-level associations. It might be useful to do a more extensive simulation.

We only investigated the methods based on a completely randomized setting of cluster randomization trials. Stratified and matched-pair designs can also be used for cluster randomization trials (Donner and Klar, 2000). However, the completely randomized design is the simplest and a wide number of statistical methods can be used for analysis.

We used a cluster-specific model to generate the pretest-posttest binary data. Again, we generated the data assuming that the cluster-specific random effects follow a normal distribution. Thus it is difficult to say from this study whether the random effects generated from other distributions also perform well.

Generating data using a cluster-specific model only specifies the measures of ICC based on latent variables not based on manifest variables. The ICC based on manifest variables is always less than the ICC based on latent variables (Rodríguez and Elo, 2003). However, when the between-cluster variation is small, both measures are similar.

We considered equal numbers of subjects per cluster in our simulation study. This design helps us to understand the performance of these methods in simple scenarios. Future research involving a more general setting such as unbalanced cluster sizes is required to assess our findings.

For a small number of clusters, the generalized estimating equations (GEE) (Liang and Zeger, 1986) approach has a number of known limitations. For example, Wald test statistics are known to yield overly liberal type I error rates. Several

methods have been developed to avoid these difficulties including degrees of freedom correction (Mancl and DeRouen, 2001; McCaffrey and Bell, 2006; Pan and Wall, 2002). Moreover, a score type test can be used instead of a Wald test as it has better small sample properties (Guo *et al.*, 2005). Future research is required exploring the extension of GEE for trials involving a small number of clusters.

Twisk and Proper (2004) investigated the nominal logistic regression approach for analysing change from baseline to follow-up measurements. They showed that a categorical variable with four categories can be created based on the change in pretest and posttest dichotomous measurements. This categorical variable can be analysed using nominal logistic regression. It would be interesting to extend this study using nominal logistic regression for analysing change in pretest and posttest binary observations in the context of cluster randomization trials.

6.4   Summary

In conclusion, in this study we examined different statistical methods for assessing the effect of intervention using pretest-posttest binary measurements in the context of cluster randomization trials. Empirical power of these methods was marginally affected by different individual- and cluster-level associations. The LDA approach yielded the lowest power (approximately a minimum of 15% lower except when number of clusters 30 and cluster size 100 and number of clusters 50 and cluster size 50) for testing the intervention effect among the competing methods. Population-averaged (GEE) methods are generally not appropriate when the number of clusters is small (e.g. 15).

BIBLIOGRAPHY

Agresti, A. (2002). *Categorical Data Analysis.* John Wiley and Sons New York.

Altman, D. and Dore, C. (1990). Randomisation and baseline comparisons in clinical trials. *Lancet* **335**, 149.

Austin, P. C. (2010). A comparison of the statistical power of different methods for the analysis of repeated cross-sectional cluster randomization trials with binary outcomes. *The International Journal of Biostatistics* **6** (1), 11.

Bellamy, S. L., Gibberd, R., Hancock, L., Howley, P., Kennedy, B., Klar, N., Lipsitz, S. and Ryan, L. (2000). Analysis of dichotomous outcome data for community intervention studies. *Statistical Methods in Medical Research* **9**, 135–159.

Bland, J. (2004). Cluster randomised trials in the medical literature: Two bibliometric surveys. *BMC Medical Research Methodology* **4**, 21–27.

Bonate, P. (2000). *Analysis of Pretest-Posttest Designs.* Chapman and Hall USA.

Burton, A., Altman, D. G., Royston, P. and Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine* **25**, 4279–4292.

Cameron, R., Brown, S., Best, A., Pelkman, C., Madill, C., Manske, S. and Payne, M. (1999). Effectiveness of a social influences smoking prevention program as a function of provider type, training method, and school risk. *American Journal of Public Health* **89**, 1827–1831.

Cronbach, L. and Furby, L. (1970). How should we measure change or should we? *Psychological Bulletin* **74**, 68–80.

Diggle, P., Heagerty, P., Liang, K.-Y. and Zeger, S. (2002). *Analysis of Longitudinal Data.* Oxford University Press New York.

Donner, A. and Klar, N. (2000). *Design and Analysis of Cluster Randomization Trials in Health Research.* Arnold London.

Donner, A. and Zou, G. Y. (2007). Methods for the statistical analysis of binary data in split-mouth designs with baseline measurements. *Statistics in Medicine* **26**, 3476–3486.

Fitzmaurice, G. M., Larid, N. M. and Ware, J. H. (2004). *Applied Longitudinal Analysis.* John Wiley and Sons New York.

