Fall 11-13-2018

The compensation for few clusters in clustered randomized trials with binary outcomes

Lily Stalter
University of New Mexico - Main Campus

Lily Stalter

Candidate

Mathematics and Statistics

Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Fares Qeadan ,Chairperson

Helen Wearing

Yan Lu

THE COMPENSATION FOR FEW CLUSTERS IN CLUSTERED RANDOMIZED TRIALS WITH BINARY OUTCOMES

BY

LILY STALTER

BA, MATHEMATICS, STATE UNIVERSITY OF NEW YORK AT PLATTSBURGH

THESIS

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

Mathematics

The University of New Mexico
Albuquerque, New Mexico

December, 2018

DEDICATION

To my family, Sam, Michael, and anyone else who has been a part of my support system over the past couple of years.

Acknowledgements

A special thank you to Dr. Qeadan, whose knowledge and enthusiasm spurred this project.

I would also like to thank Dr. Lu and Dr. Wearing, for their time and expertise.

THE COMPENSATION FOR FEW CLUSTERS IN CLUSTERED RANDOMIZED TRIALS WITH BINARY OUTCOMES

by

Lily Stalter

MS, Mathematics, University of New Mexico, 2018
BA, Mathematics, State University of New York at Plattsburgh

ABSTRACT

Cluster randomized trials are increasingly popular in epidemiological and medical research. When analyzing the data from such studies it is imperative that the hierarchical structure of the data be taken into account. Multilevel logistic regression is used to analyze clustered data with binary outcomes. Previous literature shows that a greater number of clusters is more important than a large number of subjects per cluster. This paper investigates if it is possible to compensate for the increased bias found for parameter estimates when the number of clusters is decreased. A simulation study was conducted where the absolute percent relative bias for each parameter estimate with 5 to 49 clusters and 10, 20, 30, 60, 90, 120, 150, 180, and 210 subjects per cluster were compared to the bias found for corresponding parameter estimates when the number of clusters was 50 with 10 subjects per cluster. Maximum Likelihood, Restricted Maximum Likelihood, and Generalized Estimating Equation methods, with multiple Intraclass Correlation Coefficients were examined. For Maximum Likelihood estimates, results show that it is possible to account for the effects of few clusters by increased sample size when examining fixed effect parameter estimates. For variance components, it was not possible to fully compensate under all conditions, but in general, the trend found was that increasing the number of subjects per cluster either results in decreased bias or the bias plateaued after a certain sample size. Further investigation is needed on Restricted Maximum Likelihood and Generalized Estimating Equation estimates, but results show that they do not perform well when the number of subjects per cluster is few. The results of this study are very informative for researchers who are limited to few clusters.

Contents

1 Introduction **1**
1.1 Background . . . 1
1.2 Objective . . . 2
1.3 Linear Regression . . . 2
1.4 Logistic Regression . . . 3
1.5 Multilevel Logistic Regression . . . 4
1.6 Estimation Methods . . . 6
1.6.1 Maximum Likelihood . . . 6
1.6.2 Generalized Estimating Equations . . . 8
1.7 Optimization . . . 10
1.8 Literature Review . . . 12

2 Methods **16**
2.0.1 Simulation Conditions . . . 17
2.0.2 Maximum Likelihood . . . 18
2.0.3 Restricted Maximum Likelihood . . . 19
2.0.4 Generalized Estimating Equations . . . 19

3 Results **21**
3.1 Maximum Likelihood . . . 21
3.1.1 Convergence and Inadmissible Solutions . . . 21
3.1.2 Parameter Estimates for $\rho = 0.17$ . . . 22
3.1.3 Parameter Estimates for $\rho = 0.04$ . . . 25
3.1.4 Parameter Estimates for $\rho = 0.38$ . . . 28
3.1.5 Standard Errors . . . 30
3.2 Restricted Maximum Likelihood . . . 32
3.2.1 Convergence . . . 32
3.2.2 Parameter Estimates for $\rho = 0.17$ . . . 33
3.2.3 Parameter Estimates for $\rho = 0.04$ . . . 36
3.2.4 Parameter Estimates for $\rho = 0.38$ . . . 38
3.2.5 Coverage of the 95% Confidence Intervals . . . 41
3.3 Generalized Estimating Equations . . . 43
3.3.1 Convergence . . . 43
3.3.2 Parameter Estimates . . . 43
3.3.3 Coverage of the 95% Confidence Intervals . . . 51

4 Discussion **53**
4.0.1 Complete Compensation versus Incomplete Compensation . 53
4.0.2 Comparison of Estimation Methods 55
4.0.3 Impact of ICC 57

5 Conclusion **59**

List of Figures

3.1 Number of clusters needed for compensation, MLE with $\rho = 0.17$. 24
3.2 Number of clusters needed for compensation, MLE with $\rho = 0.04$. 26
3.3 Number of clusters needed for compensation, MLE with $\rho = 0.38$. 30
3.4 Number of clusters needed for compensation, REML with $\rho = 0.17$ 35
3.5 Number of clusters needed for compensation, REML with $\rho = 0.04$ 36
3.6 Number of clusters needed for compensation, REML with $\rho = 0.38$ 41
3.7 Number of clusters needed for compensation, GEE with $\rho = 0.17$. 44
3.8 Number of clusters needed for compensation, GEE with $\rho = 0.04$. 49
3.9 Number of clusters needed for compensation, GEE with $\rho = 0.38$. 50

List of Tables

3.1 Absolute percent relative bias of MLE estimates with $\rho = 0.17$ . . . 23
3.2 Absolute percent relative bias of MLE estimates with $\rho = 0.04$ . . . 27
3.3 Absolute percent relative bias of MLE estimates with $\rho = 0.38$ . . . 29
3.4 Noncoverage of the 95% CI by ICC for MLE 31
3.5 Noncoverage of the 95% CI by number of clusters for MLE 31
3.6 Noncoverage of the 95% CI by cluster size for MLE 31
3.7 Absolute percent relative bias of REML estimates with $\rho = 0.17$ . . 34
3.8 Absolute percent relative bias of REML estimates with $\rho = 0.04$ . . 37
3.9 Absolute percent relative bias of REML estimates with $\rho = 0.38$ . . 40
3.10 Noncoverage of the 95% CI by ICC for REML 42
3.11 Noncoverage of the 95% CI by number of clusters for REML 42
3.12 Noncoverage of the 95% CI by cluster size for REML 43
3.13 Absolute percent relative bias of GEE estimates with $\rho = 0.17$ . . . 46
3.14 Absolute percent relative bias of GEE estimates with $\rho = 0.04$ . . . 47
3.15 Absolute percent relative bias of GEE estimates with $\rho = 0.38$ . . . 48
3.16 Noncoverage of the 95% CI by ICC, GEE 51
3.17 Noncoverage of the 95% CI by number of clusters for GEE 52
3.18 Noncoverage of the 95% CI by cluster size for GEE 52

4.1 Absolute percent relative bias with 1000 subjects per cluster 54
4.2 Absolute percent relative bias for REML estimates with 10 subjects per cluster 57
4.3 Absolute percent relative bias for GEE estimates with 10 subjects per cluster 57

Chapter 1

Introduction

1.1 Background

Over the past twenty years, there has been an increase in the number of studies that use hierarchical or clustered data structure (Austin and Merlo, 2017; Hayes and Moulton, 2017; Eldridge and Kerry, 2012; Wang, Xie and Fisher, 2012). Clusters frequently occur throughout the medical and epidemiological fields, where binary outcomes are common. Examples of clustered data include patients nested within hospitals, children nested within families, students nested within schools, and individuals nested within communities. In Cluster Randomized Trials (CRTs), clusters rather than individuals are randomly placed in the control or intervention arm of the study. Responses are measured at the subject level but the unit of randomization is the cluster (Austin 2007). CRTs prevent treatment contamination and may be necessary for certain intervention methods (van Breukelen and Candel, 2012).

The randomization of clusters rather than individuals increases the complexity of the design and analysis, since the independence assumption is often violated by subjects within the same cluster (Rutterford et al., 2015). Therefore when analyzing data from CRTs, it is imperative that the hierarchical structure of the data be taken into account. Failing to do so may result in underestimation of standard errors and an inflation of type-I error rate for significance tests of regression coefficients (McNeish and Stapleton, 2014). The Intraclass Correlation Coefficient (ICC) is a measure of how much subjects within a cluster are correlated (Wang, Xie, and Fisher, 2012). Methods that can account for the dependence of individuals within clusters are known as multilevel models (MLMs), hierarchical linear models (HLMs), or mixed-effect models (Raudenbush and Bryk, 2002). In this paper, we will refer to them as multilevel models. For binary outcomes, multilevel logistic models are used to estimate the odds that an event will occur while taking the dependency of data into account.

1.2 Objective

This study will investigate if increasing cluster size can compensate for the increased bias experienced when few clusters are used in a multilevel logistic model. We hypothesize that this may be true for some fixed parameter estimates, but not for random effects. A review of multilevel logistic models, an overview of the estimation methods used, and a review of the current literature are presented. A simulation study is conducted that tests this hypothesis.

1.3 Linear Regression

To understand multilevel logistic models, we must first briefly review simple linear regression and logistic regression. Unless otherwise noted, sections 1.3, 1.4, and 1.5 are based on Sommet and Morselli (2017).

In linear regression the mean value of an outcome variable is predicted for a particular value of the predictor variable. The equation for linear regression is given by

$$Y_i = \beta_0 + \beta_1 * X_i + \varepsilon_i \tag{1.1}$$

...where Y_i is the observed value of the outcome variable for subject i;

...X_i is the observed value of the predictor variable for subject i;

...β_0 is the value of Y_i when $X_i = 0$, known as the intercept;

...β_1 is the coefficient for the independent variable X_i, known as the slope;

...ε_i is the residual, which is the difference between what is predicted for subject i and what is actally observed for subject i.

Interpretation of a simple linear regression model is straightforward; for every one unit increase in X_i, on average one can expect to see Y_i increase by β_i units. If β_1 is not significantly different from zero, then the null hypothesis (H_0), that there is no significant relationship between the predictor variable and the outcome, cannot be rejected. If β_1 is significantly different from zero, then the null hypothesis is rejected.

1.4 Logistic Regression

When the outcome of interest is binary, logistic regression must be used to predict the conditional probability that an event of interest occurs for a particular value of the predictor variable. The logistic function is represented by the equation

$$P(Y_i = 1|X_i) = \frac{\exp(\beta_0 + \beta_1 * X_i)}{1 + \exp(\beta_0 + \beta_1 * X_i)} \tag{1.2}$$

...where $P(Y_i = 1|X_i)$ is the conditional probability that the outcome variable equals one for a given value X_i;

...exp is the exponent function;

...X_i, β_0, and β_1 are defined the same as in equation 1.1.

The logit transformation can be used to convert this s-shaped curve into a straight line and allow for easier interpretation of results. The log-odds is the logit of the conditional probability that the outcome variable equals one over the probability that it equals zero. The logit function is the natural logarithm of the odds, with the post-logit transformation logistic regression equation being

$$Logit(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 * X_i \tag{1.3}$$

...where $p = P(Y_i = 1|X_i)$ (Wang, Xie, and Fisher, p 117).

To interpret β_1, we use the Odds Ratio (OR), where $OR_i = exp(\beta_i)$. The odds ratio is defined as the amount by which the odds of the outcome increase or decrease when the value of the predicting variable, X_i, increases by one unit. For example, suppose $\beta_1 = 2$, then $OR_1 = exp(\beta_1) = exp(2) \approx 7.4$, meaning that for every one unit increase in the variable X_1, we should expect to see the odds of $Y = 1$ increase by 7.4. In the case of an odds ratio less than 1, the odds of $Y = 1$ decrease. Note that in logistic regression the notion of a residual is not necessary as the distance between the observed and predicted values can only take on two values, 0 or 1. If the OR is not significantly different from 1, one cannot reject H_0, where H_0 is that the $OR = 1$. If the OR is significantly different from 1, one rejects H_0.

1.5 Multilevel Logistic Regression

When data has a hierarchical structure, a standard logistic regression model will not produce accurate results because it does not take the dependency of individuals into account. In this case, multilevel models must be used. Level-1 variables refer to individual characteristics such as patient's age and level-2 variables refer to cluster characteristics such as number of individuals that attend a particular clinic. Level-1 characteristics can change within clusters, whereas level-2 characteristics are the same for all subjects within a cluster. The odds that the outcome variable equals one instead of zero are allowed to vary between clusters. The functional form for multilevel logistic regression is given by

$$Logit(p) = \beta_{00} + (\beta_{10} + u_{1j}) * x_{ij} + u_{0j} \tag{1.4}$$

...where x_{ij} is the observed value of the predictor variable for subject i in cluster j;

...β_{00} is the fixed effect intercept or the average log-odds that the outcome variable equals one in the overall sample;

...u_{0j} is the deviation of the cluster-specific intercept from the fixed effect intercept, also known as the level-2 residual;

...β_{10} is the fixed effect slope or the average effect of the level-1 variable in the overall sample;

...u_{1j} is the deviation of the cluster-specific slope from the fixed effect slope, or the residual term associated with the level-1 variable.

$\beta_{00} + \beta_{10} * x_{ij}$ is known as the fixed effect part of the model and $u_{0j} + u_{ij} * x_{ij}$ is the random effect part. The level-2 residual u_{0j} is the amount the intercept varies between clusters. The mean of these deviations is assumed to be zero and the variance is known as the random intercept variance, $var(u_{0j})$. The higher the random intercept variance, the larger the variation of the log-odds from one cluster to another. Interpretation of β_{10} is similar to that for logistic regression. The residual term associated with the level-1 predictor, u_{1j}, corresponds to the deviation of the effects of the level-1 variable x_{ij} on a given cluster from the overall effect of the level-1 variable, x_{ij}, across all clusters. Here the mean is also assumed to be zero and the variance component is the random slope variance, $var(u_{1j})$. The higher the random slope variance, the larger the variation of the effect of x_{ij} from one cluster to another. A non-significant random slope variance means that the variation of the effect of x_{ij} is very close to zero, so β_{10} is effectively the same in all clusters.

The ICC for a multilevel model is

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \tag{1.5}$$

...where σ_w^2 is the within group variance and σ_b^2 is the between group variance (Wang, Xie, and Fisher, 2012). The ICC may fall between 0 and 1, with $\rho = 0$ meaning that there is no within group homogeneity, so the model is reduced to a fixed effect model (Wang, Xie, and Fisher, 2012).

1.6 Estimation Methods

Estimation is complex in multilevel models due to the need to compute both regression coefficients and variance components (Wang, Xie, and Fisher, 2012). There are several methods used to estimate parameters in multilevel logistic regression. Here we give an overview of the methods used in this study.

1.6.1 Maximum Likelihood

The following section is referenced from Wang, Xie, and Fisher (2012), unless otherwise specified. In multilevel models, two matrices, G and R, must be reasonably estimated.

$$G = \begin{bmatrix} \sigma^2_{u0} & \sigma^2_{u01} \\ \sigma^2_{u01} & \sigma^2_{u1} \end{bmatrix} \tag{1.6}$$

is the variance/covariance matrix for the level-2 residuals, shown here for a model with a level-1 random intercept and a random slope coefficient.

$$R = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} \tag{1.7}$$

is the variance/covariance matrix for the level-1 residuals.

Maximum Likelihood (ML) estimation is commonly used to estimate the matrices G and R. This type of estimation maximizes the Likelihood Function

$$L(\beta, \theta, Y) = \int f(Y \mid u)p(u)du \tag{1.8}$$

...where β denoted the fixed effects;

...θ denotes the unknown parameters of variances and covariances;

...$f(y \mid u)$ is the distribution of the outcomes measure conditional on the random effect U;

...and $p(u)$ is the distribution of the random effects.

Estimates are obtained by integrating over the distribution of the random effects. Parameter estimates are those that maximize this marginal likelihood function. In a linear model, this integral can be solved in closed form, while in nonlinear multilevel models, this integral has an open form and hence must be solved numerically. Two methods to do so are linearization which approximates the integrated likelihood function using techniques such as Taylor series expansion or integral approximation with numerical methods.

ML estimation is an iterative process, where initial starting values for parameter estimates are generated and these become the starting values for the next iteration. This process is repeated until the estimates stabilize from one iteration to the next, such that the specified convergence criterion is satisfied. If convergence is not achieved this can indicate poor model specification or an inadequate sample size.

ML estimates are consistent and asymptotically normal, meaning the maximum likelihood estimate will have an approximate normal distribution centered on the true parameter value, thus making significance testing for parameters possible.

Two commonly used likelihood functions that can be used in multilevel modeling are the Full Maximum Likelihood (FML) and Restricted Maximum Likelihood (REML), also sometimes referred to as residual Maximum Likelihood (Hox, 2010; McNeish and Stapleton, 2016; Browne, 1998). In this paper we will refer to FML simply as Maximum Likelihood Estimators (MLE). In the MLE method, both regression coefficients and the variance components are included in the likelihood function, therefore all parameters in the model are estimated simultaneously (Hox, 2010). In REML, only the variance components are included in the likelihood function and the regression coefficients are estimated in the second estimation step (Hox, 2010).

For a small number of groups, MLE and REML produce similar level-1 residual variance or R matrix estimates, but REML has been shown to provide less biased level-2 residual variances/covariances or the G matrix. MLE estimates of variance components are known to be biased downwards unless there are a large number of clusters (Eldridge and Kerry, 2012). REML, which has been developed more recently, tends to provide less biased estimates and is also less computationally intensive (Eldridge and Kerry, 2012). Little et al. (1996) discourages MLE estimation because the variance components are often biased downward, resulting in narrow confidence intervals.

