2012-04-29

The Implications of Centering in a Three-Level Multilevel Model

Ahnalee Brincks
University of Miami, ahnabanana@gmail.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation
https://scholarlyrepository.miami.edu/oa_dissertations/743

This Embargoed is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.
UNIVERSITY OF MIAMI

THE IMPLICATIONS OF CENTERING IN A THREE-LEVEL MULTILEVEL MODEL

By

Ahnalee M. Brincks

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

May 2012
THE IMPLICATIONS OF CENTERING IN A THREE-LEVEL MULTILEVEL MODEL

Ahnalee M. Brincks

Approved:

Daniel Feaster, Ph.D.
Associate Professor of Epidemiology

Terri A. Scandura, Ph.D.
Dean of the Graduate School

Maria Llabre, Ph.D.
Professor of Psychology

Guillermo Prado, Ph.D.
Associate Professor of Epidemiology

Craig Enders, Ph.D.
Professor of Psychology
Arizona State University

Victoria Mitrani, Ph.D.
Professor of Nursing
Hierarchical data are becoming increasingly complex, often involving more than two levels. This study investigated the implications of centering within context (CWC) and grand mean centering (CGM) in three distinct three-level models. The goals were to (1) determine equivalencies in the means and variances across the centering options, (2) identify the algebraic relationships between the three-level contextual models, and (3) clarify the interpretation of the estimated parameters. Artificial datasets were used for illustration. Centering decisions in multilevel models are closely tied to substantive hypotheses and require researchers to be clear and cautious about their choices. This work is designed to assist the researcher in making centering decisions for analysis of three-level hierarchical data.
ACKNOWLEDGEMENTS

There are many people who helped me throughout the process of completing this work. I could not have reached this goal without the unwavering confidence and support from my husband, Nick Myers. In so many ways, he made my success possible.

I am extremely grateful to my advisor and mentor, Dr. Daniel Feaster, for his energy, interest and encouragement. Dr. Maria Llabre was also instrumental to my success in this, and so many other, endeavors. I would also like to thank my committee members, Dr. Guillermo Prado, Dr. Victoria Mitrani, and Dr. Craig Enders.

My deepest appreciation, however, goes to Ms. Yanet and Ms. Juana at South Miami Lutheran Church School. These two women provided a warm, loving environment in which my young son, Benjamin, could learn and grow. Without them, none of this would have been possible.
# TABLE OF CONTENTS

| LIST OF FIGURES | vi |
| LIST OF TABLES | vii |

## Chapter

1. **INTRODUCTION** 1
   - Theoretical Framework 3
   - Multilevel Models 7
   - Centering 9
   - Statement of the Problem 12
   - Research Questions 14
   - Limitations 15
   - Notation 16
   - Summary 18

2. **LITERATURE REVIEW** 19
   - Historical Overview 19
   - Centering Literature 20
   - Three-Level Models 29
   - Summary 29

3. **METHODS** 31
   - Research Methodology 31
   - Data Generation 33
   - Software 35
   - Motivating Example 35

4. **RESULTS** 36
   - Research Questions 36
   - Centering Options in Three-Level Models 37
   - Means and Correlations 39
   - Equivalencies Between Centered and Raw Score Models 41
   - Algebraic Relationships Between CWC and RAS in Model 2 and Model 3... 48
   - Interpretation of Model 1 Estimates 52
   - Interpretation of Model 2 Estimates 60
   - Interpretation of Model 3 Estimates 67
   - Model Misspecification 78

iv
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISCUSSION</td>
<td>81</td>
</tr>
<tr>
<td>Implications for Researchers</td>
<td>83</td>
</tr>
<tr>
<td>Variances and Covariances</td>
<td>90</td>
</tr>
<tr>
<td>Contextual Variables</td>
<td>91</td>
</tr>
<tr>
<td>Limitations</td>
<td>92</td>
</tr>
<tr>
<td>Summary</td>
<td>92</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>93</td>
</tr>
<tr>
<td>TABLES</td>
<td>97</td>
</tr>
<tr>
<td>FIGURES</td>
<td>112</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>116</td>
</tr>
<tr>
<td>APPENDIX B</td>
<td>124</td>
</tr>
<tr>
<td>APPENDIX C</td>
<td>127</td>
</tr>
<tr>
<td>APPENDIX D</td>
<td>130</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1: Two-level hierarchical data structure…………………………………  112
Figure 2: Example of longitudinal data represented in two-level model………..  113
Figure 3: Three-level hierarchical data structure…………………………………  114
Figure 4: Intercepts under random slope and fixed slope models………………  115
LIST OF TABLES

Table 1: Summary of the equivalencies from Kreft, de Leeuw and Aiken (1995) .. 97
Table 2: Model 1, Model 2 and Model 3 raw score population parameter........... 98
Table 3: Intraclass correlations for Model 1, Model 2 and Model 3 ................. 99
Table 4: Model 3 data neighborhood and city means .................................. 100
Table 5: Model 3 data bivariate correlations............................................. 101
Table 6: Variance-covariance matrix for three-level model with random intercept and random level-1 slope................................................................. 102
Table 7: Summary of the equivalencies...................................................... 104
Table 8: Model 1 parameter estimates averaged across ten datasets............... 105
Table 9: Model 1 RAS and CGM equivalencies using estimates from Table 8.... 106
Table 10: Model 2 parameter estimates averaged across ten datasets............ 107
Table 11: Model 2 RAS and CGM equivalencies using estimates from Table 10... 108
Table 12: Model 3 parameter estimates averaged across ten datasets............ 109
Table 13: Model 3 RAS, CGM and CWC equivalencies using estimates from Table 12........................................................................................................... 110
Table 14: Model 1 estimates resulting from analysis on Model 3 data............ 111
CHAPTER 1: INTRODUCTION

Hierarchical data structures are abundant across many research disciplines. Examples from education are common, such as students nested within classrooms, classrooms nested within schools, or schools nested within districts. In psychology, examples might include clients nested within therapists or individuals nested within families. In epidemiology and public health, residents nested within neighborhoods or individuals nested within health care delivery systems might be the examples. Multilevel models provide researchers with a tool to analyze these data while accounting for dependencies due to the nesting, resulting in better estimation of parameters and standard errors.

Centering variables (deviating the raw score from a specified value), is a useful strategy in multilevel models because it provides a meaningful zero point and makes it possible to disentangle person-level effects from group-level effects (Raudenbush & Bryk, 2002). In a two-level model, the predictor variable can be centered on the overall mean (termed “grand mean centering” or CGM) or centered on the cluster mean (termed “centering within cluster” or CWC). A third option is not to center at all and using the variable in its original metric, which is referred to as employing the “raw score” (RAS). Decisions about whether to center at all, or which centering option to use, must be made with care as different choices can result in dramatically different interpretations of the parameter estimates.

The impact of centering in two-level multilevel models has been addressed and the results are presented in detail in Chapter 2. However, the implications of centering in three-level models have yet to be elucidated. A public health example of a three-level
model is residents nested in neighborhoods, and neighborhoods further nested in cities. When a third level of nesting is introduced, additional centering options become available that were not present in the two-level model. For instance, the Level-2 variable, which previously could only be centered on the grand mean (CGM) or used in its original metric (RAS) now has a cluster above it, making it possible to use CWC. As researchers work with increasingly complex hierarchical data structures, the methodological work is needed to maintain our understanding of the implications of centering in these models.

Central to the issue of centering is the question of whether to include cluster means as predictors in the multilevel models. As will be discussed in the next section, researchers might be interested in the effects of these upper-level means controlling for the influence of the lower-level predictor, or have an interest in the influence of both the upper-level mean and the lower-level predictor. There may also be substantive interest in whether the two (or three, in the case of three-level models) effects differ significantly. When these upper-level means are not only used as points around which to center lower-level variables, but are also included in the model as predictors, the interpretation of the resulting coefficients is significantly influenced by the centering decision made at the lower level.

Despite the existing literature clarifying the implications of centering in HLM, examples of unclear or incorrect centering decisions persist. In the health literature, a recent study of contextual effects on family planning behaviors exemplifies this (Paek, Lee, Salmon, & Witte, 2007). Although clear about the use of Level-1 predictors and inclusion of the aggregate means of these predictors at Level-2, the authors fail to indicate whether the Level-1 predictors were centered or left in raw score, making
interpretation of the Level-2 effects impossible to ascertain. If, in fact, the Level-1 predictors were left in their raw score format, as might be assumed, the reported interpretations of the Level-2 coefficients are incorrect. Another study in the health literature examines the influence of individual and school religiosity on substance use among adolescents (Wallace, Yamaguchi, Bachman, O'Malley, Schulenberg, & Johnston, 2007). As in the previous example, the authors include both individual and school mean religiosity as predictor variables at Level-1 and Level-2 respectively, however they misinterpret the Level-2 coefficient to be the effect of school religiosity rather than the contextual effect, or difference between the effect of school religiosity and individual religiosity.

This study investigates the implications of centering in three-level models. The primary goal was to clarify the interpretations of the coefficients and variance terms under centering within context (CWC), grand mean centering (CGM) and raw scores (RAS). The abbreviations CWC, CGM and RAS will be used throughout this dissertation, consistent with previous literature in this area (Kreft, de Leeuw, & Aiken, 1995).

Theoretical Framework

Although there are many disciplines in which multilevel models can be considered and discussed, this dissertation will focus within the context of epidemiology. Epidemiology is in many ways a discipline defined by the recognition of population-level phenomena. The very practice of disease surveillance involves the tracking of cases within groups (hospitals, cities, states, countries), as well as calculations of disease
prevalence and incidence within these groups. However, Diez-Roux (2000) pointed out a major problem with the modern theory of epidemiology:

“Populations (or groups) are thought of as collections of independent individuals, rather than entities with properties that may affect individuals within them. Consequently, there is generally little interest in examining group-to-group variation per se. Although there has been abundant discussion in the epidemiologic literature of the fallacy inherent in using data at one level to draw inferences at another level (specifically of the ecological fallacy), until recently there has been relatively little discussion of the substantive problem of ignoring potentially important variables that are best conceptualized and measured at the group level” (p. 172).

In some ways, the adoption of multilevel models among epidemiology researchers has realigned the field to once again consider the influence of population-based risk factors. Multilevel analysis gives researchers the tools they need to include both person-level and group-level variables in their models. Thus, population-based risk factors can be simultaneously considered with individual risk factors.

Group-level variables can be conceptualized in two ways. The first includes those variables that can only be measured at the group level, sometimes referred to as global or integral variables (Morgenstern, 1998; Susser, 1994). An example is a count of the number of liquor stores within a given neighborhood. The second way of conceptualizing a group variable includes those variables that provide a summative measure of individual characteristics, sometimes called aggregate or contextual variables (Morgenstern, 1998; Susser, 1994). For instance, taking the average stress score from scores obtained by administering a stress survey to individual members of a family is one way to measure “family stress” (Brincks, Feaster, & Mitrani, 2010). It is these aggregated or contextual variables that will be the focus of this dissertation.
Inclusion of group-level variables requires careful consideration. Diez-Roux (1998) cautions about being intentional when adding either global or contextual variables to the analytic model:

“The inclusion of group- or macro-level variables (together with individual-level variables) in public health research is quite challenging from a methodological point of view but is even more challenging from a theoretical point of view. If it is to be meaningful (and not reduced to the mere addition of yet another set of variables to the "web of causation"), it requires the development of models of disease causation (and testable hypotheses) that extend across levels and explain how individual- and group-level variables jointly shape health and disease” (p. 217).

This point is particularly salient with respect to group variables that are derived from individual measures, which are prominent in the models selected for this dissertation. When including these variables in a multilevel framework, the researcher needs to consider whether the individual measure and the aggregate measure are in fact representing separate constructs (Diez-Rouz, 1998). In the family stress example, aggregating the individual stress scores is done to create a measure of family stress, designed to represent a measure of the family environment. This is theoretically different from the measure of individual stress at the person-level. It is also important to use theory to carefully determine the level of aggregation (Blakely & Woodward, 2000). For instance, when deciding on a measure of contextual stress, a researcher may choose to aggregate scores to the family level rather than the neighborhood level given the stronger influence of stress among individuals who live in the same household, as opposed to the same neighborhood. As with decisions about whether to include any variable in a model, the decision to include these aggregated variables in the multilevel model should be soundly based in theory.
There is also the question of whether inclusion of group-level variables in multilevel models accurately portrays current theories of health and health behavior. Stokols (1992) laid the groundwork for what he terms a “social ecology of health promotion” in which he called for analyses that address the physical and social environment in addition to individual characteristics that predict healthy behavior. Taylor, Repetti, & Seeman (1997) made a strong case for considering the social contexts of health predictors and stated that “psychological predictors of health outcomes do not occur and should not be studied in an economic, racial, developmental, and social vacuum” (p. 439). There are theoretical models stemming from social psychology and focused primarily on healthy behaviors that support inclusion of group-level variables. For instance, the Theory of Reasoned Action (Fishbein & Ajzen, 1975) includes “subjective norm” as a predictor of behavioral intention. This subjective norm might be conceptualized as a social pressure resulting from the individual’s awareness of a high proportion of people in his or her environment that have adopted a specific pro-health behavior. The Health Belief Model (Champion & Skinner, 2008) also considers perceived susceptibility as a predictor of the likelihood of adopting a certain healthy behavior and this could be operationalized as the incidence or prevalence in a community and/or the visibility of this incidence or prevalence among members of the community.

The ecologic fallacy, making potentially inaccurate inferences about individuals based on group-level data, is commonly discussed in epidemiology. Less well-known, but equally important are the atomistic, psychologistic and sociologistic fallacies (Diez-Roux, 1998). The atomistic fallacy is the opposite of the ecologic fallacy, that is, inferences made about groups based on individual-level data may be inaccurate. The
psychologistic and sociologistic fallacies have direct implications to this discussion. The psychologistic fallacy assumes that only person-level covariates are needed to explain individual outcomes, essentially leaving group-level variables out of the model. The sociologistic fallacy focuses on group-level covariates, but fails to include important person-level covariates. An example of the psychologistic fallacy provided by Diez-Roux (1998) is a sample of immigrants that show higher rates of depression, but for which the racial composition of the individuals’ communities was omitted. In fact, the higher rates of depression were only found for individuals who lived in communities where the immigrants were a minority.

This dissertation seeks to address a very specific question with respect to centering decisions for aggregated group-level. Models in which these aggregated upper-level variables are included will be referred to as “contextual models.”

**Multilevel models**

The nesting of hierarchical data can be diagrammed as a flow chart as depicted in Figure 1. The nature of the data in hierarchical models requires researchers to talk about different “levels” of the data (and analyses). One way of doing this is through the use of numbers. The lowest level is often referred to as “Level-1,” the next higher level “Level-2,” and so forth. In the pictured example, “Level-1” refers to the residents and “Level-2” refers to the neighborhoods. Another way of describing the levels is to use descriptive words. In the public health example, this could mean “resident level” versus “neighborhood level.” Nested data are often not limited to just two levels. Extending the public health example, residents might be nested within neighborhoods and those
neighborhoods nested within cities, resulting in a three-level nesting structure as demonstrated in Figure 3.

