Impact of Design Features for Cross-Classified Logistic Models When the Cross-Classification Structure Is Ignored

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Weijia Ren, M.A.

Graduate Program in Education

The Ohio State University

2011

Dissertation Committee:

Ann O’Connell, Ph.D., Advisor
Richard G. Lomax, Ph.D.
Dorinda J. Gallant, Ph.D.
Abstract

In education and other social science research, researchers often encounter data with multilevel structures. Over the past two decades, hierarchical linear models (HLMs) have rapidly developed and been applied to model these structures. However, many datasets involve a cross-classified structure, instead of a purely hierarchical structure for which an HLM may not be appropriate. Given that cross-classified data structures are often ignored in many studies, this proposal intends to: (1) explore the consequences of ignoring a cross-classified structure in a cross-classified dataset with a dichotomous outcome; (2) examine differences due to estimation method; and (3) clarify the conditions under which there is a need to use the cross-classified random effects logistic model. Also, this proposal provides a demonstration of the cross-classified random effects logistic model and focuses on investigating possible recidivism predictors for incarcerated youth.

This methodological study will contribute to the literature on hierarchical generalized linear modeling and cross-classified random-effect logistic modeling. In addition, it will explore the importance of incarcerated youth’s reading ability and the effectiveness of an in-prison reading program on their recidivism. The effect of the county where the youth come from and the facility where the youth are retained on incarcerated youth’s recidivism will also be addressed.
Acknowledgments

I would like to express my sincerest thanks to everyone who has helped and supported me throughout my academic career.

First, I would like to thank my advisor, Dr. Ann A. O’Connell for her ideas, immense wealth of knowledge, and continuous encouragement and guidance during the past four years. Without her invaluable support, neither finishing the dissertation nor the degree would be possible.

I am truly grateful for Dr. William Loadman for his support during the past four years. He taught and guided me through the project evaluation process. I wish to thank him for being considerate and supportive the whole time! I owe many thanks to Dr. Raeal Moore for her guidance and support during the past four years. She is a great supervisor and friend, and I have learned a lot from her.

My grateful appreciation also goes to my dissertation committee members, Dr. Richard G. Lomax and Dr. Dorinda J. Gallant, for their advice and encouragement. Thank you also to Dr. Jerome D’Agostino for his support and for sitting on my candidacy examination committee.
I also wish to acknowledge and thank the administrators and staff in the School of Educational Policy and Leadership in the College of Education and Human Ecology for their support, specifically Barbara Heinlein and Deborah Zabloudiv.

In addition, I would like to thank all of my colleagues and friends here at The Ohio State University for their support. Thank you all for your encouragement and help!

Last but not least, I would like to thank my husband Yongfang for his support to me during the past eight years. I wish to thank my parents, my parents-in-law, and my grandparents for their constant support. I love you all!
Vita

April 1983 ............................................................... Born in Sichuan, China

2001 ........................................................................ Chongqing No. 1 High School, China

2005 ................................................................................. B.A., English, Sichuan International Studies University, Chongqing, China

2007 ................................................................................ M.A., Social & Philosophical Studies in Education, University of Kentucky

2007 to present ........................................ Graduate Research Associate, The Ohio State University

Publications


Fields of Study

Major Field: Education
Table of Contents

Abstract ........................................................................................................................................... ii

Acknowledgments .......................................................................................................................... iii

Vita .................................................................................................................................................... v

List of Tables .................................................................................................................................. ix

List of Figures .................................................................................................................................. x

Chapter 1: Introduction ................................................................................................................. 1

1.1 Background ................................................................................................................................. 1

1.2 Purpose of the Study .................................................................................................................... 4

1.3 Research Questions ..................................................................................................................... 5

1.4 Significance of the Study ............................................................................................................ 5

1.5 Organization of the Dissertation ............................................................................................... 6

Chapter 2: Literature Review ......................................................................................................... 8

2.1 Hierarchical Generalized Linear Modeling (HGLM) ............................................................... 8

2.1.1 Multilevel Data Structure with Binary Outcome ................................................................. 8

2.1.2 Hierarchical Generalized Linear Models .............................................................................. 9

2.1.3 HGLM Estimation Methods ................................................................................................. 12
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.2</td>
<td>Random Effect Estimates</td>
<td>72</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Fit Indices</td>
<td>79</td>
</tr>
<tr>
<td>4.1.4</td>
<td>Convergence rate</td>
<td>80</td>
</tr>
<tr>
<td>4.2</td>
<td>Study 2: Empirical Data Analysis</td>
<td>83</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Descriptive Statistics</td>
<td>83</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Results</td>
<td>84</td>
</tr>
<tr>
<td>Chapter 5: Discussion</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Summary</td>
<td>88</td>
</tr>
<tr>
<td>5.2</td>
<td>Limitations</td>
<td>94</td>
</tr>
<tr>
<td>5.3</td>
<td>Recommendations</td>
<td>96</td>
</tr>
<tr>
<td>5.4</td>
<td>Future Research</td>
<td>97</td>
</tr>
<tr>
<td>References</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>Appendix: SAS Syntax</td>
<td>106</td>
<td></td>
</tr>
</tbody>
</table>
List of Tables

Table 1. Pure three-level clustering of incarcerated youth within counties within facilities (modified from Beretvas, 2010) .................................................................................................................. 15
Table 2. Cross-classified data structure of incarcerated youth cross-classified by counties and facilities (modified from Beretvas, 2010) ............................................................................. 17
Table 3. Simulation conditions ...................................................................................................................... 32
Table 4. Population parameter for simulation.................................................................................................. 35
Table 5. Three facilities feeding to each county ................................................................................................. 37
Table 6. Four facilities feeding to each county ..................................................................................................... 38
Table 7. Bias statistics for the fixed-effect coefficient read score ....................................................................... 54
Table 8. Bias statistics for the fixed-effect coefficient treatment ......................................................................... 60
Table 9. Bias statistics for the fixed-effect level-2 coefficient poverty ................................................................. 66
Table 10. Bias statistics for the random effect coefficient county ....................................................................... 73
Table 11. Pseudo-BIC fit indices of percentage CCREM was preferred to HGLM .................. 80
Table 12. Convergence rate of CCREM ............................................................................................................... 81
Table 13. Descriptive statistics .......................................................................................................................... 84
Table 14. Real data study’s parameter and standard error estimates ............................................................... 85
List of Figures

Figure 1. Network graph of pure three-level clustering of incarcerated youth within counties within facilities (modified from Beretvas, 2010) ............................................. 16

Figure 2. Network graph of cross-classified data structure of incarcerated youth cross-classified by counties and facilities (modified from Beretvas, 2010) .............................. 18

Figure 3. The data generation procedure for the Monte Carlo study .................................. 34

Figure 4. Sample flow chart of the empirical recidivism study ......................................... 48

Figure 5. Graph of EST and METH for the read score estimate relative bias ...................... 56

Figure 6. Graph of EST and CORR for the read score estimate relative bias ...................... 56

Figure 7. Graph of EST for the read score standard error relative bias .............................. 57

Figure 8. Graph of EST and CORR for the treatment estimate relative bias ......................... 62

Figure 9. Graph of EST for the treatment standard error relative bias ............................... 63

Figure 10. Graph of EST and CORR for the treatment estimate RMSE ............................ 64

Figure 11. Graph of EST and IUCC for the treatment estimate RMSE .............................. 65

Figure 12. Graph of EST, METH, and CORR for the poverty estimate relative bias ............. 68

Figure 13. Graph of EST and CORR for the poverty standard error relative bias .................. 69

Figure 14. Graph of EST, METH, and CORR for the poverty estimate RMSE .................... 71

Figure 15. Graph of EST and IUCC for the county estimate relative bias ............................ 75

Figure 16. Graph of EST and METH for the county estimate relative bias .......................... 75
Figure 17. Graph of EST, METH, and FEEDER for the county standard error relative bias ........................................................................................................................................................................ 77

Figure 18. Graph of EST and METH for the county estimate RMSE ................................ 79

Figure 19. Graph of EST and N for the convergence rate ............................................. 82

Figure 20. Graph of EST and IUCC for the convergence rate ........................................ 83

Figure 21. Graph of CORR for the convergence rate .................................................... 83
Chapter 1: Introduction

Chapter 1 establishes the research purpose of this study. This chapter introduces some cross-classified random-effect modeling background information, brings in the purpose of the study, identifies the research questions, addresses the significance of the study, and states the organization of this dissertation.

1.1 Background

In education and other social science research, researchers usually encounter data with multilevel structures, including cross-sectional structures (e.g., students nested within schools) and longitudinal structures (e.g., test scores over time nested within students). Over the past two decades, hierarchical linear models (HLMs), also known as multilevel models, mixed-effects models, random effects models, random coefficient regression models, or covariance component models, have rapidly developed and been applied to research in various fields, including education, social and behavioral sciences, psychology, and medical studies, to analyze multilevel datasets (Raudenbush & Bryk, 2002). In many educational or medical studies, however, the outcome variable is not continuous, but rather discrete (dichotomous, nominal, or ordinal), such as in depression studies or studies of whether students pass or fail a class. During recent years, social
science and education studies have yielded data that are both non-normally distributed and have an inherent hierarchical structure. To study datasets with such complex structures and discrete outcomes, hierarchical generalized linear models (HGLMs), also called generalized linear mixed models (GLMMs), were developed to expand the treatment of single-level generalized linear models to multilevel contexts (Brown & Prescott, 2006; Fielding, 2003; McCulloch & Searle, 2001).

With the development of multilevel analysis techniques during the past decade, researchers realized that instead of a pure hierarchy, some data structures involved cross-classified data which could not fit into HLMs. A common example is that students are usually nested within schools as well as in neighborhoods, yet schools can neither be considered as nested within neighborhoods, nor vice versa. In this case, students are cross-classified by school and neighborhood. To model data with a cross-classified structure, HLMs are no longer appropriate. Instead, cross-classified random effects models (CCREMs) have been developed to provide more accurate model estimations.

In recent years, CCREMs have been applied to various fields of study, including education research (e.g., students from different schools and neighborhoods), sociology (e.g., the effect of women’s place of birth and current place of residence on the choice of bearing a second child), as well as in business (e.g., business segment performance) and medical science (e.g., patients nested in different hospitals and neighborhoods).

Although applications of and methodological research on CCREMs have increased in recent years, there have been very limited studies using CCREMs to analyze data with dichotomous outcomes or methodological studies focusing on such models.
Some examples of previous studies include investigating the effect of ignoring the cross-classified structure in data with continuous outcome (Meyers & Beretvas, 2006), issues with missing data in cross-classified data structure (Alemdar, 2009), and the impact of omitting interaction among cross factors in CCREMs (Shi et al., 2010). However, these studies are all about data with continuous outcome, and there is a lack of investigation on the effect of ignoring the cross-classified structure in a cross-classified dataset with a dichotomous outcome.
1.2 Purpose of the Study

Given that many studies involve cross-classified data structures, but often ignore such structures and treat the data as purely hierarchical or even non-nested, it is essential for researchers to understand the consequences of ignoring cross-classified structures. Although previous studies have investigated the consequences of ignoring a cross-classified structure in a dataset with continuous outcome, there is a lack of literature on investigating the potential factors affecting the need to use a cross-classified random effects logistic model, as well as the impact of ignoring the cross-classification structure in a dataset with dichotomous outcome. To address this problem, this study expands on the work of Meyers and Beretvas (2006), which simulated a continuous outcome, and intends to: (1) explore the consequences of ignoring the cross-classified structure in a cross-classified dataset with a dichotomous outcome; (2) examine differences due to estimation method; and (3) clarify the conditions under which there is a need to use CCREM.

A secondary purpose of this study is to provide a demonstration of CCREM for dichotomous outcomes in an empirical dataset and focus on investigating possible predictors of recidivism (whether a person returns to the retained facility again after he/she is released) for incarcerated youth. In particular, this study examined the youth’s reading ability and the effect of a reading intervention called Read180 relative to traditional English instruction. Read180 provides 90 minutes of daily highly structured instruction, broken down into five components—whole group, individualized learning, computer activities, small group, and wrap up. It uses the “Read. Write. React.” books
(rBooks), which contain lessons in reading comprehension, vocabulary, writing, and grammar for the nine workshops that make up a year of Read180 instruction (Scholastic, Inc., 2005). The traditional English class is 45 minutes for a given day and has less time a week allocated to the class compared to Read180 classes. The students in the traditional classroom receive regular classroom instruction, using techniques such as paragraph and essay writing, etc. This research also aims to study the effects of environmental discrepancy, particularly the poverty rate of the county the youth come from, on their recidivism.

1.3 Research Questions

The present study is conducted to address the following overall research questions:

Given the various experimental conditions, what is the impact of ignoring the cross-classified structure for cross-classified data with dichotomous outcomes in terms of the (1) relative bias of fixed and random parameter estimates, (2) relative bias of standard error estimates, (3) relative bias of the root-mean-square error (RMSE), and (4) fit indices?

1.4 Significance of the Study

This methodological study will contribute to the literature on HGLM and CCREM by investigating the consequences of ignoring cross-classified structures in cross-classified data with a dichotomous outcome in terms of various parameter estimation and fit indices. Furthermore, this study addresses the difference between the estimation methods in cross-classified model estimation and results. It is expected to
provide useful methodological guidance as it simulates realistic unbalanced datasets for estimation, which are frequently encountered by education researchers. Relevant findings from the simulation studies may shed light on findings from similarly structured studies in the future.

Another aspect of this study is the importance of incarcerated youth’s reading ability and investigating the relationship between the reading program received in correctional facilities and the rate of recidivism to the department of youth services or adult detention facilities. In addition, identifying the effects of their environment on incarcerated youth’s recidivism is crucial for future relevant studies and contributes greatly to public safety and well-being by examining the connection between environmental poverty and youth’s recidivism. It is hoped that with consideration of the environmental effect, more effective reading programs can be provided to enhance the reading proficiency of incarcerated youth, with the intention of limiting the incidence of recidivism and educating them to become productive members after their re-entry into society. However, it should be noted that this study is an oversimplified demonstration of CCREM, thus it is hard to draw the conclusion that which predictors prevent recidivism.

1.5 Organization of the Dissertation

This dissertation consists of five chapters. Chapter 1 briefly introduces the research background, purpose, and research questions of this study. Chapter 2 focuses on introducing HGLM, estimation methods, cross-classified data structure, and CCREM, as well as reviewing other studies on inappropriate modeling of data with cross-classified
structures. Chapter 3 describes the simulation criteria and conditions, data analysis plan, and models used in the study, as well as the empirical data sample, predictors, and analysis models. Chapter 4 presents the main findings for the research questions. Chapter 5 provides the summary of major research findings, discussion of findings with respect to relevant literature, limitations of the study, and recommendations for the future.
Chapter 2: Literature Review

Chapter 2 is mainly devoted to methodological issues in HGLM and cross-classified random effect modeling CCREM. The HGLM models and estimation methods were first introduced. Then the CCREM literature was brought in, including previous studies on the inappropriate modeling of cross-classified data structures.

2.1 Hierarchical Generalized Linear Modeling (HGLM)

2.1.1 Multilevel Data Structure with Binary Outcome

With an increasing number of education and social science studies favoring more complex and nested designs, the HLM has been widely applied in various fields of research. Although many response variables of interest in the HLMs are on a continuous scale, they may also be non-normal in nature. Most often, these outcome variables will be binary. For example, in education, researchers may be interested in studying students’ graduation from high school (\(Y = 1\) if a student graduates from high school, or \(Y = 0\) if a student doesn’t graduate from high school). In medical research, scientists often want to study the presence of a disease (\(Y = 1\) if the disease is present; \(Y = 0\) if the disease is absent). In criminology, recidivism is usually studied (\(Y = 1\) if he/she returns during a given time interval; \(Y = 0\) if he/she doesn’t return). In each of these examples, there is
also the presence of a nested data structure: students nested in schools, patients nested in hospitals, and prisoners nested in facilities/buildings.

Although the nested data structure promotes the use of hierarchical modeling, HLM would not be appropriate as Raudenbush & Bryk (2002) mentioned in their book that: (a) the predicted values of level 1 have constrained intervals; (b) the level-1 random effects are not normally distributed; and (c) the level-1 random effects do not have homogeneous variance. In this case, the HLM assumptions of the linearity of the expected outcome at each level and the regression coefficients as well as the normality of the random effects at each level will no longer be held. To handle this type of data appropriately, researchers proposed the use of HGLMs.

2.1.2 Hierarchical Generalized Linear Models

HGLMs, referred to as multilevel analyses for binary, ordinal, count, or other types of non-normal data, are also known as generalized linear mixed models (Fielding, 2003; McCulloch & Searle, 2001; Breslow & Clayton, 1993) or generalized linear models with random effects (Schall, 1991). The application of generalized linear models in a multilevel framework parallels their use in single-level research design and the comparison between HLM and HGLM parallels the comparison in single-level models between the standard linear regression model and the generalized linear model (O’Connell, Goldstein, Rogers, & Peng, 2008; Raudenbush & Bryk, 2002).

In the single-level generalized linear models (GLMs), there are three related model components: (1) a random component, for which the dependent variable Y follows
one of the distributions from the exponential family, such as the normal, binomial, or inverse Gaussian; (2) a linear component, which describes how a function, $\eta$, of the mean of the dependent variable $Y$ depends on a collection of $Q$ explanatory variables; and (3) a link function, which describes the transformation of the expected value of the dependent variable $Y$ to $\eta$ (O’Connell, Goldstein, Rogers & Peng, 2008; Fox, 1997). GLMs use distributions drawn from the exponential family (Fox, 1997; Liao, 1994). The most commonly used link function for dichotomous data is the logit link function, which offers a connection between odds and probability. The logit, or log-odds, is defined as the natural log of the odds of success. The probability of success for the $i^{th}$ person, conditional on a collection of $q$ predictors, is $P(Y = 1 | X_{i1}, X_{i2}, ..., X_{iq}) = \pi_i(\chi)$. The logistic regression model predicts the logit for the $i^{th}$ person and the resulting structural model is linear in the parameters:

$$
\eta_i = \text{logit}[\pi_i(\chi)] = \ln\left(\frac{\pi_i(\chi)}{1 - \pi_i(\chi)}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_Q X_{iq}
$$

The parameter estimates (regression coefficients) in the above equation represent the expected change in the log-odds of success as each $X_q$ ($q = 1$ to $Q$) increases by one unit, holding the other predictors constant (O’Connell et al., 2010).

The sampling distribution of the error terms in the above model is Bernoulli instead of normal. In a Bernoulli distribution, the mean or expected value of the dichotomous outcome is the proportion of observations possessing the characteristic of interest (i.e., the “success” outcome) conditional on a specific set of values for the explanatory variables, and the variance of a Bernoulli distribution is a function of its
mean, $\pi_i(y)\times(1 - \pi_i(y))$, making the error variance heteroscedastic, since the variance itself depends on the conditional predicted values (O’Connell et al., 2010).

An HGLM extends the single-level generalized linear model to include the hierarchical nature in the dataset. A simple example would be of incarcerated youth sampled from within various counties in a state. The goals for the analysis would be to:

(a) estimate the contribution of the $q^{th}$ predictor at the youth-level (e.g., age, reading ability) to the log-odds of success (here success mean recidivism) and the odds of success;
(b) to investigate whether or not these variable effects differ across counties; and (c), if they do differ, to examine the contribution of county-level variables as predictors of this.

The most general form for the multilevel logistic model, based on the logit transformation, is:

$$
\text{Level 1: } \eta_{ij} = \logit[\pi_{ij}(x)] = \ln\left[\frac{\pi_{ij}(x)}{1 - \pi_{ij}(x)}\right] = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \ldots + \beta_{qj}x_{qij}
$$

$$
\text{Level 2: } \beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs}W_{sj} + u_{qj}
$$

Here, the collection of intercepts and regression slope estimates are allowed to vary across the $j$ counties. Each of these level-1 parameter estimates can be estimated by a collection of $S_q$ level-2 predictors, $W_{sj}$, where $s = 1$ to $S_q$. The number and nature of predictors in each level-2 equation can vary as well (O’Connell, 2010).
2.1.3 HGLM Estimation Methods

In many multilevel studies with continuous outcomes, maximum likelihood estimation (MLE) method is the most commonly used method. More specifically, two types of maximum likelihood estimation methods are usually referred to: full maximum likelihood (FEML) and restricted or residual maximum likelihood (REML). Researchers have noticed the difference between the two methods as well as their advantages and limitations (Raudenbush & Bryk, 2002).