MLE for binary outcomes can implemented by PROC NLMIXED when using SAS (SAS/STAT(R) 14.1 User's Guide: The NLMIXED Procedure). A psuedo-maximum likelihood can also be implemented by designating METHOD=MSPL in PROC GLIMMIX (SAS/STAT(R) 14.1 User's Guide: The GLIMMIX Procedure).

For binary outcomes, a REML analogue known as Residual Pseudo-Likelihood (RSPL) is used (Austin, 2010; SAS/STAT(R) 14.1 User's Guide: The GLIMMIX Procedure). RSPL in PROC GLIMMIX is identical to the REML method used in PROC MIXED, a procedure for modeling continuous data (SAS/STAT(R) 14.1 User's Guide: Comparing the GLIMMIX and MIXED Procedure).

1.6.2 Generalized Estimating Equations

The Generalized Estimating Equations (GEE) method was developed by Liang and Zeger (1986) and fits a population average model for cluster randomized trials (Eldridge and Kerry, 2012). In population average models, the mean population response rather than the individual response is used for estimation, therefore they do not have G-side effects (Hox, 2002; SAS/STAT(R) 14.1 User's Guide: Proc GLIMMIX Contrasted with Other SAS Procedures). The variance and covariance of the random effect part of the multilevel model are estimated directly from the residuals, which in the case of binary outcomes have a binomial distribution on the linear scale (Hox, 2010; Eldridge and Kerry, 2012). After GEE estimates are obtained for the variance components, an iterative Generalized Least Squares (GLS) method is used to estimate the parameters (Hox, 2010). Within clusters, correlation is treated as a nuisance parameter, which is estimated from the residuals and used to correct estimates (Eldridge and Kerry, 2012).

A correlation structure must be specified in the modeling (Eldridge and Kerry, 2012). Variance estimators based on a specified correlation structure are known as "model-based" (Eldridge and Kerry, 2012). If the correct correlation structure is specified, model-based estimators are more efficient. Sandwich estimators, where the estimator is written as an approximate correlation matrix "sandwiched" between two similar expressions of matrix algebra, are more robust to misspecification of the correlation matrix (Eldridge and Kerry, 2012). Most researchers use these robust estimators to analyze clustered data (Eldridge and Kerry, 2012). However, when a small number of clusters are used, the sandwich variance estimator can inflate type 1 error (Hayes and Moulton, 2017).

GEE estimates are faster to compute than ML estimates (Hox, 2002), but since GEE models do not correspond to the full probability model for the data, the likelihood of the data cannot be defined and therefore likelihood ratio tests cannot be conducted (Hayes and Moulton, 2017).

In SAS, GEE models can be computed using PROC GENMOD, but this only gives the fixed effect parameters (SAS/STAT(R) 14.1 User's Guide: Model Fitting in PROC GENMOD). For multilevel logistic mixed models with random effects, it is only possible to do a "GEE type" estimation when using SAS (SAS/STAT(R) 14.1 User's Guide: Fitting a Marginal (GEE-Type) Model). This is done through PROC GLIMMIX.

1.7 Optimization

The above estimation techniques require evaluation of open form problems to estimate parameters. To resolve this, an optimization technique must be used. An optimization algorithm finds the global optimum for a general nonlinear minimization problem (SAS/STAT(R) 9.22 User's Guide: Optimization Algorithms). These techniques are iterative and require repeated computations of the function value, the gradient vector, and the Hessian matrix for some techniques.

The following section references Chong and Zak (2001). Let $f : \mathbb{R}^n \to \mathbb{R}$ be the real-valued function that we wish to minimize, known as the objective or cost function. Since a maximum of f is a minimum of $-f$, we must only discuss minimization. The vector $\mathbf{x} = [x_1, x_2, ..., x_n]^T \in \mathbb{R}^n$ is known as the minimizer of f over Ω, where Ω is a subset of $\mathbb{R}^n$ known as the constraint set. If $\Omega = \mathbb{R}^n$, the problem is referred to as an unconstrained optimization problem. Let $\mathbf{d} \in R^n$, then $\mathbf{d} \neq 0$ is known as a feasible direction at $\mathbf{x} \in \Omega$ if there exists $\alpha_0 > 0$ such that $\mathbf{x} + \alpha\mathbf{d} \in \Omega$ for all $\alpha \in [0, \alpha_0]$. Assume $f \in C^2$, that is f is twice differentiable with continuous first and second derivatives. If $\mathbf{x}^*$ is a local minimizer of f over Ω, then for any feasible direction $\mathbf{d}$ at $\mathbf{x}^*$, we have

$$\mathbf{d}^T \nabla \mathbf{f}(\mathbf{x}^*) \geq 0 \tag{1.9}$$

where the gradient vector

$$\nabla \mathbf{f}(\mathbf{x}) = \mathbf{f}'(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_p} \right)^T. \tag{1.10}$$

If $\mathbf{x}^*$ is an interior point of Ω, then $\nabla \mathbf{f}(\mathbf{x}^*) = \mathbf{0}$. Additionally, a second necessary condition for the solution to $\mathbf{x}^*$ to be a minimum is that

$$\mathbf{d}^T H(\mathbf{x}^*)\mathbf{d} \geq 0, \tag{1.11}$$

where the matrix H is the Hessian of f:

$$H = \nabla^2 \mathbf{f}(x) = \mathbf{f}''(x) = \left(\frac{\partial^2 f}{\partial_{x_i} \partial_{x_j}} \right)_{p \times p}. \tag{1.12}$$

In the case that $\mathbf{x}^*$ is an interior point of Ω, $H(\mathbf{x}^*)$ is positive definite.

When either the gradient vector or the Hessian matrix can not be computed analytically they must be approximated discretely. This can be done using one of a variety of search methods, such as steepest descent, the Newton-Raphson algorithm, quasi-Newton methods, conjugate gradient, or trust region methods. These iterative methods employ one or more convergence criteria that determine when the algorithm has converged.

The optimization method used in this study was the trust region method, implemented in SAS by TECH=TRUREG in the syntax for the regression model. The trust region algorithm generates a sequence of points x_k (Schultz, Schnabel, Byrd, 2018). At the kth iteration, the quadratic model

$$\psi_k(w) = f_k + \nabla f_k p + \frac{1}{2} p^T B_k p \tag{1.13}$$

...where f_k is the function evaluated at x_k;

...∇f_k, is the gradient evaluated at x_k;

...B_k is the Hessian evaluated at x_k or some approximation of it;

...and $p \in \mathbb{R}^n$ is the candidate step (Schultz, Schnabel, Byrd, 2018).

An initial value for the trust radius, Δ_k, is found. In SAS this region is hyperelliptical (SAS/STAT(R) 9.22 User's Guide: Optimzation Algorithms). The size of the trust region is updated at each step. Each "minor iteration" uses the current trust radius Δ_k to compute a step

$$p_k(\Delta_k) = p(\nabla f_k, B_k, \Delta_k) \tag{1.14}$$

and then compares the actual reduction of the objective function to the predicted reduction by the quadratic model (Schultz, Schnabel, Byrd, 2018):

$$p_k = \frac{f_k - f(x_k + p_k(\Delta_k))}{f_k - \psi_k(p_k(\Delta_k))} = \frac{\text{actual reduction of } f}{\text{predicted model reduction of } f}. \tag{1.15}$$

If the step p_k is close to 1, then the quadratic model is a good predictor and the trust region can be increased (Nocedal and Wright, 1999). If p_k is small, the the trust region is reduced and the minor iteration is repeated (Shultz, Schnabel, and Byrd, 2018). Trust region methods choose the step size and direction simultaneously (Nocedal and Wright, 1999).

The algorithm for the trust region method, as presented by Shultz, Schnabel, and Byrd (1985), is as follows:

Algorithm 1 Trust Region

(0) Choose $\gamma_1, \eta_1, \eta_2 \in (0, 1)$, $x_1 \in \mathbb{R}^n$, $\Delta_0 > 0$, and let $k = 1$.
(1) Compute $f_k = f(x_k)$, $\nabla_k = \nabla(x_k)$, symmetric $B_k \in \mathbb{R}^{n \times n}$.
(2) Find Δ_k and compute $p_k = p_k(\Delta_k)$ satisfying:
$||p_k|| \leq \Delta_k$ and
(a) $\frac{\text{actual reduction of } f_k}{\text{predicted model reduction of } f_k} \geq \eta_1$ and
(b) either $\Delta_k \geq \delta_{k-1}$, or
$\Delta_k \geq ||B_k^{-1}\nabla_{k-1}||$ with B_{k-1} positive definite, or
for some $\Delta \leq \frac{1}{\gamma_1}\Delta_k$, $\frac{\text{actual reduction of } f_k}{\text{predicted model reduction of } f_k} < \eta_2$ or
$\frac{\text{actual reduction of } f_{k-1}}{\text{predicted model reduction of } f_{k-1}} < \eta_2$
(3) Let $x_{k+1} = x_k + p_k$ and $k = k + 1$.
(4) Go to (1)

1.8 Literature Review

Several "rules of thumb" have been recommended for the necessary number of clusters and cluster size in CRTs. These are a sample size of 30 or greater at either level for precise estimation, 50 groups each with 20 individuals if the cross-level interaction is the parameter of interest, or 100 groups with 10 individuals per group if the variance and covariance components are of interest (Hox, Moerbeek, and van de Schoot, 2017; Hox, 2010). In reality, a small number of available clusters or budget constraints often limit the number of clusters available to researchers. Few studies have examined the impact of sample size for CRTs with binary data and no studies have examined if it is possible to compensate when using few clusters.

Several authors have presented formulas for calculating sample size in CRTs (Rutterford et al., 2015; van Breukelen and Candel, 2012; Hemming et al., 2017; Hayes and Bennett, 1999; Hemming and Taljaard, 2016; Hemming et al., 2011). Formulas generally use either a fixed number of clusters, a fixed number of subjects, or the ICC. Using the ICC can be difficult for researchers because it is usually not known in the design phase of a study. Formulas for binary outcomes are given in Rutterford et al. (2015) and Hemming et al. (2011). Rutterford et al. (2015) states that for designs with few clusters, many of these calculations will likely result in a underestimation of the required sample size.

For continuous outcomes, it has been shown that a larger number of clusters is preferable over a large number of individuals per cluster (Maas and Hox, 2004; Mok, 1995). Maas and Hox (2004), found that when 50 or fewer clusters were used it led to bias estimates of the second level standard errors. Mok (1995) found that when total sample size was fixed, smaller bias was achieved when more groups with fewer individuals per group were used. Hemming et al. (2017) suggest that the total sample size required can be greatly decreased by a slight increase in the number of clusters.

Studies that have investigated binary outcomes have shown that, not only are a large number of clusters better, but small cluster size can also cause greater bias for the fixed effect point estimation and associated standard error estimation. Austin (2010), Clark (2008), and Moineddin et al. (2007) all found that in circumstances where 5 or fewer subjects per cluster were used there was an increase in bias, regardless of the number of clusters. Moineddin et al. (2007) found bias to be between 4% and 12% for level-1 fixed effect parameter estimates and 6% and 16% for level-2 fixed effect parameter estimates in their smallest sample size tested, 30 clusters with 5 subjects per cluster. When the cluster size was increased to 30, this bias became negligible. Even with 50 clusters and 5 subjects per cluster significant bias was found.

Moineddin found level-2 variance estimates were not biased when following the 30-30 rule, however when cluster size was below 30 estimates were positively biased, especially for those where the number of clusters was few. Similarly, Clark (2008) found level-2 variance components to have bias exceeding 100% even with more than 200 clusters, for cluster size of 2 and 5.

Model convergence also becomes more problematic when dealing with binary outcomes (Moineddin et al., 2007; Paccagnella, 2011). In a study with unbalanced groups, Paccagnella (2011) found conditions with fewer clusters had the greatest rate of non-convergence. Moineddin's et al. (2007) findings show that a small number of clusters can affect convergence, but small cluster size has a greater impact.

Lower convergence rates were also found in models with a lower prevalence (McNeish and Stapleton, 2016). Moineddin et al. (2007) recommend a minimum of 100 groups and 50 individuals per group for multilevel logistic regression models where the outcome has a low prevalence to produce valid estimates. They also found greater bias for all estimates when the prevalence of the outcome was 10% as compared to 45%.

Few studies have examined the difference in estimation methods for binary outcomes. Both Moineddin et al (2007) and Clark (2008) used MLE. Austin (2010) studied several estimation methods and found meaningful bias for level-2 variance components in simulations with less than 30 clusters, when using MLE. REML analogs, or Residual Pseudo-Likelihood (RSPL), still provided unbiased variance component estimates even with 10 clusters and either 10 or 15 subjects within each cluster (Austin 2010).

While more clusters may be preferable, real life scenarios often limit the number of available clusters, making it difficult to achieve the desired level of power (Hox, 2010). Decreasing the number of groups can also inflate the standard error estimates for both fixed and random effects (Theal et al., 2010). The few studies that have investigated sample size for studies with binary outcomes show that binary outcomes can be even more difficult to deal with than continuous outcomes.

Chapter 2

Methods

For all simulations conducted in this study, a multilevel logistic model with one level-1 explanatory variable and one level-2 explanatory variable was used. Therefore the level one equation for this model is:

$$log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{0j} + \beta_{1j}x_{1ij}. \tag{2.1}$$

The level 2 equations are:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}z_{1j} + u_{0j}, \tag{2.2}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}z_{1j} + u_{1j}. \tag{2.3}$$

When written as one model, this becomes:

$$log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}z_{1j} + \gamma_{11}x_{ij}z_{1j} + u_{0j} + u_{1j}x_{ij}, \tag{2.4}$$

where

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}\right) \tag{2.5}$$

and p_{ij} is the probability that individual i in group j will experience the outcome,

x_{ij} is the level-1 explanatory variable, and z_{1j} is the level-2 explanatory variable (Wang, Xie, and Fisher p 120). $\gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}z_{1j} + \gamma_{11}x_{ij}z_{1j}$ is referred to as the fixed effect part of the model and $u_{0j} + u_{1j}x_{ij}$ is the random effect part. $\gamma_{11}x_{ij}z_{1j}$ is a cross level interaction term where the coefficient γ_{11} shows how the slope of equation 2.1, β_{1j}, varies with $z_1 j$.

2.0.1 Simulation Conditions

Simulation conditions were chosen to be the same as the studies conducted by Maas and Hox (2005) and Moineddin et al. (2007). For all simulations, the fixed effect parameter estimates were set as $\gamma_{00} = -1$, $\gamma_{01} = 0.3$, $\gamma_{10} = 0.3$, and $\gamma_{11} = 0.3$.

The ICC for a logistic model is

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \tag{2.6}$$

...where $\sigma_e^2 = \frac{\pi^2}{3}$ and σ_u^2 is the variance of the random intercept, $u_{0j} \sim N(0, \sigma_u^2)$ (Eldridge and Kerry, 2012).

Three ICC conditions were tested, $\rho = 0.04$, $\rho = 0.17$, and $\rho = 0.38$. $\rho = 0.17$ falls within the range of commonly found ICCs described in Gulliford, Ukoumunne and Chinn (1999), while $\rho = 0.04$ and $\rho = 0.38$ were included to test whether an extreme ICC will further affect the accuracy of the estimates. The cluster random components u_{0j} and u_{1j} are independent normal variables with mean 0 and standard deviations σ_0 and σ_1, respectively. Equation 2.6 was used to calculate the variances of the random intercept, resulting in the values of 0.13, 0.67 and 2.0, corresponding to ICCs of 0.04, 0.17 and 0.38, respectively. The standard deviation σ_0 also follows from the ICC, with $\sigma_0 = 0.36$ for $\rho = 0.04$, $\sigma_0 = 0.82$ for $\rho = 0.17$ and $\sigma_0 = 1.42$ for $\rho = 0.38$. For all simulations $\sigma_1 = 1$.

The level-1 and level-2 explanatory variables were generated from the standard normal distribution, using the RANNOR(SEED) call in SAS. A Bernoulli distribution with probability

$$p_{ij} = \frac{\exp(\beta_{0j} + \beta_{1j}x_{ij})}{1 + \exp(\beta_{0j} + \beta_{1j}x_{ij})} \tag{2.7}$$

was used to generate the outcome. The prevalence of the outcome was approximately 30% in all simulations.

The number of groups was initially set to 50 clusters, with 10 subjects each. This was used as the reference point to assess relative accuracy of parameter estimates. For cluster sizes 5 through 49, the number of subjects per cluster was set to 10, 20, 30, 60, 90, 120, 150, 180, and 210. For simplicity, cluster size was not variable within a model, assuming balance design.

For the population parameter θ, let $\hat{\theta}$ be the associated parameter estimate, then $\frac{\hat{\theta}-\theta}{\theta} \times 100$ is the percentage relative bias for θ. A 95% confidence interval was created using the asymptotic standard normal distribution; therefore for parameter θ the confidence interval was $\theta \pm 1.96 * SE$. This was used to access the accuracy of the standard error of each parameter estimate.

All simulations were conducted in SAS 9.4 (SAS Institute, North Carolina, US). 1000 data sets for each combination were generated.