Longitudinal studies in which individuals are measured at multiple timepoints can also be viewed as nested data such that repeated measures are nested within person (and persons may be nested in upper-level clusters). A diagram of this model is in Figure 2. In these models, it is common to center the Level-1 time covariate around a fixed value as opposed to centering on the mean. For this reason, the focus of this dissertation will be on cross-sectional models.

Multilevel models offer methods for analyzing hierarchical data and enable better estimation of individual effects (Raudenbush & Bryk, 2002). For example, a researcher might be interested only in the effects at Level-1 (the resident level). He or she might be tempted to ignore the nested nature of the data and use a regression analysis that treats each resident’s score as independent. The trouble with this approach is that residents who come from the same neighborhood likely have more in common than residents who come from different neighborhoods. Thus, the independence assumption of regression analysis is violated. Models which do not account for this violation will have inefficient parameter estimates and biased estimates of the standard error. Multilevel models allow researchers to analyze data while accounting for the nested nature of the data, and result in better estimation of parameters and standard errors. Raudenbush and Bryk (2002) note that multilevel models are able to borrow strength from a larger pool of data to estimate regression equations for samples that are too small to provide strong estimates on their own (e.g. small numbers of minority students across 25 universities). Standard error estimates are improved in multilevel models by the modeling of separate random errors.
for each Level-2 unit, thus addressing the dependence among individuals from the same Level-2 group.

Multilevel models also make it possible to test cross-level hypotheses (Raudenbush & Bryk, 2002). In the public health example, a researcher might be interested in how a characteristic of the neighborhood (such as density of fast food restaurants) influences an effect at the resident level (such as the effect of diet on unhealthy weight). These types of research questions are not easily addressed in traditional models. Multilevel models, however, make it possible to estimate the effect of an interaction between Level-2 variables and Level-1 variables (as in the fast food example).

Finally, multilevel models allow researchers to deconstruct the variance of the outcome variable to better understand the source of the variability (Raudenbush & Bryk, 2002). In the public health example, a researcher might be interested in whether most of the variability in a health outcome is the result of differences between neighborhoods or between residents. Multilevel models provide variance estimates at each level of the model, enabling the researcher to better understand where most of the differences can be found.

**Centering**

Researchers sometimes consider “centering” a predictor variable (X) when using regression analysis. Centered variables are those that have been deviated from a particular value. An example is a raw score variable ($X_{ij}$) that is deviated from the overall mean ($\bar{X}$.) of the sample:
\( X_{CGM} = X_{ij} - \bar{X}. \)  

(1)

Other values besides the sample mean can also be used. In regression analyses on data that are not hierarchically nested, centering is sometimes used to create a meaningful zero point for a measure that does not have one. A classic example is when the \( X \) measures a person’s height or weight. In this case, interpretation of the intercept would be the expected value of \( Y \) when \( X \) is zero. Since neither height nor weight are ever zero, this is not particularly useful. However, by centering \( X \) on the mean of the sample, the interpretation of the intercept is the expected \( Y \) for an individual whose height or weight is at the sample mean of \( X \).

In two-level models there are two different options for mean centering of the Level-1 predictor variable. The first option is to center on the mean of the whole sample. This is referred to as “grand mean centering” or “centering grand mean” (CGM). In the public health example above, this would translate to taking the mean of \( X \) across all residents, regardless of where they live (neighborhood or city), and subtracting that value from the observed \( X \) for each resident:

\( X_{CGM} = X_{ij} - \bar{X}. \)  

(2)

The second option is to center the Level-1 predictor variable on the mean of that variable for the Level-2 group. This is referred to as “centering within context” (CWC). In the public health example, each neighborhood has its own mean for \( X \) which is the aggregated mean of \( X \) for the residents in that particular neighborhood. The \( X \) that is
centered within context would be the $X$ for an individual resident minus the mean for the neighborhood in which that person resides:

$$X_{CWC} = X_{ij} - \bar{X}_j$$

(3)

In the three-level model, the centering options expand. In addition to the CWC option in (3), the Level-1 $X$ variable can be centered on Level-3 mean. I note this form of centering within context as CWC-L3 to show that it is being centered on the Level-3 mean.

Although an option for centering, there is not a strong substantive reason why this form of centering would be useful in a three-level model. Thus, this form of centering is not considered further in this dissertation.

$$X_{CWC-L3} = X_{ijk} - \bar{X}_k$$

(4)

In addition to providing a meaningful zero point, Kreft, de Leuuw and Aiken (1995) suggest centering in multilevel models can be helpful in terms of “computational ease and stability.” They note that wide differences in scales may also be a valid reason to consider centering. Centering in multilevel models can also clarify parameter estimates making it easier to disentangle person-level effects from group-level effects. Finally, centering can produce orthogonal within and between coefficients, which improve interpretation and decreases collinearity in random coefficient models (Raudenbush, 1989a). From a practical perspective, within any particular study, centering can improve the likelihood of converging to a stable solution. The orthogonal coefficients produced
when centering within context allows a researcher to isolate the relationship at each level (e.g., the Level-1 effect only and/or the Level-2 effect only) devoid of potential confounding effects from relationships at other levels (e.g., a composite coefficient at Level-1 that is influenced by the relationship at Level-1 and at Level-2 but may not represent either effect accurately).

Centering in multilevel models presents a unique challenge because different centering choices have a significant impact on how the parameter estimate is interpreted. Raudenbush and Bryk (2002) illustrated this in a two-level model where the group mean of X was included at Level-2 as a predictor of the intercept. The authors showed that using group mean centering (CWC) at Level-1 resulted in direct estimation of the group-level effect at Level-2 (beta-between; $\beta_b$) and direct estimation of the individual-level effect at Level-1 (beta-within; $\beta_w$). When grand mean centering (CGM) is used at Level-1, the individual-level effect is still directly estimated at Level-1 ($\beta_w$) but instead of estimating $\beta_b$, the Level-2 parameter estimated is the contextual effect ($\beta_c$, or the difference between $\beta_w$ and $\beta_b$). In some cases, hypotheses around this contextual effect may be of substantive importance to the researcher. More importantly, inattention to the centering choices can result in misinterpretation of the Level-2 parameter estimate.

**Statement of the Problem**

When centering decisions for aggregated variables are not carefully made and explicitly stated, the ability to accurately interpret the resulting parameter estimates is compromised and can result in erroneous inferences. For example, when CGM is used at Level-1, the coefficient on the Level-2 mean may incorrectly be interpreted as the Level-2 effect rather than the Level-2 *contextual* effect. The purpose of this dissertation is
to clarify the interpretation of the fixed parameters and variance components in three-
level hierarchical models when centering within context (CWC), grand mean centering
(CGM) and uncentered raw scores (RAS) were used. Model 1 has a single predictor at
Level-1. Model 2 adds the Level-2 (neighborhood) mean to Model 1. Model 3 adds the
Level-3 (city) mean to Model 2. The full RAS equations for these models are:

Model 1: \( Y_{ijk} = \gamma_{000} + \gamma_{100}X_{ijk} + e_{ijk} + r_{0jk} + r_{1jk}X_{ijk} + u_{00k} \)  \( (5) \)

Model 2: \( Y_{ijk} = \gamma_{000} + \gamma_{010}\bar{X}_{jk} + \gamma_{100}X_{ijk} + e_{ijk} + r_{0jk} + r_{1jk}X_{ijk} + u_{00k} \)  \( (6) \)

Model 3: \( Y_{ijk} = \gamma_{000} + \gamma_{001}\bar{X}_{..k} + \gamma_{010}\bar{X}_{jk} + \gamma_{100}X_{ijk} + e_{ijk} + r_{0jk} + r_{1jk}X_{ijk} + u_{00k} \)  \( (7) \)

Model 1 and Model 3 are three-level extensions of the two-level models bearing the
subscripts 1 and 2, respectively, in Kreft, de Leeuw, and Aiken (1995). In all three of
these models, \( Y_{ijk} \) is the Y score for person \( i \) in neighborhood \( j \) in city \( k \); \( X_{ijk} \) is the X
score for person \( i \) in neighborhood \( j \) in city \( k \); \( \bar{X}_{jk} \) is the neighborhood mean of
neighborhood \( j \) in city \( k \); \( \bar{X}_{..k} \) is the city mean for city \( k \); \( \bar{X}_{..} \) is the overall mean. The error
term \( e_{ijk} \) represents that which is unique to person \( i \) in neighborhood \( j \) and city \( k \). In each
of these models, each neighborhood and city are allowed to have their own intercept. The
error terms \( r_{0jk} \) and \( u_{00k} \) reflect the uniqueness associated with the neighborhood and
city, respectively. All three models also allow each neighborhood to have its own Level-1
slope, resulting in the error term \( r_{1jk} \).
Model 2 and Model 3 are contextual models in that they include the group mean, aggregated from the Level-1 data, as a predictor at Level-2 and/or Level-3. As discussed earlier, these models are important because they allow examination of the effects of aggregated variables on the individual’s outcome separate from the person-level effect, which can have important substantive implications in many disciplines. It is in these contextual models that the interpretation of the coefficients is most affected by the choice of centering.

Significant contextual effects may or may not be present. As will be explored in the discussion, decisions about whether to model a contextual effect should ultimately be driven by the research hypothesis. However, in this dissertation, I will explore both situations regarding contextual effects. For the first research question, I will explore the interpretation of coefficients resulting from centering decisions when the model fits the data. The data for these models will be generated such that if contextual effects will be included in the model, then the contextual effects will be present in the data. For the second research question, I will explore the implications for interpretation of the coefficients when the model does not fit the data. That is, when contextual effects are present in the data, but not included in the model.

**Research Questions**

This study addressed two research questions:

1. What is the interpretation of the coefficients and variance terms in each model under RAS, CWC and CGM when the model fits the data?

2. How do the coefficients and variance terms under RAS, CWC and CGM differ when the model does not fit the data?
Given the extensive work done in this area using two-level models, I anticipate that for Model 1 and Model 3 in particular, some of the findings will be quite similar to those reported by Kreft, de Leeuw and Aiken (1995) and Enders and Tofighi (2007). Model 2 is unique to the three-level models and will be a key extension to the literature in this area.

**Limitations**

There are a vast number of combinations of centering options and models that could be conceived in the three-level framework; this study will address a limited number of these models chosen for their direct association with previous literature (Kreft, de Leeuw, & Aiken, 1995). Any number of global variables can be included at the three levels in these models; this dissertation focuses on a Level-1 variable that is aggregated at Level-2 and Level-3 because the centering choices directly affect the interpretation of the coefficients associated with these aggregated terms. As noted above, although hierarchical linear models are useful for longitudinal data as well as cross-sectional data; this study will only be considering the cross-sectional approach. Centering decisions in longitudinal models take on a different role than in cross-sectional models, and often involve centering around a constant rather than group or overall mean values (Singer & Willett, 2003; Hedeker & Gibbons, 2006). Finally, the use of artificially generated data may not accurately mimic idiosyncracies found in data naturally occurring in the field.
Notation

Multilevel models employ a unique notation. For this study, the following notation was used:

<table>
<thead>
<tr>
<th>Level-1</th>
<th>Variables</th>
<th>$X_{ijk}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficients</td>
<td>$\pi$</td>
</tr>
<tr>
<td></td>
<td>Error Term</td>
<td>$e$</td>
</tr>
<tr>
<td></td>
<td>Error Variances</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level-2</th>
<th>Aggregated Level-1 Variables</th>
<th>$\bar{X}_{jk}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficients</td>
<td>$\beta$</td>
</tr>
<tr>
<td></td>
<td>Error Terms</td>
<td>$r$</td>
</tr>
<tr>
<td></td>
<td>Error Variances</td>
<td>$\tau_{\pi}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level-3</th>
<th>Aggregated Level-1 Variables</th>
<th>$\bar{X}_{.k}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficients</td>
<td>$\gamma$</td>
</tr>
<tr>
<td></td>
<td>Error Terms</td>
<td>$u$</td>
</tr>
<tr>
<td></td>
<td>Error Variances</td>
<td>$\tau_{\beta}$</td>
</tr>
</tbody>
</table>

Subscripts are used to help identify the levels in multilevel models. Returning to the public health example, and extending it to a three level model, we can consider residents (Level-1) nested in neighborhoods (Level-2) nested in cities (Level-3).

The independent variables at Level-1 are written as $X_{ijk}$, where $i$ indicates Level-1 (resident), $j$ indicates Level-2 (neighborhood) and $k$ indicates Level-3 (city). Level-2 variables in this dissertation are measured at Level-1 and then aggregated at Level-2 and this is denoted using a period in the subscript for the level across which the term is
aggregated, and a bar above the X to indicate this variable is a mean. So, a Level-1 variable (X) averaged across Level-2 units is noted as $\bar{X}_{jk}$. Finally, the grand mean of that Level-1 variable is noted as $\bar{X}_{..}$ indicating the mean of X is taken across all three levels. Independent variables at Level-3 follow the same conventions as those at Level-2.

The coefficients follow a similar notation with a few exceptions. The coefficients of independent variables at Level-1 are written as $\pi_{ijk}$, but now $i$ indicates the intercept or unique predictor such as $\pi_{0jk}, \pi_{1jk}, \ldots, \pi_{njk}$. At Level-2, the coefficients are written as $\beta_{ijk}$, where $i$ indicates which Level-1 coefficient the variable is specified to predict (e.g., $\beta_{0jk}$ is predicting the Level-1 intercept, $\beta_{1jk}$ is predicting the slope of the Level-1 X variable, etc.) and $j$ indicates the intercept or unique Level-2 predictor (similar to $i$ at Level-1). Coefficients at Level-3 follow similar conventions as those at Level-2.

As an example, consider a three-level model with a single predictor at Level-1 (Model 1):

\[
Y_{ijk} = \pi_{0jk} + \pi_{1jk}(X_{ijk}) + e_{ijk} \quad \text{Level-1 (L1) model} \tag{8}
\]
\[
\pi_{0jk} = \beta_{00k} + r_{0jk} \quad \text{L2 model for the L1 intercept} \tag{9}
\]
\[
\pi_{1jk} = \beta_{10k} + r_{1jk} \quad \text{L2 model for the L1 slope} \tag{10}
\]
\[
\beta_{00k} = \gamma_{000} + u_{00k} \quad \text{L3 model for L2 intercept of L1 intercept} \tag{11}
\]
\[
\beta_{10k} = \gamma_{100} \quad \text{L3 model for L2 intercept of L1 slope} \tag{12}
\]

Substituting across these equations produces the full model:

\[
Y_{ijk} = \gamma_{000} + \gamma_{100}(X_{ijk}) + e_{ijk} + r_{0jk} + r_{1jk}(X_{ijk}) + u_{00k} \tag{13}
\]
In this model, and the models that follow, the \( \gamma \) terms are “fixed effects” and the \( e, r, \) and \( u \) terms are “random effects”. Fixed effects, indicated by regression coefficients, represent the average relationship between \( X \) and \( Y \) across the sample (Snijders, 2005). When it is possible, or hypothesized, that a mean (in the case of the intercept) or relationship (in the case of the slopes) might differ within clusters (i.e. the effect is different in different families or schools or neighborhoods), a term is added to the model to capture the variability in these differing effects. These are represented by the \( e, r, \) and \( u \) terms.