However, when estimating the parameters in HGLM with dichotomous outcomes, the exact likelihood function usually involves an intractable high-dimensional integration and is hard to compute (Jang & Lim, 2006). To solve the problem, researchers proposed several approximations (i.e. quasi-likelihood procedures, higher-order Laplace approximations and numerical integration methods) to the likelihood function and approximate maximum likelihood estimators (Schall, 1991; Breslow & Clayton, 1993; McCulloch & Searle, 2001; Diaz, 2007). Among them, the most commonly used strategy is the penalized quasi-likelihood (PQL) proposed as an approximation to MLE optimizing a quasi-likelihood with a penalty term on the random effects. Instead of specifying a probability distribution for the data, the PQL only specifies the relationship between the mean and the variance.

Another estimation approach is the Laplace approximation, which can be approximated by careful Taylor expansion of the log-integrand (Tierney & Kadane, 1986). Laplace method is found to yield more accurate estimates than PQL, but slower to
converge (Bolker et al., 2008). This approximation also works well within large samples (O’Connell et al., 2010).

The other approach used in HGLM estimation is the numerical integration method achieved through Gauss-Hermite quadrature methods, in which the users have the options to specify the number of quadrature points and to choose to use a first or a second derivative approximation (Pinheiro & Bates, 2006). Two quadrature methods are usually used, ordinary quadrature (OQ) and adaptive quadrature (AQ). The OQ method is essentially a deterministic version of simple Monte Carlo integration and tends to yield better results if the procedure is adapted to center the process under the most peaked area of the response distribution (Rabe-Hesketh, Skrondal & Pickles, 2005; O’Connell et al., 2010). The AQ method, as an improvement over OQ, approximates the integral of a function using static quadrature rules on adaptively refined subintervals of the integration domain, but becomes progressively more difficult as the number of random effects per cluster increases (Raudenbush et al., 2000; McCulloch & Searle, 2001; Rabe-Hesketh, Skrondal & Pickles, 2005). It has been noted that the first-order Laplace approximation was the same as a one-point AQ solution (Pinheiro & Bates, 2006).

Researchers have noticed the difference between these estimation methods, in terms of estimates approximations, model comparison, as well as convergence rate. Studies have shown that although PQL is simple and fast, the variances of the random effects in HGLM yield a considerable downward bias when the PQL technique is used, especially when the random variance is large or the cluster proportions are close to zero or one (Raudenbush et al., 2000). Comparatively, better approximations for the
quadrature methods and Laplace’s method have been reported, especially as the number of quadrature points increases (O’Connell et al., 2010). Another benefit to the Laplace strategy and the numerical quadrature methods is that the deviance statistic produced by these procedures can be used in the chi-square difference test for model comparison purposes, while the PQL methods do not yield a reliable deviance statistic for likelihood-based comparisons across competing models (Snijders & Bosker, 1999). Finally, although the estimation through Laplace transformation overcomes the bias from PQL, it may have convergence problems for some datasets, and the numerical integration (i.e., quadrature) procedures are even slower than Laplace (O’Connell, 2010; Vonesh, 1996).

2.2 Cross-Classified Random Effect Models

2.2.1 Cross-Classified Data Structure

In many multilevel studies, data were treated as purely hierarchical. In such cases, for example, incarcerated youth were nested only within counties. Nonetheless, in practice, the hierarchical structure of the data is not always purely nested and simple. In reality, incarcerated youth could be nested within the facilities which they attend and within the counties from which they come. Sometimes the youth from the same counties go to the same facilities, but often this is not the case. Youth retained in the same facility usually come from different counties, depending on the severity of the crime committed, their age and gender, etc. Therefore, counties and facilities are said to be crossed with each other. Examples will be given below to explain and demonstrate the distinction between pure hierarchical data structure and cross-classified data structure by using
tables and network graphs. The table represents each youth’s affiliation with facilities and counties. The network graph labels each unit of each variable (youth, facility, and county) and represents lines as the connections (clusters) among elements across clustered and clustering variables. Alternative graphical technique was introduced by Goldstein (2003) for describing a dataset’s structure. Beretvas (2010) introduced network graphs as a better way for identifying the type of structure than Goldstein’s graphs. Depiction of a dataset’s structure using a table or a network graph can help identify whether data are purely clustered or cross-classified. For data to be considered purely clustered, clusters of elements from one level all belong to single elements of a higher level. If it is possible to construct a table with either exactly one cell per row or exactly one cell per column (but not both rows and columns) containing elements in it, or a network graph in which the lines do not cross, then the data’s structure can be considered a pure hierarchy (Rasbash & Browne, 2001; Beretvas, 2010). Table 1 lists a dataset demonstrating the pure hierarchical data structure with youth retained in the same facility come from the same county, in which columns represent one of the clustering variables (i.e., counties) and rows represent levels of the other clustering variables (i.e., facilities).

<table>
<thead>
<tr>
<th>Facility</th>
<th>County</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>A,B,C, D,E</td>
</tr>
<tr>
<td>II</td>
<td>F,G,H,I, J,K</td>
</tr>
<tr>
<td>III</td>
<td>L,M,N, O,P</td>
</tr>
<tr>
<td>IV</td>
<td>Q,R,S,T, U,V</td>
</tr>
</tbody>
</table>

Table 1. Pure three-level clustering of incarcerated youth within counties within facilities (modified from Beretvas, 2010)
In the above table, each column only has one cell with elements in it. For example, county 1 only has elements in the cell that corresponds to facility I. On the other hand, all rows except facility IV contain more than one cell with elements in them. For example, there are youth from county columns 1 and 2 who are retained in facility I. If only one cell per column has all of that column’s elements in it, then the column variable (i.e., county) is purely nested within the row variable (i.e., facility), and vice versa. It won’t affect the data structure if the column variable and row variable are reversed. Therefore, the above table indicates a three-level hierarchy within which youth are nested in counties, which are purely clustered within facilities.

Figure 1 contains a network graph of the same data structure contained in Table 1. For purely clustered data, it is possible to line up each (clustered or clustering) variable’s units such that no lines on the graph cross each other (Beretvas, 2010). In this figure, none of the lines that connect level-1 (youth) units with those of level-2 units and the lines connecting elements for level-2 (counties) with level-3 (facilities) cross.

Figure 1. Network graph of pure three-level clustering of incarcerated youth within counties within facilities (modified from Beretvas, 2010)
However, the pure hierarchy shown in Table 1 and Figure 1 is usually unrealistic. Usually, clustered data structures are not purely hierarchical. In the context of the current example, it is more likely that youth at the same facility do not come from the same county. Instead, the data might look more like the structure in Table 2. In Table 2, youth are nested within facilities and within counties as evidenced by some cells in the table containing more than one youth. The facility and county are both clustering variables for youth. Yet this differs from the above example, because counties are not nested within facilities, and facilities are not nested within counties either. This can be inferred from the table because not a single cell in rows contains all of that row’s elements, and neither do any in the columns. In this situation, we can say that the pattern of the data depicted in Table 2 involves a cross-classification of youth by county and facility. Thus, one set of units (i.e., youth) is nested within each of the classification variables (i.e., facility and county), but neither of the classified variables is purely nested within each other (Beretvas, 2010).

<table>
<thead>
<tr>
<th>Facility</th>
<th>County 1</th>
<th>County 2</th>
<th>County 3</th>
<th>County 4</th>
<th>County 5</th>
<th>County 6</th>
<th>County 7</th>
<th>County 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>A, B</td>
<td>D, E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>C</td>
<td>F, G</td>
<td>J, K</td>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td></td>
<td>H, I</td>
<td>M, N</td>
<td>O, P</td>
<td>Q, R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td></td>
<td></td>
<td>S, T</td>
<td>U, V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Cross-classified data structure of incarcerated youth cross-classified by counties and facilities (modified from Beretvas, 2010)
In addition to the table, the network graph in Figure 2 also indicates that youth are cross-classified by facility and county. The crossed lines in the figure supporting the cross-classification of youth nested within facility and county, but neither facility nor county are nested within each other. However, researchers need to be cautious that in a purely hierarchical data structure as above in Figure 1, if we change the order of the youth, the lines in the network graph could be crossed. However, that doesn’t indicate a cross-classified data structure (Beretvas, 2010).

Figure 2. Network graph of cross-classified data structure of incarcerated youth cross-classified by counties and facilities (modified from Beretvas, 2010)

The cross-classified structure could be more complicated with more than one level of clustering occurring. An example could be that in addition to the cross-classification of youth by facility and county, the clustering of youth within the classes in each facility might affect the outcome of interest. In that case, youth are purely nested within classes and facilities and cross-classified by county. Beretvas (2010) gives
detailed examples of the more complicated structures of cross-classified data, such as a three-level data structure in which the cross-classification occurs either at the highest level (level-3), at the middle level (level-2), or at the lowest level (level-1).

2.2.2 Cross-Classified Random Effect Modeling in the Literature

The cross-classified data structure is commonly encountered in all realms of social science, as well as in many other types of research. For cross-classified multilevel data, CCREM was used to investigate the relationships among variables within a given level and across levels (Rasbash & Goldstein, 1994; Raudenbush, 1993). Several multilevel textbooks provide chapters describing cross-classified models (Beretvas, 2010, 2008; Goldstein, 2003; Hox, 2002; Raudenbush & Bryk, 2002; Rasbash & Browne, 2001; Snijders & Bosker, 1999). Hox’s and Beretvas’s chapters included demonstrations of various multilevel software packages for estimating cross-classified models (Beretvas, 2010; Hox, 2002). A report by Fielding and Goldstein (2006) also provides a comprehensive resource offering several examples of cross-classified models (both estimation methods and software options).

Although cross-classified data structures are always encountered in social science, the use of CCREM is far from commonplace because of the complexity in the estimation and interpretation of CCREMs. In a search for published journal articles which applied the CCREM during the past two decades, only a limited number of studies were found, including the following fields: business (Hough, 2006), medicine (Ecob et al., 2004; Rasbash & Browne, 2001), political science (Park & Jensen, 2007), sociology (Seghieri,
Desantis, & Tanturri, 2006; Zaccarin & Rivellini, 2002; Teitler & Weiss, 2000), and education (Fielding, 2002; Goldstein & Sammons, 1997; Raudenbush, 1993). In these articles, the levels of hierarchy varied from three levels (Hough, 2006; Jayasinghe, Marsh, & Bond, 2003) to as many as five levels (Marsh, Martin, & Cheng, 2008). Also, the number of levels of each cross-classified factor varied greatly as well, from as few as eight (Van Tubergen, 2006) to thousands (Simonite & Browne, 2003; Fielding, 2002).

In addition to empirical studies, researchers also investigated the derivation of the estimation algorithms used (Browne, Goldstein, & Rasbash, 2001; Clayton & Rasbash, 1999), the impact of missing data in CCREM (Alemdar, 2009), as well as the link between the CCREM and the multilevel measurement model (MMM) (called the CCMMM model), and used the CCMMM model to validate instruments (Beretvas, Meyers, & Rodriguez, 2005; VandenNoortgate, DeBoeck, & Meulders, 2003). Sample size was also mentioned as an issue in cross-classified models. Researchers should be cautious when their average sample size of each level-2 unit is too small (i.e., an average sample size of six per level-2 unit) or when the degree of missing cells (empty cells in the cross-classified data structure) was large (Meyers & Beretvas, 2006; Goldstein & Sammons, 1997; Garner & Raudenbush, 1991).

2.2.3 Inappropriate Modeling of Cross-Classified Data Structures

Given the prevalence of cross-classified data structure in all fields of studies and the limited studies conducted using this method, there are a large number of studies ignoring the cross-classified data structure by treating it as purely hierarchical or even
nonhierarchical. However, as mentioned previously, just as ignoring the pure clustering inherent in the social science datasets could reduce the validity of associated statistical inferences, so would ignoring the cross-classified structure. Many researchers have conducted analyses disregarding the cross-classified structure of the data, usually by simply ignoring one of the cross-classified clustering variables or deleting the datasets which prevent the structure from being purely hierarchical (Meyers & Beretvas, 2006).

Researchers have found that ignoring a cross-classified factor will not lead to biased nor inconsistent fixed-effect estimates (Fielding & Goldstein, 2006); however, it usually will produce an under-specified model, which could result in negatively biased standard error estimates (Fielding & Goldstein, 2006; Meyers & Beretvas, 2006; Raudenbush & Bryk, 2002; Rasbash & Browne, 2001). The type I error rate of associated statistical tests, therefore, would be inflated given the deflated standard errors. Ignoring a cross-classified factor would also lead to inaccurate variance component estimation (Fielding & Goldstein, 2006; Rasbash & Browne, 2001). Meyers and Beretvas (2006) found that the variance component associated with the ignored cross-classified factor was instead associated with the non-ignored cross-classified factor, and thus the non-ignored cross-classified factor’s variance component would be overestimated. However, Fielding (2002) found that ignoring the crossed factor resulted in underestimated variance component of the remaining crossed factor. Luo and Kwok (2009) did further studies and found that with balanced design data structure, ignoring a crossed factor would cause overestimation of the variance components of adjacent levels and underestimation of the variance component of the remaining crossed factor.
Moreover, ignoring a crossed factor at the $k^{th}$ level would cause underestimation of the standard error of the regression coefficient of the predictor associated with the ignored factor and overestimation of the standard error of the regression coefficient of the predictor at the $(k-1)^{th}$ level. One year later, Shi et al. (2010) investigated the impact of the omission of the interaction effect on parameter estimates and standard errors and found that both coefficient estimates and accompanied standard error estimates for fixed effects were not biased. For random effects, in contrast, results were affected at level 2 but not at level 1 by the presence of an interaction variance and/or a correlation between the residual of level 2 factors.

If researchers decide to delete from the analysis the cross-classified data points, the sample size might be greatly decreased. After deletion, the data could be considered a purely hierarchical structure. However, this deletion strategy will result in less power and reduce the generalizability of the results (Meyers & Beretvas, 2006).

Either data deletion or ignoring the cross-classified factor strategy results in a loss of information about the ignored factor. Beretvas (2010) pointed out that while relationships among characteristics of that ignored the cross-classified factor and the outcome of interest could be modeled, the precision of estimation of this relationship would be biased if the variance component structure was inappropriately modeled. While it is possible to use corrections to correct for variance components that are not being explicitly modeled, this restricts the associated research questions that can be explored.
2.2.4 Cross-Classified Random Effects Modeling for Dichotomous Outcomes

Although more and more researchers have begun to apply CCREM in their studies, most of these studies use continuous responses. However, as described at the beginning of this chapter, there are many studies for which the response variables are not continuous (e.g., binary, counts, nominal and ordered categorizations, and the like).

As the modeling of cross-classified hierarchical structures parallels the modeling of purely hierarchical structures, CCREM for dichotomous outcomes parallels HGLM as well. The following simple model of a cross-classified data structure in which youth are cross-classified by facility and county will be used to present the formulation of the CCREM. The research question is to investigate the recidivism rate of these youth, thus the outcome variable would be dichotomous, and each $Y_{ij}$ would be a 0 or 1 (no or yes) for the $i^{th}$ student in the $j^{th}$ county and $k^{th}$ facility. Similar to that in HGLM, logit transformation will be the most common transformation method for the outcome variable.

$$
\eta_{i(jk)} = \logit[\pi_{i(jk)}(x)] = \ln \left[ \frac{\pi_{i(jk)}(x)}{1-\pi_{i(jk)}(x)} \right]
$$

Using the simplest two cross-classified factors, a two-level CCREM unconditional model includes no predictor variables in either level-1 or level-2, and estimates the variance between facilities, between counties, and within cells. Cells are the combination of the two cross-classified factors. The models are represented as:

**Level 1:**
$$
\eta_{i(jk)} = \beta_{0(jk)}
$$

**Level 2:**
$$
\beta_{0(jk)} = \gamma_{000} + u_{0j0} + u_{00k} + u_{0jk}
$$

**Combined:**
$$
\eta_{i(jk)} = \gamma_{000} + u_{0j0} + u_{00k} + u_{0jk}
$$
where $\eta_{i(jk)}$ is the log-odds (logits) of the recidivism probability rate of a youth $i$ in county $j$ and facility $k$. With no explanatory variables in the model, $\beta_{0(jk)}$ represents the log-odds of the recidivism rate for youth from the $j^{th}$ county in the $k^{th}$ facility. The average log-odds of recidivism rate of all youth is represented by $\gamma_{000}$. The level-2 model partitions the variation between cells into three components due to facility effects, county effects, and the interaction between the two factors. $u_{0j0}$ is the random effect of county $j$. This is assumed to be normally distributed with a mean of 0 and a constant variance of $\tau_{0j0}$. Similarly, $u_{00k}$ is the random effect of facility $k$, which is assumed to be normally distributed with a mean of 0 and a constant variance of $\tau_{00k}$. Finally, $u_{0jk}$ is the random interaction effect, which is assumed to be normally distributed with a mean of 0 and a constant variance of $\tau_{0jk}$. In many cases, the within-cell sample sizes are too small to distinguish the variance associated with the interaction effect, $\tau_{0jk}$. Thus the interaction effect is often dropped from the model, and the covariance among residuals is typically assumed to be zero (Meyers & Beretvas, 2006; Raudenbush & Bryk, 2002; Beretvas, 2010).

As with purely clustered multilevel generalized models, variance component estimation under the unconditional model can be used to calculate a form of intra-unit correlation coefficient (IUCC). As Snijders and Bosker (1999) discussed, there are two approaches to estimate the IUCC for dichotomous outcomes. The first approach involves fitting a multilevel linear probability model (that is, specifying an empty model and applying standard HLM to the dichotomous (0, 1) outcomes). The IUCC for two youth
who come from the same county and are retained at the same facility can be determined from the following equation:

\[ \rho_{0jk} = \frac{\tau_{0j0} + \tau_{00k} + \tau_{0jk}}{\tau_{0j0} + \tau_{00k} + \tau_{0jk} + \sigma^2} \]

For youth who come from the same county but stay in different facilities use:

\[ \rho_{0j0} = \frac{\tau_{0j0}}{\tau_{0j0} + \tau_{00k} + \tau_{0jk} + \sigma^2} \]

For youth who stay in the same facility but are from different counties use:

\[ \rho_{00k} = \frac{\tau_{00k}}{\tau_{0j0} + \tau_{00k} + \tau_{0jk} + \sigma^2} \]

Alternatively, the outcome, \( Y \), can be considered as a dichotomization of an unknown latent continuous variable, \( \tilde{Y} \), with a level-1 residual that follows the logistic distribution (Snijders & Bosker, 1999). The mean and variance of the logistic distribution are 0 and \( \frac{\pi^2}{3} \) (i.e., 3.29), respectively (Evans, Hastings, & Peacock, 2000). Accordingly, the \( \sigma^2 \) in the above equations will be replaced by 3.29 in the IUCC equation in the second method. In general, the results should be similar in the presence of a strong clustering effect, and under both approaches the variability attributed to groups is expected to be larger than zero (O’Connell et al, 2010).

A more complicated model is a random intercept model expressed as below. At level-1, one independent variable \( X \) is presented, and at level-2, two independent variables appear (\( W \) is the county-level factor, \( Z \) is the facility-level factor).
In the equations, the intercept $\beta_{0(jk)}$ is the log-odds of incarcerated youth’s recidivism rate when $X = 0$ for youth stay at facility $k$ and are from county $j$. The regression coefficient $\beta_{1(jk)}$ is the predicted change in the log-odds of youth’s recidivism rate associated with a one-point increase in $X$. $\gamma_{00}$ represents the expected log-odds of a youth’s recidivism rate with $X = 0$, $W = 0$ and $Z = 0$. The county effect on the intercept ($\gamma_{010}$) represents the increase (or decrease) in the estimated log-odds of recidivism as $W$ increases by one unit, holding other variables constant. Similar to the facility effect $\gamma_{001}$, the effect of $X$ is assumed to be constant across all counties and facilities. In the above cross-classified examples, the crossing occurred at the second level of the two-level model. It should be noted that cross-classifications can occur at any level in the data hierarchy.

2.2.5 CCREM Estimation and Software Options

In cross-classified random effect modeling for continuous variables, the ML estimation method was used. However, for dichotomous outcomes, ML is not appropriate. Specifically, the integration required for a solution in terms of maximizing the likelihood function becomes intractable and involves a considerable computational
load, involving extensive numerical integration techniques (Fielding & Goldstein, 2006; Hedeker & Gibbons, 1994). Thus, similar to HGLMs, alternative strategies, including quasi-likelihood procedures, numerical integration methods, and higher-order Laplace approximations, are generally used.

As mentioned earlier, quasi-likelihood functions are designed to have properties similar to true likelihood functions and yield an approximation to ML. The three major estimation methods: PQL, AQ, and Laplace are still appropriate to use in estimating the CCREM with a dichotomous outcome. Yet, to the author’s knowledge, no study has been conducted to evaluate the estimation procedures in CCREMs for data with dichotomous outcomes.