2.0.2 Maximum Likelihood

The SAS procedure PROC NLMIXED was used for MLE estimation. The optimization technique used was the trust region method (TRUREG), as it is well suited for small problems and considered more stable than other similar techniques (SAS/STAT(R) 9.22 User's Guide: Optimization Algorithms). The maximum number of iterations was 50, which is the default for the trust region method when using PROC NLMIXED. The integration method was Adaptive Gaussian Quadrature (AGQ), the default in PROC NLMIXED, where the number of quadrature points is selected adaptively (SAS/STAT(R) 9.22 User's Guide: The NLMIXED Procedure). The following SAS code was used:

```
proc nlmixed data=test tech=trureg;
by rep nregion n;
ods listing close;
parms g00=&g00 g10=&g10 g01=&g01 g11=&g11 s0=&s0 s1=&s1
s01=0;
p0j=g00+g01*zj+u0j;
p1j=g10+g11*zj+u1j;
eta=p0j+p1j*xij;
p = exp(eta)/(1+exp(eta));
model yij   binary(p);
random u0j u1j normal([0,0],[s0,s01,s1]) subject=region;
run;
```

2.0.3 Restricted Maximum Likelihood

To fit a REML analogue or Residual Pseudo-Likelihood model, we used PROC GLIMMIX with METHOD=RSPL, which is the same as REML when using PROC MIXED (Austin, 2010; SAS/STAT(R) 9.22 User's Guide: Comparing the GLIMMIX and MIXED Procedure). Trust region optimization was used, with default settings that are the same as those for PROC NLMIXED. The RANDOM statement invokes the random slope and the random intercept. The following SAS code was used:

```
proc glimmix data=test method=rspl;
by rep nregion n;
class rep nregion n;
model yij=xij zj xij*zj/dist=bin link=logit solution s;
random int xij/subject=region type=un gcorr ;
nloptions tech=trureg;
run;
```

2.0.4 Generalized Estimating Equations

PROC GLIMMIX was used to fit a GEE-type model. The EMPIRICAL option was used to invoke the sandwich estimator and the RANDOM statement was used to incorporate random effects (SAS/STAT(R) 9.22 User's Guide: Fitting a Marginal (GEE-type) Model). Trust region optimization, with the default number of maximum iterations of 50, was implemented. An unstructured correlation structure was specified. The following SAS code was used:

```
proc glimmix data=test empirical;
by rep nregion n;
class rep nregion n;
model yij=xij zj xij*zj/dist=bin link=logit solution s;
random int xij/subject=region type=un gcorr ;
nloptions tech=trureg;
run;
```

Chapter 3

Results

For all simulated conditions, we investigated the cluster size necessary to decrease the absolute relative bias to within 0.25% of the reference bias for each parameter. If increasing the cluster size decreased the absolute relative bias to within 0.25% of the reference bias this was deemed "Complete Compensation." In the case that complete compensation was not found we looked for a plateau point in the graph of absolute percent relative bias for each parameter and cluster size, or a point for which bias is no longer increasing or decreasing as the cluster size is increasing. If a plateau point was found this was deemed "Incomplete Compensation."

We also examined the convergence rate for various conditions and the accuracy of the standard error estimate for each estimation technique.

3.1 Maximum Likelihood

3.1.1 Convergence and Inadmissible Solutions

The overall rates of convergence for $\rho = 0.04$, $\rho = 0.17$, and $\rho = 0.38$ were 97.3%, 98.9%, 99.3%, respectively. For each ICC, the lowest convergence rate was approximately 72% which was found in models with 5 clusters and 10 subjects per cluster - the smallest sample size tested. For both $\rho = 0.17$ and $\rho = 0.38$, models with as few as 8 clusters, but large number of subjects per cluster had a 100% convergence rate. All non-converging models were excluded from the following analysis.

In the case that the estimation procedure produces a negative variance estimate, these solutions are deemed inadmissible and it is general practice to constrain these values to 0 (Maas and Hox, 2005). For all three ICC conditions only admissible solutions were produced.

3.1.2 Parameter Estimates for $\rho = 0.17$

Table 3.1 shows the percent relative bias for selected simulation conditions, when $\rho = 0.17$.

For fixed effect parameters, 87.62% of estimates fall within the guideline set by Hoogland and Boomsma (1998) of point estimates lower than 5% being acceptable. For the reference conditions of 50 groups with 10 subjects per group, the absolute relative bias was 0.24% for γ_{00}, 3.22% for γ_{01}, 0.68% for γ_{10}, and 1.18% for γ_{11}. The largest bias for each parameter was 4.7% for γ_{00}, 16.8% for γ_{01}, 23.0% for γ_{10}, and 13.8% for γ_{11}. These were all found when number of groups was less than 30 (10, 6, 28, and 13 respectively) and when group size was 20 or less. The highest bias of 23.0%, was found when the number of groups was 28 and group size was 20, but for the same number of clusters, when group size was increased to 60 the bias fell to below 11%. As shown by Table 3.1, the general trend for fixed parameters, when $\rho = 0.17$, was that for a fixed number of clusters, increasing sample size decreased the absolute relative bias.

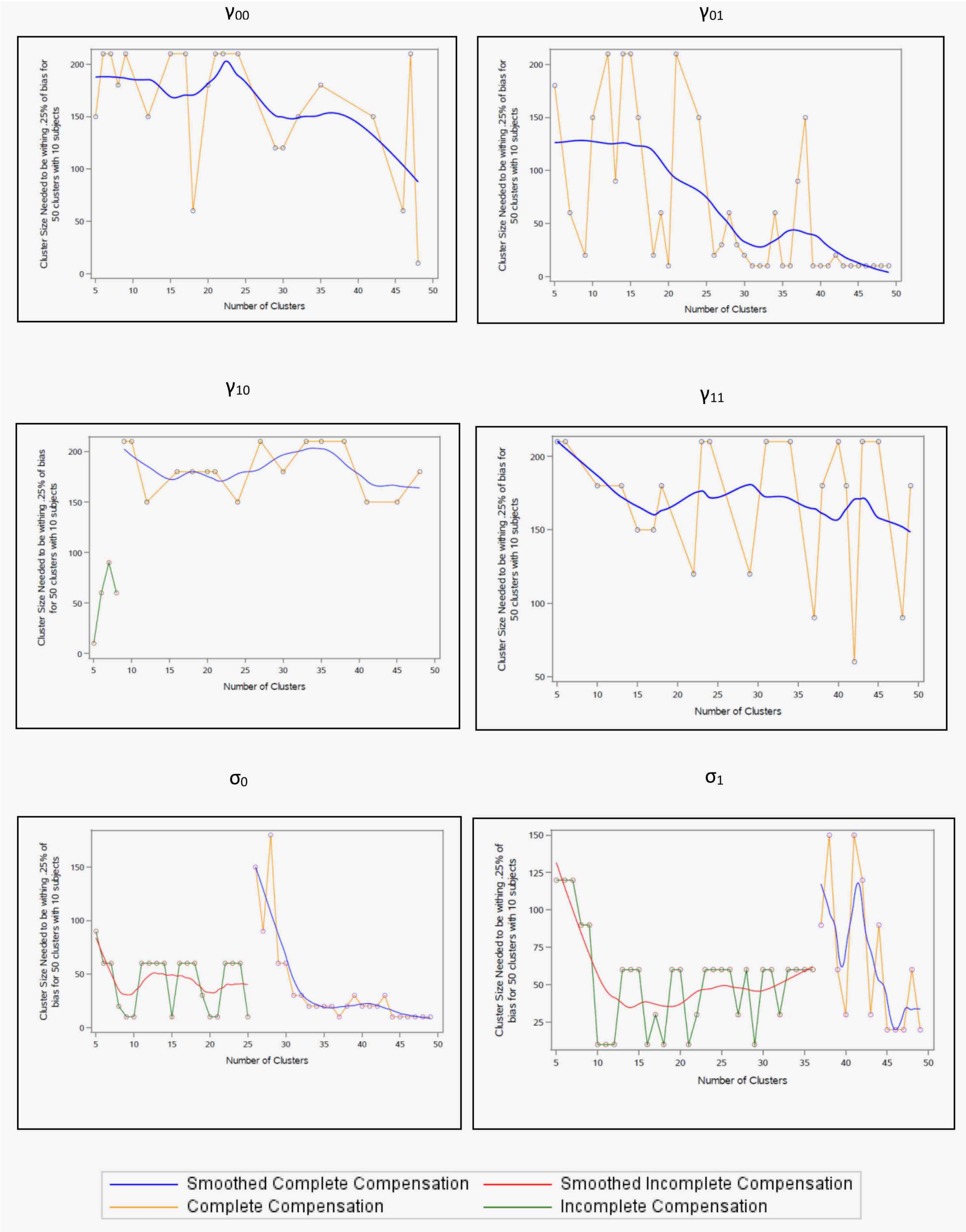
Figure 3.1 shows the necessary cluster size needed to achieve complete or incomplete compensation for each parameter and various number of clusters. For fixed effect parameters, even for a very few number of clusters, complete compensation was achieved by taking a large number of subjects per cluster. For 5 clusters, the smallest simulated, absolute relative bias within 0.25% of the reference bias was achieved for γ_{00} with group size 150, γ_{01} with group size 180, and γ_{11}

Table 3.1: Absolute percent relative bias of MLE estimates with $\rho = 0.17$

Number of Clusters	Cluster Size	% Converged	γ_{00}	γ_{10}	γ_{01}	γ_{11}	σ_0	σ_1
50	10	100	0.245	0.687	3.225	1.178	10.277	6.981
40	10	99	0.876	0.391	2.117	5.315	12.551	9.098
	20	100	0.121	1.396	1.649	5.147	8.959	8.915
	30	100	0.057	0.605	1.334	1.732	9.322	6.233
	60	100	1.566	0.916	2.071	0.477	8.177	5.639
	90	100	0.504	1.513	2.087	0.342	7.601	7.053
	120	100	0.548	1.822	2.318	0.678	6.428	6.397
	150	100	0.01	2.2	2.701	0.299	7.641	6.599
	180	100	0.483	0.002	0.998	2.209	6.6	5.899
	210	100	0.798	2.364	1.201	1.177	6.792	6.025
30	10	99	0.236	2.474	3.832	0.573	15.945	14.965
	20	100	1.169	1.017	2.736	3.188	12.567	12.719
	30	100	0.611	2.196	1.983	2.049	12.946	9.867
	60	100	0.424	0.8	0.589	6.097	9.206	7.238
	90	100	0.91	3.494	1.256	3.7	8.537	8.084
	120	100	0.089	1.361	0.186	4.09	10.271	8.003
	150	100	0.046	1.135	1.15	3.272	9.893	7.876
	180	100	0.386	0.173	1.259	4.28	9.236	8.246
	210	100	0.197	0.766	0.791	4.766	7.744	7.499
20	10	96	0.612	2.158	0.584	6.608	17.406	16.143
	20	100	0.377	2.138	0.079	0.451	18.543	13.685
	30	100	0.883	3.025	0.414	2.896	17.031	13.962
	60	100	0.271	0.519	1.056	6.841	14.935	9.943
	90	100	0.329	1.163	0.482	0.728	13.866	11.662
	120	100	0.015	0.919	0.696	0.309	14.695	13.125
	150	100	0.597	1.055	1.095	2.339	14.274	12.518
	180	100	0.081	0.591	0.551	5.094	14.439	13.541
	210	100	0.163	0.09	0.739	2.663	14.858	12.087
10	10	87	4.735	2.121	12.194	3.437	29.407	24.544
	20	96	2.262	7.939	8.593	2.378	29.93	23.907
	30	98	1.452	3.721	5.671	2.942	26.164	24.687
	60	100	1.444	1.283	2.034	1.908	30.546	25.849
	90	100	1.712	1.792	3.368	0.037	26.864	25.743
	110	100	1.358	1.746	0.387	2.205	26.54	28.121
	120	100	1.611	2.999	4.996	0.623	25.596	27.081
	150	100	1.386	0.575	3.357	3.516	24.268	24.36
	180	100	1.162	1.049	0.939	0.003	25.247	26.097
	210	100	0.837	0.396	2.417	1.052	26.374	25.811
5	10	72	1.161	2.579	6.414	3.046	36.467	23.819
	20	80	1.379	1.958	5.87	5.98	40.23	25.98
	30	85	1.071	3.781	0.784	1.942	35.289	27.072
	60	90	0.367	0.062	0.199	4.654	45.119	38.783
	90	93	0.346	3.966	2.782	1.141	46.272	41.775
	110	94	0.598	9.633	2.342	1.141	48.434	47.914
	120	95	0.911	2.61	0.727	0.621	46.03	48.055
	150	96	0.298	4.272	3.733	2.21	45.991	48.178
	180	95	0.486	11.215	0.543	4.314	48.524	49.374
	210	96	0.325	6.222	1.103	0.605	49.343	48.938

with group size 210. For γ_{10}, with 5 groups, only incomplete compensation was achieved, but for 9 groups and 210 subjects, complete compensation was achieved.

Figure 3.1: Number of clusters needed for compensation, MLE with $\rho = 0.17$



For the random variance components, overall much larger absolute relative biases were found. Only 1.46% were lower than 5%, across all number of clusters and group sizes. The smallest biases were found for a large number of groups and a large number of subjects per cluster. With 50 clusters and 10 subjects per cluster, the absolute relative bias for σ_0 was 10.28% and for σ_1 was 6.98%. As shown in Figure 3.1, for the random intercept variance, complete compensation was only achieved for 26 or more clusters. For the random slope variance, 37 or more clusters were needed for complete compensation.

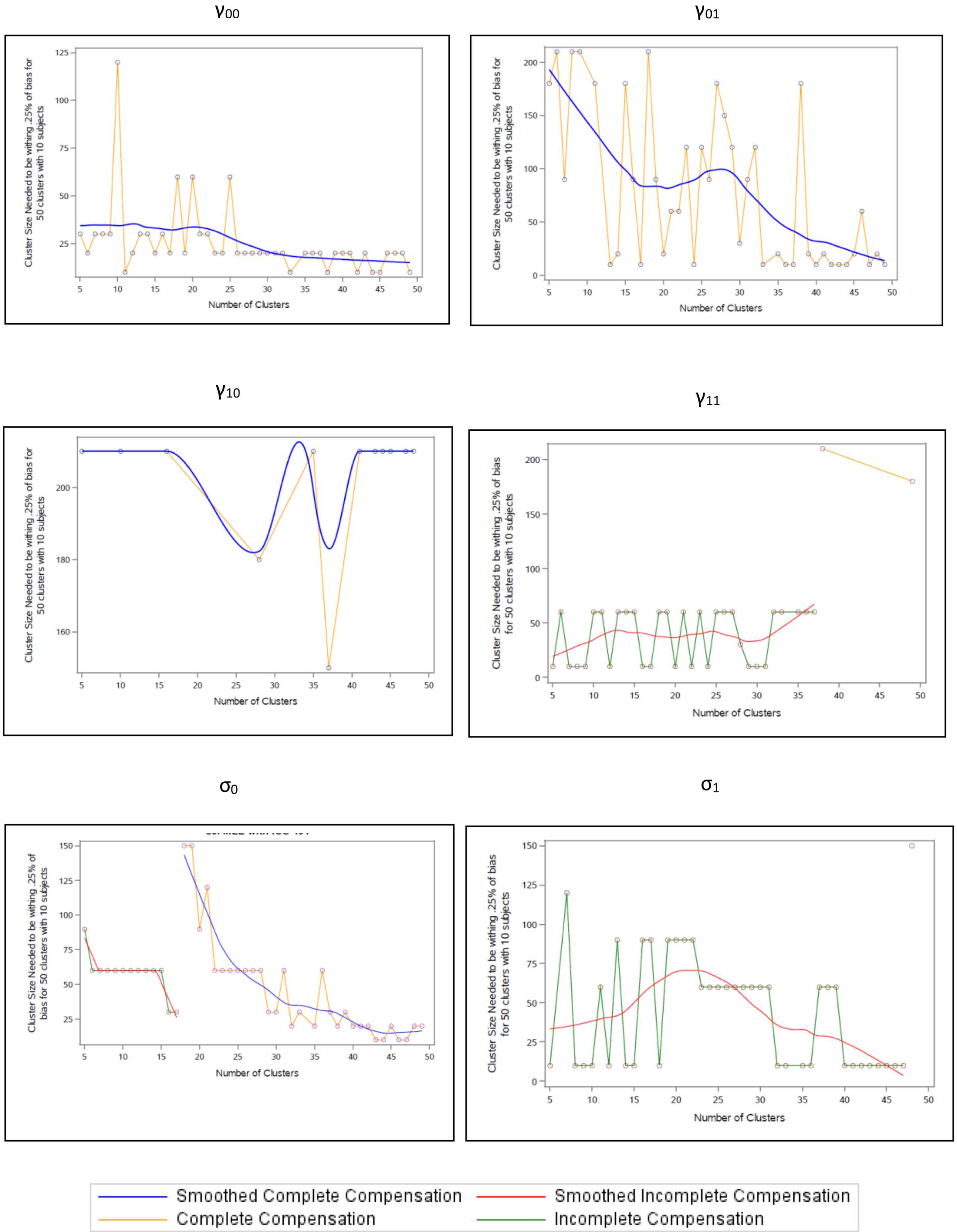
Incomplete compensation was achieved for even 5 clusters with 90 subjects per cluster for both random variance components. Although a plateau was found, the highest absolute relative biases, those greater than 30%, were found in simulations with less than 10 clusters. For a fixed number of subjects per cluster decreasing the number of clusters resulted in an increase in bias. When the number of clusters was fixed, the trend for both variance components was that the bias decreased as cluster size increased. The exceptions to this were in conditions where very few clusters were used, such as 5 clusters, where increasing cluster size actually resulted in an increase in bias that eventually plateaued.