**Summary**

Epidemiology is ripe with opportunities to utilize multilevel models given the complex nested nature of data in the field. This study seeks to clarify the interpretation of coefficients in centered (CGM or CWC) and uncentered (RAS) three-level models. The focus will be on three-level cross-sectional models and the motivating example will come from public health with residents nested within neighborhoods nested within cities.
CHAPTER 2: LITERATURE REVIEW

Historical Overview

Multilevel models have a broad history beginning in the late 1980’s, and have included several texts (e.g., Bock, 1989; Bryk & Raudenbush, 1992; Goldstein, 1987). Snijders & Bosker (1999) discuss the history of multilevel models as the merging of two areas: contextual analysis and mixed effects models. Because the literature to date focused on two-level hierarchical models, this chapter will also be focused on the two-level model.

Contextual analysis focuses on how social context influences individual behavior. The ‘ecological fallacy’ discussed in the introduction is an example. In multilevel models, it is possible to consider the relationship between a predictor and an outcome at both the person level and the neighborhood level through the use of aggregated (a.k.a. contextual) variables. Of interest to many researchers are the effects of these aggregated variables on the individual’s outcome, $Y_{ij}$, particularly if the effect is different from the effect of $X_{ij}$ on $Y_{ij}$ (Feaster, Brincks, Robbins, & Szapocznik, 2011). Chan (1998) used the term "additive composition models" to describe these models. Firebaugh (1978) discussed cross-level bias, or “the difference, in a population, between the aggregate-level regression coefficient obtained and the individual-level coefficient of interest” (p. 560). Blalock (1984) describes the models as contextual-effects models and identifies as their primary feature “an allowance for macro processes that are presumed to have an impact on the individual actor over and above the effects of any individual-level variables that may be operating” (p. 354). Raudenbush and Bryk (2002) also termed these “contextual effects” and identified them as something that occurs when “the aggregate of
a person-level characteristic, $X_{ij}$, is related to the outcome $Y_{ij}$, even after controlling for the effect of the individual characteristic, $X_{ij}$” (p. 139).

Mixed effects models differ from traditional regression models in that some of the coefficients can be modeled as random while others are modeled as fixed. When considered in conjunction with contextual modeling, multilevel models quickly moved in the direction of separate variance estimates for individual and contextual predictors, both modeled as random effects.

Centering Literature

As discussed in Chapter 1, the work of Raudenbush (1989a) established the usefulness of centering within context for identifying contextual effects, estimating cross-level interactions, and identifying slopes for random regressions in two-level models. One major statistical advantage to centering within context is the resulting orthogonality of the centered Level-1 variable and the Level-2 mean which leads to more accurate parameter estimates (Raudenbush, 1989a; Paccagnella, 2006). Kreft, de Leeuw, and Aiken (1995) algebraically compared different centering options (RAS, GMC, CWC) for Level-1 predictors and identified those techniques which resulted in theoretically equivalent models. Kreft, de Leeuw, and Aiken (1995) defined equivalence as equality in both (a) the model implied expected values of the outcome, $\hat{Y}_{ij}$, (i.e., expectations) and (b) the model implied variance of the outcome, $Var(Y_{ij})$, (i.e., dispersions).

There have been few critics of centering (Plewis, 1989; Longford, 1989a, 1989b). Plewis (1989) was concerned that the correlation between $X_{ij}$ and $\bar{X}_j$ may not always be large enough to pose a problem, an issue for which Raudenbush (1989b) provided significant contrary evidence. Plewis also pointed out that contextual effects could be
measured using statistics other than the mean (for instance, the median), which is true, but which does not produce the orthogonality to the group mean that CWC produces.

Longford critiqued the accuracy of the group mean as a measure of context, particularly in studies with small sample sizes within Level-1 groups (Longford, 1989a). He was also concerned about the change in specification of the random effects, particularly in random slope models, when centering within context was used (Longford, 1989b), again because of the question about how well the Level-2 mean captures “context” when the group size is small.

**Centering in Two-Level Models**

Enders and Tofighi (2007) examined centering in a cross-sectional design, considering a model with a single Level-1 predictor in which both the intercept and slope were allowed to vary at Level-2:

\[
Y_{ij} = \pi_0 + \pi_1 (X_{ij}) + e_{ij} \quad \text{the Level-1 model} \quad (14)
\]

\[
\pi_0 = \beta_{00} + r_{0j} \quad \text{the Level-2 model for the intercept} \quad (15)
\]

\[
\pi_1 = \beta_{10} + r_{1j} \quad \text{the Level-2 model for the slope} \quad (16)
\]

\[
Y_{ij} = \beta_{00} + \beta_{10} (X_{ij}) + e_{ij} + r_{0j} + r_{1j} (X_{ij}) \quad \text{the full model} \quad (17)
\]

Grand mean centering (CGM) deviates the raw score \((X_{ij})\) from the overall mean across all individuals in the sample \((\bar{X})\) to produce \((X_{ij} - \bar{X})\). The authors noted that centering along the grand mean creates a composite score that is made up of a deviation of the raw score from the group mean \((X_{ij} - \bar{X}_j)\) and the deviation of the group mean from the grand mean \((\bar{X}_j - \bar{X})\):
\[ (X_{ij} - \bar{X}) = (X_{ij} - \bar{X}_j) + (\bar{X}_j - \bar{X}) \]  

(18)

Partitioning the CGM X variable in this way highlights the two sources of variability in X: within-group variability \((X_{ij} - \bar{X}_j)\) and between-group variability \((\bar{X}_j - \bar{X})\).

CWC, sometimes referred to as group mean centering, deviates the raw score \(X_{ij}\) from the mean across individuals in the same group \((\bar{X}_j)\), producing \((X_{ij} - \bar{X}_j)\). As a result, the mean of the centered scores for every group is zero, and there is no between-group variation.

A fundamental difference between CGM and CWC is that CGM produces scores that are correlated with variables at both Level-1 and Level-2, while CWC produces scores that are uncorrelated with variables at Level-2. Another way to say this is that with CGM, the between group variance that was in \((X_{ij})\) is still present in \((\bar{X}_j - \bar{X})\) while the same is not true for CWC. In this way, centering within context (CWC), “fundamentally alters the mean and correlation structure of the data” (Enders & Tofighi, 2007, p. 127).

**Effects of CWC and CGM on Level-1 intercepts.** Centering choices have differing effects on the interpretation of the intercept term. The \(j\) subscript on the intercept term \((\pi_{0j})\) in equation (15) indicates that this term represents the intercept for group \(j\). When no centering is employed, and the raw score is used, the interpretation of \(\pi_{0j}\) is the expected value of Y for an individual whose score on X equals zero.

When CWC is employed, the equation changes to:

\[ Y_{ij} = \pi_{0j} + \pi_{1j}(X_{ij} - \bar{X}_j) + e_{ij} \]  

(19)
The interpretation of $\pi_{0j}$ here is the expected value of $Y$ for an individual whose score on $X$ is identical to the mean score of $X$ for their group ($X_{ij} = \bar{X}_j$). Taking the expectation of (20) within each group, we find the intercept is equal to the mean for group $j$:

$$\pi_{0j} = \mu_{yj}. \quad (20)$$

$\pi_{0j}$ is then interpreted as the unadjusted mean for group $j$. $\beta_{00}$ is the average unadjusted cluster mean, averaged across all of the clusters.

When CGM is employed, the intercept is a little more complex. The Level-1 equation is:

$$Y_{ij} = \pi_{0j} + \pi_{1j}(X_{ij} - \bar{X}_j) + e_{ij} \quad (21)$$

$\pi_{0j}$ is the expected value of $Y$ for the person whose $X$ score is equal to the grand mean of $X$ ($X_{ij} = \bar{X}_\cdot$). Taking the expectation of (21) within each cluster, we find that the intercept is the expected $Y$ for cluster $j$ minus an additional term made up of a regression slope ($\pi_{1j}$) and the deviation of $X_{ij}$ from $\bar{X}_\cdot$ ($X_{ij} - \bar{X}_\cdot$):

$$\pi_{0j} = \mu_{yj} - \pi_{1j}(\bar{X}_j - \bar{X}_\cdot) \quad (22)$$
When CGM is used, the Level-1 intercept, \( \pi_{0j} \), is no longer the mean for group \( j \), but is instead an adjusted mean for group \( j \) similar to analysis of covariance (ANCOVA). Enders and Tofighi (2007) refer to this as “the expected \( Y \) for a case belonging to cluster \( j \) after ‘equating’ the clusters on average \( X \)” (p. 125). The implication of this for the combined model is that \( \beta_{00} \) is an average adjusted mean intercept.

**Effects of CWC and CGM on Level-1 Slopes.** In CWC, the mean of \( (X_{ij} - \bar{X}_j) \) is zero for every group, thus there is no between-cluster variation in \( X \). Under CWC, \( \pi_{1j} \) is the pooled within-cluster regression of \( Y \) on \( X \). CGM, by contrast, does not result in the group mean of \( X \) being equal to zero, so there is still between-group variation in the \( X \) term. This makes it difficult to interpret the slope coefficient as it is not a simple pooled within-cluster regression coefficient, but is instead a combination of the within and between regression coefficients. Raudenbush and Bryk (2002) show that the regression coefficient produced under CGM in a two-level random intercept model with one predictor at Level-1 (and no predictors at Level-2) is a weighted average such that:

\[
\gamma_{10} = \frac{WGT_1 \beta_w + WGT_2 \beta_b}{WGT_1 + WGT_2} \quad (23)
\]

where \( \beta_w \) represents the within-cluster regression and \( \beta_b \) represents the between-cluster regression. In the words of the authors, the weights, \( WGT_1 \) and \( WGT_2 \), “are quite complex for the general case” (p. 138).

**Effects of CWC and CGM on Variances and Covariances.** Centering can also have an effect on the variances and covariances between the intercepts and slopes in this model. Remember in a random coefficient model each group is allowed to have its own
intercept and its own slope, thus each of these parameters has variability, and there is the potential for covariance between them. The variance of the intercept is frequently named $\tau_{00}$ and the variance of the slope $\tau_{11}$. The covariance between the intercept and the slope is $\tau_{10}$ which equals $\tau_{01}$.

**Variance of the Level-1 intercepts ($\tau_{\pi_{00}}$).** When CWC is employed, $\tau_{00}$ refers to the between-group variation in the intercepts. Because the Level-1 intercept ($\pi_{0j}$) in a CWC model is the group mean, this is similar to saying that $\tau_{00}$ represents the between-group variation in $Y$. As Enders and Tofighi (2007) point out “$\tau_{00}$ from CWC should be nearly equivalent to $\tau_{00}$ from an unconditional model with no predictors” (p. 127). When CGM is used, the intercept is the group mean adjusted by the regression coefficient and deviation of the group mean from the grand mean described earlier. So, $\tau_{00}$ in the context of CGM represents variation in the adjusted group mean.

Centering at Level-1 influences the estimated variability of the Level-1 intercept (Snijders & Bosker, 1999, p 69). If the slope of $Y$ on $X$ varies across groups, the location of the intercept can mean different things for the variability of the random intercept ($\tau_{\pi_{00}}$) and the covariances between the random intercept and the random slope ($\tau_{\pi_{01}}$). In the figure on the left in Figure 4, $X_1$ and $X_2$ are two possible locations for the intercept. If $X_1$ represents the intercept, there appears to be no variability in the intercepts. However, if $X_2$ represents the intercept, there appears to be significant variability in the intercepts. In the figure on the right in Figure 4, the slopes are fixed, therefore location of the intercept does not influence the estimated variability of the intercept.

**Covariance of $r_{0jk}$ and $r_{1jk}$ ($\tau_{\pi_{01}}$).** Although assumed not to covary with residuals at other levels, the two residual terms at Level-2 are expected to covary with
one another. $\tau_{01}$ represents the covariance between these residuals. Snijders and Bosker (1999, p. 69) discourage forcing the value of $\tau_{01}$ to be zero in the presence of random slopes. Interpretation of a negative correlation between slope and intercept implies that neighborhoods with a higher mean $Y$ for a resident with an $X$ equal to his/her average neighborhood $X$ (in CWC model) or equal to the average population $X$ (for CGM model) have a lower within-neighborhood effect of $X$ on $Y$.

**Variance of the Level-1 slopes ($\tau_{11}$).** For the CWC model, the interpretation of $\tau_{11}$ is straightforward. Because there is no between-group variation in the slopes, they represent the pooled within-group regression of $Y$ on $X$. Thus the variance of the slopes represents the variance in these regressions.

Variance of the Level-1 slopes in the CGM model is more complicated. Hox (2002) explains that conducting separate within-group regressions for each Level-2 group using ordinary least squares produces slope estimates with more variability than the variability seen when the regressions are done in a multilevel framework. This is because multilevel modeling uses a method of estimating that shrinks individual slope estimates toward the overall mean of the individual slope estimates. The amount of shrinkage applied is influenced by the reliability of the group’s data. Estimates of a group’s intercept and slope are shrunk toward the grand mean for groups that have small sample sizes or whose estimate is significantly different from the overall estimate. Because estimates from small groups are less reliable than estimates from large groups, estimates from small groups experience more shrinkage. Similarly, when a group estimate is very far from the overall estimate, the reliability is assumed to be small and the group estimate is shrunk toward the overall estimate. Although the shrinkage in these new estimates
(referred to as empirical Bayes estimates) causes them to be biased, they are also more precise. Raudenbush and Bryk (2002) further explain that the empirical Bayes estimation procedures in the CGM model result in estimates of the variability of the slopes ($\tau_{\pi_{11}}$) that appear artificially minimized. This is the result of the estimate of $\tau_{\pi_{11}}$ being based on the square of the empirical Bayes residual, and the fact that under CGM and CWC these residuals are defined differently (Raudenbush & Bryk, 2002, p. 145).

In the CWC model, the estimates of the intercept term will experience some shrinkage (relative to the simple OLS estimates), however this only affects the slope estimate if there is a strong correlation between the intercept and slope. Under CWC, the slope estimates will also experience shrinkage, but this is relatively unaffected by the shrinkage in the intercept estimates. This is a result of the intercept in the CWC model being an unadjusted mean; unlike in the CGM model, where the intercept is influenced by the slope as seen in the deviations of the expectations earlier in this chapter (see equation 22).