Possible software packages for cross-classified random effects generalized modeling include MIXIR and MIXNO, which were specifically designed for non-continuous response (Hedeker, 1999; Hedeker & Gibbons, 1996a, 1996b). Within STATA 11 (StataCorp, 2009), there are also advances in the use of GLAMM procedures, which use the improved method of numerical integration—that is, AQ (Rabe-Hesketh, Skrondal, & Pickles, 2004; Skrondal & Rabe-Hesketh, 2003). Practical illustrations of this use of STATA are provided by Rabe-Hesketh et al. (2005). MLwiN (Rasbash et. al., 2009; Goldstein et. al., 1998) adapts its iterative generalized least squares (IGLS) procedure and Monte Carlo Markov Chain (MCMC) procedure in a variety of ways to do this and makes possible the fitting of complex structures (Goldstein & Rasbash, 1992; Goldstein, 1986). The GLIMMX and NLMIX macro packages in SAS software (SAS/STAT, 2009) also use an approximate approach to fit CCREMs with various
estimation methods. In the most recent HLM software version 7 (Raudenbush et. al., 2011), cross-classified models can also be handled. While there are many software packages capable of fitting cross-classified random effects generalized modeling, these packages differ in terms of the models that can be selected and the estimation procedures used. Zhou, Perkins, and Hui (1999) compared several statistical packages for HGLM, and O’Connell et al. (2010) then did further study to compare five statistical packages (HLM, SAS, SUPERMIX, STATS, and R) in fitting HGLM models.
Chapter 3: Methodology

This study was comprised of two sections. The first section was a Monte Carlo simulation study designed to investigate the impact of correctly modeling versus neglecting to model cross-classified data under a variety of conditions, using different estimation methods. The second section was conducted as a follow-up to the simulation study. A CCREM was applied to an evaluation dataset to determine the county and facility effect as well as the relationship between youth reading ability, the effect of a reading program, and youth recidivism. This study extended Meyers and Beretvas’ (2006) work on CCREM for continuous outcomes and examined the effects of ignoring the data’s cross-classified structure for data with dichotomous outcomes.

3.1 Study 1: Monte Carlo Simulation

3.1.1 Conditions

Table 3 details the combinations of conditions that were manipulated in this simulation study. The conditions that were manipulated included the correlation of the residuals at level 2 (two conditions: 0.00 and 0.40), the number of juvenile incarceration facilities feeding into each county (two conditions: 3 and 4), the number of levels of each cross-classified factor (i.e., number of county and facility (two conditions: 7 facilities and
28 counties, and 7 facilities and 49 counties respectively), and the proportion of the total variance between facilities and counties (IUCC: two conditions: 10% and 15%). The fully crossed study involved 16 combinations, and under each of the 16 conditions, two estimation methods were used. Estimation was assessed for each of the 32 possible combinations. The total sample size in all conditions was set to be around 1000.

**Correlation of the Residuals at Level 2**

With real data, it is often (but not always) the case that the factors being crossed are somewhat related. For example, it is possible that youth from the same area (i.e., county) committed similar crimes and then were sent to the same juvenile correctional facility. In other words, there is usually some pattern as to which levels of cross-classified factors can be combined. To mimic reality, two conditions were used to represent the correlation of the residuals at level 2: one in which the facility and county conditional residuals were correlated moderately ($\rho = 0.40$) and the other in which there was no correlation ($\rho = 0.00$).

**Number of Facility Feeders**

As an indicator of the number of empty cells in the dataset, the number of facilities feeding into each county was manipulated by using the simulation technique that Meyers and Beretvas used (2006). In the empirical dataset, it was found that typically either three or four facilities “fed into” a given county. Therefore, the simulation conditions were set to three or four. The fewer the number of facility feeders,
the more empty cells exist in the dataset. Thus, there are more empty cells in the three-feeder facilities than in the four-feeder facilities.

**Number of Levels of Cross-Classified Factors**

To manipulate the overall sample size, which is around 1000 in the empirical dataset, two combinations of facility size and neighborhood size were used: 7 facilities and 28 counties, or 7 facilities and 49 counties. In previous cross-classified model studies, the two cross-classified factors were usually set to be equal (Meyers & Beretvas, 2006). However, that was not necessarily the case in real life. Actually, the cross-classified factors typically have different numbers of levels. In the juvenile recidivism data used in the empirical study, the number of facilities is 7, and the number of counties is 37. The number of feeder facilities and the number of levels of cross-classified factors yield the degree of cross-classification. Degree of cross-classification (DCC), the percent of non-empty cells, is defined as the percentage of non-empty cells out of the total number of cells:

\[
DCC = \frac{\text{Number of nonempty cells}}{\text{Cell size}} = \frac{\text{Number of feeder counties} \times \text{number of levels of cross-classified factor}}{\text{Number of levels of cross-classified factor} \times \text{number of levels of cross-classified factor}} = \frac{\text{Number of feeder counties}}{\text{Number of levels of cross-classified factors}}
\]
Intra-unit Correlation Coefficients (IUCCs)

Previous research found that among the studies involving the use of CCREM, as well as in multilevel modeling textbook examples, the actual observed estimates for IUCCs ranged from .009% to 24%, with a mean of 6.6% and standard deviation of .0682 (Meyers & Beretvas, 2006). However, more zero random effect estimates may emerge as the IUCC gets smaller. This study, therefore, uses the values 10% and 15% to represent moderate and large IUCCs for each C-C factor.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Correlation of residuals</th>
<th>Number of feeder facilities</th>
<th>Level of cross-classified factors (facility, neighborhood)</th>
<th>IUCC per cross-classified factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.00</td>
<td>3</td>
<td>7,28</td>
<td>.10</td>
</tr>
<tr>
<td>2</td>
<td>.00</td>
<td>3</td>
<td>7,28</td>
<td>.15</td>
</tr>
<tr>
<td>3</td>
<td>.00</td>
<td>3</td>
<td>7,49</td>
<td>.10</td>
</tr>
<tr>
<td>4</td>
<td>.00</td>
<td>3</td>
<td>7,49</td>
<td>.15</td>
</tr>
<tr>
<td>5</td>
<td>.00</td>
<td>4</td>
<td>7,28</td>
<td>.10</td>
</tr>
<tr>
<td>6</td>
<td>.00</td>
<td>4</td>
<td>7,28</td>
<td>.15</td>
</tr>
<tr>
<td>7</td>
<td>.00</td>
<td>4</td>
<td>7,49</td>
<td>.10</td>
</tr>
<tr>
<td>8</td>
<td>.00</td>
<td>4</td>
<td>7,49</td>
<td>.15</td>
</tr>
<tr>
<td>9</td>
<td>.40</td>
<td>3</td>
<td>7,28</td>
<td>.10</td>
</tr>
<tr>
<td>10</td>
<td>.40</td>
<td>3</td>
<td>7,28</td>
<td>.15</td>
</tr>
<tr>
<td>11</td>
<td>.40</td>
<td>3</td>
<td>7,49</td>
<td>.10</td>
</tr>
<tr>
<td>12</td>
<td>.40</td>
<td>3</td>
<td>7,49</td>
<td>.15</td>
</tr>
<tr>
<td>13</td>
<td>.40</td>
<td>4</td>
<td>7,28</td>
<td>.10</td>
</tr>
<tr>
<td>14</td>
<td>.40</td>
<td>4</td>
<td>7,28</td>
<td>.15</td>
</tr>
<tr>
<td>15</td>
<td>.40</td>
<td>4</td>
<td>7,49</td>
<td>.10</td>
</tr>
<tr>
<td>16</td>
<td>.40</td>
<td>4</td>
<td>7,49</td>
<td>.15</td>
</tr>
</tbody>
</table>

Table 3. Simulation conditions (There are two estimation methods used under each of the 16 conditions, so together there are 32 conditions)
Estimation Methods

Given that there exist different estimation methods but limited research has been conducted to compare these methods, especially under HGLM and CCREM, this study will compare the difference between the estimation methods within each of the simulated databases. Two estimation methods will be compared, PQL and AQ. As an approximation to the ML function, PQL is the most commonly used estimation method and is the default method in most statistical software. On the other hand, AQ was found to be preferred to other estimation methods in estimating parameters from previous research (O’Connell et al., 2010). Therefore, the simulated datasets under each of the conditions will be estimated by two models (HGLM and CCREM), and under each model, both estimation methods will be used.

3.1.2 Data Generation

This study used SAS software (SAS/STAT, 2009) to generate and analyze the data for each of the 500 replications per condition. In generating the two-level cross-classified random effects generalized model, responses of dichotomous outcome (recidivism) for each incarcerated youth will be generated, with youth at level 1 and facility and county crossed at level 2.

The following figure presents the data generation procedure flow chart. The first step of the procedure was to draw the population parameter values, including variable means and standard deviations for both level-1 and level-2 variables from the empirical dataset (Table 4). Then a correlated residual matrix of the two level-2 factors was
constructed by using SAS PROC IML with two conditions, uncorrelated ($\rho = 0.00$) and moderately correlated ($\rho = 0.40$). Later, the cross-classified data structure of the level-2 factors were generated in SAS with two conditions (7 facilities and 28 counties, 7 facilities and 49 counties). After that, the predictors were created by using parameters generated from the empirical dataset, and the outcome variable was generated by using SAS RANBIN command. This procedure was repeated 500 times.

Figure 3. The data generation procedure for the Monte Carlo study
A total of three predictors were included in the cross-classified data structure. Specifically, the level-1 predictor, \(X_1\) (Read Score) and \(X_2\) (Treatment), and the county predictor, \(W_1\) (Poverty), were included in this model. The parameter values used to generate the data were from the empirical dataset and are listed in Table 4.

<table>
<thead>
<tr>
<th>Fixed effect</th>
<th>Parameter value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\gamma_{000}), Intercept</td>
<td>-0.54</td>
</tr>
<tr>
<td>(\gamma_{100}), Read Score</td>
<td>-0.10</td>
</tr>
<tr>
<td>(\gamma_{200}), Treatment</td>
<td>0.26</td>
</tr>
<tr>
<td>(\gamma_{010}), Poverty</td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effect</th>
<th>Parameter value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\tau_{0j0}), County</td>
<td>0.6413 (0.8396)</td>
</tr>
<tr>
<td>(\tau_{00k}), Facility</td>
<td>0.6413 (0.8396)</td>
</tr>
</tbody>
</table>

Table 4. Population parameter for simulation

Using the notation introduced in Chapter 2, and substituting the generating parameter values results in the cross-classified random effect logistic modeling, we get:

**level 1:**
\[
\eta_{i(jk)} = \beta_{0(jk)} + \beta_{1(jk)}X_{1i(jk)} + \beta_{2(jk)}X_{2i(jk)}
\]

\[
\begin{align*}
\beta_{0(jk)} &= -0.54 + 0.05W_{10j0} + u_{0j0} + u_{00k} \\
\beta_{1(jk)} &= -0.10 \\
\beta_{2(jk)} &= 0.26
\end{align*}
\]

**level 2:**

**Combined:**
\[
\eta_{i(jk)} = (-0.54) + 0.05W_{10j0} + (-0.10)X_{1i(jk)} + 0.26X_{2i(jk)} + u_{0j0} + u_{00k}
\]
where $\eta_{i(jk)}$ is the log-odds of the probability of a recidivism for youth $i$ within the cross-classification of county $j$ and facility $k$. $X_{1(ijk)}$ is a continuous variable measuring the baseline reading CAT score for each incarcerated youth; and $X_{2(ijk)}$ is a dummy-coded variable that is used to refer to the treatment group which the incarcerated youth are in (either READ180 class or traditional English class). $W_{10j0}$ is a continuous variable at the county level, indicating the percentage of population below poverty level in that particular county.

As shown in Table 4, the generating values of the variances of each of $u_{0j0}$ and $u_{00k}$ were either 0.6413 (representing a conditional IUCC of 0.10) or 0.8396 (representing a conditional IUCC of 0.15). To get the two values for $u_{0j0}$ and $u_{00k}$ under the condition of IUCC = 0.10, the IUCC equation $\rho_{0j0} = \frac{\tau_{0j0}}{\tau_{0j0} + \tau_{00k} + \sigma^2}$ and $\rho_{00k} = \frac{\tau_{00k}}{\tau_{0j0} + \tau_{00k} + \sigma^2}$ were solved, where $\tau_{0j0} = \tau_{00k}$, $\sigma^2 = 3.29$, and $\rho_{0j0} = \rho_{00k} = 0.10$. Similarly, for the condition of IUCC = 0.15, the equation $\rho_{0j0} = \rho_{00k} = 0.15$ should be solved instead.

The probability of recidivism would be generated first by using following equation:

$$p = \exp\left(\frac{(-0.54) + 0.05W_{10j0} + (-0.10)X_{1(ijk)} + 0.26X_{2(ijk)} + u_{0j0} + u_{00k}}{1 + [(-0.54) + 0.05W_{10j0} + (-0.10)X_{1(ijk)} + 0.26X_{2(ijk)} + u_{0j0} + u_{00k}]}ight)$$

Then the recidivism variable would be generated in SAS software by calling the “ranbin” command plugging in the probability above.
Combination of county and facility cross-classified data structure

To set up the cross-classified structure for the combination of county and facility is an important step in this study. In reality, it is unlikely that there are youth in every possible combination of facility and county. In the empirical dataset, it was found that generally, there are three or four facilities feeding into each county. Following Meyers and Beretvas’ (2006) step, a matrix of facility and county residuals was generated to correlate on average either to 0.40 or 0.00. Then the matrix was sorted by county residual in ascending order. According to the conditions in Table 3, the cell combinations were generated. Because the two cross-classified factor were generated to be unequal number (7 and 28, or 7 and 49), the following two tables displayed the method for data structure construction.

<table>
<thead>
<tr>
<th>Facility</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>County</td>
<td>1</td>
<td>33%</td>
<td></td>
<td></td>
<td>33%</td>
<td></td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>33%</td>
<td>33%</td>
<td></td>
<td></td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>33%</td>
<td>33%</td>
<td>33%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>33%</td>
<td>33%</td>
<td>33%</td>
<td>33%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>33%</td>
<td>33%</td>
<td>33%</td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>33%</td>
<td>33%</td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33%</td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td>33%</td>
<td>33%</td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>33%</td>
<td></td>
<td></td>
<td>33%</td>
<td></td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>33%</td>
<td>33%</td>
<td></td>
<td></td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td>33%</td>
<td>33%</td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td>33%</td>
<td>33%</td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td>33%</td>
<td>33%</td>
<td>33%</td>
</tr>
</tbody>
</table>

Table 5. Three facilities feeding to each county
When there were three facility feeders, as shown in Table 5, the youth in county one consisted of youth from three facilities (i.e., youth in county one were assigned to facilities one, six, and seven), and each facility fed the same number of youth into one county (33% each). Similarly, the youth in county two were from facilities one, two, and seven. This pattern continued until county 28 or county 49, where youth were assigned to facilities five, six, and seven. Because there were only seven facilities but more than seven counties, youth from one facility could correspond to multiple counties. The feeding pattern wrapped around for the first and last facility in the dataset. So, for

<table>
<thead>
<tr>
<th>County</th>
<th>Facility 1</th>
<th>Facility 2</th>
<th>Facility 3</th>
<th>Facility 4</th>
<th>Facility 5</th>
<th>Facility 6</th>
<th>Facility 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>25%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>25%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>25%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25%</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25%</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Four facilities feeding to each county
example, county eight would consist of youth from facilities one, six, and seven again, the same as for county one.

Similarly, as shown in Table 6, each of the four feeding facilities fed 25% of youth into their county of origin. Because the total sample size was fixed at approximately 1000, the generated sample size for each county varied depending on the number of feeder facilities and number of county levels. For example, for a condition with seven facilities and 28 counties, and with three facility feeder structures, each 33% in the table cell represents around 12 youth; however, for the same feeder number, but with seven facilities and 49 counties, each 33% in the table cell represents only around seven youth.

The above process was conducted again in the conditions for which the correlation between residuals was 0, except that no ranking was done. Therefore, counties were randomly paired with facilities.

3.1.3 Data Analysis

The models in both studies were constructed in SAS software using the SAS PROC GLIMMIX macro package. The use of SAS in the analysis provides more control over the options for model estimation. GLIMMIX fits GLMMs, which assume that the random effects are Gaussian, and the response data can have any distribution within the exponential family, conditional on these random effects. The distributions available within GLIMMIX include the normal (Gaussian), binary, binomial, Poisson, negative binomial, gamma, geometric, inverse Gaussian, lognormal, multinomial, and the
Student’s t. Six estimation methods are available in SASv9.2 for GLIMMIX. Four of these methods are pseudo-likelihood methods (rspl, mspl, rmpl, mmpl), with the default estimation method set to be the subject-specific residual pseudo-likelihood (rspl). The two newer methods are Laplace (laplace), and adaptive quadrature (quad). O’Connell et al. (2010) compared the estimation methods in HGLM and concluded that the adaptive quadrature method was preferred to other estimation methods.

For each of the 16 conditions, two models were estimated. Model 1 was a conditional cross-classified random effects model with two level-1 explanatory variables and one explanatory variable for one cross-classified factor “county.” The modeling procedure matched the simulating procedure. The following models were fit accordingly:

\[
\begin{align*}
\text{level 1:} & \quad \eta_{i(jk)} = \beta_{0(jk)} + \beta_{1(jk)} X_{1i(jk)} + \beta_{2(jk)} X_{2i(jk)} \\
\text{level 2:} & \quad \begin{cases} 
\beta_{0(jk)} = \gamma_{000} + \gamma_{010} W_{10j0} + u_{0j0} + u_{00k} \\
\beta_{1(jk)} = \gamma_{100} \\
\beta_{2(jk)} = \gamma_{200}
\end{cases} \\
\text{Combined:} & \quad \eta_{i(jk)} = \gamma_{000} + \gamma_{010} W_{10j0} + \gamma_{100} X_{1i(jk)} + \gamma_{200} X_{2i(jk)} + u_{0j0} + u_{00k}
\end{align*}
\]

where \(i\) indexes youth, \(j\) indexes counties, and \(k\) indexes facilities.

In this model, only the intercept term at level-1 (\(\beta_{0(jk)}\)) was allowed to vary. The cross-classified interaction term was not modeled, and the slope effects of the two level-1 predictors (\(\beta_{1(jk)} \beta_{2(jk)}\)) were fixed for simplicity.

Model 2 was a conditional hierarchical generalized linear model with county as the level-2 factor. In this case, the cross-classified data structure was ignored. The following model would be used to model the data:
where \( i \) indexes youth and \( j \) indexes counties.

In both models, two estimation methods were used for the analysis: penalized quasi-likelihood (PQL) procedure and adaptive quadrature (AQ).

The values of the fixed parameter estimates and random effect estimates of both models were summarized across the 500 replications for each of the 16 combinations of conditions. Also, the estimates of the standard errors associated with each coefficient were compared across models for each condition. These values then were compared to the known generating values, and the average differences were calculated. In total, four evaluation criteria were used: relative bias of estimates, relative bias of standard errors, RMSE, and fit indices.

### Relative Bias of Estimates

Relative bias has been used as a criterion to estimate the average difference between the parameter estimate and the true population parameter value over the true population value for each fixed and random effect estimate across replications (Meyers & Beretvas, 2006; Hoogland & Boomsma, 1998).

\[
B(\hat{\theta}_i) = \frac{\bar{\theta}_i - \theta_i}{\theta_i}, \; i = 1, 2, \ldots, t,
\]

---

**level 1:**  
\[ \eta_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} \]

**level 2:**  
\[
\begin{align*}
\beta_{0j} &= \gamma_{00} + \gamma_{01}W_{10j} + u_{0j} \\
\beta_{1j} &= \gamma_{10} \\
\beta_{2j} &= \gamma_{20}
\end{align*}
\]

**Combined:**  
\[ \eta_{ij} = \gamma_{00} + \gamma_{01}W_{10j} + \gamma_{10}X_{1ij} + \gamma_{20}X_{2ij} + u_{0j} \]

where \( i \) indexes youth and \( j \) indexes counties.
where $\theta_i$ is the value of the $i^{th}$ parameter, $\bar{\theta}_i$ is the mean of the $i^{th}$ parameter estimates across the 500 replications, and $t$ is the number of parameters to be estimated.