3.1.3 Parameter Estimates for $\rho = 0.04$

For $\rho = 0.04$, 93.53% of fixed parameter estimates and 2.22% of random parameter estimates had percent relative bias less than 5%.

For fixed parameter estimates, the reference bias was 0.9% for γ_{00}, 1.9% for γ_{01}, 0.3% for γ_{10}, and 0.1% for γ_{11}. For all except γ_{11}, complete compensation was found with as few as 5 clusters, shown in Figure 3.2. For the interaction term γ_{11}, complete compensation was only achieved with 38 clusters, with 210 subjects per cluster and 49 clusters and 180 subjects. The largest bias for each parameter was 5.1% γ_{00}, 12.3% for γ_{01}, 12.5% for γ_{10}, and 10.5% for γ_{11}.

Figure 3.2: Number of clusters needed for compensation, MLE with $\rho = 0.04$



For the random slope variance and the random intercept variance, for most conditions increasing the number of subjects per cluster resulted in an increase in bias, shown in Table 3.2. This resulted in complete compensation not being found for most clusters sizes, shown in Figure 3.2. For σ_0, complete compensation was achieved with 18 clusters with 150 subjects per cluster and for cluster sizes greater

Table 3.2: Absolute percent relative bias of MLE estimates with $\rho = 0.04$

Number of Clusters	Cluster Size	% Converged	γ_{00}	γ_{10}	γ_{01}	γ_{11}	σ_0	σ_1
50	10	96	0.875	0.32	1.891	0.045	20.191	4.805
40	10	95	1.643	0.773	1.982	0.02	21.698	6.95
	20	99	0.21	2.354	1.431	1.865	19.747	6.084
	30	100	0.322	1.46	0.767	3.219	14.329	6.5
	60	100	0.324	0.749	0.347	2.572	9.775	8.292
	90	100	0.388	0.634	0.16	2.78	10.598	7.886
	120	100	0.076	2.665	0.52	0.254	8.456	6.777
	150	100	0.079	1.209	0.178	2.352	8.921	7.349
	180	100	0.399	1.701	0.293	0.695	9.423	7.889
	210	100	0.406	1.718	1.177	0.537	5.426	6.227
30	10	92	1.959	1.218	3.801	5.647	32.664	5.416
	20	98	0.819	0.466	2.862	5.165	22.617	8.016
	30	99	0.603	1.426	1.181	5.434	17.62	7.724
	60	100	0.902	4.309	0.533	4.101	12.902	9.814
	90	100	0.502	1.793	1.289	2.738	13.652	8.97
	120	100	0.534	1.998	1.251	1.06	10.734	9.063
	150	100	0.622	0.583	1.249	2.054	11.445	9.502
	180	100	0.475	1.563	0.423	1.578	10.79	9.134
	210	100	0.57	2.647	0.82	2.353	10.039	8.125
20	10	90	2.273	1.36	3.141	1.43	29.961	6.224
	20	95	0.026	2.074	1.709	4.142	28.966	7.414
	30	97	1.567	1.652	1.137	1.098	29.194	8.208
	60	99	0.203	1.339	0.74	3.582	22.064	12.535
	90	100	0.466	1.571	0.095	2.07	18.252	14.934
	120	100	0.146	2.807	0.421	1.56	15.939	14.078
	150	100	0.085	3.203	1.459	0.087	16.336	12.662
	180	100	0.568	4.475	0.698	2.858	17.725	14.111
	210	100	0.028	0.801	0.941	2.52	14.342	13.324
10	10	88	2.87	2.923	10.339	10.515	266.263	24.56
	20	96	1.275	2.922	7.127	7.038	252.98	24.662
	30	97	0.967	0.312	1.522	1.35	249.318	25.559
	60	100	0.268	3.067	4.49	8.389	260.082	24.878
	90	100	1.653	1.318	0.582	4.348	269.97	25.399
	120	100	0.948	0.7	6.342	0.903	263.422	25.358
	150	100	0.422	1.394	4.496	8.513	269.431	25.93
	180	100	0.142	2.063	5.764	3.921	277.166	25.446
	210	100	0.272	0.194	6.083	6.375	274.484	24.715
5	10	73	3.567	0.142	3.286	1.614	46.881	5.79
	20	75	1.72	0.292	3.462	8.005	51.978	6.181
	30	80	1.081	0.43	1.493	1.343	51.518	6.015
	60	79	0.417	3.339	0.022	3.457	60.829	8.48
	90	81	0.37	0.561	0.17	1.362	58.173	15.732
	110	85	0.675	6.764	0.781	1.995	59.42	22.681
	120	84	0.219	0.971	1.142	2.483	55.809	21.199
	150	87	0.015	2.181	3.191	0.489	50.599	15.292
	180	90	0.98	1.249	1.513	2.023	59.489	22.302
	210	91	0.457	0.408	0.212	0.82	52.955	28.69

than 18. For σ_1, complete compensation was only achieved for 48 clusters and 150 subjects per cluster. The absolute relative biases were more extreme for $\rho = 0.04$, with the highest bias for σ_0 and σ_1 being 297.9% and 32.5%, respectively. For σ_0, this is much higher than the highest bias found when $\rho = 0.17$, of 49.3%. These results show that in the case of a small ICC, increasing sample size may actually be detrimental to the accuracy of the variance components when using MLE.

3.1.4 Parameter Estimates for $\rho = 0.38$

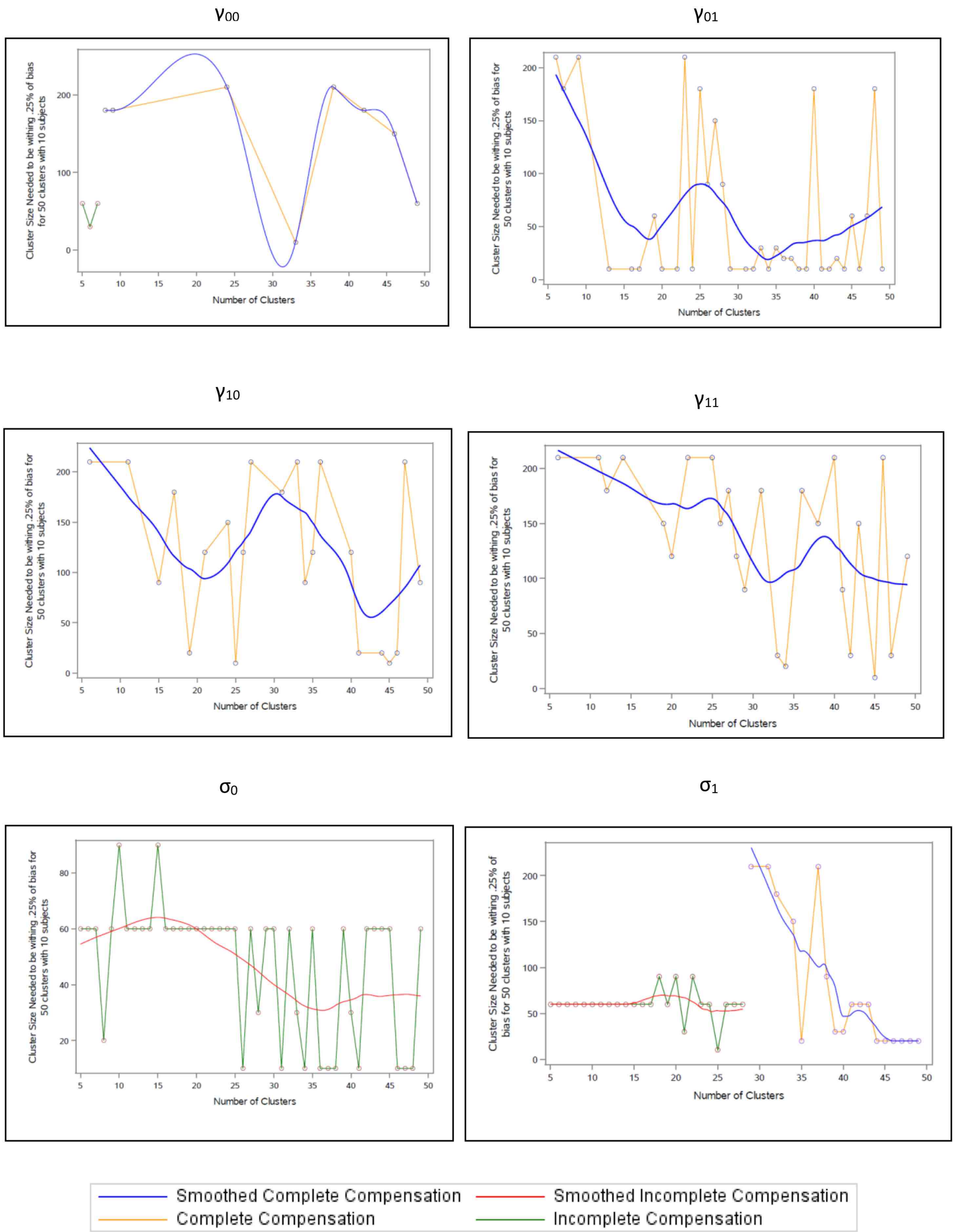
For $\rho = 0.38$, 86.5% of fixed parameter estimates and 2.96% of random parameter estimates had absolute percent relative bias less than 5%. The reference bias was 0.3% for γ_{00}, 4.7% for γ_{01}, 1.6% for γ_{10}, 1.8% for γ_{11}, 4.4% for σ_0 and 7.6% for σ_1. For γ_{00} complete compensation was achieved with as few as 8 clusters and for all other fixed effect parameter estimates it was achieved with as few as 6 by using 210 subjects per cluster, see Figure 3.3. The largest biases for fixed parameter estimates were 4.8%, 18.0%, 13.0%, and 17.1%, for γ_{00}, γ_{01}, γ_{10}, γ_{11}, were again found in conditions with 10 or fewer clusters, as shown in Table 3.3.

The trends seen when $\rho = 0.38$ are similar those seen when $\rho = 0.17$. For the variance components: as number of clusters decreased, bias increased. Increasing the number of subjects per cluster either decreased the bias or resulted in a plateau. Again, the exception to this was when less than 10 clusters were used. In this case the bias initially increased and then plateaued as sample size increased. For σ_1, complete compensation was initially achieved when 29 clusters with 210 subjects per cluster were used. For σ_0, only incomplete compensation was found. The highest absolute bias for the random components were found in conditions with the fewest clusters, with the highest bias for σ_0 and σ_1 being 50.2% and 52.1%, respectively.

Table 3.3: Absolute percent relative bias of MLE estimates with $\rho = 0.38$

Number of Clusters	Cluster Size	% Converged	γ_{00}	γ_{10}	γ_{01}	γ_{11}	σ_0	σ_1
50	10	100	0.339	1.574	4.729	1.757	4.433	7.614
40	10	100	0.761	0.175	3.724	1.936	9.484	8.869
	20	100	0.208	0.602	4.312	3.849	7.91	8.94
	30	100	0.91	1.86	2.817	1.237	8.329	6.995
	60	100	0.963	0.573	0.923	0.353	6.753	6.396
	90	100	0.136	2.54	4.781	0.311	7.417	6.56
	120	100	0.015	1.438	2.381	0.466	7.113	6.569
	150	100	0.053	1.299	5.64	2.817	7.092	6.328
	180	100	0.594	1.145	4.092	2.404	6.151	5.116
	210	100	1.208	1.134	3.655	1.286	7.18	5.979
30	10	100	0.615	3.898	5.214	2.383	9.65	14.652
	20	100	0.493	2.264	0.202	1.57	8.42	10.112
	30	100	0.792	2.685	2.972	0.314	7.841	10.682
	60	100	1.225	1.481	4.858	0.118	7.401	8.278
	90	100	0.625	0.363	3.425	0.328	7.452	9.243
	120	100	0.755	0.218	3.133	3.305	8.127	9.52
	150	100	1.138	1.753	3.558	3.964	7.043	8.741
	180	100	1.799	1.277	4.293	1.39	6.373	9.845
	210	100	1.028	2.013	5.726	2.087	8.003	8.576
20	10	98	3.182	3.608	2.722	7.496	18.316	20.866
	20	100	1.001	3.457	4.625	5.182	13.212	14.766
	30	100	1.453	0.071	0.703	2.288	13.005	17.304
	60	100	0.351	0.694	0.499	0.195	12.552	12.159
	90	100	1.082	1.145	2.831	3.287	12.472	13.269
	120	100	0.043	2.004	1.436	0.466	11.68	14.576
	150	100	0.524	0.711	1.005	0.44	12.457	12.988
	180	100	1.398	1.908	4.649	0.196	13.06	12.764
	210	100	0.839	2.093	1.892	0.291	12.954	12.643
10	10	91	0.6	3.155	7.66	5.314	33.91	38.977
	20	97	1.337	1.459	13.977	0.496	28.516	31.578
	30	99	0.411	0.213	17.96	6.527	29.75	30.055
	60	100	3.05	4.507	9.384	5.693	25.353	27.609
	90	100	1.65	6.236	14.349	0.349	25.867	27.574
	120	100	0.482	3.657	12.044	0.752	25.539	27.157
	150	100	2.778	4.585	11.631	2.064	23.701	27.913
	180	100	3.397	2.804	8.632	2.202	26.461	27.517
	210	100	2.756	2.636	11.892	3.519	25.623	26.868
5	10	72	2.712	4.527	4.546	8.134	42.229	43.837
	20	81	1.822	10.064	0.524	10.138	41.895	48.23
	30	86	0.457	5.749	10.738	2.363	43.309	50.238
	60	92	0.584	5.151	5.988	4.554	44.965	49.793
	90	94	0.396	11.645	6.253	7.343	44.413	52.072
	120	96	0.815	1.708	7.226	6.391	46.795	51.92
	150	97	0.261	4.029	9.157	7.891	49.11	50.75
	180	96	0.775	1.936	2.003	5.381	50.244	50.342
	210	97	2.541	7.948	6.38	2.604	49.073	50.701

Figure 3.3: Number of clusters needed for compensation, MLE with $\rho = 0.38$



3.1.5 Standard Errors

Table 3.4: Noncoverage of the 95% CI by ICC for MLE

Parameter	ICC		
	0.04	0.17	0.38
γ_{00}	0.094	0.089	0.084
γ_{01}	0.094	0.088	0.086
γ_{10}	0.097	0.088	0.084
γ_{11}	0.095	0.088	0.084
σ_0	0.188	0.17	0.165
σ_1	0.167	0.169	0.121

Table 3.5: Noncoverage of the 95% CI by number of clusters for MLE

Parameter	Number of Clusters				
	5	10	20	30	40
γ_{00}	0.24	0.125	0.086	0.073	0.061
γ_{01}	0.231	0.125	0.083	0.069	0.072
γ_{10}	0.239	0.119	0.078	0.076	0.069
γ_{11}	0.246	0.12	0.077	0.07	0.064
σ_0	0.459	0.181	0.174	0.138	0.115
σ_1	0.443	0.253	0.164	0.128	0.108

Table 3.6: Noncoverage of the 95% CI by cluster size for MLE

Parameter	Group Size								
	10	20	30	60	90	120	150	180	210
γ_{00}	0.119	0.092	0.084	0.084	0.084	0.084	0.083	0.083	0.083
γ_{01}	0.118	0.093	0.082	0.085	0.082	0.084	0.083	0.085	0.084
γ_{10}	0.118	0.092	0.087	0.085	0.083	0.094	0.083	0.083	0.084
γ_{11}	0.117	0.093	0.084	0.085	0.085	0.083	0.082	0.083	0.082
σ_0	0.209	0.183	0.165	0.165	0.165	0.17	0.127	0.167	0.167
σ_1	0.172	0.155	0.155	0.164	0.163	0.166	0.167	0.17	0.168

The coverage of a 95% Wald confidence interval was used to assess the accuracy of standard error estimates. Whether or not the confidence interval contained the true parameter was assessed for each parameter. The nominal non-coverage rate is 5%. Table 3.5 shows the effect of several number of clusters on the noncoverage of the 95% confidence interval. Table 3.6 shows the effect of cluster size on the noncoverage and the effects of ICC on noncoverage are shown in Table 3.4.

Table 3.5 shows that for the fixed effect parameter estimates, number of groups can impact the accuracy of the standard errors. For 40 groups, the noncoverage rate for all fixed parameter estimates was 6.1%, but with 5 groups it was 24%. The noncoverage rates are even greater for variance components. From this we can see that when few clusters are used the 95% confidence interval is too narrow, indicating that the estimated standard errors are too small (Maas and Hox, 2005).

Table 3.6 shows that standard error estimates are more accurate for a larger number of subjects per cluster, but the difference is less pronounced. For σ_1, with both 10 subjects per group and with 180 subjects per group, the noncoverage rate was 17% for both. Similarly to Table 3.5 it can be seen that the noncoverage rate is higher for the variances components than the fixed effect parameter estimates.

Table 3.4 again shows that the coverage rates are better for fixed effect parameter estimates.

A 95% confidence interval coverage of 92.5% or less shows that the standard errors were underestimated and a coverage of 97.5% or greater shows that the standard errors were overestimated (Bradley, 1978). The findings shown here indicate that for both fixed effect parameter estimates and the variances, the standard errors were too small.