In the CGM model, the intercept is an adjusted mean and influenced by the slope (see equation 22). The problem is that when this adjusted mean is created, the value may fall outside the range of the actual data collected. This is particularly a problem when the upper level cluster mean is very different from the grand mean. Because these estimates of the intercept are then very far from the mean estimate of the intercept, they are treated as less reliable and are attenuated toward the mean estimate for the empirical Bayes estimate. Under CGM, a consequence of this extensive shrinkage is a significant alteration of the slope estimate. The result is less variability among the slope estimates and an underestimation of $\tau_{\pi_{11}}$. Therefore the recommendation from Raudenbush and
Bryk (2002) is to use CWC to accurately estimate the variance of the Level-1 slopes (p. 143).

Raudenbush and Bryk (2002) show for balanced data that the empirical Bayes estimate for the intercept in a two-level, intercept-only model can be represented as:

\[
\hat{\beta}_{0j}^{EB} = \lambda_j \hat{\beta}_{0j}^{OLS} + (1 - \lambda_j) \hat{\gamma}_{00}
\]

In this equation, \( \hat{\gamma}_{00} \) is the overall intercept and \( \lambda_j \) represents the reliability of the OLS estimated beta as an estimate of the true beta. The equation for \( \lambda_j \) is:

\[
\lambda_j = \frac{\tau_{00}}{\tau_{00} + \frac{\sigma^2}{n_j}}
\]

Here, \( \tau_{00} \) represents variance of the Level-2 residuals, \( \sigma^2 \) represents variance of the Level-1 residuals and \( n_j \) represents sample size for group j. \( \lambda_j \) is thus the variance of the intercept divided by the intercept variance plus the error variance. When group sizes are large, \( \lambda_j \) approaches 1.0. \( \lambda_j \) can also approach 1.0 when there is large variability of the intercepts across groups (\( \tau_{00} \)).

**Equivalence between different forms of centering and raw scores.** Theoretical equivalence between models is met when two models generate the same expectations and dispersions. More simply stated, in equivalent models the estimates produced by one model can be recalculated from the estimates produced by another model. Kreft, de Leeuw, and Aiken (1995) illustrated the equivalence of raw scores (RAS) and CGM in
fixed effects models, random coefficient models, random coefficient models in which the group mean predicts the random intercept, and in random coefficient models in which the group mean predicts both the random intercept and the random slope. CWC models with random coefficients are not equivalent to either RAS or CGM models. However, CWC models in which the group mean predicts the random intercept are equivalent in their fixed parts, but not variance parts, to the same models employing RAS or CGM. When the slopes are fixed, CWC models with the group means predicting the random intercept are equivalent to RAS and CGM models. Table 1 summarizes these findings.

Three-Level Models

Many of the concepts for two-level models translate to three-level models. Variance partitioning is a good example. In the two-level model, the intraclass correlation (ICC) can be defined in two ways: 1) the proportion of variance within Level-2 clusters (within neighborhood) and 2) the proportion of variance between Level-2 clusters (between neighborhoods). In the three-level model, the definition of the ICC at Level-2 changes to the proportion of variance between neighborhoods within cities. Because of the third level of clustering, an additional definition of the ICC is also possible: the proportion of variance between Level-3 clusters (between cities). The equations in the three-level model will be shown in Chapter 4 and are very similar to those for the two-level model. Similarly, Raudenbush and Bryk (2002) note “testing hypotheses in three-level models is directly analogous to the procedures … for two-level models” (p.234).

Summary

As summarized in this chapter, significant work has been done to establish the usefulness and implications of centering in the two-level framework. Kreft, de Leeuw and
Aiken (1995) provided a foundation for comparing CWC models to CGM models in two levels. Enders and Tofighi (2007) extended this work to clarify the implications of CWC and CGM in two-level models and provided an important link between centering decisions and research questions. However, the consequences of centering in the three-level model have not been elucidated. This study seeks to extend this previous work into cross-sectional three-level hierarchical linear models.
CHAPTER 3: METHODS

Research Methodology

This study investigated the implications of centering in three-level models, focusing on three models. The raw score equations at each level for Model 1 are:

\[ Y_{ijk} = \pi_{0jk} + \pi_{1jk}(X_{ijk}) + e_{ijk} \] (26)
\[ \pi_{0jk} = \beta_{00k} + r_{0jk} \] (27)
\[ \pi_{1jk} = \beta_{10k} + r_{1jk} \] (28)
\[ \beta_{00k} = \gamma_{000} + u_{00k} \] (29)
\[ \beta_{10k} = \gamma_{100} \] (30)

Substituting across the levels results in the full model for Model 1:

\[ Y_{ijk} = \gamma_{000} + \gamma_{100}(X_{ijk}) + e_{ijk} + r_{0jk} + r_{1jk}(X_{ijk}) + u_{00k} \] (31)

The raw score equations for each level of Model 2 are:

\[ Y_{ijk} = \pi_{0jk} + \pi_{1jk}(X_{ijk}) + e_{ijk} \] (32)
\[ \pi_{0jk} = \beta_{00k} + \beta_{01k}(\bar{X}_{jk}) + r_{0jk} \] (33)
\[ \pi_{1jk} = \beta_{10k} + r_{1jk} \] (34)
\[ \beta_{00k} = \gamma_{000} + u_{00k} \] (35)
\[ \beta_{01k} = \gamma_{010} \] (36)
\[ \beta_{10k} = \gamma_{100} \] (37)
Substituting across the levels results in the full model:

\[ Y_{ijk} = \gamma_{000} + \gamma_{010}(\bar{X}_{jk}) + \gamma_{100}(X_{ijk}) + u_{00k} + r_{0jk} + r_{1jk}(X_{ijk}) + e_{ijk} \]  

(38)

The raw score equations for each level of Model 3 are:

\[ Y_{ijk} = \pi_{0jk} + \pi_{1jk}(X_{ijk}) + e_{ijk} \]  

(39)

\[ \pi_{0jk} = \beta_{00k} + \beta_{01k}(\bar{X}_{jk}) + r_{0jk} \]  

(40)

\[ \pi_{1jk} = \beta_{10k} + r_{1jk} \]  

(41)

\[ \beta_{00k} = \gamma_{000} + \gamma_{001}(\bar{X}_{.k}) + u_{00k} \]  

(42)

\[ \beta_{01k} = \gamma_{010} \]  

(43)

\[ \beta_{10k} = \gamma_{100} \]  

(44)

Substituting across the levels results in the full model for Model 3:

\[ Y_{ijk} = \gamma_{000} + \gamma_{001}(\bar{X}_{.k}) + \gamma_{010}(\bar{X}_{.jk}) + \gamma_{100}(X_{ijk}) + u_{00k} + r_{0jk} + r_{1jk}(X_{ijk}) + e_{ijk} \]  

(45)

As discussed in Chapter 1, these models were chosen for consistency with previous work on centering in multilevel models (Kreft, de Leeuw & Aiken, 1995). Taken together, these three models represent the most likely choices for centering and inclusion of upper-level means.
The study was undertaken in four steps. **Step 1** built upon the work by Kreft, de Leeuw and Aiken (1995) to determine equivalencies in the means and variances across the centering options in each model. This required finding the first (expectations) and second (variances) moments of the full equation for each model within each centering option. These moments were then equated between the centering options to determine whether the centered and uncentered models were theoretically equivalent.

**Step 2** clarified the relationships between the three-level RAS model and the CWC model for the two contextual models (Model 2 and Model 3). Using the work of Raudenbush (1989a) as a guide, this step required algebraic manipulation of the fixed effects to highlight under which circumstances the coefficient was estimating a contextual effect (the difference between an upper level and lower level regression) and when the coefficient was estimating a direct effect.

**Step 3** addressed the first research question by utilizing the equivalencies derived in Step 1 and the algebraic relationships identified in Step 2 to determine the interpretation of the terms in each of the models. These interpretations were then demonstrated using artificially generated datasets. The artificial datasets were also used to verify the equivalencies mapped out in Step 1.

**Step 4** addressed the second research question by estimating Model 1 using Model 3 data. This was done to highlight the importance of testing for contextual effects, and how lack of caution in choice of centering can lead to incorrect inferences.

**Data Generation**

Population values for data generation were initially taken from a multilevel, publicly available dataset (Study of Instructional Improvement). Unfortunately, the
parameter values that were estimated from this data were not suitable for further use in this dissertation. The primary problems were lack of significant contextual effects at the upper levels in Model 3 and very small intraclass correlations.

Using the parameter estimates from the Study of Instructional Improvement data as a starting point, I made significant alterations to the population parameters for the Level-1, Level-2 and Level-3 slopes. My first priority was that these effects be statistically significant for both the centered and uncentered models. My second priority was that the direction of the effects fit the motivating example. I also altered the variances so that the ICC’s at the upper levels were larger. The resulting parameter estimates are in Table 2.

Using the population values, thirty artificial datasets were generated, 10 datasets for each of the three models. Multiple datasets were generated for each model so that sampling anomalies did not influence the results. First, 300 random numbers for use as seeds throughout the project were created using the following command in STATA10: gen random=1 + int(100000*uniform()). These seeds were then used as part of the SAS command files to generate the artificial datasets. The artificial datasets were balanced and included 30 Level-3 units with 15 Level-2 units in each Level-3 unit, and 15 Level-1 units in each Level-2 unit (N = 6,750). SAS command files used for data generation are in Appendix A.

For each model, the ten artificial datasets were imported to HLM as part of the multiple imputation feature. Analyses were run on all ten datasets, and parameter estimates were averaged across the ten runs. The intraclass correlations for each artificial dataset are in Table 3.
Software

Parameter estimates were obtained using HLM 6.08 (Raudenbush, Bryk & Congdon, 2004). The HLM software was selected for its capability to model three-level, cross-sectional, hierarchical models. Artificial datasets were created using SAS 9.2.1 (SAS Institute Inc., 2008).

Motivating Example

To improve readability, the example from epidemiology will continue to be used throughout the results. In the example, Level-1 units are individual residents, Level-2 units are neighborhoods and Level-3 units are cities. The predictor variable, X, is walking habits measured in hours of walking per week. The outcome variable, Y, is a continuous measure of overall health for which larger, positive values equal better health. Assuming that the more hours of walking a person does per week, the better his or her health score will be, the population values for the regression coefficients at each level are all positive. Also assuming that the effect of an individual’s walking habits has a greater impact on his or her health than the effect of the walking habits of an individual’s neighbors, the raw score population value for the effect of the neighborhood mean $X (\bar{X}_{jk})$ is negative indicating the individual resident effect is larger than the neighborhood effect. Similarly, the raw score population value for the effect of the city mean $X (\bar{X}_{.k})$ is negative, indicating the neighborhood effect is larger than the city effect.
CHAPTER 4: RESULTS

This study investigated the implications of centering in three-level models. The primary goal was to identify the interpretations of the coefficients and variance terms in each model under CWC, CGM and RAS. The goal of Step 1 was to determine equivalencies in the means and variances across the centering options in each model. The goal of Step 2 was to identify the algebraic relationships between the centering options in the two contextual models, Model 2 and Model 3. The goal of Step 3 was to determine the interpretations of the terms in each of the models using artificially generated datasets to provide empirical examples. The goal of Step 4 was to demonstrate the implications of model misspecification in the three-level framework.

Research Questions

The purpose of this dissertation was to examine the effects of centering variables in three-level models. This dissertation was guided by two main research questions:

1. What are the interpretations of the coefficients and variance terms in each model under RAS, CWC and CGM when the model fits the data?

2. How do the coefficients and variance terms under RAS, CWC and CGM differ when the model does not fit the data?

This chapter is arranged as follows: First, the centering options in the three-level framework are discussed. Next, means and correlations between centered and uncentered variables in three levels are discussed and demonstrated using an artificial dataset. Third, equivalencies between the centered and raw score estimates are derived for the three models (Step 1). Fourth, the algebraic equivalencies of the centered and uncentered terms in the contextual model are presented (Step 2). Fifth, the parameter estimates are
discussed, and results from the artificial datasets are presented (Step 3). Finally, parameter estimates resulting from model misspecification are presented (Step 4).

**Centering Options in Three-Level Models**

At Level-1, grand mean centering (CGM) deviates the person’s raw score \(X_{ijk}\) from the grand mean, which is the mean across all individuals in the sample \(\bar{X}_{...}\), to produce \(X_{ijk} - \bar{X}_{...}\). Similar to what Enders and Tofighi (2007) demonstrated in the two-level model, centering on the grand mean in a three-level model creates a composite score that can be expressed as the sum of three terms: (1) a deviation of the raw score from the Level-2 (neighborhood) group mean \(X_{ijk} - \bar{X}_{jk}\), (2) a deviation of the Level-2 (neighborhood) group mean from the Level-3 group mean (city) \(\bar{X}_{jk} - \bar{X}_{..k}\), and (3) a deviation of the Level-3 (city) group mean from the grand mean \(\bar{X}_{..k} - \bar{X}_{...}\):

\[
(X_{ijk} - \bar{X}_{...}) = (X_{ijk} - \bar{X}_{jk}) + (\bar{X}_{jk} - \bar{X}_{..k}) + (\bar{X}_{..k} - \bar{X}_{...}) \tag{46}
\]

Consistent with notation used by Enders (in press) these three segments can be referred to as \(W_1\), \(B_2\), and \(B_3\) (a new term, unique to the Level-3 model) such that:

\[
(X_{ijk} - \bar{X}_{jk}) = W_1 \tag{47}
\]
\[
(\bar{X}_{jk} - \bar{X}_{..k}) = B_2 \tag{48}
\]
\[
(\bar{X}_{..k} - \bar{X}_{...}) = B_3 \tag{49}
\]
This notation highlights the within-neighborhood variability in X among residents ($W_1$), the between neighborhood variability in X ($B_2$) and the between-city variability in X ($B_3$). The CGM variable thus has three sources of variability.

By contrast, centering within context (CWC) at Level-1 deviates the raw score ($X_{ijk}$) from the mean across individuals in the neighborhood ($\bar{X}_{jk}$), to produce ($X_{ijk} - \bar{X}_{jk}$). In a three-level framework, there are two ways of conceptualizing CWC for the Level-1 variable. One is to center around the mean of the Level-2 group (neighborhood; $\bar{X}_{jk}$); another is to center around the mean of the Level-3 group (city; $\bar{X}_{..k}$). As discussed in Chapter 1, this latter form of centering at Level-1 ($X_{ijk} - \bar{X}_{..k}$) is not explored further in this dissertation because this form of centering is not likely in the presence of significant variability in the outcome due to neighborhood clustering.