**Relative Bias of Standard Errors.** To summarize the standard error estimates, relative bias was calculated for those estimates as well. This required the calculation of two types of standard error estimates. First, for each coefficient estimate, the standard deviation of the estimates across the 500 replications was calculated, yielding the empirical standard error, $\hat{se}_\theta$. In addition, the mean of the standard error estimates associated with each coefficient was calculated, yielding the mean standard error estimate, $\hat{se}_\theta$. The relative bias of the standard error (Hoogland & Boomsma, 1998) is defined as:

$$B(\hat{se}_\theta) = \frac{\overline{\hat{se}_\theta} - \hat{se}_\theta}{\hat{se}_\theta}, \quad i = 1, 2, \ldots, t,$$

**Root-Mean-Square Error (RMSE).** A useful measure of overall precision is the root-mean-square error, which is the square root of the average squared difference between the estimate and the true parameter value (Collins et al., 2001) as follows.

$$\text{RMSE} = \sqrt{\frac{1}{R-1} \sum_{r=1}^{R} (\hat{\theta}_r - \theta)^2}$$

where $\hat{\theta}_r$ = the parameter estimate obtained from $r^{th}$ replication within each cell, $\theta$ = the corresponding true population parameter value, and $R$ = the number of replications within each cell.
This measure combines the concepts of bias and variability of estimation (Enders, 2001) since the MSE of an estimate is the sum of the squared bias and its variance. It is desirable that an estimate is more accurate (i.e., less biased) and has minimal variance. If parameter estimates are unbiased, MSE quantifies the sampling variance of an estimate which could be regarded as the precision aspect of the estimation. If the estimates are biased, MSE serves to quantify the overall bias and precision of estimation into a single numeric term.

**Fit Indices.** Finally, the pseudo-Bayesian Information Criterion (pseudo-BIC) was used to identify the best fitting model. As a commonly used fit index, pseudo-BIC is usually used when the models being compared are not nested, as is the case with this study. Smaller values for the pseudo-BIC indicate a better fit. Because the data were generated to fit a CCREM, a tally was kept describing the proportion of the 500 replications for which the CCREM was chosen over the HGLM to represent the proportion of correct model identifications.

All analyses were conducted using analyses of variance (2 × 2 × 2 × 2 × 2 ANOVAs) in the SPSS GLM package. The ANOVAs included four between-subjects factors (correlation of residuals at level-2, number of feeder facilities, level of cross-classified factors, and IUCC) and two within-subjects factors (modeling method and estimation method). The dependent variables were the relative bias of the estimates, standard error relative bias, and the RMSE for each fixed and random effect estimate. To simplify the complex structure, only the main effects of the between-subject factors were
estimated. For the within-subject factors, the main effects of the two within factors, the interaction between the within factors, and the interactions between the within factors and the between factors were all included. The statistical significance criterion was set to be $\alpha = 0.05$, following the previous research (Meyers & Beretvas, 2006). However, this alpha level may be too liberal due to the large number of replications for each condition. Therefore, in addition, for factors that resulted in statistical significance, $\eta^2$ effect sizes were computed as a measure of practical significance. Eta-squared is calculated by dividing the sum of squares for the effect by the total sum of squares.

3.2 Study 2: Empirical Data Analysis

3.2.1 Settings

Juvenile delinquency continues to be a serious problem across the United States. The number of youth incarcerated in correctional facilities has increased dramatically over the past 30 years. In the United States, estimates of the number of youth released annually from correctional placements range from roughly 100,000 (Griffin, 2005; Snyder, 2004) to up to 200,000 (Mears & Travis, 2004). Most of them were older, male, and disabled (Sickmund, Sladky, Kang, & Puzzanchera, 2008). The cost of youth detention is much higher than adult detention and costs approximately $88,000 per person per year (American Correctional Association, 2008; Pew Center on the States, 2008). Despite the high costs of juvenile incarceration, research has consistently found that the majority (more than 50%) of youth who spend time in juvenile correctional facilities will recidivate into the juvenile and/or adult penal system within a few years of
release (California Juvenile Justice Reentry, 2007; Trulson, Marquart, Mullings, & Caeti, 2005). Among incarcerated youth, repeat offenders or recidivists account for the majority of delinquency, and state-level juvenile recidivism rates as high as 55% have been reported (Snyder & Sickmund, 2006; Spencer & Jones-Walker, 2004). Reducing the likelihood that juvenile offenders will commit future offenses is a primary goal of the juvenile justice system.

While few studies have examined neighborhood-level predictors of juvenile recidivism, the effects of environmental forces have played a leading role in the development of criminological theory and juvenile justice policy. The investigation of the effects of neighborhoods on reoffending is very new in the field of criminology, with several studies finding that space does influence the likelihood of adult recidivism (Mears et al., 2008; Kubrin et al., 2007; Kubrin & Stewart 2006). This conclusion is supported by the even more sparse research available on the effects of spatial factors on juvenile recidivism (LeBaron, 2002; Simmons, 2001) and can be contrasted with a number of studies examining ecological explanations of delinquency (Sampson et al., 1999; Sampson & Groves, 1989; Bursik, 1988).

Furthermore, previous studies have detected that prevention programs targeting school engagement and academic success, treatment services during incarceration, post-release follow-up, and support in a comprehensive community-based care system have been effective in reducing recidivism (Rodriguez, 2007; Carney & Buttell, 2003; Bullis, Yovanoff, Mueller, & Havel, 2002). Among many of the treatment programs for incarcerated youth, reading programs have been found in many studies to be effective in
improving their reading skills and therefore reducing their recidivism (Brunner, 1993). However, few studies have been conducted to detect which education programs, especially reading programs, would be most efficient in preventing incarcerated youth from returning to incarceration after they have been released.

Thus, the goal of the study is to investigate the effect of a reading program, Read180, compared with a traditional English program, in preventing incarcerated youth from recidivism, as well as other potential individual-level factors and neighborhood-level factors which may influence their recidivism rates, in a hierarchical context with youth nested within counties and facilities. Read180, which provides 90 minutes of daily highly structured instruction, was developed by Scholastic as a reading intervention program that focuses on improving reading ability for youth who are reading below their proficiency level in grades 4 to 12.

3.2.2 Data

*Population and sample.* The target population in the used in the empirical study was all incarcerated youth in the department of youth services in a large Midwestern state (*n* = 6,218) within a period of four years (2006-2010). The sample in this study was restricted to youth who were: (a) assigned to the facilities for more than six months; (b) determined to be “below proficient” but above “below basic” in reading levels as assessed by the Scholastic Reading Inventory (SRI) when they entered the facility the first time; (c) non-high school graduates; and (d) intended to receive two or more quarters of treatment. The eligible youth were then randomly assigned to the treatment group

46
(Read180 class) and the comparison group (traditional English class). Across the four-year longitudinal dataset, there was a total of 6,218 youth. Of these, 1,824 youth were identified as eligible for the targeted intervention, with 977 (53%) assigned to Read180 classes, and 847 (47%) assigned to the traditional English classes. The sample size for the current study with recidivism as the outcome variable, however, was 1,058 (562 in Read180 classes, and 496 in traditional English classes). Since Scholastic makes the argument that only youth with two or more quarters exposure to Read180 should be included in any impact analyses, youth who were not supposed to have any Read180 treatment (e.g., they were in school for less than five weeks at any time during the first three years of the project) or who were supposed to have only one quarter of treatment were omitted from the analyses.

Figure 4 shows the flowchart for the population and sample size.
Figure 4. Sample flow chart of the empirical recidivism study
3.2.3 Variables

Study 2 was a preliminary study investigating the cross-classified random effects generalized linear modeling. Thus only a simple model was conducted, using the dichotomous variable recidivism (0 = not returned; 1 = returned) as the dependent variable, and baseline reading CAT scores, treatment groups, and county poverty level as predictor variables. The recidivism rate was modeled as being nested within the cross-classification of facility (n = 7) and county (n = 37).

Recidivism. Repeat offenders or recidivists account for the majority of incarcerated youth. A recidivist is typically defined as a youth whose second incarceration occurs within three years after the first incarceration (Katsiyannis & Archwamety, 1997). In this study, recidivism for a youth was defined by either of the following three conditions within the four-year range: 1) youth released from the department of youth services (DYS) and readmitted to the DYS; 2) youth released from the DYS and admitted to the adult prison; and 3) youth released from the DYS and on parole but put in juvenile/adult facility in a county. This definition of recidivism was made given the state’s intention to reduce the DYS admission rate (including the recidivism rate), as well as the closure of two facilities. When the youth returned at least one time to the DYS, adult facilities, or county facilities, they would be assigned “1” in the recidivism variable; otherwise, they were assigned “0” for the variable.
**Predictor variables.** At the level-1 model, only two youth-level variables (**ReadCAT, Treatment**) were used. **ReadCAT** was a continuous variable measuring the youth’s baseline reading ability when they first entered the DYS. Treatment was a dichotomous variable classifying which treatment group the youth were assigned to (0 = Read180 classes, 1 = traditional English classes). Both variables were adapted to measure whether reading ability had an effect on incarcerated youth’s recidivism behavior and whether there were significant differences between the two different reading programs after controlling for other factors. At the level-2 model, only one county-level variable (**Poverty**) was used to characterizing the percentage of people under the poverty line in that county. None of the variables were used at the facility level because: 1) the characteristics of facilities changed over time, and it was hard to find a stable variable to describe the facilities; and 2) there were only seven facilities, and any additional variable might bias the results.

3.2.4 Data Analysis

The original dataset was cleaned and prepared in IBM SPSS v. 19 (IBM Corporation, 2010). At level-1, youth with missing data in variable reading scores were listwise deleted from the dataset, and counties with fewer than five youth were also listwise deleted.

The analysis was conducted by using SAS software version 9.2 (SAS/STAT, 2009). Two HGLMs which ignored the cross-classified data structure, and one CCREM were fit in SAS software. Model 1 treated the youth as purely nested in counties and
ignored the facility factor. It involved recidivism as the dependent variable, \textit{ReadCAT} and \textit{Treatment} as the level-1 predictor variables, and \textit{Poverty} as the level-2 predictor variable. The model equations were as follows:

\textit{level 1:} \quad \eta_{ij} = \beta_{0j} + \beta_{1j} \text{ReadScore}_{ij} + \beta_{2j} \text{Treatment}_{ij}

\begin{align*}
\beta_{0j} &= \gamma_{00} + \gamma_{01} \text{Poverty}_{0j} + u_{0j} \\
\beta_{1j} &= \gamma_{10} \\
\beta_{2j} &= \gamma_{20}
\end{align*}

\textit{level 2:} \quad \eta_{ij} = \gamma_{00} + \gamma_{01} \text{Poverty}_{0j} + \gamma_{10} \text{ReadScore}_{ij} + \gamma_{20} \text{Treatment}_{ij} + u_{0j}

Combined: \quad \eta_{ij} = \gamma_{00} + \gamma_{01} \text{Poverty}_{0j} + \gamma_{10} \text{ReadScore}_{ij} + \gamma_{20} \text{Treatment}_{ij} + u_{0j}

In this model, only the intercept was allowed to vary, and all the slopes were fixed.

Model 2 also ignored the cross-classification factor, but it included the dummy-coded facility variables as predictors at level-1. In this case, youth were treated as only nested within counties.

\textit{level 1:} \quad \eta_{ij} = \beta_{0j} + \beta_{1j} \text{ReadScore}_{ij} + \beta_{2j} \text{Treatment}_{ij} + \beta_{3j} F_2 + \beta_{4j} F_3 \\
&\quad + \beta_{5j} F_4 + \beta_{6j} F_5 + \beta_{7j} F_6 + \beta_{8j} F_7

\begin{align*}
\beta_{0j} &= \gamma_{00} + \gamma_{01} \text{Poverty}_{0j} + u_{0j} \\
\beta_{1j} &= \gamma_{10} \\
\beta_{2j} &= \gamma_{20} \\
\beta_{3j} &= \gamma_{30} \\
\beta_{4j} &= \gamma_{40} \\
\beta_{5j} &= \gamma_{50} \\
\beta_{6j} &= \gamma_{60} \\
\beta_{7j} &= \gamma_{70} \\
\beta_{8j} &= \gamma_{80}
\end{align*}

\textit{level 2:} \quad \eta_{ij} = \gamma_{00} + \gamma_{01} \text{Poverty}_{0j} + \gamma_{10} \text{ReadScore}_{ij} + \gamma_{20} \text{Treatment}_{ij}

Combined: \quad \eta_{ij} = \gamma_{00} + \gamma_{01} \text{Poverty}_{0j} + \gamma_{10} \text{ReadScore}_{ij} + \gamma_{20} \text{Treatment}_{ij} + u_{0j}

+ \gamma_{30} F_2 + \gamma_{40} F_3 + \gamma_{50} F_4 + \gamma_{60} F_5 + \gamma_{70} F_6 + \gamma_{80} F_7 + u_{0j}
Where facility 1 was set to be the reference and only the intercept was allowed to vary.

Model 3 used the cross-classified generalized model with youth nested within both facility and county. The equations were as follows:

\[
\text{level 1:} \quad \eta_{i(jk)} = \beta_{0(jk)} + \beta_{1(jk)} \text{ReadScore}_{i(jk)} + \beta_{2(jk)} \text{Treatment}_{i(jk)}
\]

\[
\begin{align*}
\beta_{0(jk)} &= \gamma_{000} + \gamma_{010} \text{Poverty}_{0j0} + u_{0j0} + u_{00k} \\
\beta_{1(jk)} &= \gamma_{100} \\
\beta_{2(jk)} &= \gamma_{200}
\end{align*}
\]

\[
\text{Combined:} \quad \eta_{i(jk)} = \gamma_{000} + \gamma_{010} \text{Poverty}_{0j0} + \gamma_{100} \text{ReadScore}_{i(jk)} + \\
&\quad \gamma_{200} \text{Treatment}_{i(jk)} + u_{0j0} + u_{00k}
\]

It should be noted that the factor interaction term \(u_{0j0}\) was not modeled, given the complexity needed to estimate it as mentioned earlier.

The results generated from models 1, 2, and 3 were compared in terms of the estimates, standard errors, and the statistical significance of the fixed effects, as well as the estimates of the random effect components of the counties. In addition, the fit indices (i.e., pseudo-BIC) of the three models were compared to identify which model fit the data best.
Chapter 4: Results

This chapter provides results focused on investigating the consequences of ignoring the cross-classified structure in a cross-classified dataset with a dichotomous outcome, which estimation method yields more appropriate results, and under which conditions is there the need to use cross-classified random effect modeling. The results of both the simulation study and the empirical data study are presented.

4.1 Study 1: Monte Carlo Simulation

4.1.1 Fixed-Effect Estimates

\(X_j (\text{ReadCAT}): \text{Level-1 continuous predictor}

Relative bias of the fixed-effect estimate. The values of the fixed parameter estimates of the level-1 continuous predictor (\(\gamma_{100}\) for CCREM and \(\gamma_{10}\) for HGLM) were averaged across the 500 replications, in combinations of 16 conditions, two modeling methods (CCREM and HGLM), and two estimation methods (PQL, AQ) of the simulation study. The relative bias of the estimate statistics, the relative bias of the standard errors, and the RMSEs were generated by the equations provided in Chapter 3. All the estimate statistics were rounded to four decimal places. The relative bias of the estimates, the relative bias of the standard errors, and the RMSEs are provided in Table 7.
<table>
<thead>
<tr>
<th>Condition</th>
<th>Est. relative bias</th>
<th>SE relative bias</th>
<th>RMSE</th>
<th>Est. relative bias</th>
<th>SE relative bias</th>
<th>RMSE</th>
<th>Est. relative bias</th>
<th>SE relative bias</th>
<th>RMSE</th>
<th>Est. relative bias</th>
<th>SE relative bias</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.1084</td>
<td>-0.0063</td>
<td>0.0164</td>
<td>-0.1045</td>
<td>-0.0120</td>
<td>0.0163</td>
<td>-0.1086</td>
<td>-0.0067</td>
<td>0.0164</td>
<td>-0.1036</td>
<td>-0.0095</td>
<td>0.0162</td>
</tr>
<tr>
<td>2</td>
<td>-0.0458</td>
<td>-0.0149</td>
<td>0.0209</td>
<td>-0.0357</td>
<td>-0.0200</td>
<td>0.0210</td>
<td>-0.0459</td>
<td>-0.0144</td>
<td>0.0209</td>
<td>-0.0347</td>
<td>-0.0198</td>
<td>0.0210</td>
</tr>
<tr>
<td>3</td>
<td>-0.0816</td>
<td>-0.0216</td>
<td>0.0234</td>
<td>-0.0816</td>
<td>-0.0216</td>
<td>0.0234</td>
<td>-0.0819</td>
<td>-0.0203</td>
<td>0.0234</td>
<td>-0.0706</td>
<td>-0.0258</td>
<td>0.0234</td>
</tr>
<tr>
<td>4</td>
<td>-0.0417</td>
<td>0.0141</td>
<td>0.0292</td>
<td>-0.0417</td>
<td>0.0141</td>
<td>0.0292</td>
<td>-0.0422</td>
<td>0.0145</td>
<td>0.0292</td>
<td>-0.0217</td>
<td>0.0046</td>
<td>0.0296</td>
</tr>
<tr>
<td>5</td>
<td>-0.0743</td>
<td>0.0068</td>
<td>0.0247</td>
<td>-0.0680</td>
<td>0.0352</td>
<td>0.0241</td>
<td>-0.0745</td>
<td>0.0076</td>
<td>0.0246</td>
<td>-0.0601</td>
<td>0.0005</td>
<td>0.0247</td>
</tr>
<tr>
<td>6</td>
<td>-0.0716</td>
<td>-0.0064</td>
<td>0.0185</td>
<td>-0.0716</td>
<td>-0.0064</td>
<td>0.0185</td>
<td>-0.0716</td>
<td>-0.0062</td>
<td>0.0185</td>
<td>-0.0636</td>
<td>-0.0099</td>
<td>0.0184</td>
</tr>
<tr>
<td>7</td>
<td>-0.0417</td>
<td>0.0141</td>
<td>0.0292</td>
<td>-0.0417</td>
<td>0.0141</td>
<td>0.0292</td>
<td>-0.0422</td>
<td>0.0145</td>
<td>0.0292</td>
<td>-0.0217</td>
<td>0.0046</td>
<td>0.0296</td>
</tr>
<tr>
<td>8</td>
<td>-0.0709</td>
<td>-0.0089</td>
<td>0.0214</td>
<td>-0.0586</td>
<td>0.0216</td>
<td>0.0216</td>
<td>-0.0715</td>
<td>-0.0096</td>
<td>0.0214</td>
<td>-0.0596</td>
<td>-0.0156</td>
<td>0.0214</td>
</tr>
<tr>
<td>9</td>
<td>0.0024</td>
<td>0.0487</td>
<td>0.0244</td>
<td>0.0159</td>
<td>0.0447</td>
<td>0.0248</td>
<td>0.0017</td>
<td>0.0483</td>
<td>0.0244</td>
<td>0.0196</td>
<td>0.0390</td>
<td>0.0249</td>
</tr>
<tr>
<td>10</td>
<td>-0.0096</td>
<td>0.0305</td>
<td>0.0255</td>
<td>0.0038</td>
<td>0.0311</td>
<td>0.0258</td>
<td>-0.0009</td>
<td>0.0319</td>
<td>0.0254</td>
<td>0.0100</td>
<td>0.0223</td>
<td>0.0260</td>
</tr>
<tr>
<td>11</td>
<td>-0.0431</td>
<td>0.0450</td>
<td>0.0332</td>
<td>-0.0208</td>
<td>0.0417</td>
<td>0.0336</td>
<td>-0.0444</td>
<td>0.0476</td>
<td>0.0331</td>
<td>-0.0197</td>
<td>0.0353</td>
<td>0.0338</td>
</tr>
<tr>
<td>12</td>
<td>-0.0996</td>
<td>0.0131</td>
<td>0.0178</td>
<td>-0.0939</td>
<td>0.0266</td>
<td>0.0174</td>
<td>-0.1000</td>
<td>0.0137</td>
<td>0.0178</td>
<td>-0.0929</td>
<td>0.0101</td>
<td>0.0176</td>
</tr>
<tr>
<td>13</td>
<td>-0.0336</td>
<td>-0.0332</td>
<td>0.0362</td>
<td>-0.0107</td>
<td>-0.0512</td>
<td>0.0374</td>
<td>-0.0348</td>
<td>-0.0323</td>
<td>0.0362</td>
<td>-0.0091</td>
<td>-0.0431</td>
<td>0.0370</td>
</tr>
<tr>
<td>14</td>
<td>0.0224</td>
<td>0.0243</td>
<td>0.0348</td>
<td>0.0533</td>
<td>0.0025</td>
<td>0.0368</td>
<td>0.0215</td>
<td>0.0226</td>
<td>0.0348</td>
<td>0.0547</td>
<td>0.0073</td>
<td>0.0363</td>
</tr>
<tr>
<td>15</td>
<td>-0.0197</td>
<td>-0.0111</td>
<td>0.0299</td>
<td>-0.0027</td>
<td>-0.0288</td>
<td>0.0375</td>
<td>-0.0204</td>
<td>-0.0084</td>
<td>0.0298</td>
<td>0.0010</td>
<td>-0.0191</td>
<td>0.0304</td>
</tr>
<tr>
<td>16</td>
<td>-0.0804</td>
<td>0.0054</td>
<td>0.0191</td>
<td>-0.0728</td>
<td>0.0037</td>
<td>0.0190</td>
<td>-0.0808</td>
<td>0.0063</td>
<td>0.0191</td>
<td>-0.0713</td>
<td>0.0018</td>
<td>0.0190</td>
</tr>
</tbody>
</table>

Table 7. Bias statistics for the fixed-effect coefficient read score (see Table 3 for condition descriptions)
The relative bias of the CCREM estimates using the PQL method ranged from -0.1084 to 0.0224, with a mean of -0.0485 and a standard deviation (SD) of 0.0375; while the relative bias of the HGLM estimates using the PQL method had almost the same range (from -0.1086 to 0.0215, with a mean of -0.0490 and an SD of 0.0373). When the AQ method was adopted, however, the relative bias had a larger range and was more positive than using the PQL method (CCREM had a range from -0.1045 to 0.0533, with mean of -0.0381 and standard deviation of 0.0438; while HGLM had a range from -0.1036 to 0.0547, with mean of -0.3196 and standard deviation of 0.0447).