3.2 Restricted Maximum Likelihood

3.2.1 Convergence

For REML with $\rho = 0.17$, the convergence rate for all models was 99.34%. When $\rho = 0.04$ the convergence rate was 97% and when $\rho = 0.38$ there was 99.44% convergence of models. Lower convergence rates were found in models with a small total sample size. Nonconverging models were excluded from the following analysis.

PROC GLIMMIX applies a lower boundary of 0 to all variance components, so there were no inadmissible solutions (Kiernan, Tao, and Gibs, 2009).

3.2.2 Parameter Estimates for $\rho = 0.17$

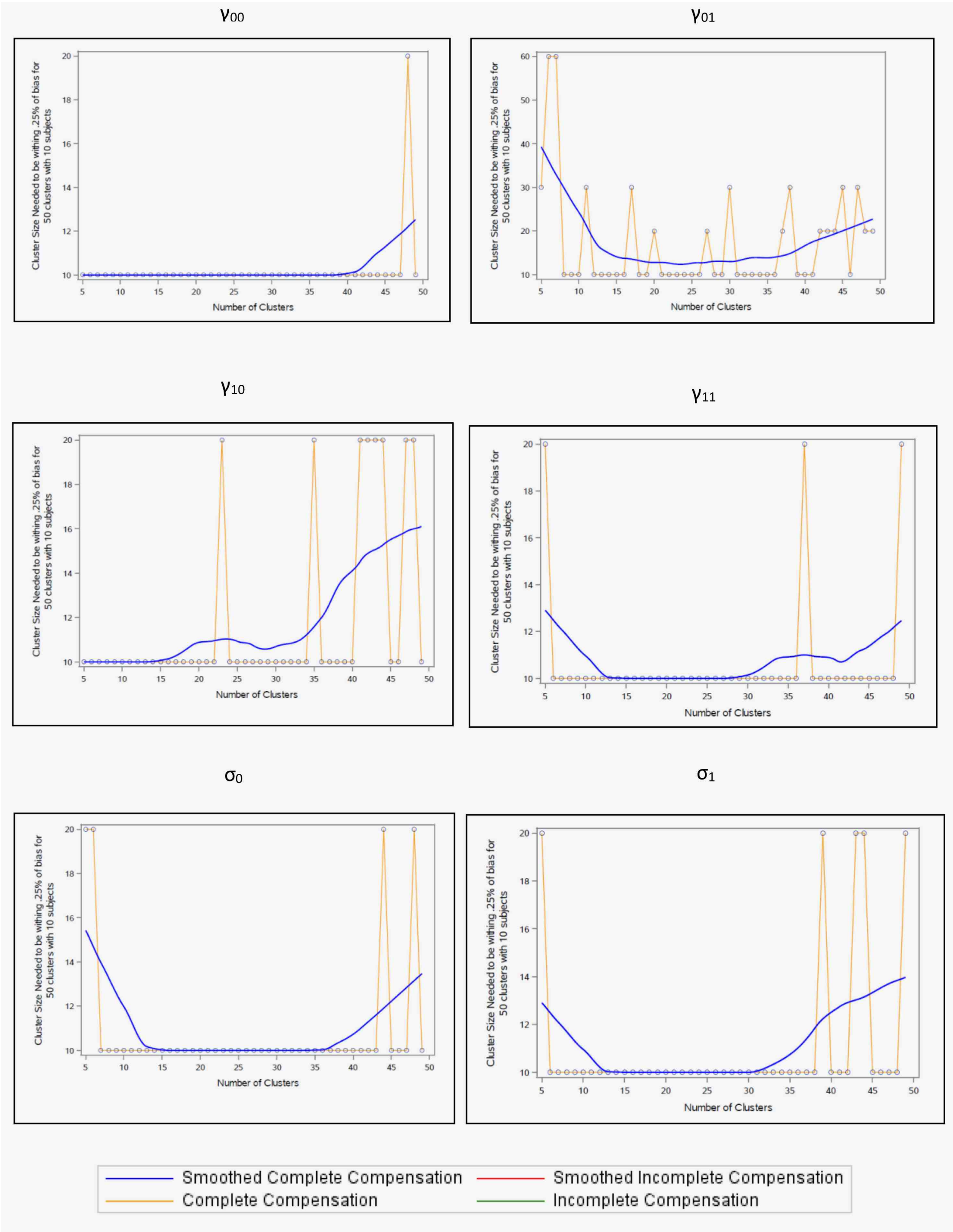
About 61% of fixed effect parameter estimates were below 5% for $\rho = 0.17$. For the reference simulation, the average absolute relative bias for γ_{00} was 12.21%, for γ_{01} was 10.71%, for γ_{10} was 19.22%, and for γ_{11} was 16.10%. The highest absolute percent relative bias found for each fixed parameter estimate was as follows: 12.53% for γ_{00} with 48 clusters and 10 subjects per cluster, 21.87% for γ_{10} with 35 clusters and 10 subjects per cluster, 20.45% for γ_{01} with 5 clusters and 20 subjects per group, and 36.07% for γ_{11} with 5 clusters and 10 subjects per group. Low absolute relative biases were found in conditions with a large number of subjects per cluster, regardless of the number of clusters, as shown by Table 3.7. As shown in Figure 3.4, complete compensation was achieved for each fixed effect parameter with a low number of subjects per cluster.

For the random effect variance components, only 7.02% of estimates had absolute relative bias less than 5%. The reference bias for σ_0 was 34.91% and 43.03% for σ_1. For a fixed number of subjects per clusters, 10, decreasing the number of clusters resulted in an decrease in bias until the number of clusters reached 10, where the bias then increased as clusters decreased. Increasing the sample size within each cluster resulted in an even further decrease in the bias. The lowest absolute percent relative bias for σ_0 was 0.33% found for 6 clusters and 20 subjects per cluster. For σ_1 the lowest bias, 0.77%, was found for 5 clusters, with 30 subjects per cluster. However, in general, smaller bias was found in conditions with a large total sample size (refer to Table 3.7). Complete compensation was achieved for all number of clusters simulated with a cluster size of twenty or less,

Table 3.7: Absolute percent relative bias of REML estimates with $\rho = 0.17$

Number of Clusters	Cluster Size	% Converged	γ_{00}	γ_{10}	γ_{01}	γ_{11}	σ_0	σ_1
50	10	100	12.209	19.216	10.702	16.095	34.907	43.028
40	10	100	11.249	18.091	10.731	16.73	33.886	42.949
	20	100	8.371	9.431	10.174	10.174	21.433	27.033
	30	100	5.768	7.223	6.875	8.55	16.905	21.059
	60	100	3.363	6.044	3.796	2.261	10.413	13.794
	90	100	2.532	5.005	3.043	0.559	7.505	10.085
	120	100	1.858	3.263	3.047	2.201	5.985	7.976
	150	100	1.271	4.134	2.431	1.566	5.696	7.486
	180	100	0.997	1.249	2.441	0.466	4.288	6.828
	210	100	0.269	2.893	2.796	0.531	2.914	4.853
30	10	99	8.764	17.159	10.298	12.005	31.496	41.268
	20	100	6.78	10.957	13.085	10.66	20.683	27.758
	30	100	5.187	6.674	8.966	4.844	14.882	20.231
	60	100	2.335	10.164	6.889	0.943	9.799	12.128
	90	100	1.424	6.464	4.733	0.699	6.879	10.184
	120	100	0.648	4.262	3.817	0.61	6.961	8.2
	150	100	0.066	3.369	4.865	0.149	5.39	7.508
	180	100	0.329	3.13	4.098	0.755	3.563	6.27
	210	100	0.468	5.172	4.02	0.381	3.687	5.617
20	10	98	8.84	16.331	11.363	12.662	27.564	33.07
	20	100	6.791	6.763	7.716	13.903	21.307	28.172
	30	100	4.257	8.066	4.752	5.202	17.228	20.468
	60	100	3.554	5.202	3.582	6.057	10.612	11.679
	90	100	2.194	4.31	3.198	0.158	6.616	9.051
	120	100	2.301	1.302	3.238	3.919	6.208	9.184
	150	100	1.916	1.276	0.678	0.425	6.352	6.446
	180	100	0.778	1.082	2.904	2.374	7.421	6.958
	210	100	0.416	2.197	4.525	2.497	5.694	7.014
10	10	90	6.915	17.336	1.028	14.225	11.392	14.702
	20	97	4.193	6.686	1.52	9.903	13.828	17.637
	30	99	2.683	4.628	8.749	7.074	15.986	15.271
	60	100	0.721	0.187	1.357	8.585	13.279	13.288
	90	100	1.006	4.965	3.653	9.009	11.981	11.353
	120	100	1.371	2.16	3.137	6.69	13.78	11.361
	150	100	0.46	1.228	4.979	5.356	9.265	12.805
	180	100	0.579	2.94	4.713	4.11	9.66	10.205
	210	100	0.215	2.46	5.624	1.832	9.409	8.574
5	10	63	11.341	6.464	16.391	36.073	44.721	46.661
	20	81	9.068	11.295	20.451	13.029	17.141	12.554
	30	86	3.775	8.176	1.881	2.697	6.698	0.775
	60	93	0.903	7.184	2.898	3.688	12.437	12.795
	90	95	1.931	8.216	1.599	1.58	14.931	14.14
	120	96	2.631	4.197	2.38	0.647	17.78	19.861
	150	97	0.685	3.695	1.867	1.969	17.661	16.753
	180	97	1.492	2.608	5.367	3.566	21.654	18.502
	210	98	2.719	3.184	4.576	3.829	20.789	18.777

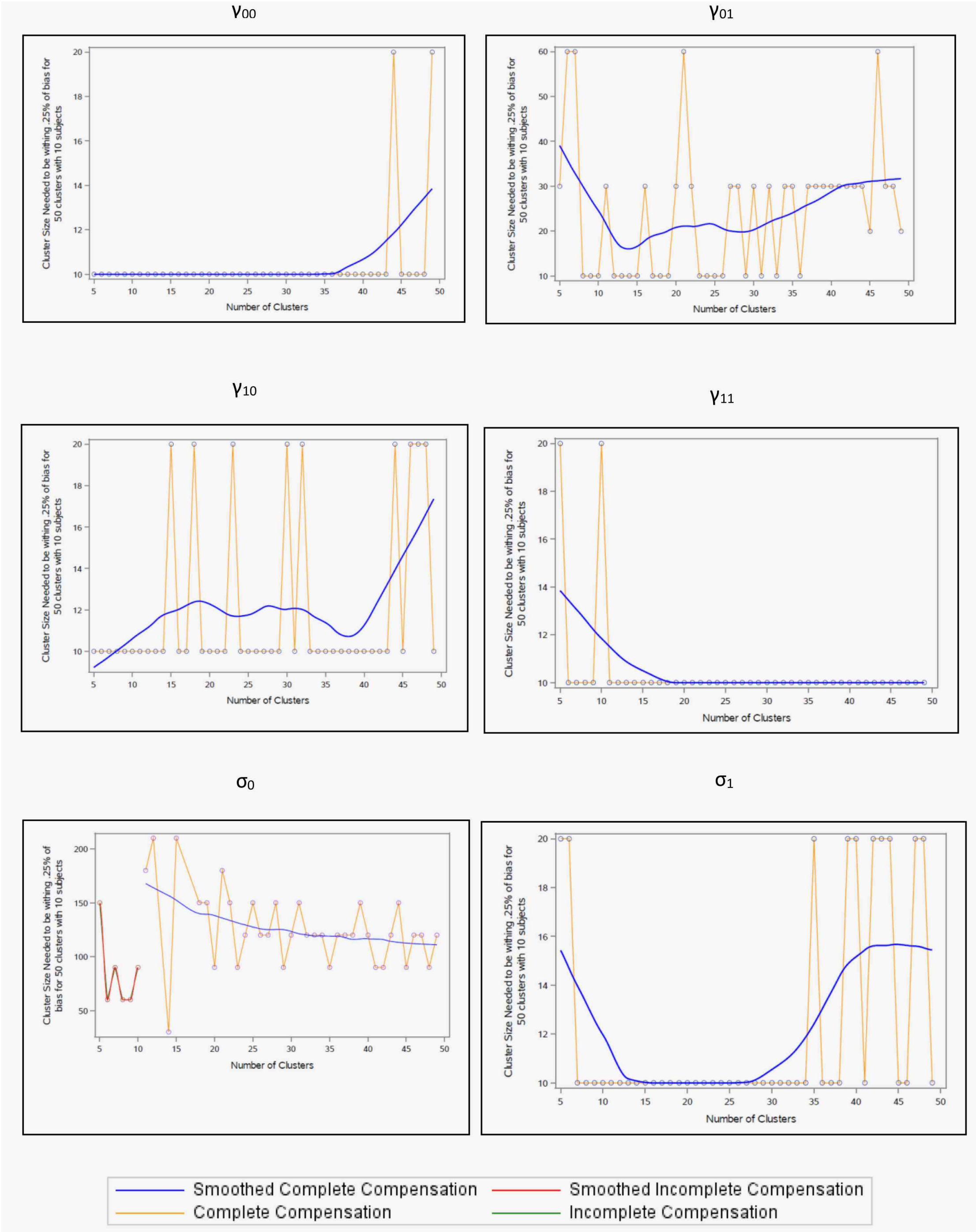
Figure 3.4: Number of clusters needed for compensation, REML with $\rho = 0.17$



most likely due to the high biases found for the reference conditions.

3.2.3 Parameter Estimates for $\rho = 0.04$

Figure 3.5: Number of clusters needed for compensation, REML with $\rho = 0.04$



When the ICC was small, the trend for all REML estimates was similar to those seen when $\rho = 0.17$, but with slightly lower convergence rates. 73.5% of fixed effect estimates and 16.4% of random effect variance estimates had bias less

Table 3.8: Absolute percent relative bias of REML estimates with $\rho = 0.04$

Number of Clusters	Cluster Size	% Converged	γ_{00}	γ_{10}	γ_{01}	γ_{11}	σ_0	σ_1
50	10	95	6.774	15.793	5.566	13.724	7.136	33.808
40	10	95	5.895	13.24	6.838	12.31	6.934	34.341
	20	98	5.038	7.347	6.702	6.924	18.78	22.797
	30	99	4.086	5.674	5.683	6.005	13.811	16.634
	60	99	2.367	4.63	2.604	3.555	8.665	12.139
	90	97	1.558	3.647	2.65	1.073	7.419	9.373
	120	97	1.481	0.254	1.037	3.176	4.895	7.588
	150	98	1.295	3.471	1.826	0.045	4.964	6.953
	180	97	0.902	0.746	1.675	2.038	4.293	6.943
	210	97	0.562	1.825	2.64	2.012	0.714	5.5
30	10	94	4.534	16.224	5.59	8.611	0.902	30.839
	20	98	4.246	9.526	8.235	3.764	10.916	22.39
	30	99	3.24	8.727	5.677	2.068	11.992	16.029
	60	99	1.826	9.26	2.968	1.147	7.537	11.758
	90	99	1.533	5.045	3.118	0.014	8.754	8.993
	120	98	0.997	5.42	2.904	1.681	4.635	7.135
	150	98	0.798	2.994	2.931	0.632	5.035	6.974
	180	98	0.739	3.428	0.827	0.15	4.16	6.306
	210	98	0.534	4.259	2.131	0.295	3.483	5.24
20	10	92	4.983	10.988	0.716	8.953	41.748	26.902
	20	96	4.776	7.575	6.488	11.838	4.089	19.603
	30	98	2.307	8.903	4.352	8.245	8.43	14.942
	60	99	2.418	3.649	3.296	7.66	9.718	11.166
	90	99	1.296	2.564	2.47	2.61	7.15	10.316
	120	99	1.653	0.471	0.889	4.415	5.592	8.769
	150	99	1.375	1.346	0.306	1.443	5.721	6.094
	180	98	0.457	2.935	0.695	4.332	6.485	8.311
	210	99	0.981	0.035	1.374	3.528	3.52	6.699
10	10	83	3.594	11.173	1.753	16.138	103.547	3.658
	20	92	1.92	9.106	2.15	7.213	36.787	12.74
	30	95	0.507	2.884	0.512	6.333	21.952	11.595
	60	98	1.482	4.887	0.747	6.226	0.331	14.715
	90	99	1.039	7.193	0.367	7.957	7.171	10.414
	120	99	1.444	4.303	0.931	5.773	9.205	11.523
	150	99	0.327	4.695	2.357	2.149	7.475	9.732
	180	99	0.313	2.989	3.869	5.905	6.252	10.165
	210	99	0.003	2.461	3.984	7.124	9.516	8.755
5	10	56	0.21	2.219	5.691	19.825	361.119	53.543
	20	73	1.348	8.761	8.954	2.509	171.745	5.885
	30	80	2.119	7.349	2.925	13.798	78.527	0.481
	60	87	0.356	1.408	2.472	5.049	22.521	13.207
	90	90	1.042	14.623	1.953	5.646	5.124	17.705
	120	91	1.606	3.719	3.115	0.286	6.289	19.812
	150	92	1.029	6.832	0.421	4.158	0.171	17.878
	180	93	0.121	4.921	2.403	2.198	11.019	17.486
	210	94	1.676	2.628	0.663	1.66	9.761	17.029

than 5%.

For fixed effect parameter estimates, biases tended to decrease as cluster size increased. The highest biases were found when there were 10 subjects per cluster, regardless of the number of clusters. Similar to REML estimates with $\rho = 0.17$ simulations, the reference bias was comparatively high for fixed effect parameter estimates, resulting in complete compensation being found for most simulated conditions.