For this dissertation, Level-1 variables that are centered within context along the Level-2 group (neighborhood) mean will be referred to as CWC-L2. Level-2 variables that are centered within context around the Level-3 group (city) mean will be referred to as CWC-L3. Rearranging the terms in (46) to solve for the CWC-L2 Level-1 term, demonstrates how between-cluster variation (both $B_2$ and $B_3$) is removed and therefore not present in the CWC-L2 ($W_1$) term:

$$
(X_{ijk} - \bar{X}_{jk}) = (X_{ijk} - \bar{X}_{..}) - (\bar{X}_{jk} - \bar{X}_{..}) - (\bar{X}_{..k} - \bar{X}_{..})
$$

$$
= (X_{ijk} - \bar{X}_{jk})
$$

Thus, the estimates produced when CWC-L2 is used will only address associations with the Level-1 cluster variation in X. In fact, as will be seen shortly, employing CWC-L2 on
the Level-1 variable removes the correlation between the Level-1 variable and variables at both Level-2 and Level-3. Similarly, solving (46) for the CWC-L3 Level-2 term results in a variable that contains only Level-2 ($B_2$) between-neighborhood variation:

\[
\begin{align*}
(X_{jk} - \bar{X}_{..}) = (X_{ijk} - \bar{X}_{jk}) - (X_{ijk} - \bar{X}_{..}) - (\bar{X}_{..} - \bar{X}_{...}) \\
= (X_{jk} - \bar{X}_{..})
\end{align*}
\]  

(51)

Empirical examples of the implications of CGM and CWC-L2 for the Level-1 variable, and CGM and CWC-L3 for the Level-2 variable are presented in the next section.

The combination of centering options in a three-level framework can become quite complex. For this dissertation, when CWC is employed at Level-1, CWC is also employed at Level-2. When CGM is employed at Level-1, CGM is also employed at Level-2. Alternate centering combinations will be addressed in the discussion.

Means and Correlations

Centering within context (CWC) affects the mean and correlation structure among variables. A critical difference in the choice between CGM and CWC in two-level models is that CGM results in scores that are correlated with variables at both Level-1 and Level-2, while CWC produces scores that are not correlated with variables at Level-2 (Enders & Tofighi, 2007). In this section I demonstrate how CGM and CWC differ in their means and correlations with variables across the levels in three-level models using the Model 3 dataset described in Chapter 3 which includes $X_{ijk}$, $\bar{X}_{jk}$, and $\bar{X}_{..}$ as significant predictors.
Table 4 reports the means for raw scores, Level-1 variables centered on the Level-2 mean (CWC-L2), Level-1 variables centered on the grand mean (CGM), Level-2 variables centered on the Level-3 mean (CWC-L3), and Level-2 variables centered on the grand mean (CGM). The table values demonstrate empirically how CWC-L2 on Level-1 variables results in a mean of zero for all neighborhoods and for all cities. This differs from the raw score means and the CGM means. Notice that when CWC-L3 is employed on Level-2 variables, the city means are all zero. This is a mathematical consequence of centering on the mean – the resulting variable has a cluster mean of zero.

Table 5 contains correlations between the centered and raw score variables. Letters were assigned to each variable to simplify the discussion. Variables A, B, C, and D are all Level-1 variables. Variables E, F and G are Level-2 variables. Variable H is a Level-3 variable. Notice that the raw score for X, \((\bar{X}_{ijk})\), and the CGM score for X, \((D; (X_{ijk} - \bar{X}_{..k}))\), are each correlated with all other variables and the value of the correlation is identical because removing a constant (the grand mean) does not affect the relationship between the two variables. Thus, RAS and CGM produces scores that are correlated with variables at other levels within the multilevel model, consistent with Enders and Tofighi (2007). When CWC-L2 is employed on the Level-1 variable (C; \((X_{ijk} - \bar{X}_{..jk}))\) notice there is no correlation with the Level-2 mean (E; \(\bar{X}_{jk}\)), nor the Level-2 variable that is CWC-L3 (F; \((\bar{X}_{jk} - \bar{X}_{..k})\)), nor the Level-2 variable that is CGM (G; \((\bar{X}_{jk} - \bar{X}_{..})\)), nor the Level-3 mean (H; \(\bar{X}_{..k}\)).

It is important to recognize there are no correlations between the CWC-L2 Level-1 variable (C; \((X_{ijk} - \bar{X}_{..jk}))\) and any of the Level-2 or Level-3 variables, specifically the Level-2 mean (E; \(\bar{X}_{jk}\)), the CWC-L3 Level-2 variable (F; \((\bar{X}_{jk} - \bar{X}_{..k})\)), the CGM Level-
2 variable (G; \( \bar{X}_{jk} - \bar{X}_{..} \)), nor the Level-3 mean (H; \( \bar{X}_{..k} \)). Similarly, there is no correlation between the CWC-L3 Level-2 variable (F; \( \bar{X}_{jk} - \bar{X}_{..k} \)), and the Level-3 mean (H; \( \bar{X}_{..k} \)). This suggests that in the presence of significant upper-level effects, a model in which the Level-1 variable is CWC-L2, and the Level-2 variable is CWC-L3 produces estimates that are not subject to confounding across the multilevel structure. As we will see, this is not true if CGM is used instead of CWC-L2 at Level-1.

**Equivalencies Between Centered and Raw Score Models**

**Model 1 equivalencies.** Model 1 is similar to the models subscripted with a \( 1 \) by Kreft, de Leeuw, and Aiken (1995), but extended to three levels. The raw score equations for each level are:

\[
Y_{ijk} = \pi_{0jk} + \pi_{1jk}X_{ijk} + e_{ijk} \tag{52}
\]

\[
\pi_{0jk} = \beta_{00k} + r_{0jk} \tag{53}
\]

\[
\pi_{1jk} = \beta_{10k} + r_{1jk} \tag{54}
\]

\[
\beta_{00k} = \gamma_{000} + u_{00k} \tag{55}
\]

\[
\beta_{10k} = \gamma_{100} \tag{56}
\]

Substituting across the levels results in the full model:

\[
Y_{ijk} = \gamma_{000} + \gamma_{100}X_{ijk} + e_{ijk} + r_{0jk} + r_{1jk}X_{ijk} + u_{00k} \tag{57}
\]

The expectations of the model under RAS, CGM and CWC are below. Consistent with the notation by Kreft, de Leeuw, and Aiken (1995), the CGM terms are indicated with a
single *, CWC terms are indicated with a double **, and RAS terms are indicated with no asterisk.

RAS: \[ E(Y_{ijk}) = \gamma_{000} + \gamma_{100}(X_{ijk}) \] (58)

CGM: \[ E(Y_{ijk}) = \gamma_{000}^* + \gamma_{100}^*(X_{ijk} - \bar{X}_\ldots) \] (59)

CWC: \[ E(Y_{ijk}) = \gamma_{000}^{**} + \gamma_{100}^{**}(X_{ijk} - \bar{X}_{jk}) \] (60)

The dispersions are:

RAS: \[ \text{Var}(Y_{ijk}) = \sigma^2 + \tau_{\pi_{00}} + 2\tau_{\pi_{01}}X_{ijk} + \tau_{\pi_{11}}X_{ijk}^2 + \tau_{\beta_{00}} \] (61)

CGM: \[ \text{Var}(Y_{ijk}) = \sigma_{0}^2 + \tau_{\pi_{00}}^* + 2\tau_{\pi_{01}}^*(X_{ijk} - \bar{X}_\ldots) + \tau_{\pi_{11}}^*(X_{ijk}^2 - 2X_{ijk}\bar{X}_\ldots + \bar{X}_{\ldots}^2) + \tau_{\beta_{00}}^* \] (62)

CWC: \[ \text{Var}(Y_{ijk}) = \sigma_{0}^{**} + \tau_{\pi_{00}}^{**} + 2\tau_{\pi_{01}}^{**}(X_{ijk} - \bar{X}_{jk}) + \tau_{\pi_{11}}^{**}(X_{ijk}^2 - 2X_{ijk}\bar{X}_{jk} + \bar{X}_{jk}^2) + \tau_{\beta_{00}}^{**} \] (63)

Where \( \sigma^2 \) is the variance of \( e_{ijk} \); \( \tau_{\pi_{00}} \) is the variance of \( r_{0jk} \); \( \tau_{\pi_{11}} \) is the variance of \( r_{1jk} \); \( \tau_{\pi_{01}} \) is the covariance of \( r_{0jk} \) and \( r_{1jk} \); and \( \tau_{\beta_{00}} \) is the variance of \( u_{00k} \). Table 6 provides a sample variance-covariance matrix based on the specifications of the models in this dissertation. Because all three models have a random intercept at Level-1 and Level-2 and a random Level-1 slope, the variance-covariance matrix is the same for all three models in this dissertation.
The findings in Model 1 are similar to those reported by Kreft, de Leeuw, and Aiken (1995): equivalencies can be demonstrated between the RAS and CGM expectations, but as a result of the unique \( \bar{X}_{jk} \) term in the CWC model, no equivalencies exist between RAS and CWC expectations, nor CGM and CWC expectations. RAS and CGM have equivalent dispersions, but neither RAS nor CGM is equivalent in dispersions to CWC. The coefficients for RAS in terms of the CGM coefficients are:

\[
\begin{align*}
\gamma_{000} &= \gamma_{000}^* - \gamma_{100}^* (\bar{X}_{..}) \\
\gamma_{100} &= \gamma_{100}^* \\
\sigma^2 &= \sigma_{2*}^2 \\
\tau_{\pi_{00}} &= \tau_{\pi_{00}}^* - 2\tau_{\pi_{01}}^* (\bar{X}_{..}) + \tau_{\pi_{11}}^* (\bar{X}_{..}^2) \\
\tau_{\pi_{01}} &= \tau_{\pi_{01}}^* - \tau_{\pi_{11}}^* (\bar{X}_{..}) \\
\tau_{\pi_{11}} &= \tau_{\pi_{11}}^* \\
\tau_{\beta_{00}} &= \tau_{\beta_{00}}^* 
\end{align*}
\]

The algebraic work demonstrating these equivalencies is available in Appendix B.

**Model 2 equivalencies.** An extension to the work of Kreft, de Leeuw and Aiken (1995) in the three-level model is the possibility of including predictors at Level-1 and Level-2, but not Level-3. Model 2 builds on Model 1 by adding the neighborhood mean as a predictor of the intercept at Level-2. The raw score equations are as follows:

\[
\begin{align*}
Y_{ijk} &= \pi_{ojk} + \pi_{1jk} (X_{ijk}) + e_{ijk} \\
\pi_{ojk} &= \beta_{00k} + \beta_{01k} (\bar{X}_{jk}) + r_{0jk}
\end{align*}
\]
\[ \pi_{1jk} = \beta_{10k} + r_{1jk} \]  
(73)

\[ \beta_{00k} = \gamma_{000} + u_{00k} \]  
(74)

\[ \beta_{01k} = \gamma_{010} \]  
(75)

\[ \beta_{10k} = \gamma_{100} \]  
(76)

Substituting across the levels results in the full model:

\[ Y_{ijk} = \gamma_{000} + \gamma_{010}(\bar{X}_{jk}) + \gamma_{100}(X_{ijk}) + u_{00k} + r_{0jk} + r_{1jk}(X_{ijk}) + e_{ijk} \]  
(77)

The expectations of the model are as follows:

RAS:  
\[ E(Y_{ijk}) = \gamma_{000} + \gamma_{010}(\bar{X}_{jk}) + \gamma_{100}(X_{ijk}) \]  
(78)

CGM:  
\[ E(Y_{ijk}) = \gamma_{000} + \gamma_{010}^*(\bar{X}_{jk} - \bar{X}_{..}) + \gamma_{100}^*(X_{ijk} - \bar{X}_{..}) \]  
(79)

CWC:  
\[ E(Y_{ijk}) = \gamma_{000}^{**} + \gamma_{010}^{**}(\bar{X}_{jk} - \bar{X}_{..}) + \gamma_{100}^{**}(X_{ijk} - \bar{X}_{..}) \]  
(80)

The dispersions are:

RAS:  
\[ \text{Var}(Y_{ijk}) = \sigma^2 + \tau_{\pi_{00}} + 2\tau_{\pi_{01}}X_{ijk} + \tau_{\pi_{11}}X_{ijk}^2 + \tau_{\beta_{00}} \]  
(81)

CGM:  
\[ \text{Var}(Y_{ijk}) = \sigma^{2*} + \tau_{\pi_{00}}^{*} + 2\tau_{\pi_{01}}^{*}(X_{ijk} - \bar{X}_{..}) + \tau_{\pi_{11}}^{*}(X_{ijk}^2 - 2X_{ijk}\bar{X}_{..} + \bar{X}_{..}^2) + \tau_{\beta_{00}}^{*} \]  
(82)

CWC:  
\[ \text{Var}(Y_{ijk}) = \sigma^{2**} + \tau_{\pi_{00}}^{**} + 2\tau_{\pi_{01}}^{**}(X_{ijk} - \bar{X}_{..}) + \tau_{\pi_{11}}^{**}(X_{ijk}^2 - 2X_{ijk}\bar{X}_{..} + \bar{X}_{..}^2) + \tau_{\beta_{00}}^{**} \]  
(83)

In the three-level model, RAS and CGM are equivalent in expectations and dispersions, but neither are equivalent in expectations nor dispersions to CWC. The reason for this is
the CWC-L3 Level-2 variable \((\bar{X}_{jk} - \bar{X}_{..})\) in the CWC model. The coefficients for RAS in terms of the CGM coefficients for Model 2 are:

\[
\gamma_{000} = \gamma_{000}^* - \gamma_{010}^* (\bar{X}_{..}) - \gamma_{100}^* (\bar{X}_{..}) \tag{84}
\]

\[
\gamma_{010} = \gamma_{010}^* \tag{85}
\]

\[
\gamma_{100} = \gamma_{100}^* \tag{86}
\]

\[
\sigma^2 = \sigma^{2*} \tag{87}
\]

\[
\tau_{\pi_{00}} = \tau_{\pi_{00}}^* - 2\tau_{\pi_{01}}^* (\bar{X}_{..}) + \tau_{\pi_{11}}^* (\bar{X}_{..}^2) \tag{88}
\]

\[
\tau_{\pi_{01}} = \tau_{\pi_{01}}^* - \tau_{\pi_{11}}^* (\bar{X}_{..}) \tag{89}
\]

\[
\tau_{\pi_{11}} = \tau_{\pi_{11}}^* \tag{90}
\]

\[
\tau_{\beta_{00}} = \tau_{\beta_{00}}^* \tag{91}
\]

If there is no group-mean centering at Level-2, this model is quite similar to the model with a subscript of 2 proposed by Kreft, de Leeuw, and Aiken (1995), and RAS = CGM = CWC in the expectations. The algebraic work demonstrating these equivalencies is in Appendix C.