Flay, B., Miller, T., Hedeker, D., Siddiqui, O., Britton, B., Johson, C., Hansen, W., Sussman, S. and Dent, C. (1995). The television, school, and family smoking prevention and cessation project. VII student outcomes and mediating variables. *Preventive Medicine* **24**, 29–40.

Flay, B. R. (1985). Psychological approaches to smoking prevention: A review of findings. *Statistics in Medicine* **4** (5), 449–488.

Flay, B. R. (1987). Mass media and smoking cessation: A critical review. *Amercian Journal of Public Health* **77** (2), 153–160.

Glynn, T. J. (1990). School programs to prevent smoking: The national cancer institute guide to strategies that succeed. Technical Report NIH publication 990-500 National Cancer Institute.

Guo, X., Pan, W., Connett, J. E., Hannan, P. J. and French, S. A. (2005). Small-sample performance of the robust score test and its modifications in generalized estimating equations. *Statistics in Medicine* **24**, 3479–3495.

Hedeker, D. (1999). Mixno: A computer program for mixed-effects nominal logistic regression. *Journal of Statistical Software* **4** (5), 1–92.

Hedeker, D. and Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics* **50**, 933–944.

Hedeker, D., Gibbons, R. D. and Flay, B. R. (1994). Random-effects regression models for clustered data with an example from smoking prevention research. *Journal of Consulting and Clinical Psychology* **62** (4), 757–765.

Heo, M. and Leon, A. C. (2005). Performance of a mixed effects logistic regression model for binary outcomes with unequal cluster size. *Journal of Biopharmaceutical Statistics* **15** (3), 513–526.

Hollingworth, W., Cohen, D., Hawkins, J. Hughes, R. A., Moore, L. A. R., Holliday, J. C., Audrey, S., Starkey, F. and Campbell, R. (2012). Reducing smoking in adolescents: Cost-effectiveness results from the cluster randomized ASSIST (A Stop Smoking In School Trial). *Nicotine and Tobacco Research* **14** (2), 161–168.

Klar, N. and Darlington, G. (2004). Methods of modeling change in cluster randomization trials. *Statistics in Medicine* **22**, 2341–2357.

Klar, N. and Donner, A. (2001). Current and future challenges in the design and analysis of cluster randomization trials. *Statistics in Medicine* **20**, 3729–3740.

Liang, K. and Zeger, S. (2000). Longitudinal data analysis of continuous and discrete responses for pre-post designs. *Sankhya: The Indian Journal of Statistics* **62**, 134–148.

Liang, K.-Y. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Localio, A. R., Berlin, J. A. and Ten Have, T. R. (2006). Longitudinal and repeated cross-sectional cluster-randomization designs using mixed effects regression for binary outcomes: bias and coverage of frequentist and bayesian methods. *Statistics in Medicine* **25**, 2720–2736.

Mancl, L. A. and DeRouen, T. A. (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics* **57** (1), 126–134.

McCaffrey, D. F. and Bell, R. M. (2006). Improved hypothesis testing for coefficients in generalized estimating equations with small samples of clsuters. *Statistics in Medicine* **25** (23), 4081–4098.

Neuhaus, J. and Jewell, N. (1993). A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrika* **80**, 807–815.

Neuhaus, J., Kalbfleisch, J. and Hauck, W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review* **59** (1), 25–35.

Nixon, R. M. and Thompson, S. G. (2003). Baseline adjustments for binary data in repeated cross-sectional cluster randomization trials. *Statistics in Medicine* **22**, 2673–2692.

Pan, W. and Wall, M. M. (2002). Small-sample adjustments in using the sandwich estimator in generalized estimating equation. *Statistics in Medicine* **21**, 1429–1441.

Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the loglikelihood function in the nonlinear mixed effects model. *Journal of Computational and Graphical Statistics* **4**, 12–35.

Plewis, I. (1985). *Analysing Change: Measurement and Explanation Using Longitudinal Data*. Wiley Chicester.

Prentice, R. L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of The American Statistical Association* **81**, 321–327.

Rodríguez, G. and Elo, I. (2003). Intra-class correlation in random-effects models for binary data. *The Stata Journal* **3** (1), 32–46.

SAS Institute Inc. (2011). *SAS 9.3 Help and Documentation.* SAS Institute Inc.

Sashegyi, A., Brown, S. and Farell, P. (2000). Estimation in empirical bayes model for longitudinal and cross-sectionally clustered binary data. *The Canadian Journal of Statistics* **28** (1), 45–63.

Senn, S. (1994). Testing for baseline balance in clinical trials. *Statistics in Medicine* **13**, 1715–1726.