For σ_0, bias increased as number of clusters decreased. For a fixed number of clusters the bias for σ_0 decreased as sample size increased, except in the cases of very few clusters (10 or less). For σ_1, the general trend seen is that as cluster size increased the bias decreased, again with the exception of 10 or less clusters. The highest bias for both σ_0 and σ_1 were for 5 clusters with 10 subjects per cluster. In such cases, an eventual plateau was seen in the bias, so it was designated as incomplete compensation. Complete compensation was found for σ_0 with 11 clusters and 180 subjects per cluster. For σ_1 complete compensation was found with 5 clusters and 20 subjects per cluster.

3.2.4 Parameter Estimates for $\rho = 0.38$

For a large ICC, we see similar trends as when using REML estimation with $\rho = 0.17$, but with slightly higher convergence rates. Approximately 53% of fixed effect estimates and 3.45% of random effect estimates were within the acceptable limit of 5%. For fixed effect parameter estimates, the bias tended to be highest when the number of subjects was equal to ten and decreased as the cluster size increased. Similar to REML estimates with smaller ICCs, the lowest biases were found when the number of subjects per cluster was large. The relatively high bias found for the reference bias, resulted in compete compensation being found for all fixed effect estimates with a small number of subjects per cluster.

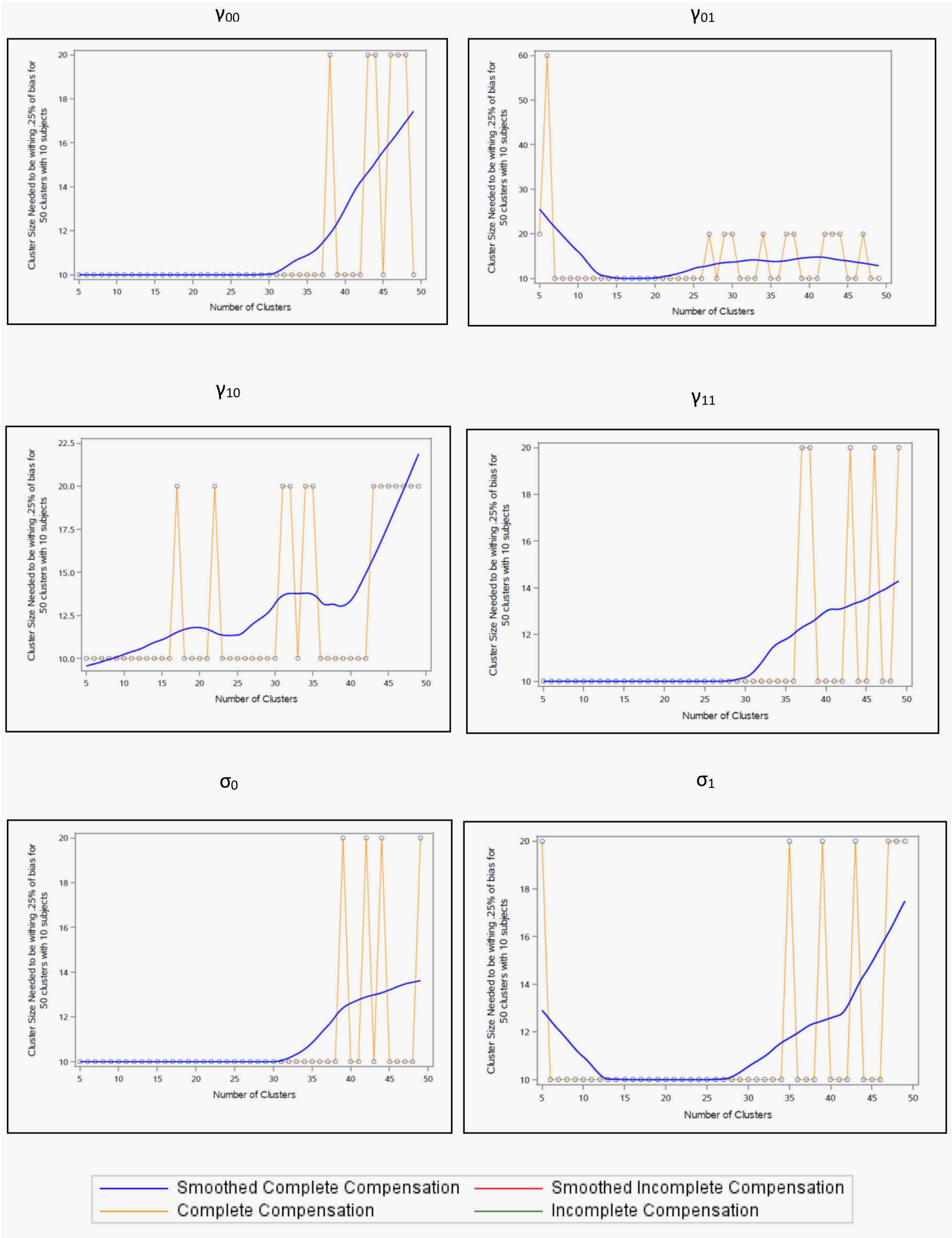
The trend towards high bias when the number of subjects per cluster was also found for the random variance components. It can be seen in Table 3.9

that increasing the sample size within each cluster resulted in a decrease in bias. As with REML estimates with small ICCs, for 10 subjects per cluster the bias decreased as number of clusters decreased. With fewer than 10 clusters, the bias increased again with cluster size fixed. High biases were found with a low number of subjects per cluster, regardless of number of clusters. Complete compensation was found for both variances components for all number of clusters simulated, which is shown in Figure 3.6.

Table 3.9: Absolute percent relative bias of REML estimates with $\rho = 0.38$

Number of Clusters	Cluster Size	% Converged	γ_{00}	γ_{10}	γ_{01}	γ_{11}	σ_0	σ_1
50	10	100	16.793	21.732	16.008	19.899	37.62	51.335
40	10	99	15.451	20.488	15.865	17.63	36.261	51.551
	20	100	11.447	14.336	12.631	13.592	25.293	30.553
	30	100	8.452	10.263	7.729	7.2	19.807	24.513
	60	100	4.441	7.545	6.993	3.062	12.445	14.411
	90	100	2.55	4.555	4.426	0.927	9.887	11.603
	120	100	1.884	2.014	3.637	0.369	8.129	8.328
	150	100	1.108	5.105	2.818	0.562	6.547	7.324
	180	100	0.603	1.339	4.949	0.497	5.136	7.11
	210	100	0.433	1.49	1.585	1.332	4.944	5.263
30	10	99	13.989	18.226	17.22	18.046	35.163	50.497
	20	100	10.238	14.314	14.317	9.014	24.655	30.791
	30	100	6.568	7.866	15.551	4.839	18.996	23.416
	60	100	3.061	8.561	9.956	4.917	12.138	13.076
	90	100	0.591	6.382	7.352	0.086	9.502	10.559
	120	100	0.226	3.722	7.905	1.985	7.894	8.268
	150	100	0.298	4.69	7.648	1.783	6.431	6.722
	180	100	1.073	2.552	4.701	1.062	5.095	6.271
	210	100	0.43	4.571	6.221	1.159	5.5	4.674
20	10	96	13.158	21.094	13.654	12.296	32.584	44.826
	20	100	10.811	8.896	6.931	15.391	24.348	29.055
	30	100	7.031	9.786	8.753	7.187	19.374	24.374
	60	100	5.193	5.995	4.552	5.228	13.122	12.709
	90	100	2.692	2.921	8.616	0.53	10.045	10.483
	120	100	2.663	1.121	2.435	2.893	8.025	10.017
	150	100	1.978	2.115	1.39	2.244	7.466	7.37
	180	100	0.926	0.257	6.435	2.077	7.434	6.819
	210	100	0.868	0.097	4.695	2.359	7	6.111
10	10	87	12.511	18.105	5.482	10.505	25.368	21.962
	20	96	9.223	11.161	5.702	6.149	18.834	20.777
	30	98	5.01	4.058	10.667	13.864	19.228	18.399
	60	100	1.165	0.493	5.405	9.018	12.867	13.893
	90	100	1.553	2.784	9.404	5.303	11.924	13.127
	120	100	1.456	1.564	8.622	1.76	10.713	11.653
	150	100	0.878	1.992	9.771	4.858	8.401	11.56
	180	100	2.08	1.473	7.007	4.218	11.293	10.876
	210	100	1.6	1.146	11.846	4.231	9.407	10.504
5	10	100	10.712	0.5	24.184	1.008	0.59	57.241
	20	72	12.783	12.693	2.851	19.836	11.065	19.351
	30	80	8.494	19.571	0.669	8.444	18.693	3.435
	60	85	3.133	0.195	0.731	6.439	18.067	8.853
	90	90	3.368	10.957	2.75	7.746	21.212	12.812
	120	93	4.283	2.412	0.375	4.408	16.923	18.779
	150	94	3.937	4.265	6.859	1.308	22.602	15.214
	180	95	1.833	1.249	3.94	4.1	23.512	16.484
	210	96	5.262	5.316	9.9	2.854	20.257	15.685

Figure 3.6: Number of clusters needed for compensation, REML with $\rho = 0.38$



3.2.5 Coverage of the 95% Confidence Intervals

The 95% Wald Confidence Interval was again used to assess the accuracy of the standard errors for REML estimates. Table 3.10 shows the noncoverage rates by ICC. The fixed effect standard errors were mostly within the acceptable range set by Bradley (1978) across all ICCs. The noncoverage rates were higher for the variance component standard errors.

Table 3.11 shows the noncoverage rate by the number of clusters. For all estimates, fixed and random, noncoverage rates were highest when only 5 clusters were used.

The noncoverage rate by group size is shown in Table 3.12. The variance component standard errors were underestimated across all cluster sizes, but non-coverage was much worse for simulations with fewer clusters.

It appears that for REML estimates many standard error estimates were still too small, especially for conditions with few clusters or small group size.

Table 3.10: Noncoverage of the 95% CI by ICC for REML

Parameter	ICC		
	0.04	0.17	0.38
γ_{00}	0.073	0.078	0.076
γ_{01}	0.061	0.066	0.067
γ_{10}	0.067	0.068	0.069
γ_{11}	0.062	0.063	0.064
σ_0	0.187	0.154	0.185
σ_1	0.187	0.173	0.183

Table 3.11: Noncoverage of the 95% CI by number of clusters for REML

Parameter	Number of Clusters				
	5	10	20	30	40
γ_{00}	0.124	0.088	0.072	0.069	0.068
γ_{01}	0.111	0.074	0.061	0.057	0.057
γ_{10}	0.115	0.081	0.058	0.075	0.074
γ_{11}	0.127	0.076	0.053	0.058	0.060
σ_0	0.392	0.190	0.156	0.147	0.162
σ_1	0.385	0.183	0.165	0.155	0.188

Table 3.12: Noncoverage of the 95% CI by cluster size for REML

Parameter	Group Size								
	10	20	30	60	90	120	150	180	210
γ_{00}	0.096	0.086	0.079	0.074	0.071	0.069	0.069	0.068	0.070
γ_{01}	0.059	0.063	0.061	0.066	0.064	0.066	0.067	0.069	0.067
γ_{10}	0.070	0.066	0.069	0.068	0.070	0.068	0.068	0.068	0.069
γ_{11}	0.062	0.063	0.062	0.064	0.064	0.063	0.064	0.064	0.063
σ_0	0.334	0.223	0.181	0.151	0.144	0.143	0.135	0.133	0.130
σ_1	0.333	0.235	0.201	0.161	0.146	0.144	0.139	0.138	0.131

3.3 Generalized Estimating Equations

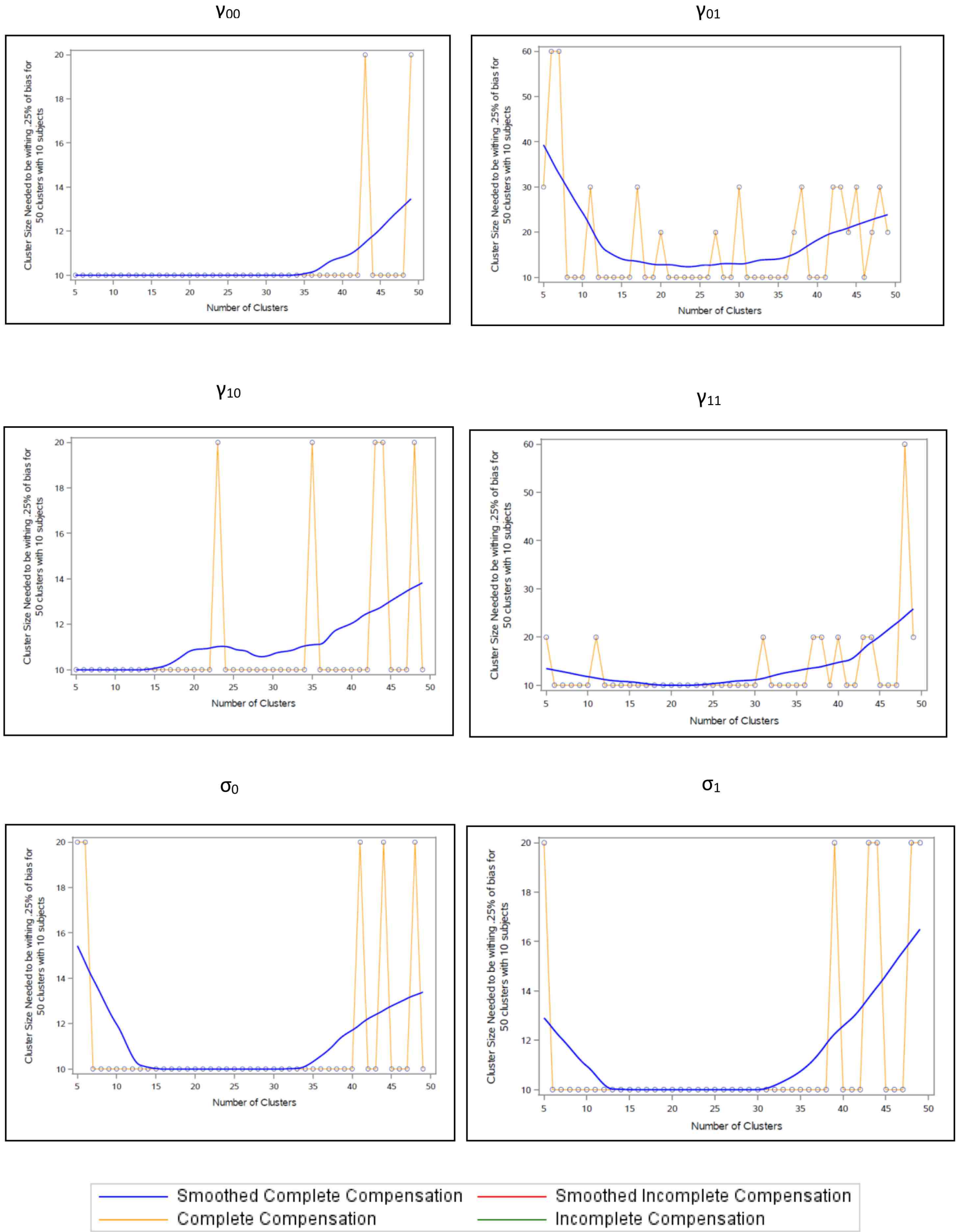
3.3.1 Convergence

The overall convergence rates for GEE were 99.28% when $\rho = 0.17$, 96.97% when $\rho = 0.04$, and 99.41% when $\rho = 0.38$. The lowest convergence rate was seen for models with $\rho = 0.04$ and 5 clusters with 10 subjects per cluster. For large total sample sizes, the rate of convergence was 100%. This data is shown in Table 3.13 for selected conditions.

3.3.2 Parameter Estimates

For GEE with $\rho = 0.17$, only 60.84% of fixed effect parameter estimates fell below 5% in absolute value. For the reference simulation, the average absolute relative bias for γ_{00} was 12.21%, for γ_{01} was 10.71%, for γ_{10} was 19.22%, and for γ_{11} was 16.10%. For almost all simulated conditions, the absolute relative bias was below the reference bias, resulting in complete compensation, shown in Figure 3.7. The largest bias found, 36.073%, was found in the smallest sample size tested, 5 clusters, with ten subjects per cluster, for the interaction term γ_{11}.

Figure 3.7: Number of clusters needed for compensation, GEE with $\rho = 0.17$



Other comparatively large biases were found in simulations with ten or twenty subjects per cluster, even if the number of clusters was larger. The smallest biases were found for conditions with a large number of subjects per cluster, regardless of number of cluster.

For random effects, only 8.13% were below 5%. The reference bias for σ_0 was

34.91% and σ_1 was 43.03%. The highest absolute relative bias for each σ_0 and σ_1 was again found for 5 clusters, with ten subjects per cluster. Smaller biases were found when the number of clusters was greater than 40 and the cluster size was larger than 150. As shown in Figure 3.7, complete compensation was achieved for all number of clusters, with a cluster size of twenty or less.

The similarities of Table 3.13 and Table 3.7 are noted and discussed in a later section.

For $\rho = 0.04$ and $\rho = 0.38$, the results are again quite similar to the corresponding ICCs when using REML estimation. This can be seen by comparing Table 3.14 to Table 3.8 and Table 3.15 to Table 3.9. For this reason these results are not presented in detail.