**Model 3 equivalencies.** Model 3 is similar to the model denoted by a subscripted 2 by Kreft, de Leeuw, and Aiken (1995), but again extended to three levels. The raw score equations for each level of the model are:

\[
Y_{ijk} = \pi_{0jk} + \pi_{1jk}(X_{ijk}) + e_{ijk} \tag{92}
\]

\[
\pi_{0jk} = \beta_{00k} + \beta_{01k}(\bar{X}_{jk}) + r_{0jk} \tag{93}
\]

\[
\pi_{1jk} = \beta_{10k} + r_{1jk} \tag{94}
\]
\[ \beta_{00k} = \gamma_{000} + \gamma_{001}(\bar{X}_{\cdot k}) + u_{00k} \]  
(95)

\[ \beta_{01k} = \gamma_{010} \]  
(96)

\[ \beta_{10k} = \gamma_{100} \]  
(97)

Substituting across the levels results in the full model:

\[ Y_{ijk} = \gamma_{000} + \gamma_{001}(\bar{X}_{\cdot k}) + \gamma_{010}(X_{ijk}) + \gamma_{100}(X_{ijk}) + u_{00k} + r_{0jk} + r_{1jk}(X_{ijk}) + e_{ijk} \]  
(98)

The expectations of the model are as follows:

**RAS:**
\[ E(Y_{ijk}) = \gamma_{000} + \gamma_{001}(\bar{X}_{\cdot k}) + \gamma_{010}(\bar{X}_{\cdot k}) + \gamma_{100}(X_{ijk}) \]  
(99)

**CGM:**
\[ E(Y_{ijk}) = \gamma_{000}^* + \gamma_{001}^*(\bar{X}_{\cdot k} - \bar{X}_{\cdot \cdot}) + \gamma_{010}^*(\bar{X}_{\cdot k} - \bar{X}_{\cdot \cdot}) + \gamma_{100}^*(X_{ijk} - \bar{X}_{\cdot \cdot}) \]  
(100)

**CWC:**
\[ E(Y_{ijk}) = \gamma_{000}^{**} + \gamma_{001}^{**}(\bar{X}_{\cdot k} - \bar{X}_{\cdot \cdot}) + \gamma_{010}^{**}(\bar{X}_{\cdot k} - \bar{X}_{\cdot \cdot}) + \gamma_{100}^{**}(X_{ijk} - \bar{X}_{\cdot \cdot}) \]  
(101)

The dispersions are:

**RAS:**
\[ \text{Var}(Y_{ijk}) = \sigma^2 + \tau_{\pi_{00}} + 2\tau_{\pi_{01}}X_{ijk} + \tau_{\pi_{11}}X_{ijk}^2 + \tau_{\beta_{00}} \]  
(102)

**CGM:**
\[ \text{Var}(Y_{ijk}) = \sigma^{2*} + \tau^{*}_{\pi_{00}} + 2\tau^{*}_{\pi_{01}}(X_{ijk} - \bar{X}_{\cdot \cdot}) + \tau^{*}_{\pi_{11}}(X_{ijk}^2 - 2X_{ijk}\bar{X}_{\cdot \cdot} + \bar{X}_{\cdot \cdot}^2) + \tau^{*}_{\beta_{00}} \]  
(103)

**CWC:**
\[ \text{Var}(Y_{ijk}) = \sigma^{2**} + \tau^{**}_{\pi_{00}} + 2\tau^{**}_{\pi_{01}}(X_{ijk} - \bar{X}_{\cdot \cdot}) + \tau^{**}_{\pi_{11}}(X_{ijk}^2 - 2X_{ijk}\bar{X}_{\cdot \cdot} + \bar{X}_{\cdot \cdot}^2) + \bar{X}_{\cdot \cdot}^2 + \tau^{**}_{\beta_{00}} \]  
(104)
Where $\sigma^2$ is the variance of $e_{ijk}$; $\tau_{\pi_{00}}$ is the variance of $r_{0jk}$; $\tau_{\pi_{11}}$ is the variance of $r_{1jk}$; $\tau_{\pi_{01}}$ is the covariance of $r_{0jk}$ and $r_{1jk}$; and $\tau_{\beta_{00}}$ is the variance of $u_{00k}$.

The findings here are similar to those reported by Kreft, de Leeuw, and Aiken (1995). In the two-level model $RAS = CGM = CWC$ in expectations, but only $RAS = CGM$ in dispersions. The same is true in the three-level model. The coefficients for $RAS$ in terms of the CGM coefficients are:

\begin{align*}
\gamma_{000} &= \gamma_{000}^* - \gamma_{010}^* (\bar{X}...) - \gamma_{100}^* (\bar{X}...) - \gamma_{001}^* (\bar{X}...) \\
\gamma_{010} &= \gamma_{010}^* \\
\gamma_{100} &= \gamma_{100}^* \\
\gamma_{001} &= \gamma_{001}^* \\
\sigma^2 &= \sigma^{2*} \\
\tau_{\pi_{00}} &= \tau_{\pi_{00}}^* - 2\tau_{\pi_{01}}^* (\bar{X}...) + \tau_{\pi_{11}}^* (\bar{X}^2...) \\
\tau_{\pi_{01}} &= \tau_{\pi_{01}}^* - \tau_{\pi_{11}}^* (\bar{X}...) \\
\tau_{\pi_{11}} &= \tau_{\pi_{11}}^* \\
\tau_{\beta_{00}} &= \tau_{\beta_{00}}^* \\
\end{align*}

The coefficients for $RAS$ in terms of the CWC coefficients are:

\begin{align*}
\gamma_{000} &= \gamma_{000}^{**} \\
\gamma_{001} &= \gamma_{001}^{**} - \gamma_{010}^{**} \\
\gamma_{010} &= \gamma_{010}^{**} - \gamma_{100}^{**} \\
\gamma_{100} &= \gamma_{100}^{**} \\
\end{align*}
The algebraic work demonstrating these equivalencies is available in Appendix D. A summary of these equivalencies can be found in Table 7. In Model 1 and Model 2 RAS is equal to CGM in expectations and dispersions, however neither RAS nor CWC are equal to CWC. In Model 3, all three centering options produce equivalent expectations, however CWC is again different from CGM and RAS in dispersions. The findings for Model 1 and Model 3 are consistent with the models and findings of Kreft, de Leeuw and Aiken (1995) but extended to three levels. The unique centering options available in Model 2 resulting from the three-level model add to this literature.

**Algebraic Relationships Between CWC and RAS in Model 2 and Model 3**

Contextual models are models in which lower-level variables are aggregated and included as predictors at upper-levels. As discussed in Chapter 2, Raudenbush (1989) demonstrated the algebraic relationships between RAS and CWC in contextual models. He showed that by centering the Level-1 variable on the Level-2 group mean in a two-level contextual model, the interpretation of the Level-2 slope coefficient changes from a contextual effect (the difference between the within-group effect and the between-group effect) in the uncentered model to a direct effect (the between-group effect) in the CWC model. Using his work as a guide, this can be extended to a contextual model with three-levels.

First, consider a two-level model in which $\bar{X}_j$ is included as a predictor at Level-2:
\[ Y_{ij} = \beta_{0j} + \beta_p(X_{ij}) + \gamma(\bar{X}_j) + u_{0j} + e_{ij} \] (118)

In this model, \( Y_{ij} \) is the outcome for person i in neighborhood j, \( \bar{X}_j \) is the neighborhood mean, \( X_{ij} \) is the score for person i in neighborhood j, \( u_{0j} \) is the random effect associated with neighborhood j and \( e_{ij} \) is the random effect at the person-level. Raudenbush showed that in this model, \( \beta_p \) is the person-level effect of X on Y (the subscript \( p \) indicates person-level effect) and \( \gamma \) is the contextual effect, that is, the difference between the neighborhood-level effect and the person-level effect of X on Y (\( \beta_n - \beta_p \)) (the subscript \( n \) indicates neighborhood-level effect). He further showed that the person-level effect and the neighborhood-level effect could be directly estimated by employing CWC:

\[ Y_{ij} = \mu_j + \beta_p(X_{ij} - \bar{X}_j) + \beta_n(\bar{X}_j) + u_{0j} + e_{ij}. \] (119)

In this centered model, \( \beta_p \) is again the person-level effect of X on Y and the coefficient on the neighborhood mean is \( \beta_n \), the neighborhood-level effect of X on Y, rather than the difference between the neighborhood-level effect and the person-level effect (previously noted as \( \gamma \)). So equation (120) shows that using CWC and including the Level-2 mean as a predictor directly estimates the person-level effect and the neighborhood-level effect. As discussed in Chapter 1, this formulation of the two-level model does not have the collinearity problems found in the raw score model because as we saw earlier in this chapter, when CWC-L2 is employed on the Level-1 variable, there is no longer a correlation with any upper-level variables.
Raudenbush points out that the centered equation (120) is a reparameterization of the raw score equation (119). This can be shown algebraically:

1. Begin with the RAS model and substitute \((\beta_n - \beta_p)\) for \(\gamma\):
\[
Y_{ij} = \beta_0 + \beta_p X_{ij} + (\beta_n - \beta_p)(\bar{X}_j) + u_0 + e_{ij}
\]  
(120)

2. Distribute terms:
\[
Y_{ij} = \beta_0 + \beta_p X_{ij} + (\beta_n)(\bar{X}_j) - (\beta_p)(\bar{X}_j) + u_0 + e_{ij}
\]  
(121)

3. Rearrange terms to uncover the CWC model:
\[
Y_{ij} = \beta_0 + \beta_p (X_{ij} - \bar{X}_j) + (\beta_n)(\bar{X}_j) + u_0 + e_{ij}
\]  
(122)

The decomposition of the three-level model can be approached in the same way for Model 3 because the CWC and RAS models are equivalent in their expectations (as shown in the previous section). Recognizing, as in the two-level model, that the raw score model contains contextual effects, the effects can be decomposed algebraically beginning with the CWC model. The derivation is as follows:

1. Begin with the CWC model (only fixed effects are shown here):
\[
Y_{ijk} = \beta_{0jk} + \beta_c (\bar{X}_{.k}) + \beta_n (\bar{X}_{jk} - \bar{X}_{.k}) + \beta_p (X_{ijk} - \bar{X}_{jk})
\]  
(123)

2. Distribute:
\[
Y_{ijk} = \beta_{0jk} + \beta_c (\bar{X}_{.k}) + \beta_n (\bar{X}_{jk} - \bar{X}_{.k}) + \beta_p (X_{ijk} - \bar{X}_{jk})
\]  
(124)

3. Regroup to uncover the raw score contextual model:
\[
Y_{ijk} = \beta_{0jk} + (\beta_c - \beta_n)(\bar{X}_{.k}) + (\beta_n - \beta_p)(\bar{X}_{jk}) + \beta_p (X_{ijk})
\]  
(125)
In these equations, $\beta_c$ is the effect of the city mean, $\beta_n$ is the effect of the neighborhood mean and $\beta_p$ is the effect of the person score, $\bar{X}_{..k}$ is the city mean, $\bar{X}_{.jk}$ is the neighborhood mean and $X_{ijk}$ is the person score. As in the two-level model, CWC estimates the person, neighborhood and city effects directly while use of the raw score estimates the person effect and two contextual effects: (1) the difference between the city and neighborhood effects, $(\beta_c - \beta_n)$, and (2) the difference between the neighborhood and person effects, $(\beta_n - \beta_p)$.

As shown in the previous section, the CWC and RAS models for Model 2 are not equivalent in their expectations. Thus, the fixed effects between these two models are not simple reparameterizations of one another. Note, however, that if CWC is employed only at Level-1, and the Level-2 mean is left in RAS form, the same issues do not exist because the CWC and RAS models are equivalent in their expectations in this model. Thus, the reparameterization works:

1. Begin with the CWC model with CWC at Level-1 (only fixed effects are shown here; note the Level-2 mean is RAS, not CWC):

$$Y_{ijk} = \beta_{0jk} + \beta_n(\bar{X}_{.jk}) + \beta_p(X_{ijk} - \bar{X}_{.jk})$$  \hspace{1cm} (126)

2. Distribute:

$$Y_{ijk} = \beta_{0jk} + \beta_n(\bar{X}_{.jk}) + \beta_p(X_{ijk}) - \beta_p(\bar{X}_{.jk})$$  \hspace{1cm} (127)

3. Regroup to uncover the raw score contextual model:

$$Y_{ijk} = \beta_{0jk} + (\beta_n - \beta_p)(\bar{X}_{.jk}) + \beta_p(X_{ijk})$$  \hspace{1cm} (128)
This looks very similar to the two-level case. It is important to note, however, that the RAS Level-2 mean ($\bar{X}_{jk}$) still contains between-city variation. As a result, if the effect of the city mean and the effect of the neighborhood mean are not equal (that is, if a contextual effect is present at Level-3), then the coefficient associated with the uncentered (RAS) Level-2 mean in this model will be a blend of the Level-2 effect and the Level-3 effect, complicating interpretation of this coefficient.

**Interpretation of Model 1 Estimates**

Model 1 has just one predictor variable, residing at Level-1. The RAS equations for Model 1 are again as follows:

\[
Y_{ijk} = \pi_{ojk} + \pi_{1jk}(X_{ijk}) + e_{ijk} \tag{129}
\]

\[
\pi_{ojk} = \beta_{00k} + r_{0jk} \tag{130}
\]

\[
\pi_{1jk} = \beta_{10k} + r_{1jk} \tag{131}
\]

\[
\beta_{00k} = \gamma_{000} + u_{00k} \tag{132}
\]

\[
\beta_{10k} = \gamma_{100} \tag{133}
\]

Substituting across the levels results in the full model:

\[
Y_{ijk} = \gamma_{00} + \gamma_{100}(X_{ijk}) + e_{ijk} + r_{0jk} + r_{1jk}(X_{ijk}) + u_{00k} \tag{134}
\]

The results in this section will be discussed in terms of the hypothetical example in Chapter 3. That is, the predictor variable is walking habits measured in hours per week and the outcome variable is health. Although the regression coefficients of greatest
interest are the gammas ($\gamma_{100}$ in Model 1; $\gamma_{100}$ and $\gamma_{010}$ in Model 2; $\gamma_{100}$ and $\gamma_{010}$ and $\gamma_{001}$ in Model 3), the intermediate parameters ($\pi$ and $\beta$) need to be discussed to explain how the interpretations of the gammas were derived. It is important to note that the intermediate parameters ($\pi$ and $\beta$) are not directly estimated in any of the models.