Twisk, J. and Proper, K. (2004). Evaluation of the results of randomized control trial: How to define changes between baseline and follow-up. *Journal of Clinical Epidemiology* **57**, 223–228.

Twisk, J. W. R. (2003). *Applied Longitudinal Data Analysis for Epidemiology.* Cambridge University Press New York.

Ukoumunne, O. and Thompson, S. (2001). Analysis of cluster randomized trials with repeated cross-sectional binary measurements. *Statistics in Medicine* **20**, 417–433.

Vickers, A. J. and Altman, D. G. (2001). Analysing controlled trials with baseline and follow up measurements. *British Medical Journal* **323**, 1123–1124.

Williams, D. A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics* **31**, 949–952.

Yasui, Y., Feng, Z., Diehr, P., McLerran, D., Beresford, S. A. A. and McCulloch, C. E. (2004). Evaluation of community-intervention trials via generalized linear mixed models. *Biometrics* **60**, 1043–1052.

LIST OF ACRONYMS

ANCOVA	Analysis of Covariance
CC	Social-resistance Classroom Curriculum
CS	Cluster-specific
CRT	Cluster Randomization Trials
GEE	Generalized Estimating Equations
ICC	Intracluster Correlation Coefficient
LDA	Longitudinal Data Analysis
PO	Logistic Regresion with Posttest Only
PA	Population-averaged
PQL	Penalized Quasi Likelihood
SE	Standard Error
THKS	Tobacco and Health Knowledge Scale
TVSFP	The Television, School, and Family Smoking Prevention and Cessation Project

APPENDIX:
SAS CODE TO FIT THE MODELS

In this appendix we present the SAS code used to fit the six methods described in Chapter 2. As noted previously these methods are cluster-specific and population-averaged extensions of logistic regression with posttest only, logistic ANCOVA, and longitudinal approach. We used SAS procedures PROC GLIMMIX and PROC GENMOD for cluster-specific and population-averaged methods, respectively.

Posttest Only

We used the following SAS code for cluster-specific and population-averaged methods with posttest only corresponding to models 2.1 and 2.2 respectively.

```
/* Cluster-specific with posttest only */
proc glimmix data=imldata method=quad;
class cluster;
model yijk=group/ s dist=binomial link=logit;
random intercept/subject=cluster;
run;
```

```
/* Population-averaged with posttest only */
proc genmod data=imldata descending;
class cluster;
model yijk=group/ dist=binomial link=logit;
repeated subject=cluster/type=exch;
run;
```

Logistic ANCOVA

Again, we used the following SAS code for cluster-specific (model 2.3) and population-averaged (model 2.4) methods of ANCOVA, respectively.

```
/* Cluster-specific logistic ANCOVA */
proc glimmix data=imldata method=quad;
class cluster;
model yijk=group xijk/ s dist=binomial link=logit;
random intercept/subject=cluster;
run;
```

```
/* Population-averaged logistic ANCOVA */
proc genmod data=imldata descending;
class cluster;
model yijk=group xijk/ dist=binomial link=logit;
repeated subject=cluster/type=exch;
run;
```

Longitudinal Data Analysis (LDA)

For LDA approach we created a outcome variable *ltpt* combining both pretest and posttest measurements. Also we created a variable *time* which takes value 0 for pretest measurements and 1 for posttest measurements. We used the following code to create the dataset for LDA.

```
/* LDA data creation */
data ldata;
set imldata;
ltpt=xijk; time=0; output;
ltpt=yijk; time=1; output;
run;
```

The following SAS code was used to fit the cluster-specific and population-averaged LDA corresponding to models 2.5 and 2.6 respectively.

```
/* Cluster-specific LDA */
proc glimmix data=ldata method=quad;
class cluster;
model ltpt=group time group*time / s dist=binomial link=logit;
random intercept/subject=cluster;
run;
```

```
/* Population-averaged LDA */
proc genmod data=ldata descending;
class cluster;
model ltpt=group time group*time/ dist=binomial link=logit;
repeated subject=cluster/type=exch;
run;
```

VITA

NAME	ASM Borhan
PLACE OF BIRTH	Comilla, Bangladesh
EDUCATION	Institute of Statistical Research and Training (ISRT) University of Dhaka Dhaka, Bangladesh 1998–2004, BSc in Applied Statistics Institute of Statistical Research and Training (ISRT) University of Dhaka Dhaka, Bangladesh 2005–2006, MS in Applied Statistics Department of Epidemiology & Biostatistics The University of Western Ontario London, Ontario, Canada 2010–2012, MSc in Biostatistics
EXPERIENCE	Research Assistant The University of Western Ontario London, Ontario, Canada 2010–2012