Table 3.13: Absolute percent relative bias of GEE estimates with $\rho = 0.17$

Number of Clusters	Cluster Size	% Converged	γ_{00}	γ_{10}	γ_{01}	γ_{11}	σ_0	σ_1
50	10	100	12.209	19.216	10.702	16.095	34.907	43.028
40	10	100	11.249	18.091	10.731	16.73	33.886	42.949
	20	100	8.371	9.431	10.174	10.174	21.433	27.033
	30	100	5.735	7.15	6.864	8.516	16.943	21.055
	60	100	3.363	6.044	3.996	2.261	10.413	13.826
	90	100	2.512	5.037	3.025	0.557	7.556	10.107
	120	100	1.839	3.3	3.047	2.238	5.985	7.895
	150	100	1.271	4.134	2.344	1.785	5.696	7.432
	180	100	0.946	1.267	2.569	0.671	4.154	6.718
	210	100	0.329	3.012	2.711	0.784	3.052	4.683
30	10	99	8.764	17.159	10.298	12.005	31.496	41.268
	20	100	6.78	10.957	13.085	10.66	20.683	27.758
	30	100	5.187	6.674	8.966	4.844	14.882	20.231
	60	100	2.335	10.164	6.889	0.943	9.799	12.128
	90	100	1.424	6.464	4.733	0.699	6.879	10.184
	120	100	0.648	4.262	3.817	0.61	6.961	8.2
	150	100	0.066	3.369	4.865	0.149	5.39	7.508
	180	100	0.329	3.13	4.098	0.755	3.563	6.27
	210	100	0.468	5.172	4.02	0.381	3.687	5.617
20	10	98	8.84	16.331	11.363	12.662	27.564	33.07
	20	100	6.791	6.763	7.716	13.903	21.307	28.172
	30	100	4.257	8.066	4.752	5.202	17.228	20.468
	60	100	3.554	5.202	3.582	6.057	10.612	11.679
	90	100	2.183	4.317	3.209	0.35	6.713	9.039
	120	100	2.231	1.313	3.179	3.809	6.265	9.245
	150	100	1.916	1.276	0.678	0.425	6.352	6.446
	180	100	0.776	1.106	3.008	2.36	7.443	6.931
	210	100	0.394	2.087	4.474	2.066	5.709	6.992
10	10	90	6.915	17.336	1.028	14.225	11.392	14.702
	20	97	4.193	6.686	1.52	9.903	13.828	17.637
	30	99	2.683	4.628	8.749	7.074	15.986	15.271
	60	100	0.721	0.187	1.357	8.585	13.279	13.288
	90	100	1.018	4.624	3.605	8.756	11.954	11.272
	120	100	1.371	1.468	2.932	6.594	13.78	11.413
	150	100	0.46	1.276	5.049	5.356	9.073	12.805
	180	100	0.579	2.94	4.713	4.11	9.66	10.205
	210	100	0.215	2.46	5.624	1.832	9.409	8.574
5	10	63	11.341	6.464	16.391	36.073	44.721	46.661
	20	81	9.068	11.295	20.451	13.029	17.141	12.554
	30	86	3.775	8.176	1.881	2.697	6.698	0.775
	60	93	0.903	7.184	2.898	3.688	12.437	12.795
	90	95	1.931	8.216	1.599	1.58	14.931	14.14
	120	96	2.631	4.197	2.38	0.647	17.78	19.861
	150	97	0.685	3.695	1.867	1.969	17.661	16.753
	180	97	1.492	2.608	5.367	3.566	21.654	18.502
	210	98	2.719	3.184	4.576	3.829	20.789	18.777

Table 3.14: Absolute percent relative bias of GEE estimates with $\rho = 0.04$

Number of Clusters	Cluster Size	% Converged	γ_{00}	γ_{10}	γ_{01}	γ_{11}	σ_0	σ_1
50	10	95	6.774	15.793	5.566	13.724	7.136	33.808
40	10	95	5.895	13.24	6.838	12.31	6.934	34.341
	20	98	5.038	7.33	6.611	7.226	18.78	22.797
	30	99	4.022	6.078	5.071	6.119	14.236	16.961
	60	99	2.362	4.253	2.488	3.367	8.666	12.062
	90	98	1.674	4.166	2.053	1.571	8.355	9.305
	120	97	1.451	0.593	1.112	2.96	4.877	7.573
	150	97	1.449	3.49	1.878	0.045	5.183	6.953
	180	97	0.929	0.725	1.658	2.28	4.009	6.969
	210	97	0.648	2.461	2.355	2.168	0.457	5.5
30	10	94	4.534	16.224	5.59	8.611	0.902	30.839
	20	98	4.246	9.526	8.235	3.764	10.916	22.39
	30	99	3.24	8.703	5.55	2.087	11.927	16.018
	60	99	1.813	9.274	2.968	0.929	7.371	11.772
	90	99	1.531	4.887	3.204	0.192	8.904	8.966
	120	98	1.007	4.897	2.916	1.904	4.619	6.845
	150	98	0.798	3.439	2.587	0.74	4.701	7.105
	180	98	0.66	3.37	1.48	0.444	3.525	6.036
	210	98	0.438	3.347	2.205	0.586	3.073	5.421
20	10	92	4.983	10.988	0.716	8.953	41.748	26.902
	20	97	4.731	7.473	6.493	11.875	3.998	19.597
	30	98	2.307	8.903	4.352	8.245	8.435	14.942
	60	99	2.397	3.431	3.117	7.677	9.456	11.306
	90	99	1.23	2.407	2.316	2.305	7.15	10.316
	120	99	1.676	0.471	0.876	4.132	5.684	8.792
	150	99	1.405	1.925	0.146	1.157	5.761	6.253
	180	99	0.532	2.42	0.841	4.47	6.399	8.141
	210	99	0.981	0.646	1.632	4.309	3.52	6.854
10	10	83	3.594	11.173	1.753	16.138	103.547	3.658
	20	92	1.92	9.106	2.15	7.213	36.787	12.74
	30	95	0.507	2.884	0.512	6.333	21.952	11.595
	60	98	1.482	4.887	0.747	6.226	0.331	14.715
	90	99	1.039	7.193	0.367	7.957	7.171	10.414
	120	99	1.444	4.303	0.931	5.773	9.205	11.523
	150	99	0.327	4.695	2.357	2.149	7.475	9.732
	180	99	0.313	2.989	3.869	5.905	6.252	10.165
	210	99	0.003	2.461	3.984	7.124	9.516	8.755
5	10	56	0.21	2.219	5.691	19.825	361.119	53.543
	20	73	1.348	8.761	8.954	2.509	171.745	5.885
	30	80	2.119	7.349	2.925	13.798	78.527	0.481
	60	87	0.356	1.408	2.472	5.049	22.521	13.207
	90	90	1.042	14.623	1.953	5.646	5.124	17.705
	120	91	1.606	3.719	3.115	0.286	6.289	19.812
	150	92	1.029	6.832	0.421	4.158	0.171	17.878
	180	93	0.121	4.921	2.403	2.198	11.019	17.486
	210	94	1.676	2.628	0.663	1.66	9.761	17.029

Table 3.15: Absolute percent relative bias of GEE estimates with $\rho = 0.38$

Number of Clusters	Cluster Size	% Converged	γ_{00}	γ_{10}	γ_{01}	γ_{11}	σ_0	σ_1
50	10	100	16.793	21.732	16.008	19.899	37.62	51.335
40	10	100	15.451	20.488	15.865	17.63	36.261	51.551
	20	100	11.447	14.336	12.631	13.592	25.293	30.553
	30	100	8.452	10.263	7.729	7.2	19.807	24.513
	60	100	4.441	7.545	6.993	3.062	12.445	14.411
	90	100	2.55	4.555	4.426	0.927	9.887	11.603
	120	100	1.884	2.014	3.637	0.369	8.129	8.328
	150	100	1.108	5.105	2.818	0.562	6.547	7.324
	180	100	0.603	1.339	4.949	0.497	5.136	7.11
	210	100	0.433	1.49	1.585	1.332	4.944	5.263
30	10	100	13.989	18.226	17.22	18.046	35.163	50.497
	20	100	10.238	14.314	14.317	9.014	24.655	30.791
	30	100	6.568	7.866	15.551	4.839	18.996	23.416
	60	100	3.061	8.561	9.956	4.917	12.138	13.076
	90	100	0.442	6.387	7.388	0.119	9.565	10.564
	120	100	0.205	3.642	7.871	1.798	7.832	8.213
	150	100	0.27	4.69	7.648	1.783	6.229	6.722
	180	100	1.073	2.552	4.701	1.197	5.002	6.36
	210	100	0.439	4.41	6.234	1.145	5.57	4.69
20	10	99	13.158	21.094	13.654	12.296	32.584	44.826
	20	100	10.811	8.896	6.931	15.391	24.348	29.055
	30	100	7.031	9.786	8.753	7.187	19.374	24.374
	60	100	5.193	5.995	4.552	5.228	13.122	12.709
	90	100	2.692	2.921	8.616	0.53	10.045	10.483
	120	100	2.663	1.121	2.435	2.893	8.025	10.017
	150	100	2.118	2.151	1.581	2.407	7.506	7.35
	180	100	1.02	0.443	6.345	2.123	7.441	6.753
	210	100	0.828	0.122	4.591	1.935	6.964	6.083
10	10	93	12.511	18.105	5.482	10.505	25.368	21.962
	20	98	9.223	11.161	5.702	6.149	18.834	20.777
	30	99	5.01	4.058	10.667	13.864	19.228	18.399
	60	100	1.165	0.493	5.405	9.018	12.867	13.893
	90	100	1.553	2.784	9.404	5.303	11.924	13.127
	120	100	1.5	1.765	9.055	1.76	10.763	11.824
	150	100	0.878	1.992	9.771	4.858	8.401	11.56
	180	100	2.08	1.473	7.007	4.218	11.293	10.876
	210	100	1.624	1.187	11.374	4.527	9.367	10.565
5	10	68	10.712	0.5	24.184	1.008	0.049	56.3
	20	81	12.777	12.922	3.971	20.277	11.065	19.351
	30	88	8.48	19.002	0.749	7.208	18.693	3.158
	60	94	3.041	0.016	1.791	6.814	18.254	8.984
	90	95	3.398	10.299	2.961	7.656	21.086	12.812
	120	97	4.133	2.19	1.707	5.32	16.923	18.935
	150	97	4.194	4.853	6.587	0.77	22.341	15.534
	180	98	1.744	1.248	4.112	4.155	23.495	16.279
	210	97	5.464	4.568	10.041	3.06	20.257	15.685

Figure 3.8: Number of clusters needed for compensation, GEE with $\rho = 0.04$

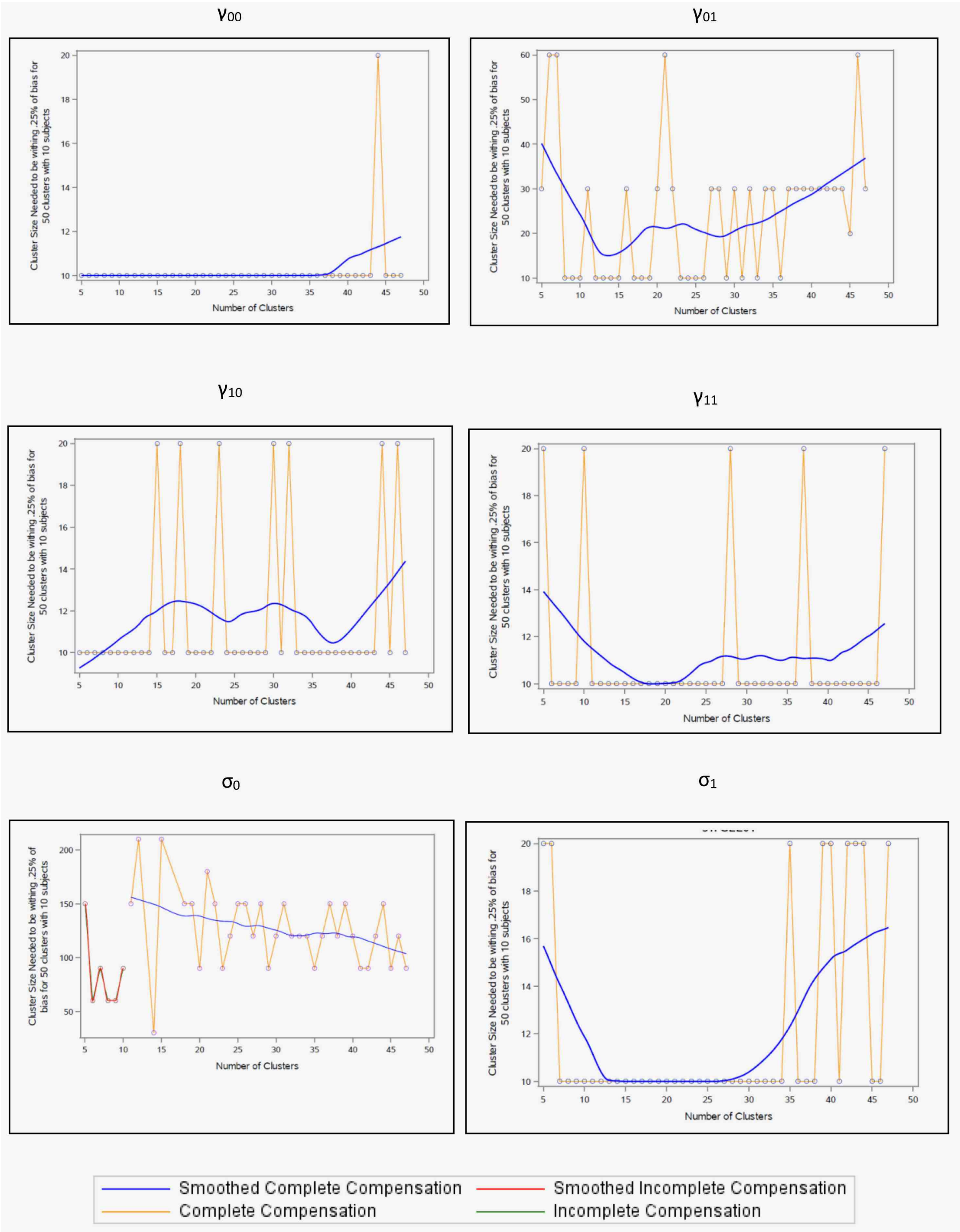
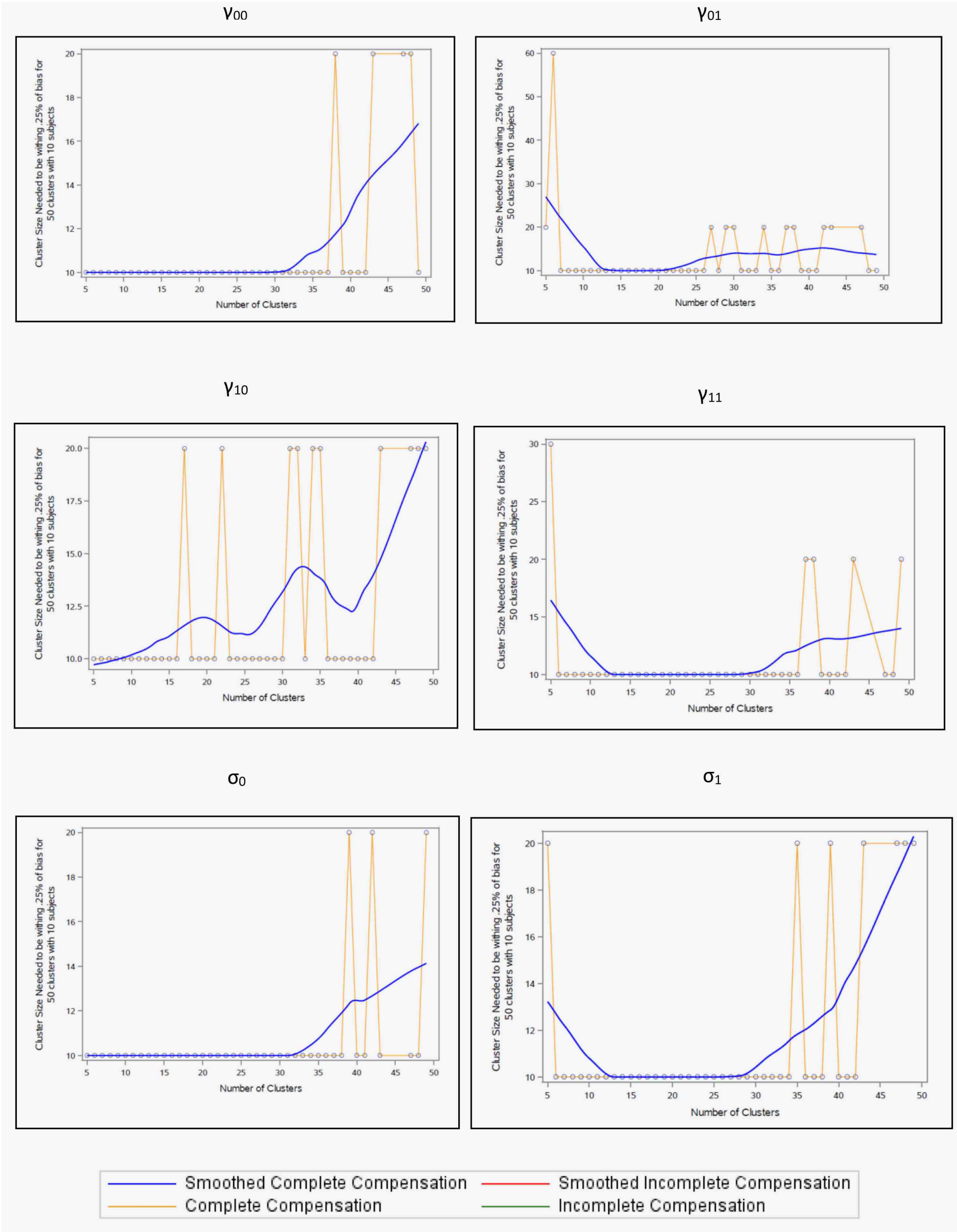


Figure 3.9: Number of clusters needed for compensation, GEE with $\rho = 0.38$



3.3.3 Coverage of the 95% Confidence Intervals

The noncoverage rates found for GEE standard error estimates were slightly higher than for either MLE or REML estimates. Table 3.16 shows that the standard errors were underestimated for all estimates across all ICCs.