**Model 1 intercept ($\gamma_{000}$).** In the RAS model, the interpretation of the intercept ($\gamma_{000}$) is the expected value of $Y_{ijk}$ when $X_{ijk} = 0$, that is the expected health for a person who walks zero hours per week. The interpretation of $\gamma_{000}$ under CGM requires consideration of the latent intercepts throughout the model. First, the full model:

$$Y_{ijk} = \gamma_{000} + \gamma_{100}(X_{ijk} - \bar{X}) + e_{ijk} + r_{0jk} + r_{1jk}(X_{ijk} - \bar{X}) + u_{00k} \quad (135)$$

As discussed in Chapter 2, the intercept in a two-level random intercept CGM model is an adjusted mean, with the adjustment being the slope coefficient for the Level-2 (neighborhood) cluster multiplied by the deviation of the Level-2 (neighborhood) cluster mean from the overall mean, ($\bar{X}_j - \bar{X}$). In the three-level random intercept CGM model, the intercept is also an adjusted mean. The meaning of the adjustment can be derived beginning with the Level-1 model:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}(X_{ijk} - \bar{X}) + e_{ijk} \quad (136)$$

Taking the expectation of $Y_{ijk}$ within each neighborhood, the expectation of $e_{ijk}$ is zero and what remains is:
\[ \mu_{Y_{jk}} = \pi_{0_{jk}} + \pi_{1_{jk}}(\bar{X}_{jk} - \bar{X}_{..}) \]  

(137)

Solving for the intercept, the Level-1 intercept is equal to the neighborhood mean of Y minus the within-neighborhood regression coefficient (\(\pi_{1_{jk}}\)) multiplied by the deviation of the neighborhood mean from the grand mean (\(\bar{X}_{.,jk} - \bar{X}_{..}\)):

\[ \pi_{0_{jk}} = \mu_{Y_{jk}} - \pi_{1_{jk}}(\bar{X}_{jk} - \bar{X}_{..}) \]  

(138)

So \(\pi_{0_{jk}}\) is interpreted as the adjusted mean of Y (adjusted for the effect of the difference between the neighborhood walking hours and the grand mean walking hours) for neighborhood j in city k, or the adjusted mean health for residents in neighborhood j of city k. The Level-2 model for the intercept is:

\[ \pi_{0_{jk}} = \beta_{00k} + r_{0jk} \]  

(139)

Taking the expectation of \(\pi_{0_{jk}}\) within each city, we get:

\[ E(\pi_{0_{jk}}) = \beta_{00k} \]  

(140)

Because \(\pi_{0_{jk}}\) is interpreted as the adjusted mean health for neighborhood j in city k, the expectation of this term, \(\beta_{00k}\), is the city average of these adjusted means. The expectation of \(\beta_{00k}\) is \(\gamma_{000}\), which is the average of the city means, which are averages of the adjusted neighborhood means. Therefore, the full model under CGM, the intercept
\( \gamma_{000} \) is an adjusted mean of \( Y \), or the adjusted mean health for all residents. The adjustment depends on the relationship between walking and health, and the difference between the average walking hours in a neighborhood compared with the average number of walking hours in the entire sample.

For CWC, the interpretation is different. When CWC-L2 is employed at Level-1, the full model is:

\[
Y_{ijk} = \gamma_{000} + \gamma_{100}(X_{ijk} - \bar{X}_{jk}) + u_{00k} + r_{1jk}(X_{ijk} - \bar{X}_{jk}) + r_{0jk} + e_{ijk} \tag{141}
\]

If we take the expectation within neighborhoods of the Level-1 equation, both the error term and the slope estimate go to zero:

\[
\mu_{Y_{jk}} = \pi_{0jk} + \pi_{1jk}(\bar{X}_{jk} - \bar{X}_{jk})
\]

\[
= \pi_{0jk}. \tag{142}
\]

Here \( \pi_{0jk} \) is interpreted as the unadjusted mean of \( Y \) for neighborhood \( j \) in city \( k \), or the mean health for a resident living in neighborhood \( j \) and city \( k \). The remainder of the derivation follows what is above, resulting in the interpretation of \( \gamma_{000} \) being the unadjusted mean of \( Y \), or the mean health of all residents (the grand mean of \( Y \)).

**Model 1 Level-1 slope (\( \gamma_{100} \)).** Recall from Chapter 2 the importance of centering when interpreting the Level-1 slope coefficient. Under CWC, the mean of \( (X_{ijk} - \bar{X}_{jk}) \) is zero for every neighborhood thus there is no between-cluster variation in \( (X_{ijk} - \bar{X}_{jk}) \) and the interpretation of \( \gamma_{100} \) is the expected change in health when a person increases
his or her walking by one hour per week. Under CGM, however, there is still between-group variation in the predictor. As discussed in Chapter 2, if the Level-1 regression of $Y$ on $X$ differs from the Level-2 and Level-3 regressions of $Y$ on $X$, the resulting Level-1 regression coefficient is not a simple pooled within-cluster regression coefficient, but is instead a mixture of the within and between regression coefficients. This complicates the interpretation of the Level-1 regression coefficient under CGM. The example dataset for this first research question was created without any significant upper-level effects, thus the between-level effect is not statistically different from the within-level effect. In this situation, CGM and CWC produce consistent estimates of $\gamma_{100}$ and the CGM (or RAS) estimate is most efficient (Raudenbush & Bryk, 2002, p. 138). Though a small difference, note that the CWC estimate of the intercept ($\gamma_{000}$) is equal to the grand mean of $Y$ (134.642) while the estimate produced by CGM is an adjusted mean (134.675). Also notice that the estimate for the slope ($\gamma_{100}$) is identical across the three centering options (5.000).

Table 8 contains the average estimates from the ten artificial dataset using RAS, CWC and CGM in Model 1.

**Model 1 estimated variances and covariances.** The full hierarchical linear model utilizing the raw score (shown above, copied here) contains fixed effects and random effects.

$$Y_{ijk} = \gamma_{000} + \gamma_{100}(X_{ijk}) + e_{ijk} + r_{0jk} + r_{1jk}(X_{ijk}) + u_{00k}$$ (143)
Centering affects the variances of the random intercepts and random slopes, as well as the covariances between them.

The error term $e_{ijk}$ represents that which is unique to person $i$ in neighborhood $j$ and city $k$. In Model 1, $e_{ijk}$ is the error in $Y_{ijk}$ after adjusting for $X_{ijk}$. The assumption in hierarchical linear modeling is that $e_{ijk}$ is normally distributed with homogenous variance across neighborhoods and cities. The variance-covariance matrix is $\Sigma$:

$$\Sigma = \sigma^2$$ (144)

The error terms $r_{0jk}$ and $r_{1jk}$ represent uniqueness resulting from residing in neighborhood $j$ in city $k$ and common to all individuals in those neighborhoods. These terms arise from the random intercept ($\pi_{0jk}$) and random slope ($\pi_{1jk}$) in Level-1. Said another way, these terms come from allowing each neighborhood to have its own mean walking habits and its own relationship between walking and health. The variance-covariance matrix for these terms is $T_\pi$:

$$T_\pi = \begin{bmatrix} \tau_{\pi_{00}} & \tau_{\pi_{01}} \\ \tau_{\pi_{10}} & \tau_{\pi_{11}} \end{bmatrix}$$ (145)

The error term $u_{00k}$ represents uniqueness resulting from residing in city $k$ and common to all individuals in that city. This term arises from the random intercept ($\beta_{00k}$) in Level-2. Again, this is a result of allowing each city to have a unique mean health value. The variance-covariance matrix is $T_\beta$: 
Model 1 variance of the Level-1 intercepts ($\tau_{\pi_{00}}$). Recall from above how centering influences the interpretation of the intercept. Under RAS, the interpretation of $\pi_{0jk}$ is the expected health of an individual who walks zero hours per week. Under CGM, the intercept ($\pi_{0jk}$) is the expected health for a resident in neighborhood $jk$ minus an additional term made up of a regression slope ($\pi_{1jk}$) and the deviation of the neighborhood mean walking hours from the overall mean walking hours ($\bar{X}_{jk} - \bar{X}_\cdot$); an adjusted mean. Under CWC, $\pi_{0jk}$ is interpreted as the unadjusted mean for group $jk$ because it does not have this dependence on $\pi_{1jk}$.

Thus, in Model 1, under RAS, $\tau_{\pi_{00}}$ is the variation in the mean neighborhood health in neighborhoods when an individual walks zero hours per week. Under CWC, $\tau_{\pi_{00}}$ refers to the between-neighborhood variation in the intercepts. Because the Level-1 intercept ($\pi_{0jk}$) in the CWC model is equal to the unadjusted neighborhood mean, this is similar to saying that $\tau_{\pi_{00}}$ represents the between-neighborhood variation in $Y$, or variance of the unadjusted neighborhood means of $Y$. When CGM is used, the intercept is an adjusted group mean. Thus, $\tau_{\pi_{00}}$ is the variance of the adjusted neighborhood means.

Under different centering options, the estimates of $\tau_{\pi_{00}}$ will be different.

Also recall from Chapter 2 that use of centering changes the location of the intercept and can have a significant effect on the variance estimate of the intercept when the slope is also random (as it is in all models of this dissertation). Under RAS, the intercept ($\pi_{0jk}$) represents the mean of $Y_{ijk}$ when $X_{ijk}$ is equal to zero. In the CWC
model, the intercept \((\pi_{0jk})\) represents the mean of \(Y_{ijk}\) when \(X_{ijk}\) is equal to the neighborhood mean of \(X\) (\(\bar{X}_{jk}\)). Finally, in the CGM model, the intercept \((\pi_{0jk})\) represents the mean of \(Y_{ijk}\) when \(X_{ijk}\) is equal to the grand mean of \(X\) (\(\bar{X}_{...}\)). Thus, when the slopes are random, and when the neighborhood mean differs from the grand mean (and both differ from zero), the estimates for the variance of the intercepts, \(\tau_{\pi_{00}}\), will differ depending on whether RAS, CWC or CGM is used because the location of the intercept is different.

**Model 1 variance of the Level-1 slopes** \((\tau_{\pi_{11}})\). When CWC is employed, \(\tau_{\pi_{11}}\) refers to the variance of the Level-1 slopes \((\pi_{1jk})\). Because the between-neighborhood variation in the slopes is removed under CWC-L2 at Level-1, the CWC slopes represent the pooled regression of \(Y_{ijk}\) on \(X_{ijk}\) within each neighborhood and \(\tau_{\pi_{11}}\) represents the variation of these neighborhood slopes. The variance of the Level-1 slopes in the CGM model is significantly less straightforward because of the dependency between slopes and intercepts (recall how the intercept is adjusted by the slope) and because CGM does not eliminate between-group variation in the slope coefficient. As discussed in Chapter 2, when the intraclass correlation is greater than zero, and in the presence of a significant upper-level effect, the interpretation of \(\pi_{1jk}\) under CGM is complicated because it is a blend of the within-cluster and between-cluster regression of \(Y\) on \(X\).

**Model 1 covariance of the Level-2 residuals** \((\tau_{\pi_{01}})\). Although assumed not to covary with residuals at other levels, the two residual terms at Level-2 \((r_{0jk}\) and \(r_{1jk}\)) are expected to covary with one another. This covariance is represented by \(\tau_{\pi_{01}}\). The influence of the random slope on the variance of the intercept described above also affects the estimates of the covariances between random intercepts and random slopes,
\[ \tau_{\pi_{01}} \]. Interpretation of a positive correlation between slope and intercept implies that neighborhoods with a higher mean health also have a higher within-neighborhood effect of walking on overall health.

**Model 1 variance of the Level-2 means (\( \tau_{\beta_{00}} \)).** At Level-2, there is just one parameter that is modeled to include a random effect (\( \beta_{00k} \)). \( \beta_{00k} \) represents the city mean of the Level-1 intercepts (\( \pi_{0jk} \)). When this term is modeled at Level-3, the equation includes an overall mean (\( \gamma_{000} \)) and the uniqueness associated with living in city \( k \) (\( u_{00k} \)). The variance of the means across cities is \( \tau_{\beta_{00}} \). Since there are no Level-3 predictors in Model 1, \( \tau_{\beta_{00}} \) refers conceptually to the between-city variation in the city means. In the CWC model the means are unadjusted, and in the CGM model the means are adjusted as described above.

Table 9 demonstrates the equivalencies in the variances and covariances calculated earlier in this chapter using the RAS and CGM estimates from the Model 1 artificial datasets.

**Interpretation of Model 2 Estimates**

Model 2 includes predictors at Level-1 and Level-2, but not at Level-3. The raw score equations for Model 2 are as follows:

\[
\begin{align*}
Y_{ijk} &= \pi_{ojk} + \pi_{1jk}(X_{ijk}) + e_{ijk} \\
\pi_{ojk} &= \beta_{00k} + \beta_{01k}(X_{jk}) + r_{ojk} \\
\pi_{1jk} &= \beta_{10k} + r_{1jk} \\
\beta_{00k} &= \gamma_{000} + u_{00k} \\
\beta_{01k} &= \gamma_{010}
\end{align*}
\]
\[ \beta_{10k} = \gamma_{100} \]  

Substituting across the levels results in the full model:

\[ Y_{ijk} = \gamma_{000} + \gamma_{010} X_{ijk} + \gamma_{100} (X_{ijk}) + u_{00k} + r_{0jk} + r_{1jk} X_{ijk} + e_{ijk} \]  

(153)

**Model 2 intercept** \((\gamma_{000})\). In the raw score model, the interpretation of the intercept is the expected value of \( Y_{ijk} \) when \( X_{ijk} = \bar{X}_{jk} = 0 \), that is the expected health for a person who walks zero hours per week and whose neighborhood average walking hours is zero hours per week. In this three-level CGM model with predictors at Level-1 and Level-2, the intercept is interpreted as the average of the city means, which are themselves averages of the adjusted neighborhood means. This derivation is very similar to the one for Model 1. At Level-1:

\[ \pi_{0jk} = \mu_{Y_{jk}} - \pi_{1jk} (\bar{X}_{jk} - \bar{X}_{..}) \]  

(154)

So \( \pi_{0jk} \) is again interpreted as the adjusted mean of \( Y \) for neighborhood \( j \) in city \( k \). The Level-2 equation for the intercept is:

\[ \pi_{0jk} = \beta_{00k} + \beta_{01k} (\bar{X}_{jk} - \bar{X}_{..}) + r_{0jk} \]  

(155)

Taking the expectation of \( \pi_{0jk} \) within cluster (city), results in:

\[ E(\pi_{0jk}) = \beta_{00k} + \beta_{01k} (\bar{X}_{.k} - \bar{X}_{..}) \]  

(156)
An again we can see that $\beta_{00k}$ is an adjusted mean intercept (and the intercept, remember, is an adjusted mean):

$$\beta_{00k} = \mu_{\pi_k} - \beta_{01k}(\bar{X}_{.k} - \bar{X}_{...}) \quad (157)$$

$\beta_{00k}$, would be interpreted as an adjusted mean of the average adjusted neighborhood mean in city k. The expectation of $\beta_{00k}$ is $\gamma_{000}$. Therefore, in the full model in which CGM is employed, the intercept $\gamma_{000}$, is an adjusted mean of Y with adjustments at both Level-2 and Level-3.