To better distinguish the difference between the relative bias of the estimates, a 2 × 2 × 2 × 2 × 2 repeated measure ANOVA was conducted with four between-subject factors: level-2 factor correlation (CORR), number of feeders (FEEDER), level of cross-classified factors (N), IUCC, and two within-subject factors: modeling method (METH) and estimation method (EST) included in the model.

The ANOVA results indicated that none of the between-subject factor main effects were significant, but the interaction between METH and EST was statistically significant ($F_{(1,11)} = 10.877, p = 0.007, \eta^2 = 0.497$), the interaction between EST and CORR was statistically significant ($F_{(1,11)} = 6.894, p = 0.024, \eta^2 = 0.385$), and the main effects of the two within-subject factors were significant as well. Figures 6 and 7 illustrate the two significant interactions. The AQ estimation method comparatively yielded more accurate estimates than the PQL method. Within the AQ estimation method, the HGLM method resulted in a smaller relative bias than CCREM. According to Figure 6, the AQ was found to yield more accurate results again, and moderately correlated
level-2 residuals resulted in less negatively biased estimates than non-correlated level-2 residuals.

Figure 5. Graph of EST and METH for the read score estimate relative bias

Figure 6. Graph of EST and CORR for the read score estimate relative bias

**Relative bias of the fixed-effect standard error.** For the relative bias of the standard errors, similarly, CCREM and HGLM yielded similar ranges when using the
PQL method (CCREM: ranged from -0.0332 to 0.0487, with a mean of -0.0048 and an SD of 0.0233; HGLM: ranged from -0.0323 to 0.0483, with a mean of -0.0054 and an SD of 0.0233). On the other hand, when the AQ method was used, CCREM generated a slightly larger range than HGLM (CCREM: ranged from -0.0512 to 0.0447, with a mean of -0.0014 and an SD of 0.0283; HGLM: ranged from -0.0431 to 0.0390, with a mean of -0.0027 and an SD of 0.0224).

The ANOVA tested the relative bias of the standard errors on the fixed effect of the level-1 continuous predictor (ReadCAT). Results indicated that neither the between-subject factor main effects nor the within-subject factors were significant, with the exception of the main effect of the estimation method ($F_{(1,11)} = 8.243, p = 0.015, \eta^2 = 0.428$). Figure 7 shows the significant within-subject factor effect. The AQ estimation method again comparatively yielded smaller but more accurate estimates than the PQL method.

![Figure 7](image.jpg)

Figure 7. Graph of EST for the read score standard error relative bias

*Root-Mean-Square Error (RMSE) of the fixed-effect estimate.* Root-mean-square error could be used as another measure of the overall consistent precision of the estimate.
Equation 35 was used to calculate the square root of the average squared difference between estimate and the true parameter. Using the same technique as for the relative bias of estimates and standard errors, RMSE was computed 500 times and averaged through the replications. Similarly, a small value of RMSE was preferred and indicated that the RMSE estimates across the 500 replications were more clustered around the true parameter. The RMSE statistics in the four within-conditions and 16 between-conditions were presented in Table 7 above.

The RMSEs of the fixed-effect continuous predictor were almost all within a 0.05 range (CCREM PQL: ranged from 0.0164~0.0362, with a mean of 0.0257 and an SD of 0.0067; CCREM AQ: ranged from 0.0163~0.0375, with a mean of 0.0264 and an SD of 0.0076; HGLM PQL: ranged from 0.0164~0.0362, with a mean of 0.0257 and an SD of 0.0067; and the HGLM AQ: ranged from 0.0162~0.0370, with a mean of 0.0260 and an SD of 0.0071). The ranges of the estimated RMSEs were very close, and an ANOVA did not indicate any significant effect existing in the within-subject effect as well as in the between-subject effect. This finding indicated that the estimates were accurate and this accuracy was consistent across the modeling method or estimation method.

$X_2 (Treatment):$ Level-1 dichotomous predictor

Relative bias of the fixed-effect estimate. Similar to the previous predictor, the values of the fixed parameter estimates of the level-1 dichotomous predictor ($\gamma_{200}$ for CCREM and $\gamma_{20}$ for HGLM) were averaged across the 500 replications. The relative bias
of the estimate statistics, the relative bias of the standard errors, and the RMSEs were generated and are provided in Table 8.
<table>
<thead>
<tr>
<th>Condition</th>
<th>Est. relative bias</th>
<th>SE relative bias</th>
<th>RMSE</th>
<th>Est. relative bias</th>
<th>SE relative bias</th>
<th>RMSE</th>
<th>Est. relative bias</th>
<th>SE relative bias</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.1223</td>
<td>-0.0150</td>
<td>0.0718</td>
<td>-0.1163</td>
<td>-0.0112</td>
<td>0.0715</td>
<td>-0.1224</td>
<td>-0.0157</td>
<td>0.0718</td>
</tr>
<tr>
<td>2</td>
<td>-0.1191</td>
<td>0.0075</td>
<td>0.1052</td>
<td>-0.1009</td>
<td>-0.0021</td>
<td>0.1063</td>
<td>-0.1190</td>
<td>0.0078</td>
<td>0.1052</td>
</tr>
<tr>
<td>3</td>
<td>-0.0668</td>
<td>-0.0458</td>
<td>0.1181</td>
<td>-0.0668</td>
<td>-0.0458</td>
<td>0.1181</td>
<td>-0.0678</td>
<td>-0.0462</td>
<td>0.1181</td>
</tr>
<tr>
<td>4</td>
<td>0.0168</td>
<td>0.0581</td>
<td>0.1715</td>
<td>0.0168</td>
<td>0.0581</td>
<td>0.1715</td>
<td>0.0171</td>
<td>0.0582</td>
<td>0.1913</td>
</tr>
<tr>
<td>5</td>
<td>-0.0789</td>
<td>0.0229</td>
<td>0.1213</td>
<td>-0.0619</td>
<td>0.0053</td>
<td>0.1240</td>
<td>-0.0797</td>
<td>0.0240</td>
<td>0.1211</td>
</tr>
<tr>
<td>6</td>
<td>-0.1324</td>
<td>0.0199</td>
<td>0.0899</td>
<td>-0.1324</td>
<td>0.0199</td>
<td>0.0899</td>
<td>-0.1323</td>
<td>0.0191</td>
<td>0.0899</td>
</tr>
<tr>
<td>7</td>
<td>-0.0156</td>
<td>-0.0115</td>
<td>0.1531</td>
<td>-0.0156</td>
<td>-0.0115</td>
<td>0.1531</td>
<td>-0.0167</td>
<td>-0.0116</td>
<td>0.1530</td>
</tr>
<tr>
<td>8</td>
<td>-0.0140</td>
<td>0.0696</td>
<td>0.0977</td>
<td>-0.0026</td>
<td>0.0562</td>
<td>0.0996</td>
<td>-0.0134</td>
<td>0.0689</td>
<td>0.0977</td>
</tr>
<tr>
<td>9</td>
<td>0.0424</td>
<td>-0.0278</td>
<td>0.1358</td>
<td>0.0495</td>
<td>-0.0399</td>
<td>0.1388</td>
<td>0.0413</td>
<td>-0.0587</td>
<td>0.1358</td>
</tr>
<tr>
<td>10</td>
<td>-0.0425</td>
<td>0.0029</td>
<td>0.1351</td>
<td>-0.0224</td>
<td>0.0032</td>
<td>0.1364</td>
<td>-0.0422</td>
<td>0.0036</td>
<td>0.1349</td>
</tr>
<tr>
<td>11</td>
<td>-0.0026</td>
<td>0.0180</td>
<td>0.1739</td>
<td>0.0355</td>
<td>0.0082</td>
<td>0.1781</td>
<td>-0.0030</td>
<td>0.0174</td>
<td>0.1738</td>
</tr>
<tr>
<td>12</td>
<td>-0.1069</td>
<td>0.0000</td>
<td>0.0823</td>
<td>-0.0972</td>
<td>-0.0032</td>
<td>0.0826</td>
<td>-0.1071</td>
<td>0.0005</td>
<td>0.0823</td>
</tr>
<tr>
<td>13</td>
<td>0.0782</td>
<td>0.0273</td>
<td>0.1750</td>
<td>0.1166</td>
<td>0.0032</td>
<td>0.1823</td>
<td>0.0777</td>
<td>0.0300</td>
<td>0.1744</td>
</tr>
<tr>
<td>14</td>
<td>-0.0795</td>
<td>0.0226</td>
<td>0.1799</td>
<td>-0.0566</td>
<td>0.0182</td>
<td>0.1837</td>
<td>-0.0777</td>
<td>0.0229</td>
<td>0.1796</td>
</tr>
<tr>
<td>15</td>
<td>-0.0013</td>
<td>-0.0455</td>
<td>0.1590</td>
<td>0.0322</td>
<td>-0.0702</td>
<td>0.1651</td>
<td>-0.0045</td>
<td>-0.0431</td>
<td>0.1585</td>
</tr>
<tr>
<td>16</td>
<td>-0.0970</td>
<td>-0.0021</td>
<td>0.0935</td>
<td>-0.0983</td>
<td>-0.0038</td>
<td>0.0942</td>
<td>-0.0971</td>
<td>-0.0013</td>
<td>0.0934</td>
</tr>
</tbody>
</table>

Table 8. Bias statistics for the fixed-effect coefficient treatment (see Table 3 for condition descriptions)
Compared to the continuous predictor, there were more biases in estimating the dichotomous predictor. The relative bias of the CCREM estimates with the PQL method ranged from -0.1324 to 0.0782, with a mean of -0.0463 and an SD of 0.0636. The relative bias of the HGLM estimates with the PQL method again had almost the same range from -0.1323 to 0.0777, with a mean of -0.0467 and an SD of 0.0632. Similar to the continuous predictor $X_1$, when the AQ method was adopted, the relative bias had a larger range and was more positive than using the PQL method (CCREM: ranged from -0.1324 to 0.1166, with a mean of -0.0325 and an SD of 0.0704; HGLM: ranged from -0.1249 to 0.1060, with a mean of -0.0295 and an SD of 0.0699).

Given the difference we found, a repeated measure ANOVA was conducted in the SPSS GLM package similar to the ANOVAs above. The ANOVA results indicated that only the interaction between EST and CORR was statistically significant ($F_{(1,11)} = 5.82$, $p = 0.034$, $\eta^2 = 0.346$), and the main effect of the within-subject factor EST was significant, too. Figure 8 illustrates the significant interaction. A similar result to those above was obtained. The AQ estimation method yielded more accurate estimates than the PQL method, and moderately correlated level-2 residuals resulted in less negatively biased estimates than non-correlated level-2 residuals.
Relative bias of the fixed-effect standard error. For the relative bias of standard errors, CCREM and HGLM yielded similar ranges, and within them, CCREM with PQL yielded the smallest range (-0.0458 to 0.0696) with a mean of 0.0063 and an SD of 0.0321, while CCREM with adaptive quadrature produced a more negative range and the largest range (-0.0702 to 0.0581) with a mean of -0.0009 and an SD of 0.0329.

The standard error relative bias ANOVA resulted in a similar conclusion. Only the main effect of the estimation method was statistically significant ($F_{(1,11)} = 4.970, p = 0.048, \eta^2 = 0.311$). Figure 9 shows the significant within-subject factor effect. The AQ estimation method was found to yield smaller but more accurate estimates than the PQL method.
Root-Mean-Square Error (RMSE) of the fixed-effect estimate. The RMSEs of this fixed-effect dichotomous predictor had a larger and more positively biased range than that of the above continuous predictor (CCREM PQL: ranged from 0.0718–0.1799, with a mean of 0.1289 and an SD of 0.0365; CCREM AQ: ranged from 0.0715–0.1837, with a mean of 0.1309 and an SD of 0.0381; HGLM PQL: ranged from 0.0718–0.1913, with a mean of 0.1300 and an SD of 0.0383; HGLM AQ: ranged from 0.0719–0.1852, with a mean of 0.1313 and an SD of 0.0383). Among them, HGLM with the PQL estimation method generated the largest range. An ANOVA test was then conducted to look into the difference.

The repeated measure ANOVA indicated that the interaction between EST and CORR was statistically significant ($F_{1,11} = 8.347, p = 0.015, \eta^2 = 0.431$), as was the interaction between EST and IUCC ($F_{1,11} = 5.693, p = 0.036, \eta^2 = 0.341$). The main effect of the within-subject factor EST was also significant. Figures 6 and 7 present the significant interaction effects. In a finding different from above, moderately correlated
level-2 residuals resulted in more positively biased estimates than non-correlated level-2 residuals, and the AQ estimation method yielded slightly more positively biased RMSEs than the PQL method. In the significant interaction of EST and IUCC, when the IUCC is larger, the RMSE is more accurate. Also, the AQ estimation method yielded more positively biased RMSEs than the PQL method.

This finding indicates that the AQ method produces more dispersed fixed-effect dichotomous estimates than the PQL method, and thus the PQL method is more stable than the AQ method. In addition, when there is no correlation between the level-2 residuals, or more variance is generated to exist between level-2 factors, the estimates were less dispersed.

![Graph of EST and CORR for the treatment estimate RMSE](image)

Figure 10. Graph of EST and CORR for the treatment estimate RMSE
Figure 11. Graph of EST and IUCC for the treatment estimate RMSE

\( W (Poverty): \text{level-2 predictor} \)

Relative bias of the fixed-effect estimate. Similarly, the values of the fixed parameter estimates of the level-2 predictor (\( \gamma_{010} \) for CCREM and \( \gamma_{01} \) for HGLM) were averaged across the 500 replications. The relative bias of the estimate statistics, the relative bias of the standard errors, and the RMSEs were generated and are provided in Table 9.
Table 9. Bias statistics for the fixed-effect level-2 coefficient poverty (see Table 3 for condition descriptions)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Est. relative bias RSPL</th>
<th>SE relative bias RSPL</th>
<th>RMSE RSPL</th>
<th>Est. relative bias AQ</th>
<th>SE relative bias AQ</th>
<th>RMSE AQ</th>
<th>Est. relative bias RSPL</th>
<th>SE relative bias RSPL</th>
<th>RMSE RSPL</th>
<th>Est. relative bias AQ</th>
<th>SE relative bias AQ</th>
<th>RMSE AQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.0823</td>
<td>-0.5559</td>
<td>0.0422</td>
<td>-0.0744</td>
<td>-0.5474</td>
<td>0.0419</td>
<td>-0.0825</td>
<td>-0.5559</td>
<td>0.0422</td>
<td>-0.0770</td>
<td>-0.5555</td>
<td>0.0423</td>
</tr>
<tr>
<td>2</td>
<td>-0.0211</td>
<td>-0.4839</td>
<td>0.0641</td>
<td>-0.0040</td>
<td>-0.4864</td>
<td>0.0653</td>
<td>-0.0218</td>
<td>-0.4839</td>
<td>0.0641</td>
<td>-0.0105</td>
<td>-0.4817</td>
<td>0.0646</td>
</tr>
<tr>
<td>3</td>
<td>-0.0128</td>
<td>-0.3489</td>
<td>0.0520</td>
<td>-0.0128</td>
<td>-0.3489</td>
<td>0.0520</td>
<td>-0.0139</td>
<td>-0.3491</td>
<td>0.0520</td>
<td>-0.0017</td>
<td>-0.3461</td>
<td>0.0525</td>
</tr>
<tr>
<td>4</td>
<td>-0.0655</td>
<td>0.0072</td>
<td>0.0893</td>
<td>-0.0655</td>
<td>0.0072</td>
<td>0.0893</td>
<td>-0.0665</td>
<td>0.0078</td>
<td>0.0893</td>
<td>-0.0280</td>
<td>-0.0286</td>
<td>0.0929</td>
</tr>
<tr>
<td>5</td>
<td>-0.0571</td>
<td>-0.3825</td>
<td>0.0609</td>
<td>-0.0484</td>
<td>-0.3728</td>
<td>0.0607</td>
<td>-0.0556</td>
<td>-0.3832</td>
<td>0.0610</td>
<td>-0.0414</td>
<td>-0.3811</td>
<td>0.0616</td>
</tr>
<tr>
<td>6</td>
<td>-0.0276</td>
<td>-0.5127</td>
<td>0.0522</td>
<td>-0.0276</td>
<td>-0.5127</td>
<td>0.0522</td>
<td>-0.0279</td>
<td>-0.5126</td>
<td>0.0521</td>
<td>-0.0196</td>
<td>-0.5110</td>
<td>0.0524</td>
</tr>
<tr>
<td>7</td>
<td>-0.0036</td>
<td>0.0082</td>
<td>0.0698</td>
<td>-0.0036</td>
<td>0.0082</td>
<td>0.0698</td>
<td>-0.0056</td>
<td>0.0089</td>
<td>0.0697</td>
<td>0.0192</td>
<td>-0.0283</td>
<td>0.0715</td>
</tr>
<tr>
<td>8</td>
<td>-0.0659</td>
<td>-0.4771</td>
<td>0.0651</td>
<td>-0.0402</td>
<td>-0.4649</td>
<td>0.0645</td>
<td>-0.0676</td>
<td>-0.4768</td>
<td>0.0651</td>
<td>-0.0554</td>
<td>-0.4760</td>
<td>0.0657</td>
</tr>
<tr>
<td>9</td>
<td>0.0172</td>
<td>-0.0869</td>
<td>0.0737</td>
<td>0.0509</td>
<td>-0.1031</td>
<td>0.0738</td>
<td>0.0178</td>
<td>-0.0857</td>
<td>0.0737</td>
<td>0.0385</td>
<td>-0.1212</td>
<td>0.0753</td>
</tr>
<tr>
<td>10</td>
<td>0.0616</td>
<td>-0.0432</td>
<td>0.0878</td>
<td>0.1101</td>
<td>-0.0745</td>
<td>0.0914</td>
<td>0.0620</td>
<td>-0.0418</td>
<td>0.0877</td>
<td>0.0890</td>
<td>-0.0790</td>
<td>0.0901</td>
</tr>
<tr>
<td>11</td>
<td>-0.0340</td>
<td>-0.0277</td>
<td>0.0758</td>
<td>-0.0079</td>
<td>-0.0761</td>
<td>0.0785</td>
<td>-0.0346</td>
<td>-0.0242</td>
<td>0.0755</td>
<td>-0.0055</td>
<td>-0.0643</td>
<td>0.0777</td>
</tr>
<tr>
<td>12</td>
<td>-0.0820</td>
<td>-0.5726</td>
<td>0.0546</td>
<td>-0.0609</td>
<td>-0.5668</td>
<td>0.0549</td>
<td>-0.0819</td>
<td>-0.5728</td>
<td>0.0547</td>
<td>-0.0732</td>
<td>-0.5721</td>
<td>0.0549</td>
</tr>
<tr>
<td>13</td>
<td>-0.0611</td>
<td>0.0464</td>
<td>0.0776</td>
<td>-0.0312</td>
<td>-0.1035</td>
<td>0.0818</td>
<td>-0.0584</td>
<td>-0.0433</td>
<td>0.0774</td>
<td>-0.0302</td>
<td>-0.0824</td>
<td>0.0797</td>
</tr>
<tr>
<td>14</td>
<td>-0.0689</td>
<td>-0.0189</td>
<td>0.0900</td>
<td>-0.0343</td>
<td>-0.0540</td>
<td>0.0938</td>
<td>-0.0695</td>
<td>-0.0179</td>
<td>0.0901</td>
<td>-0.0335</td>
<td>-0.0583</td>
<td>0.0936</td>
</tr>
<tr>
<td>15</td>
<td>0.0660</td>
<td>0.0149</td>
<td>0.0693</td>
<td>0.0820</td>
<td>-0.0231</td>
<td>0.0708</td>
<td>0.0637</td>
<td>0.0151</td>
<td>0.0694</td>
<td>0.0903</td>
<td>-0.0216</td>
<td>0.0713</td>
</tr>
<tr>
<td>16</td>
<td>-0.0379</td>
<td>-0.4897</td>
<td>0.0749</td>
<td>-0.0120</td>
<td>-0.4894</td>
<td>0.0759</td>
<td>-0.0376</td>
<td>-0.4898</td>
<td>0.0750</td>
<td>-0.0271</td>
<td>-0.4900</td>
<td>0.0755</td>
</tr>
</tbody>
</table>
The relative bias of the CCREM estimates using the PQL method ranged from -0.0823 to 0.066 (mean = -0.0297, SD = 0.0463). The relative bias of the HGLM estimates using the PQL method again had almost the same range from -0.0825 to 0.0637 (mean = -0.0299, SD = 0.046). Similar to the level-1 predictors, when the AQ method was adopted, the relative bias estimates had a larger range and were more positive than using the PQL method (CCREM: ranged from -0.0744 to 0.1101, mean = -0.0112, SD = 0.0518; HGLM: ranged from -0.077 to 0.0903, mean = -0.0104, SD = 0.0492).