In Table 3.17 it can be seen that coverage was slightly improved by an increase in the number of clusters, but the noncoverage rates for all components were much too high in conditions with 5 or 10 clusters.

Increasing cluster size appears to improve standard error estimates for variance components, as seen in Table 3.18. For fixed effect components with increased cluster size there were minimal differences in noncoverage.

As with MLE and REML, it appears that the standard errors were underestimated.

Table 3.16: Noncoverage of the 95% CI by ICC, GEE

Parameter	ICC		
	0.04	0.17	0.38
γ_{00}	0.098	0.100	0.100
γ_{01}	0.114	0.118	0.117
γ_{10}	0.088	0.089	0.092
γ_{11}	0.110	0.114	0.114
σ_0	0.188	0.155	0.184
σ_1	0.187	0.173	0.182

Table 3.17: Noncoverage of the 95% CI by number of clusters for GEE

Parameter	Number of Clusters				
	5	10	20	30	40
γ_{00}	0.233	0.132	0.089	0.080	0.080
γ_{01}	0.306	0.170	0.108	0.087	0.075
γ_{10}	0.231	0.125	0.073	0.084	0.082
γ_{11}	0.321	0.174	0.095	0.087	0.080
σ_0	0.392	0.190	0.156	0.147	0.163

Table 3.18: Noncoverage of the 95% CI by cluster size for GEE

Parameter	Group Size								
	10	20	30	60	90	120	150	180	210
γ_{00}	0.120	0.108	0.101	0.097	0.093	0.091	0.092	0.090	0.092
γ_{01}	0.107	0.114	0.111	0.118	0.115	0.121	0.121	0.123	0.121
γ_{10}	0.091	0.086	0.090	0.089	0.088	0.090	0.091	0.090	0.090
γ_{11}	0.106	0.110	0.109	0.116	0.113	0.116	0.117	0.114	0.115
σ_0	0.335	0.223	0.181	0.151	0.145	0.144	0.135	0.134	0.131
σ_1	0.328	0.231	0.199	0.161	0.146	0.139	0.140	0.140	0.133

Chapter 4

Discussion

4.0.1 Complete Compensation versus Incomplete Compensation

The aim of this study was to investigate if it was possible to compensate for the use of fewer clusters. For this reason we did not adhere to the "rules of thumb" for clusters and number of clusters when choosing our reference condition. This resulted in overall high bias for the variance components.

For conditions where complete compensation was found, increasing the number of subjects per cluster can effectively account for the increased bias found when the number of clusters was decreased. For most of the conditions tested, complete compensation was found for all fixed effect parameter estimates, but this was not the cases for the random effects.

There are several possible scenarios for why incomplete compensation was only found for some conditions. As per the definition used in this paper, a decreased absolute relative bias was only deemed complete compensation if the absolute relative bias fell to within 0.25% of the reference bias and remained within that threshold for all increased sample sizes thereafter. There were conditions for which complete compensation was achieved, but then increased cluster size resulted in a slight increase in bias to above the designated complete compensation point. In this case it was only deemed incomplete compensation.

It is also possible that complete compensation may be achieved for a number of subjects greater than those tested in this study. However, this may be true only to a point. The point of diminishing returns, or the point were observations begin to make a negligible contribution, is described by Hemming et al. (2007). This can be seen in situations where the sample size is increased but the number of clusters is not, resulting in the power of the study leveling off (Hemming et al., 2007). To examine this further we conducted one simulation, using MLE with $\rho = 0.17$, where for 5, 10, 20, 30, 40, and 50 clusters we increased the cluster size to 1,000. The results are shown in Table 4.1. When this table is compared to Table 3.1, it can be seen that the absolute percent relative bias plateaued.

Table 4.1: Absolute percent relative bias with 1000 subjects per cluster

Number of Clusters	γ_{00}	γ_{10}	γ_{01}	γ_{11}	σ_0	σ_1
40	0.026	2.428	0.184	2.093	6.854	7.247
30	0.569	0.076	1.268	1.118	9.321	7.855
20	0.383	2.566	1.864	0.795	12.552	13.350
10	0.532	2.404	1.963	4.561	26.998	24.680
5	1.310	2.306	6.400	14.547	49.872	51.478

For some conditions, it is also possible that complete compensation is not possible. For all conditions in this study where complete compensation was not found, the criterion for incomplete compensation was met. Incomplete compensation does not indicate that the bias is low, only that it plateaued as the number of subjects continued to increase. An incomplete compensation point shows that after that point no or only minimal improvements will be made in accuracy by increasing the number of subjects per group further.

4.0.2 Comparison of Estimation Methods

The results found for GEE and REML estimates, with $\rho = 0.17$, are extremely similar. The greatest difference can be seen in the coverage of the 95% confidence interval for standard error estimates, where a small number of clusters appears to result in higher noncoverage rates for GEE estimates than for REML estimates. SAS documentation on how to implement each method, along with the syntax from Austin (2010), was used to develop the syntax for each of these estimation methods (Austin, 2010; SAS/STAT(R) 9.22 User's Guide: Fitting a Marginal (GEE-type) Model). The similarity is most likely due to the fact that only a GEE-type estimator is implemented through PROC GLIMMIX, not the true GEE method. In PROC GLIMMIX parameters are estimated using likelihood-based techniques, not by the method of moments, as they would in a true GEE model (SAS/STAT(R) 9.22 User's Guide: PROC GLIMMIX Contrasted with Other SAS Procedures). Researchers should be aware of this fact when looking to use GEE for multilevel logistic regression. An option to truly find GEE estimates for binary outcomes would be to use PROC GENMOD. In this study we did not implement PROC GENMOD, as it only estimates fixed effect parameters (SAS/STAT(R) 9.22 User's Guide: Model Fitting in PROC GENMOD).

Between MLE and REML, more distinct differences were found. For fixed effect parameter estimates, on average larger percent relative biases were found for REML estimates, with only 60.96% falling below 5% in absolute value, compared to 87.62% that were below 5% MLE estimates, when $\rho = 0.17$. However, for REML complete compensation could be achieved for a much smaller number of subjects per cluster. This may be due to the fact that the reference bias for REML was relatively high compared to other simulated conditions. This follows the pattern that higher biases are found for REML estimates for conditions with small cluster sizes.

When examining the absolute relative bias for variance components in both methods ($\rho = 0.17$), it is interesting to note some differences. For REML estimates, it appears that the number of subjects per cluster is of greater importance as the largest biases were found for small clusters sizes, across all number of clusters. For MLE, the largest biases were found for conditions with a small number of clusters, regardless of number of subjects per cluster. This indicates that when using MLE estimation it is more important to have a larger number of clusters. This matches the previous literature, in which it is frequently noted that more clusters is better (Eldridge and Kerry, 2012; Moineddin et al., 2007; Clark, 2008). Most of this research, such as that done by Moineddin et al. (2007) and Clark (2008), implemented MLE. The findings from this study show, for REML estimations, the number of subjects per cluster may actually be more important.

For all methods the standard errors were underestimated. REML estimation had slightly better coverage rates for conditions where the number of groups was small. MLE standard error variance component estimates were better than REML when a small group size was used.

In general MLE estimates behaved more as expected than REML estimates. When using REML, an increase in group size resulted in a decrease in bias while the number of subjects per group was fixed. The increase in bias has been noted by Austin (2010), who also found an increase in bias as the number of clusters increased from 10 to 20 when using PROC GLIMMIX.

To investigate this behavior seen in REML and GEE estimates, we conducted another simulation where we fixed the number of subjects per cluster as 10 and tested 5, 7, 10, 13, 17, 20, 30, 40, and 50 clusters. The absolute percent bias for each estimate is shown in Table 4.2 for REML estimates and Table 4.3 for GEE estimates. In this simulation the lowest biases were found when the ratio of number to clusters to cluster size was close to 1. Further research is needed on this subject.

Table 4.2: Absolute percent relative bias for REML estimates with 10 subjects per cluster

Number of Clusters	γ_{00}	γ_{10}	γ_{01}	γ_{11}	σ_0	σ_1
50	13.27	10.68	17.49	14.87	35.04	44.36
40	12.18	13.41	15.22	12.63	36.03	41.59
30	11.18	10.19	15.49	17.41	29.61	41.24
20	8.34	7.83	2.59	7.77	25.77	34.46
17	9.31	6.50	13.12	13.85	22.34	34.05
13	7.77	1.83	7.78	5.88	20.23	22.02
10	6.07	0.05	4.82	2.39	6.70	13.02
7	3.59	6.27	8.55	4.87	18.21	20.78
5	5.40	7.28	13.76	0.07	73.64	69.48

Table 4.3: Absolute percent relative bias for GEE estimates with 10 subjects per cluster

Number of Clusters	γ_{00}	γ_{10}	γ_{01}	γ_{11}	σ_0	σ_1
50	12.97	10.22	17.31	22.77	37.29	43.64
40	12.75	7.80	20.00	12.56	35.23	39.91
30	11.27	10.90	17.65	13.10	30.60	40.13
20	12.46	11.10	14.26	9.30	29.21	36.84
17	10.32	6.00	17.75	12.20	24.26	31.27
13	7.79	8.43	12.51	7.19	15.50	21.05
10	5.72	4.93	11.73	8.48	5.50	15.09
7	4.62	6.76	0.48	3.41	11.20	13.02
5	1.08	26.41	0.24	1.04	57.46	62.44

4.0.3 Impact of ICC

For MLE estimates, increasing cluster size appears to increase bias for variance components when the ICC is small. This trend was not seen for larger ICCs.

For REML and GEE, simulations with a higher ICC had a higher convergence rate. Across all three ICCs, the trends seen for varying sample sizes were similar, but in general higher biases were found when the ICC was large. This may be due to Design Effect (DE) which is how much the standard errors are underestimated in a clustered sample compared to a simple random sample. The DE is used to calculate how much the sample size must be increased to achieve the same power as a study that used a simple random sample (Maas and Hox, 2005). To calculate DE, the following formula can be used:

$$DE = 1 + (n - 1)\rho \tag{4.1}$$

...where n is the average number of subjects per cluster (Rutterford et al., 2015). From this formula, ican be seen that a larger ICC leads to a larger design effect, resulting in the need for a larger overall sample size.

Chapter 5

Conclusion

Our findings suggest that for fixed effect parameter estimates it is possible to compensate for the use of few clusters when using MLE. For all conditions, high biases were found for the random variance components. If complete compensation was found at all it was only for a number of clusters close to the reference bias. For few clusters, incomplete compensation was found when complete compensation was not. Further study is needed on REML and GEE estimates. However, it appears that when the variance components are of interest, REML estimation should not be used when the number of subjects per group is few.

Limitations of this study include the inability to perform a true GEE model while still investigating the random effect components. The time required to run each simulation limited the number of simulations per condition to 1,000 and the variety of conditions we were able to test.

This study could be extended in several ways. A larger variety of ICCs could be tested, as could various other optimization techniques. The impact of prevalence should also be examined. Doing so would give researchers even more insight as to what sample size is necessary to compensate for the use of few clusters.

Our findings show that under certain conditions it is possible to compensate for the use of few clusters. This result is highly relevant and useful for current epidemiological and medical research.

References

Austin PC and Merlo J. (2017). Intermediate and advanced topics in multilevel logistic regression analysis. *Statistics in Medicine 36*, 3257-3277.

Austin PC. (2007). A comparison of the statistical power of different methods for the analysis of cluster randomization trials with binary outcomes. *Statistics in Medicine 26*, 3550-3565.

Austin PC. (2010). Estimating multilevel logistic regression models when the number of clusters is low: a comparison of different statistical software procedures. *The International Journal of Biostatistics 6.16.*

Bradley, JV. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31*, 144-152.

Browne WJ and Draper D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis 1*, 473-514.

Chong EKP and Zak SH. (2001). *An introduction to optimization.* 2nd edition. Danvers, Ma: John Wiley & Sons, Inc.

Clarke P. (2008). When can group level clustering be ignored? Multilevel models versus single level models with sparse data. *Journal of Epidemiology and Community Health 62*, 752-758.

Eldridge S and Kerry S. (2012). *A practical guide to cluster randomized trials in health services research.* West Sussex: John Wiley & Sons, Ltd.

Gulliford MC, Ukoumunne OC, and Chinn S. (1999) Components of variance and intraclass correlations for the design of community-based surveys and intervention studies: data from the Health Survey for England 1994. *American Journal of Epidemiology 149*, 876–883.

Hayes RJ and Bennett S. (1999). Simple sample size calculation for cluster-randomized trials. *International Journal of Epidemiology 28*, 319-326.

Hayes RJ and Moulton LH. (2017). *Cluster Randomised Trials.* Boca Raton:

Taylor & Francis Group.

Hemming K, Eldridge S, Forbes G, Weijer C, and Taljaard M. (2017) How to design efficient cluster randomised trials. *BMJ 358* j3064.

Hemming K, Girling AJ, Sitch AJ, Marsh J, and Lilford RJ. (2011). Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. *BMC Medical Research Methodology 11* 102.

Hemming K and Taljaard M. (2016). Sample size calculations for stepped wedge and cluster randomized trials: a unified approach. *Journal of Clinical Epidemiology 69*, 137-146.

Hoogland JJ and Boomsma A. (1998). Robustness studies in covariance structure modeling. An overview and a meta-analysis. *Sociological Methods & Research 26*,329-367.

Hox J. (2010). *Multilevel analyses: techniques and applications.* New York: Routledge.

Hox, Moerbeek, and van de Schoot. (2018). *Multilevel Analysis: Techniques and Application. Quantitative Methodology series.* Third Edition. New York: Routledge.

Kiernan K, Tao J, Gibbs P. (2009). Tips and strategies for mixed modeling with SAS/STAT(R) procedures. Cary, NC: SAS Institute Inc.

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

Little RC, Milliken GA, Stroup WW, and Wolfinger RD. (1996). *SAS System for Mixed Models.* Cary, NC: SAS Institute Inc.

Maas CJM and Hox JJ. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica 58*, 127-137.

Maas CJM and Hox JJ. (2005). Sufficient sample sizes for multilevel modeling. *Methodology 1(3)*, 86-92.

Mason W, Wong G, Entwisle. (1983). Multilevel models from a multiple group structural equation perspective. *Advanced Structural Equation Modelig: Issues and Techniques.* pp. 89-125. Mawah: Lawrence Erlbaum Associates, Inc.

McNeish Dm and Stapleton LM. (2016). The effect of small sample size on two-level model estimates: a review and illustration. *Educ Psychol Rev 28*, 295-314.

Moineddin R, Matheson FI, and Glazier RH. A simulation study of sample size for multilevel logistic regression models. *BMC Medical Research Methodology* 2007;7:34.

Mok M. (1995). Sample Size requirements for 2-level designs in educational research. *Modeling Newsletter 7*, 11-15.

Nocedal J and Wright SJ. (1999). *Numerical Optimization.* New York: Springer.

Paccagnella O. (2011). Sample Size and Accuracy of Estimates in Multilevel Models: New Simulation Results. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences 7*, 111-120.

Raudenbush, S.W., and Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Rutterford C, Copas A, and Eldridge S. (2015). Methods for sample size determination in cluster randomized trials. *International Journal of Epidemiology*, 1051-1067.

SAS Institute Inc., (2015). *SAS/STAT(R) 14.1 User's Guide* Cary, NC: SAS Institute Inc.

Shultz GA, Schnabel RB, Byrd RH. (1985). A Family of Trust-Region-Based algorithms for unconstrained minimization with strong global convergence properties. *SIAM Journal on Numerical Analysis 22(1)*, 47-67.

Sommet N and Moreselli D. (2017). Keep calm and learn multilevel logistic modeling: a simplified three-step procedure using Stata, R, Mplus, and SPSS. *International Review of Social Psychology 30(1).* 203-218.

Theall KP, Scribner R, Broyles S, Yu Q, Chotalia J, Simonsen N, Schonlau M, and Carlin BP. (2011). Impact of small group size on neighborhood influences in multilevel models. *Journal of Epidemiology and Community Health 65.8*, 688-695.

van Breukelen GJP and Candel MJJM. (2012). Calculating samples sizes for cluster randomized trials: we keep it simple and efficient! *Journal of Clinical Epidemiology 65*, 1212-1218.

Wang J, Xie H, and Fisher JH. (2012). *Multilevel models: applications using SAS.* Berlin/Boston: Higher Education Press and Walter de Fruyter GmbH & Co. KG.