Given the large difference among the ranges we found, a repeated measure ANOVA was conducted. The ANOVA results indicated that the three-way interaction among METH, EST, and CORR was statistically significant ($F_{(1,11)} = 5.192, p = 0.044, \eta^2 = 0.321$), as was the two-way interaction between EST and CORR, as well as the main effect of the within-subject factor EST. Figure 12 illustrates the significant three-way interaction.

Similar results as level-1 estimates were revealed. The AQ estimation method was found to yield more accurate estimates than the PQL method, and moderately correlated level-2 residuals resulted in less negatively biased estimates than non-correlated level-two residuals. For the modeling method, when the residual correlation is 0, HGLM produced less biased estimates than CCREM; however, when the residuals were moderately correlated, CCREM had more accurate estimates than HGLM. The main effect of METH indicated that there was not much difference between CCREM and HGLM in general.
Relative bias of the fixed-effect standard error. For the relative bias of the second-level predictor standard errors, the CCREM with the PQL method had the largest range from -0.5726 to 0.0464 (mean = -0.2452, SD = 0.2477). HGLM with the AQ method had the smallest range from -0.5721 to -0.0216 (mean = -0.2686, SD = 0.2227). Comparatively, the PQL estimation method yielded a larger range than the AQ method.

Figure 12. Graph of EST, METH, and CORR for the poverty estimate relative bias
The standard error relative bias ANOVA result showed that the interaction effect between EST and CORR was statistically significant ($F_{(1,11)} = 8.398$, $p = 0.014$, $\eta^2 = 0.433$), and the main effect of EST was statistically significant, too. Figure 13 shows the significant interaction effect.

Generally, the PQL estimation method yielded more accurate estimates than the AQ method, and moderately correlated level-2 residual conditions had more accurate positive estimates than non-correlated conditions. When the correlation was 0, there was not much difference between the two methods, but when the correlation was 0.4, the PQL method generated a smaller bias.

![Graph of EST and CORR for the poverty standard error relative bias](image)

Figure 13. Graph of EST and CORR for the poverty standard error relative bias

*Root-Mean-Square Error (RMSE) of the fixed-effect estimate*. The RMSEs of the level-2 predictor have a positive bias range which is around 0.05 (CCREM PQL: ranged from 0.0422–0.0900, with a mean of 0.0687, and an SD of 0.0141; CCREM AQ: ranged from 0.0419–0.0938, with a mean of 0.0698 and an SD of 0.0152; HGLM PQL: ranged
from 0.0422~0.0901, with a mean of 0.0687 and an SD of 0.0141; HGLM AQ: ranged from 0.0423~0.0936, with a mean of 0.0701 and an SD of 0.0151). When using the PQL method, CCREM and HGLM almost yield the same range; when using the AQ method, however, HGLM had a larger range than CCREM.

The repeated measure ANOVA indicated that the three-way interaction among METH, EST, and CORR was statistically significant \((F_{1,11} = 5.26, p = 0.043, \eta^2 = 0.324)\), as well as the two-way interaction between EST and CORR and the main effect of EST. Figure 14 presents the significant interaction effects.

Similar results to the level-1 estimates were revealed. The AQ estimation method was found to yield more accurate estimates than the PQL method, and moderately correlated level-2 residuals resulted in less negatively biased estimates than non-correlated level-2 residuals. For the modeling method, when the residual correlation is 0, HGLM produced less biased estimates than CCREM; however, when the residuals were moderately correlated, CCREM had more accurate estimates than HGLM.

In a similar finding to the RMSE of the level-1 dichotomous predictor finding, the PQL estimation method was found to result in more accurate estimates than the AQ method. Also, moderately correlated level-2 residuals resulted in more positively biased estimates than non-correlated level-2 residuals. For the modeling method, when the residuals were non-correlated, CCREM and HGLM had very similar RMSEs when the PQL method was used. When the AQ method was applied, CCREM had a much smaller RMSE than HGLM did. When the residuals were moderately correlated, CCREM and HGLM had similar RMSEs.
This finding confirms the previous finding that the AQ method produces more dispersed estimates than the PQL method, and the PQL method, therefore, is more stable. Also, the non-correlated residual condition produced more stable estimates than the moderately correlated condition. HGLM seems to generate more dispersed RMSEs than CCREM, especially when the level-2 residuals are uncorrelated.

Figure 14. Graph of EST, METH, and CORR for the poverty estimate RMSE
4.1.2 Random Effect Estimates

\textit{Variance between level-2 factors: counties}

\textit{Relative bias of the random-effect estimate.} For the random effect estimates, relative bias was only presented when the level-2 factor residuals were uncorrelated. When the residuals were moderately correlated, the county and facility were not randomly paired; therefore, the true value of the variance components (the IUCC) was unknown. In the CCREM model, two random effects were estimated: county and facility, while in the HGLM model, only the county random effect was kept. The generated values for the variance between counties were either 0.6413 when the IUCC condition was 0.10 or 0.8396 when the IUCC was 0.15.

The relative biases of the random effect estimates (\(\tau_{0j}\) for CCREM and \(\tau_{0j}\) for HGLM) were also averaged across the 500 replications, in the combination of 16 conditions, two modeling methods (CCREM and HGLM), and two estimation methods (pseudo quasi-likelihood [PQL], adaptive-quadrature [AQ]) in the simulation study. The relative bias of the estimates, the relative bias of the standard errors, and the RMSEs are provided in Table 10.
<table>
<thead>
<tr>
<th>Condition</th>
<th>Est. relative bias</th>
<th>SE relative bias</th>
<th>RMSE</th>
<th>Est. relative bias</th>
<th>SE relative bias</th>
<th>RMSE</th>
<th>Est. relative bias</th>
<th>SE relative bias</th>
<th>RMSE</th>
<th>Est. relative bias</th>
<th>SE relative bias</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.8245</td>
<td>-0.0669</td>
<td>0.6949</td>
<td>-0.8250</td>
<td>-0.0869</td>
<td>0.6911</td>
<td>-0.8241</td>
<td>-0.0624</td>
<td>0.6946</td>
<td>-0.8301</td>
<td>-0.0823</td>
<td>0.6996</td>
</tr>
<tr>
<td>2</td>
<td>-0.6415</td>
<td>-0.0653</td>
<td>0.4219</td>
<td>-0.6438</td>
<td>-0.1031</td>
<td>0.4263</td>
<td>-0.6399</td>
<td>-0.0674</td>
<td>0.4210</td>
<td>-0.6521</td>
<td>-0.0941</td>
<td>0.4285</td>
</tr>
<tr>
<td>3</td>
<td>-0.7063</td>
<td>0.0598</td>
<td>0.4593</td>
<td>-0.7063</td>
<td>0.0598</td>
<td>0.4593</td>
<td>-0.7054</td>
<td>0.0610</td>
<td>0.4587</td>
<td>-0.7153</td>
<td>0.0314</td>
<td>0.4650</td>
</tr>
<tr>
<td>4</td>
<td>-0.1979</td>
<td>-0.0028</td>
<td>0.3125</td>
<td>-0.1979</td>
<td>-0.0028</td>
<td>0.3125</td>
<td>-0.1977</td>
<td>-0.0057</td>
<td>0.3128</td>
<td>-0.2006</td>
<td>-0.0582</td>
<td>0.3280</td>
</tr>
<tr>
<td>5</td>
<td>-0.6072</td>
<td>0.0673</td>
<td>0.4011</td>
<td>-0.6105</td>
<td>-0.0017</td>
<td>0.4045</td>
<td>-0.6068</td>
<td>0.0640</td>
<td>0.4009</td>
<td>-0.6190</td>
<td>0.0307</td>
<td>0.4087</td>
</tr>
<tr>
<td>6</td>
<td>-0.4045</td>
<td>-0.0283</td>
<td>0.3791</td>
<td>-0.4045</td>
<td>-0.0283</td>
<td>0.3791</td>
<td>-0.4039</td>
<td>-0.0295</td>
<td>0.3785</td>
<td>-0.4103</td>
<td>-0.0518</td>
<td>0.3838</td>
</tr>
<tr>
<td>7</td>
<td>-0.3990</td>
<td>-0.0064</td>
<td>0.3005</td>
<td>-0.3990</td>
<td>-0.0064</td>
<td>0.3005</td>
<td>-0.3973</td>
<td>-0.0073</td>
<td>0.2995</td>
<td>-0.4233</td>
<td>-0.0480</td>
<td>0.3140</td>
</tr>
<tr>
<td>8</td>
<td>-0.6328</td>
<td>0.0837</td>
<td>0.5410</td>
<td>-0.6329</td>
<td>0.0694</td>
<td>0.5463</td>
<td>-0.6326</td>
<td>0.0810</td>
<td>0.5409</td>
<td>-0.6422</td>
<td>0.0503</td>
<td>0.5488</td>
</tr>
<tr>
<td>9</td>
<td>-0.3925</td>
<td>-0.0114</td>
<td>0.2923</td>
<td>-0.4179</td>
<td>-0.0451</td>
<td>0.3046</td>
<td>-0.3890</td>
<td>-0.0104</td>
<td>0.2902</td>
<td>-0.4228</td>
<td>-0.0514</td>
<td>0.3078</td>
</tr>
<tr>
<td>10</td>
<td>-0.2212</td>
<td>-0.0379</td>
<td>0.2997</td>
<td>-0.2484</td>
<td>-0.0529</td>
<td>0.3111</td>
<td>-0.2182</td>
<td>-0.0344</td>
<td>0.2973</td>
<td>-0.2477</td>
<td>-0.0775</td>
<td>0.3124</td>
</tr>
<tr>
<td>11</td>
<td>-0.4001</td>
<td>0.0457</td>
<td>0.3061</td>
<td>-0.4195</td>
<td>-0.0397</td>
<td>0.3203</td>
<td>-0.3981</td>
<td>0.0447</td>
<td>0.3048</td>
<td>-0.4327</td>
<td>-0.0062</td>
<td>0.3248</td>
</tr>
<tr>
<td>12</td>
<td>-0.7590</td>
<td>-0.1179</td>
<td>0.6427</td>
<td>-0.7643</td>
<td>-0.1353</td>
<td>0.6469</td>
<td>-0.7587</td>
<td>-0.1156</td>
<td>0.6422</td>
<td>-0.7661</td>
<td>-0.1360</td>
<td>0.6484</td>
</tr>
<tr>
<td>13</td>
<td>-0.3960</td>
<td>-0.0032</td>
<td>0.3099</td>
<td>-0.4140</td>
<td>-0.0758</td>
<td>0.3232</td>
<td>-0.3898</td>
<td>-0.0036</td>
<td>0.3064</td>
<td>-0.4238</td>
<td>-0.0522</td>
<td>0.3262</td>
</tr>
<tr>
<td>14</td>
<td>-0.2380</td>
<td>-0.0300</td>
<td>0.3318</td>
<td>-0.2470</td>
<td>-0.0704</td>
<td>0.3451</td>
<td>-0.2301</td>
<td>-0.0284</td>
<td>0.3275</td>
<td>-0.2476</td>
<td>-0.0805</td>
<td>0.3456</td>
</tr>
<tr>
<td>15</td>
<td>-0.3719</td>
<td>-0.0474</td>
<td>0.2928</td>
<td>-0.3928</td>
<td>-0.1019</td>
<td>0.3045</td>
<td>-0.3665</td>
<td>-0.0399</td>
<td>0.2893</td>
<td>-0.3923</td>
<td>-0.0788</td>
<td>0.3032</td>
</tr>
<tr>
<td>16</td>
<td>-0.6373</td>
<td>-0.0890</td>
<td>0.5471</td>
<td>-0.6447</td>
<td>-0.0973</td>
<td>0.5530</td>
<td>-0.6367</td>
<td>-0.0870</td>
<td>0.5466</td>
<td>-0.6464</td>
<td>-0.1097</td>
<td>0.5544</td>
</tr>
</tbody>
</table>

Table 10. Bias statistics for the random effect coefficient county (see Table 3 for condition descriptions)
From the table, we find that there were more biases in the random-effect estimates than in the fixed-effect estimates. All of the relative biases were negative, indicating that the estimate of the variances between counties was underestimated. Generally, in both models and estimation methods, the ranges of the relative biases were around 0.6. Among them, when using the AQ method the ranges were larger (CCREM AQ ranged from -0.8301 to -0.2006, mean = -0.5014, SD = 0.1938; HGLM AQ ranged from -0.8250 to -0.1979, mean = -0.4980, SD = 0.1931). When using the PQL method, CCREM and HGLM had smaller ranges (CCREM ranged from -0.8245 to -0.1979, mean = -0.4894, SD = 0.1978; HGLM ranged from -0.8241 to -0.1977, mean = -0.4872, SD = 0.1991).

The following ANOVA test showed that the two-way interaction between METH and EST was statistically significant ($F_{(1,4)} = 25.435$, $p = 0.007$, $\eta^2 = 0.864$) and the two-way interaction between EST and IUCC was also significant ($F_{(1,4)} = 8.429$, $p = 0.044$, $\eta^2 = 0.678$). As before, the main effects of the within-subject factor EST and METH were significant too. Figures 16 and 17 illustrate the significant two-way interactions.

Interesting results were revealed in Figure 15. For the modeling method, when the CCREM was fit, the estimate biases were similar in both estimation methods. However, when HGLM was fit, the PQL estimation method had a smaller estimate relative bias than the AQ method. In addition, CCREM had a less biased random-effect relative bias than the HGLM model. Therefore, when HGLM was fit and ignoring the clustering at the facility level, the relative bias of the county variance estimate was more substantial and unstable across estimation methods. In both models, the random effect was negatively biased. Figure 16 yielded similar results as before. Larger variance
(IUCC = 0.15) resulted in more accurate random effect estimate. The PQL method had slightly more accurate estimates than the AQ method, but there were not many differences between the two methods.

Figure 15. Graph of EST and IUCC for the county estimate relative bias

Figure 16. Graph of EST and METH for the county estimate relative bias
Relative bias of the random-effect standard error. For the relative bias of county
standard errors, CCREM with the AQ method had the largest range (from -0.1353 to
0.0694, mean = -0.0449, SD = 0.0580). HGLM with the AQ method had the smallest
range (from -0.1360 to -0.0503, mean = -0.0509, SD = 0.0527). Generally, the ranges
were very similar in both modeling methods and estimation methods.

The standard error relative bias ANOVA resulted in a statistically significant
interaction effect between METH and FEEDER ($F_{(I,4)} = 8.597, p = 0.043, \eta^2 = 0.682$),
and the main effect of EST was statistically significant, too ($F_{(I,4)} = 25.469, p = 0.007, \eta^2$
= 0.864). However, it was noticeable that the three-way interaction of METH, EST, and
FEEDER was right at the edge of significance ($F_{(I,4)} = 7.342, p = 0.054, \eta^2 = 0.647$).
Their relationships are displayed in Figure 17.

The PQL estimation method was found to yield more accurate estimates than the
AQ method except when there were four feeders. More specifically, in the four-feeder
condition, the PQL method had a similar bias to the AQ method when using CCREM, but
the AQ method performed better than the PQL method when using HGLM. HGLM was
found to have slightly more accurate standard errors than CCREM across estimation
methods and number of feeders. In addition, the four-feeder condition has a more
accurate estimate than the three-feeder condition across estimation methods and modeling
methods.
Figure 17. Graph of EST, METH, and FEEDER for the county standard error relative bias

*Root-Mean-Square Error (RMSE) of the random-effect estimate.* The RMSEs of the random-effect estimate had a positively biased range, which is approximately 0.04. When using the PQL method, the RMSEs had slightly larger range than when using the AQ method (CCREM PQL ranged from 0.2923 to ~0.6949, mean = 0.4083, SD = 0.1324; CCREM AQ ranged from 0.3005 to ~0.6991, mean = 0.4148, SD = 0.1305; HGLM PQL
ranged from 0.2893 to ~0.6946, mean = 0.4069, SD = 0.1332; HGLM AQ ranged from 0.3032 to ~0.6996, mean = 0.4187, SD = 0.1289).

The repeated measure ANOVA again indicated a statistically significant two-way interaction between METH and EST ($F_{(1,4)} = 12.445, p = 0.024, \eta^2 = 0.757$), as well as the within-factor main effect METH and EST. Figure 18 presents the significant interaction effect. The figure illustrates that the PQL estimation method resulted in more accurate estimates than the AQ method. Similar to the relative biases of estimates, when CCREM was fit, the RMSEs were similar in both estimation methods. However, when HGLM was fit, different estimation methods generated much different RMSEs, and the AQ method was more unstable. As before, the AQ method produced more dispersed estimates than the PQL method, and thus the PQL method is more stable. In terms of the modeling method, the CCREM model had fewer biased estimates than the HGLM model. This confirms the previous finding that when HGLM was fit, and by ignoring the clustering at the facility level, the estimate of the county variance was more biased and unstable across estimation methods.
4.1.3 Fit Indices

In HGLM, model fit indices were usually used to evaluate the fit of the model to the data. Two commonly used criteria are the pseudo-Akaike information criterion (pseudo-AIC) and the pseudo-Bayesian information criterion (pseudo-BIC). Previous articles revealed that Bayesian model averaging had much outperformed competing methods for model selection with both theoretical results and simulation studies (Meyers & Beretvas, 2006; Raftery & Zheng, 2003). Therefore, in this study, pseudo-BIC was used as the criterion to measure the overall statistical fit of the CCREM and HGLM model with the two estimation methods. The percentage of replications in which the CCREM model was preferred to HGLM is provided in Table 11.

The pseudo-BIC correctly identified the CCREM as the better fitting model across all conditions with 100 percent accuracy for the AQ estimation method. But it only had an approximately 60 to 80 percent accuracy rate in the PQL estimation method.
Therefore, the AQ estimation method had a much better model fit than the PQL method across all the conditions. The ANOVA results indicated that there was no significant between-subject effect explaining the pattern of variability in the proportion of correct model identifications.

<table>
<thead>
<tr>
<th>Condition</th>
<th>PQL</th>
<th>AQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>76.20</td>
<td>100.00</td>
</tr>
<tr>
<td>2</td>
<td>71.00</td>
<td>100.00</td>
</tr>
<tr>
<td>3</td>
<td>64.40</td>
<td>100.00</td>
</tr>
<tr>
<td>4</td>
<td>66.20</td>
<td>100.00</td>
</tr>
<tr>
<td>5</td>
<td>62.20</td>
<td>100.00</td>
</tr>
<tr>
<td>6</td>
<td>64.00</td>
<td>100.00</td>
</tr>
<tr>
<td>7</td>
<td>68.00</td>
<td>100.00</td>
</tr>
<tr>
<td>8</td>
<td>69.20</td>
<td>100.00</td>
</tr>
<tr>
<td>9</td>
<td>67.40</td>
<td>100.00</td>
</tr>
<tr>
<td>10</td>
<td>67.80</td>
<td>100.00</td>
</tr>
<tr>
<td>11</td>
<td>61.60</td>
<td>100.00</td>
</tr>
<tr>
<td>12</td>
<td>71.80</td>
<td>100.00</td>
</tr>
<tr>
<td>13</td>
<td>73.40</td>
<td>100.00</td>
</tr>
<tr>
<td>14</td>
<td>69.60</td>
<td>100.00</td>
</tr>
<tr>
<td>15</td>
<td>68.00</td>
<td>100.00</td>
</tr>
<tr>
<td>16</td>
<td>70.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Table 11. Pseudo-BIC fit indices of percentage CCREM was preferred to HGLM

4.1.4 Convergence rate

Previous research has found that although the AQ method tends to overcome the bias in the estimates from the PQL method for binary response models, it usually has convergence problems and is harder to converge (Vonesh, 1996; O’Connell et al., 2010).
Findings from this study also agreed with the previous studies. Table 12 showed the CCREM convergence rate for both the PQL and AQ methods. Across the 16 conditions, AQ consistently had a smaller convergence rate than the PQL method.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Convergence rate in CCREM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PQL</td>
</tr>
<tr>
<td>1</td>
<td>47.80</td>
</tr>
<tr>
<td>2</td>
<td>40.80</td>
</tr>
<tr>
<td>3</td>
<td>40.00</td>
</tr>
<tr>
<td>4</td>
<td>38.00</td>
</tr>
<tr>
<td>5</td>
<td>39.00</td>
</tr>
<tr>
<td>6</td>
<td>39.20</td>
</tr>
<tr>
<td>7</td>
<td>39.00</td>
</tr>
<tr>
<td>8</td>
<td>41.40</td>
</tr>
<tr>
<td>9</td>
<td>43.00</td>
</tr>
<tr>
<td>10</td>
<td>41.40</td>
</tr>
<tr>
<td>11</td>
<td>42.40</td>
</tr>
<tr>
<td>12</td>
<td>46.80</td>
</tr>
<tr>
<td>13</td>
<td>47.40</td>
</tr>
<tr>
<td>14</td>
<td>47.00</td>
</tr>
<tr>
<td>15</td>
<td>43.40</td>
</tr>
<tr>
<td>16</td>
<td>51.00</td>
</tr>
</tbody>
</table>

Table 12. Convergence rate of CCREM

A repeated measure ANOVA was then conducted to test the between-subject and within-subject effects on the convergence rate. The results indicated statistically significant two-way interactions between EST and N ($F_{(1,11)} = 15.487, p = 0.002, \eta^2 = 0.885$), between EST and IUCC ($F_{(1,11)} = 9.194, p = 0.011, \eta^2 = 0.627$), as well as the between-subject main effect CORR ($F_{(1,11)} = 8.150, p = 0.016, \eta^2 = 0.572$), and within-subject main effect EST. Figure 19 presents the significant interaction effect between
EST and N, Figure 20 presents the significant interaction between EST and IUCC, and Figure 21 presents the significant effect of CORR. The figures illustrate that the PQL estimation method resulted in a higher convergence rate than the AQ method. The 7 facilities and 28 counties condition had higher convergence rate than the 7 facilities and 49 counties condition. Also, larger IUCCs resulted in higher convergence rates, too. For the level-2 correlation, moderately correlated conditions had higher convergence rates than non-correlated conditions.

This finding confirms the literature review that the AQ method has more convergence problems than PQL method. In addition, the finding also illustrated that when there are fewer empty cells in the cross-classified data structure (e.g., 7 facilities and 28 counties), when the IUCC was larger (e.g., 0.15), and when the level-2 factors were moderately correlated (e.g., 0.4), the convergence rate would be higher.

Figure 19. Graph of EST and N for the convergence rate
4.2 Study 2: Empirical Data Analysis

4.2.1 Descriptive Statistics

In the real dataset, there were 1,058 youth nested within 37 counties and 7 facilities. There were an average of 28.59 youth per county and 151.14 youth per facility. Typically, youth from three or four facilities fed into a given county. The dependent variable was recidivism, in which recidivated youth were coded as 1, and unrecidivated
youth were coded as 0. Among the 1,058 youth, n=483 (46%) youth returned to the DYS facilities at least one time after their release, and n=575 (54%) youth had not returned after their release. The two level-1 independent variables were ReadCAT and Treatment Group. ReadCAT ranged from 0.6 to 12.9 with a mean of 6 and a standard deviation of 2.61. 562 (53%) youth were assigned to the Read180 class, and 496 (47%) were assigned to the traditional class. There was one county-level independent variable (poverty). Among the 37 counties, the lowest percentage of poverty rate is 5.8% and the highest is 21.9% with a mean of 14.62% and standard deviation of 2.08%. Descriptive statistics of the independent and dependent variables were included in the following Table 13.

<table>
<thead>
<tr>
<th>Variable name</th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recidivism</td>
<td>1,058</td>
<td>0</td>
<td>1</td>
<td>0.46</td>
<td>.50</td>
<td>0.18</td>
<td>-1.97</td>
</tr>
<tr>
<td>ReadCAT</td>
<td>1,058</td>
<td>0.60</td>
<td>12.90</td>
<td>6.00</td>
<td>2.61</td>
<td>0.57</td>
<td>-0.19</td>
</tr>
<tr>
<td>Treatment group</td>
<td>1,058</td>
<td>0</td>
<td>1</td>
<td>0.47</td>
<td>0.50</td>
<td>0.13</td>
<td>-1.99</td>
</tr>
<tr>
<td>Percent of people under poverty line</td>
<td>1,058</td>
<td>5.80</td>
<td>21.90</td>
<td>14.62</td>
<td>2.08</td>
<td>-0.29</td>
<td>1.68</td>
</tr>
</tbody>
</table>

Table 13. Descriptive statistics

4.2.2 Results

As described in Chapter 3, three models would be used to fit the dataset: (1) an HGLM model which ignores the facility factor; 2) an HGLM model which treats the facility as a dummy-coded predictor; and 3) a CCREM model. In addition, the simulation results indicated that the AQ estimation method yielded a more accurate estimate of both
fixed effects and random effects, so only the AQ estimation method would be used to estimate the model.

<table>
<thead>
<tr>
<th>Effects</th>
<th>HGLM (facility ignored)</th>
<th>HGLM (facility included)</th>
<th>CCREM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SD</td>
<td>Estimate</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>γ₀₀</td>
<td>-0.6385</td>
<td>0.6494</td>
</tr>
<tr>
<td>Read</td>
<td>γ₁₀</td>
<td>-0.0910**</td>
<td>0.0250</td>
</tr>
<tr>
<td>Treatment</td>
<td>γ₂₀</td>
<td>0.1900</td>
<td>0.1267</td>
</tr>
<tr>
<td>Poverty</td>
<td>γ₀₁</td>
<td>0.0616</td>
<td>0.0435</td>
</tr>
<tr>
<td>Facility2</td>
<td>γ₃₀</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Facility3</td>
<td>γ₄₀</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Facility4</td>
<td>γ₅₀</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Facility5</td>
<td>γ₆₀</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Facility6</td>
<td>γ₇₀</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Facility7</td>
<td>γ₈₀</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Random effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-county</td>
<td>τ₀j</td>
<td>0.1357</td>
<td>0.0910</td>
</tr>
<tr>
<td>Between-facility</td>
<td>τ₀₀k</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

** p < .001

Table 14. Real data study’s parameter and standard error estimates

The table indicates that the three models resulted in similar estimates for the fixed-effect estimates and standard errors. The reading score value was found to be
significant across all of the models. Therefore, as each youth’s reading score increased by 1 unit, the youth’s recidivism probability would decrease by approximately 0.09. For the HGLM model in which facilities were dummy-coded at level 1, the variables facility 4 and facility 5 were statistically significant. Since facility 1 was treated as the reference facility, facility 4 and facility 5 had significantly lower recidivism rates than facility 1.

The estimates of both treatment and poverty were positive, indicating that youth in the traditional classes or in a county with a higher poverty rate would have higher probabilities to recidivate after their release. Therefore, to reduce a youth’s recidivism rate, the most effective way would be increasing their reading ability, by using the more effective way (Read180). In addition, the neighborhood where the youth came from also has some effect on his or her recidivism.

For the random-effect estimates between counties, HGLM with the facility ignored had the largest estimate (0.1357), while HGLM with the facility included as a dummy variable had the smallest estimate (0.1144). The standard errors of the random-effect estimates had a similar trend. The pseudo-BIC of the HGLM with the facility ignored was 4574.02, the HGLM with the facility dummy-coded was 4549.22, and the CCREM was 4569.24. The HGLM with the facility dummy-coded, therefore, had the best fit among the three, and HGLM with the facility ignored had the worst fit.

Generally, in the empirical dataset in which county and facility were cross-classified, when the facility effect was ignored, the model had the worst fit. When the facility effect was considered, using the facility as a dummy-coded variable resulted in a
better fit than using CCREM. Therefore, researchers should consider second-level data structures in their studies in order to use the correct model and draw reliable conclusions.
Chapter 5: Discussion

Chapter 5 summarizes the findings of this study. Following the summary, the limitations, recommendations for practical uses are discussed. Finally, suggestions for possible future research are presented.

5.1 Summary

This study was designed to investigate the consequences of ignoring a cross-classified structure in a cross-classified dataset with a dichotomous outcome, the impact of the estimation method on fixed-effect and random-effect estimates, and the conditions (correlation of level-2 factor residuals, number of facilities fed into each county, number of level-2 cross-classified factors, IUCCs) under which there is the need to use the cross-classified logistic model. To address this purpose, this study consisted of two parts: (1) a Monte Carlo simulation study; and (2) a demonstration of cross-classified random-effect logistic model using empirical datasets.

In Study 1, datasets were generated in a way that youth were cross-classified by county and facility. Two level-1 predictors (reading score, treatment group), one level-2 predictor (county-level poverty rate), and one outcome variable (recidivism status) were simulated under 16 conditions. Under each condition, two modeling methods (CCREM
and HGLM) and two estimation methods (PQL and AQ) were estimated and replicated 500 times. The fixed-effect estimates and random-effect estimates, as well as the fit indices, were generated and compared across the four between-subject factors and two within-subject factors.

The three fixed-effect estimates (Read: $\gamma_{10}/\gamma_{100}$, Treatment $\gamma_{20}/\gamma_{200}$, Poverty $\gamma_{01}/\gamma_{010}$) were estimated and their relative biases were calculated. Final relative bias was averaged across the 500 replications. Consistently, one within-subject factor (estimation method) and one between-subject factor (residual correlation) were found to have a significant effect on the fixed-effect estimate relative bias. Evidence was provided to support the AQ method as a better estimation method to produce more accurate and precise fixed-effect estimates than the PQL method. This finding supports a previous theory that PQL was simple but might yield biased estimates for binary response models with large random effects or when the target probability is very small or very large. The AQ method usually yielded more accurate estimates than PQL, although the process was much slower and there might be more convergence problems (O’Connell et al., 2010; Bolker et al., 2008; Raudenbush et al., 2004). In addition, the correlation of the two second-level factors appeared to have an influential effect on the relative bias of fixed-effect estimation as well. For all of the three predictors, moderately correlated cross-classified factor residuals conditions had more accurate estimates than non-correlated conditions. This finding agreed with Meyers and Beretvas’s finding that given when the cross-classified factors were related, recognizing one factor might explain some of the dependency attributable to the other factor (Meyers & Beretvas, 2006). In the real
datasets, as mentioned before, it was very likely that the cross-classified factors were somewhat correlated. Therefore, it is likely that this correlated-factor condition will produce more reliable estimates. On the other hand, there was not much difference found between CCREM and HGLM. This finding supported previous researchers’ studies that ignoring a cross-classified factor usually does not lead to biased fixed-effect estimates (Fielding & Goldstein, 2006; Meyers & Beretvas, 2006).

For the fixed-effect standard error relative bias, the AQ estimation method was preferred to the PQL method for the two level-1 fixed-effect predictors, but the PQL produced smaller standard error bias at level-2 predictors. Also for the level-2 predictor poverty ($\gamma_{010}$), moderately correlated cross-classified factor conditions revealed more accurate standard error estimate than non-correlated conditions. Not much difference existed between CCREM and HGLM. It is worthwhile to note that the standard errors were negatively biased for the level-2 predictor, and this underestimation would possibly inflate the Type I error and researchers may falsely draw the conclusion that the predictor is significant while it is actually not.

RMSE was used as another indicator to estimate the bias of the fixed-effect estimates. Although AQ and moderately correlated residual conditions were found to yield more reliable estimates and standard errors, the PQL method and non-correlated conditions produced more stable estimates. Stated another way, the PQL method was a more robust estimation method. For one of the level-1 predictor treatments, a larger IUCC (0.15) had a significantly smaller RMSE than a smaller IUCC (0.10). Thus, the more variance generated, the more stable the fixed-effect estimates were. In addition,
CCREM was found to produce more reliable and stable estimates for the level-2 predictor (poverty) than HGLM. When the cross-classified data structure was ignored, the level-2 predictor estimates were not stable.

The random-effect estimates (county) were negatively biased in both models and estimation methods across the 16 generating conditions. HGLM yielded significantly more negatively biased estimates than CCREM. Therefore, when the cross-classified structure was ignored and one cross-classified factor was dropped, the variance component of the remaining crossed factor would be underestimated. This finding agrees with Fielding’s study in 2002, and Luo and Kwok’s study in 2009, but it disagrees with Meyer and Beretvas’s study in 2006. When using CCREM, the two estimation methods almost generated similar biases. However, when using HGLM, PQL was found to have more accurate estimates than AQ. Therefore, when the model was misspecified, the AQ method was much easily to be affected than the PQL method. That is to say, the PQL method was more robust in estimating the random-effect predictor. This finding was also supported by the random-effect RMSE results. CCREM was also found to have more precise and reliable results than HGLM. The PQL method still yielded more precise RMSEs than AQ. Further, the difference between PQL and AQ was much larger when using HGLM than using CCREM. In addition, a larger IUCC was found to generate more accurate random-effect estimates, too. For the random-effect standard error bias, PQL performed better in the three-feeder condition, but AQ performed better in the four-feeder condition. HGLM appeared to have slightly more accurate random-effect standard errors than CCREM. The four-feeder condition produced more accurate standard errors
than the three-feeder condition. In the four-feeder condition, there were fewer empty cells in the generated datasets (that is to say the three-feeder condition had sparser data). When there were more empty cells, the random-effect standard errors tended to be more negatively biased. In a nutshell, when the cross-classified structure was ignored, the random-effect estimates would be underestimated, and different estimation methods performed dramatically differently. Larger IUCCs produced more accurate estimates, and fewer empty-cell data designs yielded more precise standard errors.

The model fit pseudo-BIC had an AQ estimation method accuracy rate of 100 percent, but it only had approximately 60 to 80 percent accuracy with the PQL method. Therefore, the AQ estimation method had a more accurate model fit than the PQL method across all the conditions. However, none of the within-subject factors (estimation method) and the four between-subject factors had significant effects on the difference between the model fit accuracy rates.

The convergence rate analysis indicated that PQL had a higher convergence rate than AQ. In addition, if there are fewer missing cells in the cross-classified data structure, the IUCC is larger, and the residuals of the level-2 factors are somewhat correlated, the convergence rate will increase in both estimation methods.

Based on the consistent results, it is believed that AQ method is preferable to the PQL method for estimating multilevel logistic models. Whether the cross-classified factors correlated also had an effect in the parameter estimation. Specifically, moderately correlated residuals resulted in more accurate estimates. The size of the IUCC also influenced the estimates, and the larger the IUCC was, the more accurate the estimates
were. Not much difference between CCREM and HGLM was found in the fixed-effect estimates; however, CCREM was found to produce more accurate random-effect estimates, while HGLM tended to underestimate the random-effect variance component. Still, researchers should exercise caution when choosing estimation methods for their studies. Due to the convergence difficulty, it might be possible that the model could not converge if they used the AQ method. If the dataset being evaluated has a high degree of cross-classification, a high IUCC, and the residuals of the level-2 factors were correlated, there would be a larger chance for the model to converge.

In Study 2, three models (model 1: HGLM ignoring the facility effect; model 2: HGLM treating facility as dummy-coded level-1 predictor; model 3: CCREM) were used to fit the empirical youth recidivism dataset. The AQ method was chosen due to its accuracy and precision. Reading scores were found to be statistically significant across the three models. There was not much difference in the model estimates among the three models. When the reading scores increased by one unit, the recidivism rate would decreased by approximately 9 percent. Youth in the Read180 classroom and from a county with a lower poverty rate also had a lower probability of recidivism. As one would imagine, model 1 had the worst fit among the three according to the model fit statistics. Model 2, on the other hand, held the best fit among the three. This finding confirms previous research that the rates of juvenile delinquency and recidivism are highly correlated with low levels of academic performance, and the implementation of sound academic interventions, particularly in reading, can effectively reduce rates of both delinquency and recidivism (Katsiyannis et al., 2008).
5.2 Limitations

Several limitations should be addressed in this preliminary study. First, data generation in the simulation study had several restrictions:

1. The model was simplified. Only two levels were generated in the dataset, and only the intercept was allowed to vary. For simplicity, the model specified in both CCREM and HGLM had only one level-2 predictor, and this predictor was only used to predict the intercept. In addition, due to the small number of facilities, no facility-level predictor was included in the model. Thus, this study was unable to investigate the behavior of the facility-level predictor estimates.

2. For the estimation method factor, only two methods were selected, whereas there are more than two methods (RSPL, RMPL, MPSL, MMPL, Laplace, and AQ).

3. The simulation conditions were simplified. For each of the four simulation conditions, only two levels were generated, while in real life, more levels might exist.

4. For the level-2 correlation condition, when county and facility were paired up and sorted, the moderate correlation actually indicated an inherent additional dependency between the level-2 factors.

5. For the number of feeder conditions, each county feeder was designed to have the same probability (such as 33% for a three-feeder condition) to feed into the corresponding facility. It is possible that in a three-feeder condition, one county actually fed 10%, the second fed 20% and the third fed 70%. Again, three-feeder and four-feeder
conditions were chosen to resemble the empirical dataset, but this design created a lot of empty cells in the datasets, which might cause problems in estimating and converging.

6. When designing the number of level-2 factors, because the empirical dataset had an unbalanced number of facilities and counties, the level-2 factors were designed to be unbalanced. There were only limited numbers of facilities (seven) in the empirical dataset, which might occur in some other social science research. This study provided a way to explore the effect of unbalanced design with a limited level-2 sample size for future research.

7. The IUCC were chosen arbitrarily, and the values were generated to be the same for both level-2 factors for simplicity. In real-life research, however, it is not likely that the variances are the same for both factors.

8. The probability of the outcome variable recidivism in the empirical dataset was large (46%). Such a large probability might cause difficulty in the estimating process.

9. When fitting the models in the simulation study, the AQ method was found to be more difficult to converge than the PQL method. The relative bias results were generated based on successfully converged replications. Therefore, although AQ is a more accurate method, it is very possible it might not converge.

10. Study 2 was used as a demonstration, and thus the model fitted was oversimplified. Many other predictors could be added to the model to make it more meaningful. Also, as a special population, the incarcerated youth came in and left the facility at irregular times, thus it was hard to define how many treatments the youth actually had received.
5.3 Recommendations

Based on the results of the study, researchers can find that although ignoring the cross-classified data structure may not affect the fixed-effect estimates, it can cause underestimated random-effect estimates. This misspecification, therefore, will result in biased conclusions. Ultimately, in their own applied research, researchers should properly align their model design and specifications with their research interests. If they simply want to investigate the fixed-effect estimates of level-1 parameters, it may not be necessary to model the cross-classified factors. However, if they want to investigate the level-2 parameter estimates, as well as the random-effect estimates of the level-2 factors, it is important to specify the cross-classified structure appropriately, especially when the cross-classified factors are correlated, the variations existing in both level-2 factors are substantial, and the degree of cross-classification is high.

For almost all the estimates, the adaptive quadrature estimation method (AQ) is found to yield more accurate estimates than the PQL estimation method (PQL). However, PQL is more stable and much easier to converge than AQ. Therefore, although AQ is a more accurate method, it may fail to converge. However, as the degree of cross-classification increases, the correlation rate of the cross-classified factors increases, and as the IUCC increases, the convergence rate of the both PQL and AQ will increase. That is to say, if researchers have datasets with those characteristics, it is strongly
recommended that they appropriately model the cross-classification structure and use the AQ estimation method for more accurate estimations.

5.4 Future Research

In light of the findings and limitations of the study, some suggestions could be made for future research. First, the simulation conditions were simplified. Future research could manipulate additional data conditions to more realistically represent real-life situations. For example, one could design more options in the residual correlation conditions. Second, more estimation methods could be compared using various software, so that researchers could be guided to choose from different estimation methods and software to answer their specific research questions. Third, more complex models (e.g., three-level models, cross-classified ordinal regression) could be studied. In real-life research, it is not uncommon to come across a data structure with more than two levels or ordinal outcomes. Forcing them into a two-level structure or a dichotomous outcome increases bias in the estimation and interpretation. Fourth, we need further investigation on how data sparseness influenced the parameter estimation. When there were large numbers of empty cells, how would the estimates change? Furthermore, how would the small level-2 sample size influence the estimates? Finally, the empirical dataset revealed that, when the number of level-2 factors was small, it could be dummy-coded and estimated. This method provided an even better model fit than CCREM. The difference between these two models and the possibility of conducting such models in empirical study could be further investigated!
References


LeBaron, J. (2002). Examining the relative influence of community context on juvenile offender post-confinement recidivism. Doctoral Dissertation, Department of Criminal Justice, Rutgers University, Newark, NJ.


StataCorp (2009). Stata Statistical Software: Release 11. College Station, TX: StataCorp, LP.


Appendix: SAS Syntax

(Modified from Meyers & Beretvas, 2006)

/*Generating county and facility residuals based on a known correlation*/
/*first condition: correlation residual 0.4, # feeder 3, level of cc factor: 7, 28, IUCC: 0.15, #per county: 40*/

%macro generate;
proc iml;
resid_fa=normal(j(7,1,0));
resid_co=normal(j(28,1,0));/*/change this to 49 in other condition*/
cREATE sasresid_fa FROM resid_fa;
append FROM resid_fa;
cREATE sasresid_co FROM resid_co;
append FROM resid_co;
QUIT;
RUN;
data sasresid_fa;set sasresid_fa;
rename COL1=fa_res;
retain counter 0;
counter=counter+1;
fa_id=counter;
KEEP fa_id COL1; /*no facility-level predictor at this time*/
RUN;
data sasresid_co;set sasresid_co;
rename COL1=co_res;
RUN;
/*main feeder*/
data one;set sasresid_co;
retain counter 0;
counter=counter+1;
co_id=counter;
fa_id=counter;
if co_id >= 8 and co_id <= 14 then fa_id=co_id-7;
if co_id >= 15 and co_id <= 21 then fa_id=co_id-14;
if co_id >= 22 and co_id <= 28 then fa_id=co_id-21;
co_size= int((40+5*rannor(0)));
cell_size = int(.33*co_size); /*this is 3 feeders, change to 4 feeders in the other condition: .25 instead of .33*/
RUN;
proc sort data=one;by fa_id;run;
data merged;
merge one(in=A) sasresid_fa (in=B);by fa_id;
106
if A and B;
keep co_id fa_id cell_size co_res fa_res;
run;
/*correlate fa_res and co_res 0.40*/
proc iml;
use merged;
show datasets;
read all var{fa_res co_res}into x;
corr={1.00 .40 .40 1.00}; /* change to 0 in the other condition*/
corr_root=root(corr);
y=x*corr_root;
z=0.8396#y; /*change to 0.6413 to change IUCC to .10*/
create new from z;
append from z;
quit;
data new; set new;
rename COL1= fa_res;
rename COL2= co_res;
run;
data merged;set merged;
retain counter 0;
counter=counter+1;
drop fa_res co_res;
run;
data new; set new;
retain counter 0;
counter=counter+1;
run;
data mergedln;
merge merged(in=A) new(in=B); by counter;
if A and B;
keep fa_id co_id fa_res co_res;
run;
data one1; set one;
keep co_id cell_size;
run;
proc sort data=mergedln; by co_id; run;
proc sort data=one1; by co_id; run;
data merged1;
merge mergedln(in=A) one1(in=B); by co_id;
if A and B;
keep fa_id co_id fa_res co_res cell_size;
run;
data merged1; set merged1;
poverty=14.62+2.08*rannor(0);
county_effect=0.06*poverty + co_res;
run;
/*secondary feeder*/
data merged2n; set merged1;
keep co_id fa_id co_res cell_size poverty county_effect;
fa_id=fa_id+1;
if fa_id=8 then fa_id=1;
run;
data temp; set merged1;
keep fa_id fa_res;
run;
proc sort data=merged2n; by fa_id; run;
proc sort data=temp; by fa_id; run;
data merged2;
merge merged2n (in=A) temp (in=B); by fa_id;
if A and B;
keep co_id fa_id co_res fa_res cell_size poverty county_effect;
run;

/*other secondary feeder school*/
data merged3n; set merged1;
keep co_id fa_id co_res cell_size poverty county_effect;
fa_id=fa_id+2;
if fa_id=8 then fa_id=1;
if fa_id=9 then fa_id=2;
run;
proc sort data=merged3n; by fa_id; run;
data merged3;
merge merged3n (in=A) temp (in=B); by fa_id;
if A and B;
keep co_id fa_id co_res fa_res cell_size poverty county_effect;
run;

/*merge all three merged files together*/
data grand;
keep co_id fa_id cell_size co_res fa_res poverty county_effect
/*treatment size county_effect facility_effect*/;
set merged1 merged2 merged3;
run;

/*creating youth recidivism data*/
data youth;set grand;
do youth= 1 to cell_size;
read=6+2.6*rannor(0);
treatment=ranbin(19834, 1,.46);
youth_effect=(-0.09)*read+(0.17)*treatment;
P=exp((-0.68)+youth_effect+county_effect)/(1+exp((-0.68)+youth_effect+county_effect));
Y=ranbin(12345,1, P);
iter=&i;
output;
end;
run;
%mend;
%macro runall;
%do i =1 %to 500; /*number of replications*/
%generate;
proc datasets;
append base=fulldata data=youth;
%end;
%mend;
proc iml;
%runall;
quit;
ods output covp parms=random_effects;
ods output ParameterEstimates = fixed_effects;
ods output FitStatistics=fit_statistics;
/* Fitting the cross-classified model */
proc sort data=fulldata; by iter; run;
/*restricted pseudo-likelihood (RPL)-subject specific-rspl
proc glimmix data= fulldata method=rspl ic=pq noclprint noitprint ;
class co_id fa_id;
model Y (event='1') = read treatment poverty
/cl dist = binary corrb covb
   link=logit solution ddfm=bw;
random intercept/subject=co_id;
random intercept/subject=fa_id;
by iter;
run;

/*AQ*/
proc glimmix data= fulldata method=laplace ic=pq noclprint noitprint ;
class co_id fa_id;
model Y (event='1') = read treatment poverty
/cl dist = binary corrb covb
oddsratios link=logit solution ddfm=bw;
random intercept/subject=co_id;
random intercept/subject=fa_id;
by iter;
run;

/*transposing random effects*/
proc transpose data=random_effects out=random_trans;
var estimate stderr;
id subject;
by iter;
run;
data randeffs;set random_trans; by iter;
retain coest faest costd fastd coest_RMSE faest_RMSE iter;
if _NAME_ = 'Estimate' then do;
    coest=co_id;
    faest=fa_id;
    coest_RMSE=((co_id-0.8396)*(co_id-0.8396));
    faest_RMSE=((fa_id-0.8396)*(fa_id-0.8396));
end;
if _NAME_='StdErr' then do;
    costd=co_id;
    fastd=fa_id;
end;
keep coest faest costd fastd coest_RMSE faest_RMSE iter;
if last.iter;
run;

proc transpose data = fixed_effects out = fixed_trans;
var estimate StdErr tvalue Probt;
id effect;
by iter;
run;

/*renaming fixed effects*/
data fixedeff;set fixed_trans; by iter;
retain intercept_est read_est treatment_est poverty_est intercept_std
read_std
treatment_std poverty_std intercept_t read_t treatment_t poverty_t
intercept_probt
read_probt treatment_probt poverty_probt intercept_est_RMSE
read_est_RMSE treatment_est_RMSE poverty_est_RMSE iter;
if _NAME_ = 'Estimate' then do;
intercept_est = intercept;
read_est = read;
treatment_est = treatment;
poverty_est = poverty;
intercept_est_RMSE=((intercept+.68)*(intercept+.68));
read_est_RMSE=((read+.09)*(read+.09));
treatment_est_RMSE=((treatment-.17)*(treatment-.17));
poverty_est_RMSE=((poverty-.06)*(poverty-.06));
end;
if _NAME_ = 'StdErr' then do;
intercept_std = intercept;
read_std = read;
treatment_std = treatment;
poverty_std = poverty;
end;
if _NAME_ = 'tValue' then do;
intercept_t = intercept;
read_t = read;
treatment_t = treatment;
poverty_t = poverty;
end;
if _NAME_ = 'Probt' then do;
intercept_probt = intercept;
read_probt = read;
treatment_probt = treatment;
poverty_probt = poverty;
end;
keep intercept_est read_est treatment_est poverty_est intercept_std
read_std
treatment_std poverty_std intercept_t read_t treatment_t poverty_t
intercept_probt
read_probt treatment_probt poverty_probt intercept_est_RMSE
read_est_RMSE treatment_est_RMSE poverty_est_RMSE iter;
if last.iter;
run;

/*Getting Fit indices in tabular form*/
**proc transpose** data = fit_statistics out = fit_trans;
var value;
id Descr;
by iter;
run;

/*Renaming Fit Statistics*/
**data** fit_indices; **set** fit_trans; **by** iter;
_2PLL = _2_Res_Log_Pseudo_Likelihood;
PAIC = Pseudo_AIC;
PAICC = Pseudo_AICC;
PBIC = Pseudo_BIC;
PCAIC = Pseudo_CAIC;
PHQIC = Pseudo_HQIC;
keep iter _2PLL PAIC PAICC PBIC PCAIC PHQIC;
run;

**data** results;
merge randeffs
fixedeff
fit_indices;
by iter;
run;

/*Now, repeating the above for the standard HGLM */
ods output covparms = random_effects2;
ods output ParameterEstimates = fixed_effects2;
ods output FitStatistics = fit_statistics2;

/*fitting the standard HGLM model with county as the clustering unit*/
**proc sort** data=fulldata; **by** iter; **run;**

/*restricted pseudo-likelihood (RPL)-subject specific-rspl
**proc glimmix** data= fulldata method=rspl ic=pq noclprint noitprint ;
class co_id ;
model Y (event='1') = read treatment poverty
/cl dist = binary corrb covb
   link=logit solution ddfm=bw;
random intercept/subject=co_id;
by iter;
run;

/*Adaptive Gauss-Hermite quadrature*/
**proc glimmix** data= fulldata method=laplace ic=pq noclprint noitprint ;
class co_id ;
model Y (event='1') = read treatment poverty
/cl dist = binary corrb covb
   oddsratios link=logit solution ddfm=bw;
random intercept/subject=co_id;
by iter;
run;

/*Transforming random effects in tabular form */
```sas
proc transpose data = random_effects2 out = random_trans2;
var estimate StdErr ;
id subject;
by iter;
run;
/*renaming random effects*/
data randeff2;set random_trans2; by iter;
retain coest2 costd2 coest2_RMSE iter;
if _NAME_ = 'Estimate' then do;
coest2=co_id;
coest2_RMSE=(co_id-0.8396)*(co_id-0.8396);
end;
if _NAME_ = 'StdErr' then do;
costd2=co_id;
end;
keep coest2 costd2 coest2_RMSE iter;
if last.iter;
run;

/* Getting fixed effects in tabular form*/
proc transpose data = fixed_effects2 out = fixed_trans2;
var estimate StdErr tvalue Probt;
id effect;
by iter;
run;
data fixedeff2;set fixed_trans2; by iter;
retain intercept_est2 read_est2 treatment_est2 poverty_est2
intercept_std2
read_std2 treatment_std2 poverty_std2 intercept_t2 read_t2 treatment_t2
poverty_t2
intercept_probt2 read_probt2 treatment_probt2 poverty_probt2
intercept_est2_RMSE read_est2_RMSE treatment_est2_RMSE
poverty_est2_RMSE iter;
if _NAME_ = 'Estimate' then do;
intercept_est2 = intercept;
read_est2 = read;
treatment_est2 = treatment;
poverty_est2=povety;
intercept_est2_RMSE=((intercept+.68)*(intercept+.68));
read_est2_RMSE=((read+.09)*(read+.09));
treatment_est2_RMSE=((treatment-.17)*(treatment-.17));
poverty_est2_RMSE=((poverty-.06)*(poverty-.06));
end;
if _NAME_ = 'StdErr' then do;
intercept_std2 = intercept;
read_std2 = read;
treatment_std2 = treatment;
poverty_std2=povety;
end;
if _NAME_ = 'tValue' then do;
intercept_t2 = intercept;
read_t2 = read;
treatment_t2 = treatment;
poverty_t2=povety;
```
end;
if _NAME_='Probt' then do;
intercept_probt2 = intercept;
read_probt2 = read;
treatment_probt2 = treatment;
poverty_probt2=poverty;
end;

keep intercept_est2 read_est2 treatment_est2 poverty_est2
intercept_std2
read_std2 treatment_std2 poverty_std2 intercept_t2 read_t2 treatment_t2
poverty_t2
intercept_probt2 read_probt2 treatment_probt2 poverty_probt2
intercept_est2_RMSE read_est2_RMSE treatment_est2_RMSE
poverty_est2_RMSE iter;
if last.iter;
run;

proc transpose data = fit_statistics2 out = fit_trans2;
var value;
id Descr;
by iter;
run;

data fit_indices2;set fit_trans2;
_twoPLL2 = _2_Res_Log_Pseudo_Log_Likelihood;
PAIC2 = Pseudo_AIC;
PAICC2= Pseudo_AICC;
PBIC2 = Pseudo_BIC;
PCAIC2 = Pseudo_CAIC;
PHQIC2 = Pseudo_HQIC;
keep iter _2PLL2 PAIC2 PAICC2 PBIC2 PCAIC2 PHQIC2;
run;

/*Merges fixed effects, random effects, and statistical tests */
data results2;
merge randeff2
fixedeff2
fit_indices2;
by iter;
run;

/*Summarizes data across replications*/
data final;
merge results
results2;
by iter;
run;

/*Descriptive Statistics for all of the estimates*/
proc means data=final;run;

/*Proportion of time CCREM is chosen over HGLM*/
data fit;set final;
by iter;
PAIC_fit = PAIC2-PAIC;
PAICC_fit = PAICC2-PAICC;
PBIC_fit=PBIC2-PBIC;
PCAIC_fit=PCAIC2-PCAIC;
PHQIC_fit=PHQIC2-PHQIC;

if PAIC_fit ge 0 then Afit = 1;
else Afit = 0;
if PAICC_fit ge 0 then ACfit = 1;
else ACfit = 0;
if PBIC_fit ge 0 then Bfit = 1;
else Bfit = 0;
if PCAIC_fit ge 0 then Cfit = 1;
else Cfit = 0;
if PHQIC_fit ge 0 then Hfit = 1;
else Hfit = 0;

run;
proc freq;
tables Afit ACfit Bfit Cfit Hfit;
run;
data fixed_bias;set final;
by iter;
keep coest faest coest_RMSE faest_RMSE intercept_est read_est
treatment_est poverty_est intercept_est_RMSE read_est_RMSE
treatment_est_RMSE poverty_est_RMSE coest2
cost2_RMSE intercept_est2 read_est2 treatment_est2 poverty_est2
intercept_est2_RMSE read_est2_RMSE treatment_est2_RMSE
poverty_est2_RMSE;
run;
/*Calculating Relative Bias of the Parameter Estimates*/
proc means data=fixed_bias;
output out=fixed_means
mean(coest faest coest_RMSE faest_RMSE
intercept_est read_est
treatment_est poverty_est intercept_est_RMSE read_est_RMSE
treatment_est_RMSE poverty_est_RMSE coest2 cost2_RMSE intercept_est2
read_est2 treatment_est2
poverty_est2 intercept_est2_RMSE read_est2_RMSE treatment_est2_RMSE
poverty_est2_RMSE)= mean_coest mean_faest mean_coest_RMSE
mean_faest_RMSE mean_intercept_est
mean_read_est mean_treatment_est mean_poverty_est
mean_intercept_est_RMSE mean_read_est_RMSE
mean_treatment_est_RMSE mean_poverty_est_RMSE mean_coest2
mean_coest2_RMSE mean_intercept_est2
mean_read_est2 mean_treatment_est2
mean_poverty_est2
mean_intercept_est2_RMSE mean_read_est2_RMSE mean_treatment_est2_RMSE
mean_poverty_est2_RMSE;
run;
data parameter_bias;set fixed_means;
coest_bias=(mean_coest -0.8396)/0.8396;/*changes when IUCC changes*/
faest_bias=(mean_faest -0.8396)/0.8396;/*changes when IUCC changes*/
coest_RMSE_bias=sqrt(mean_coest_RMSE*(500/499));
faest_RMSE_bias=sqrt(mean_faest_RMSE*(500/499));
int_bias=(mean_intercept_est+0.68)/(-0.68);
read_bias=(mean_read_est+0.09)/(-0.09);
treatment_bias=(mean_treatment_est-.17)/(.17);
poverty_bias=(mean_poverty_est-.06)/(.06);
int_RMSE_bias=sqrt(mean_intercept_est_RMSE*(500/499));
read_RMSE_bias=sqrt(mean_read_est_RMSE*(500/499));
treatment_RMSE_bias=sqrt(mean_treatment_est_RMSE*(500/499));
poverty_RMSE_bias=sqrt(mean_poverty_est_RMSE*(500/499));
coest_bias2=(mean_coest2 -0.8396)/0.8396; /*changes when IUCC changes*/
int_bias2=(mean_intercept_est2+0.68)/(-0.68);
read_bias2=(mean_read_est2+0.09)/(-0.09);
treatment_bias2=(mean_treatment_est2-.17)/(.17);
poverty_bias2=(mean_poverty_est2-.06)/(.06);
int_RMSE_bias2=sqrt(mean_intercept_est2_RMSE*(500/499));
read_RMSE_bias2=sqrt(mean_read_est2_RMSE*(500/499));
treatment_RMSE_bias2=sqrt(mean_treatment_est2_RMSE*(500/499));
poverty_RMSE_bias2=sqrt(mean_poverty_est2_RMSE*(500/499));
run;
data parameter_bias;set parameter_bias;
  keep coest_bias faest_bias coest_RMSE_bias faest_RMSE_bias int_bias
  read_bias treatment_bias
  poverty_bias int_RMSE_bias read_RMSE_bias treatment_RMSE_bias
  coest_RMSE_bias2 int_bias2 treatment_RMSE_bias2
  poverty_bias2 int_RMSE_bias2 read_RMSE_bias2 treatment_RMSE_bias2
run;
data se_bias;set final;
  by iter;
  keep coest faest intercept_est read_est treatment_est poverty_est
costd2
  intercept_std intercept_std2 read_std read_std2 treatment_std
treatment_std2 poverty_std
run;
proc means data=se_bias;
output out=se_means mean (coest faest intercept_est read_est
treatment_est poverty_est coest2 intercept_est2 read_est2
  treatment_est2 poverty_est2)
  stddev(mean_coest
  mean_read_est mean_read_est2 mean_treatment_est
  mean_poverty_est mean_coest2 mean_intercept_est2 mean_read_est2
  mean_treatment_est2 mean_poverty_est2 mean_costd mean_fastd
  mean_costd2 mean_intercept_std
  mean_read_std mean_read_std2 mean_treatment_std
  mean_treatment_std2 mean_poverty_std
  mean_poverty_std2 stddev(coest faest intercept_est read_est
treatment_est
  read_est2 treatment_est2 poverty_est2)
run;
poverty_est cost2 intercept_est2 read_est2 treatment_est2
poverty_est2 costd
fastd costd2 intercept_std intercept_std2 read_std
read_std2 treatment_std treatment_std2 poverty_std
poverty_std2) = std_coest std_faest
std_coest2 std_intercept_est2 std_read_est2 std_treatment_est2
std_poverty_est2 std_costd std_fastd std_costd2 std_intercept_std
std_intercept_std2 std_read_std std_read_std2
std_treatment_std2 std_treatment_std2 std_poverty_std std_poverty_std2;
run;
data standard_error_bias; set se_means;
co_se_bias = (mean_costd - std_coest)/std_coest;
fa_se_bias = (mean_fastd - std_faest)/std_faest;
intercept_se_bias = (mean_intercept_std - std_intercept_est)/std_intercept_est;
read_se_bias = (mean_read_std - std_read_est)/std_read_est;
treatment_se_bias = (mean_treatment_std - std_treatment_est)/std_treatment_est;
poverty_se_bias = (mean_poverty_std - std_poverty_est)/std_poverty_est;
co_se_bias2 = (mean_costd2 - std_coest2)/std_coest2;
intercept_se_bias2 = (mean_intercept_std2 - std_intercept_est2)/std_intercept_est2;
read_se_bias2 = (mean_read_std2 - std_read_est2)/std_read_est2;
treatment_se_bias2 = (mean_treatment_std2 - std_treatment_est2)/std_treatment_est2;
poverty_se_bias2 = (mean_poverty_std2 - std_poverty_est2)/std_poverty_est2;
run;
data standard_error_bias; set standard_error_bias;
keep co_se_bias fa_se_bias intercept_se_bias
read_se_bias treatment_se_bias poverty_se_bias co_se_bias2
intercept_se_bias2 read_se_bias2 treatment_se_bias2 poverty_se_bias2;
run;
proc means data=final; run;
proc freq data=fit;
tables Afit ACfit Bfit Cfit Hfit;
run;
proc means data=fixed_bias;
proc means data=se_bias;
proc print data=parameter_bias; run;
proc print data=standard_error_bias; run;
/* more descriptive statistics */
data fulldata; set fulldata;
proc means; var co_res fa_res county_effect
cell_size;
run;
/* correlation of residuals */
proc corr; var co_res fa_res; run;