Finally, under CWC at Level-1 and Level-2, the full model is:

$$Y_{ijk} = \gamma_{000} + \gamma_{010}(\bar{X}_{.jk} - \bar{X}_{.k}) + \gamma_{100}(X_{ijk} - \bar{X}_{.jk}) + u_{00k} + r_{0jk} + r_{1jk}(X_{ijk} - \bar{X}_{.jk}) + e_{ijk}. \quad (158)$$

If we take the expectation within neighborhoods from the Level-1 equation, as above, we get:

$$\mu_{\gamma_{jk}} = \pi_{0jk} - \pi_{1jk}(\bar{X}_{.jk} - \bar{X}_{.jk})$$

$$= \pi_{0jk}. \quad (159)$$
So \( \pi_{0jk} \) in the CWC case is an unadjusted mean. Then \( \pi_{0jk} \) is interpreted as the unadjusted mean of \( Y \) for neighborhood \( j \) in city \( k \). The Level-2 equation for the intercept in the CWC model is:

\[
\pi_{0jk} = \beta_{00k} + \beta_{01k}(\bar{X}_{jk} - \bar{X}_{.k}) + r_{0jk}
\]

(160)

Taking the expectation of \( \pi_{0jk} \) within cluster (city), we get:

\[
E(\pi_{0jk}) = \beta_{00k} + \beta_{01k}(\bar{X}_{.k} - \bar{X}_{.k})
\]

\[
= \beta_{00k}
\]

(161)

Thus, \( \beta_{00k} \) is also an unadjusted mean, in this case the mean \( \pi_{0jk} \) for city \( k \). As before, the interpretation of \( \gamma_{000} \) under CWC is the unadjusted mean of \( Y \).

**Model 2 Level-2 slope (\( \gamma_{010} \)).** As discussed earlier in this chapter, in the RAS and CGM models, the interpretation of \( \gamma_{010} \) is the difference in the effect of \( X_{ijk} \) and the effect of \( X_{ijk} (\beta_n - \beta_p) \). Thus, under RAS and CGM, \( \gamma_{010} \) is a contextual effect. In the CWC model, the interpretation of \( \gamma_{010} \) is the expected change in \( Y_{ijk} \) when \( (\bar{X}_{jk} - \bar{X}_{.k}) \) changes by one unit (\( \beta_n \)).

**Model 2 Level-1 slope (\( \gamma_{100} \)).** Under all three centering options, the interpretation of \( \gamma_{100} \) is the effect of \( X_{ijk} \) on \( Y \) when all other predictors equal zero. This interpretation is intuitive for the CWC model because CWC removes the between-neighborhood and between-city variation in the predictor variable so that all that remains is within-neighborhood variability. However, this interpretation may not seem so
straightforward in the CGM and RAS models. When the upper-level means for contextual effects that exist (remember in this dataset, only a Level-2 contextual effect exists) are included in the model, the resulting regression coefficients are *partial* regression coefficients. As such, under CGM and RAS, $\gamma_{100}$ is the influence of an individual resident’s walking habits controlling for neighborhood mean walking habits, while $\gamma_{010}$ estimates the influence of neighborhood mean walking habits on the health outcome. This results in similar estimates of $\gamma_{100}$ under RAS, CGM and CWC.

Table 10 contains the gamma estimates from the artificial dataset using RAS, CWC and CGM in Model 2. Notice again that the CWC estimate of the intercept ($\gamma_{000}$) perfectly recovers the grand mean of $Y$ (86.959), while the CGM estimate is an adjusted mean (87.012). The estimate of the Level-1 slope is identical across the three centering options (5.006). The estimate of the Level-2 slope differs significantly between the centering options. CGM and RAS estimates are negative (-2.016) while the CWC estimate is positive (2.969). This is where caution is extremely important. If the researcher does not recognize that the RAS and CGM estimates are contextual effects (the difference between the neighborhood effect and the resident effect), he or she might erroneously interpret the relationship between the neighborhood mean and the health outcome as negative and conclude that increases in the average walking hours among neighborhoods result in lower health outcomes.

**Model 2 Variances and Covariances**

The full hierarchical linear model utilizing RAS (shown above, copied here) contains fixed effects and random effects:
\[ Y_{ijk} = \gamma_{000} + \gamma_{001}(\bar{X}_{jk}) + \gamma_{010}(\bar{X}_{ik}) + u_{00k} + r_{0jk} + r_{1jk}(X_{ijk}) + e_{ijk} \] (162)

The error term \( e_{ijk} \) represents that which is unique to person \( i \) in neighborhood \( j \) and city \( k \). In Model 2, \( e_{ijk} \) is the error in \( Y_{ijk} \) after adjusting for \( X_{ijk} \). The assumption in HLM is that \( e_{ijk} \) is normally distributed with homogenous variance across neighborhoods and cities. The variance-covariance matrix is \( \Sigma \):

\[ \Sigma = \sigma^2 \] (163)

The error terms \( r_{0jk} \) and \( r_{1jk} \) represent uniqueness resulting from residing in neighborhood \( j \) in city \( k \). The term \( r_{0jk} \) is the error remaining after controlling for \( X_{jk} \).

These terms arise from the random intercept (\( \pi_{0jk} \)) and random slope (\( \pi_{1jk} \)) in Level-1. The variance-covariance matrix for these terms is \( T_\pi \):

\[ T_\pi = \begin{bmatrix} \tau_{\pi_{00}} & \tau_{\pi_{01}} \\ \tau_{\pi_{10}} & \tau_{\pi_{11}} \end{bmatrix} \] (164)

The error term \( u_{00k} \) represents uniqueness resulting from residing in city \( k \). This term arises from the random intercept in Level-2. The variance-covariance matrix is \( T_\beta \):

\[ T_\beta = \tau_{\beta_{00}} \] (165)
Model 2 variance of the Level-1 intercepts ($\tau_{\pi_{00}}$). The interpretation of $\tau_{\pi_{00}}$ under RAS is the variance in the neighborhood means of $Y$ that remains after accounting for the variance explained by $\bar{X}_{jk}$ (a predictor in the model for the intercept for Model 2). Under CWC, $\tau_{\pi_{00}}$ refers to the variance of the unadjusted neighborhood means of $Y$ that remains after accounting for the variance explained by $(\bar{X}_{jk} - \bar{X}_{..})$. When CGM is used, $\tau_{\pi_{00}}$ is the variance of the adjusted neighborhood means after accounting for the variance explained by $(\bar{X}_{jk} - \bar{X}_{..})$. As in Model 1, the random Level-1 slope can also affect the variance estimate of the intercept such that in the presence of random slopes, and when the neighborhood mean differs from the grand mean (and both differ from zero), the estimates for $\tau_{\pi_{00}}$ will differ depending on whether RAW, CWC or CGM is used because of the differing location of the intercept in each model.

Model 2 variance of the Level-1 slopes ($\tau_{\pi_{11}}$). Under CWC, $\tau_{\pi_{11}}$ refers to the variance of the Level-1 slopes ($\pi_{1jk}$). Because there is no between-group variation under CWC, the CWC slopes represent the pooled regression of $Y_{ijk}$ on $X_{ijk}$ within each neighborhood and $\tau_{\pi_{11}}$ represents the variation of these neighborhood slopes. The variance of the Level-1 slopes in the CGM model is significantly less straightforward because of the dependency between slopes and intercepts described above and because CGM does not eliminate between-group variation. As discussed in Chapter 2, when the intraclass correlation is greater than zero, the interpretation of $\pi_{1jk}$ is a blend of the within-cluster and between-cluster regression of $Y$ on $X$.

Model 2 covariance of $r_{0jk}$ and $r_{1jk}$ ($\tau_{\pi_{01}}$). Although assumed not to covary with residuals at other levels, the two residual terms at Level-2 are expected to covary
with one another. \( \tau_{\pi_{01}} \) represents the covariance between these residuals. The influence of the random slope on the variance of the intercept described above also affects the estimates of the covariances between random intercepts and random slopes, \( \tau_{\pi_{01}} \).

**Model 2 variance of the Level-2 intercepts (\( \tau_{\beta_{00}} \)).** At Level-2, there is just one random effect (\( \beta_{00k} \)). \( \beta_{00k} \) represents the intercept of the model for the Level-1 intercept (\( \pi_{0jk} \)). When this term is modeled at Level-3, the equation includes the error terms \( u_{00k} \).

The variance is \( \tau_{\beta_{00}} \) and refers to the between-group (between-city) variation in the city means. When CGM is used, the intercept \( \beta_{00k} \) is adjusted on the city mean of \( X (\bar{X}_{.,k}) \).

So, \( \tau_{\beta_{00}} \) represents variation in the adjusted city means. Because Model 2 does not include \( \bar{X}_{.,k} \) as a predictor in the Level-3 model of the intercept, the interpretation of \( \beta_{00k} \) under CWC is between-city variation in the unadjusted city means.

Table 11 contains the variance and covariance estimates from the artificial dataset using RAS, CWC and CGM in Model 2.

**Interpretation of Model 3 Estimates**

Model 3 incorporates a predictor at each level. At Level-1, \( X_{ijk} \) predicts \( Y_{ijk} \). At Level-2, the neighborhood mean (\( \bar{X}_{jk} \)) predicts the Level-1 intercept. At Level-3, the city mean (\( \bar{X}_{.,k} \)) predicts the Level-2 intercept for the model of the Level-1 intercept. The raw score equations for Model 3 are as follows:

\[
Y_{ijk} = \pi_{ojk} + \pi_{1jk} (X_{ijk}) + e_{ijk} \quad (166)
\]
\[
\pi_{ojk} = \beta_{00k} + \beta_{01k} (\bar{X}_{jk}) + r_{0jk} \quad (167)
\]
\[
\pi_{1jk} = \beta_{10k} + r_{1jk} \quad (168)
\]
\[ \beta_{00k} = \gamma_{000} + \gamma_{001}(\Bar{X}_{.k}) + u_{00k} \]  

(169)

\[ \beta_{01k} = \gamma_{010} \]  

(170)

\[ \beta_{10k} = \gamma_{100} \]  

(171)

Substituting across the levels results in the full model:

\[ Y_{ijk} = \gamma_{000} + \gamma_{001}(\Bar{X}_{.k}) + \gamma_{010}(\Bar{X}_{jk}) + \gamma_{100}(X_{ijk}) + u_{00k} + r_{0jk} + r_{1jk}(X_{ijk}) + e_{ijk} \]  

(172)

In addition to modeling the effect of a person’s walking habits on his or her health, Model 3 models the effect of the walking habits of the person’s neighbors, and the walking habits of the residents in the person’s city. Though there may not be a direct connection between the walking habits of one’s neighbors and one’s health, these measures may be proxies for other measures that influence one’s health. For instance, a neighborhood or city where individuals walk a lot may indicate the presence of community enhancements that influence health such as sidewalks (reducing traffic injuries), availability of fast-food restaurants (reducing negative dietary influences on health), and community spaces for activity such as parks and bicycle trails (increasing physical activities other than walking, which influence health).

**Model 3 intercept (\( \gamma_{000} \)).** In the raw score model, the interpretation of the intercept is the expected value of \( Y_{ijk} \) when \( X_{ijk} = \Bar{X}_{.jk} = \Bar{X}_{..k} = \Bar{X}_{...} = 0 \). That is, the expected health value for an individual who walks zero hours per week, whose neighborhood average for walking is zero hours per week, and whose city average for...
walking is zero hours per week. This provides a good example of when centering might be useful to bring meaning to the intercept, as discussed in Chapter 1.

When \( X \) is grand-mean centered at Level-1, Level-2 and Level-3, the full model is:

\[
Y_{ijk} = \gamma_{000} + \gamma_{001}(X_{..k} - \bar{X}_{..}) + \gamma_{010}(X_{..jk} - \bar{X}_{..}) + \gamma_{100}(X_{ijk} - \bar{X}_{..}) + u_{00k} + r_{0jk} + r_{1jk}(X_{ijk} - \bar{X}_{..}) + e_{ijk}
\]

(173)

As discussed in Chapter 2 and above, the intercept in a random intercept CGM model is an adjusted mean. This can again be derived by considering the latent intercepts throughout this model. The Level-1 equation in Model 3 is identical to the Level-1 equation in Model 1, thus:

\[
\pi_{0jk} = \mu_{Y_{jk}} - \pi_{1jk}(\bar{X}_{..} - \bar{X}_{..}).
\]

(174)

So \( \pi_{0jk} \) is interpreted as the mean of \( Y \) for neighborhood \( j \) in city \( k \) adjusted for the effect of the difference between the neighborhood mean of \( X \) and the grand mean of \( X \). The Level-2 equation for the intercept for Model 2 includes a predictor:

\[
\pi_{0jk} = \beta_{00k} + \beta_{01k}(\bar{X}_{..} - \bar{X}_{..}) + r_{0jk} .
\]

(175)

Taking the expectation of \( \pi_{0jk} \) within cluster (city), we get:
\begin{equation}
E(\pi_{0jk}) = \beta_{00k} + \beta_{01k}(\bar{X}_{jk} - \bar{X}_{..})
\end{equation}

(176)

Rearranging these terms, we see that $\beta_{00k}$ is an adjusted mean intercept (and the intercept, remember, is an adjusted mean):

\begin{equation}
\beta_{00k} = \mu_{\pi_k} - \beta_{01k}(\bar{X}_{..} - \bar{X}_{..})
\end{equation}

(177)

$\beta_{00k}$, if it were directly estimated, would be interpreted as the mean $Y$ of the neighborhoods that make up city $k$ adjusted for the effect of the deviation of the city mean from the grand mean $X$. Model 3 includes a predictor at Level-3 as well. The Level-3 equation for the intercept in the CGM model is:

\begin{equation}
\beta_{00k} = \gamma_{000} + \gamma_{001}(\bar{X}_{..} - \bar{X}_{..}) + u_{00k}
\end{equation}

(178)

Taking the expectation of $\beta_{00k}$ we get:

\begin{equation}
E(\beta_{00k}) = \gamma_{000} + \gamma_{001}(\bar{X}_{..} - \bar{X}_{..})
\quad = \gamma_{000}
\end{equation}

(179)

The expectation of $\beta_{00k}$ is $\gamma_{000}$, which is interpreted as the average of the adjusted city means, which are averages of the adjusted averages of the neighborhood means.

Therefore, in the full CGM model, the intercept $\gamma_{000}$ is an adjusted mean of $Y$ with adjustments at Level-2 and Level-3. In terms of the example, the CGM intercept is the
expected health score adjusted for the neighborhood’s walking hours and the city’s walking hours.

Finally, when X is CWC at Level-1 and Level-2, the full model is:

\[
Y_{ijk} = \gamma_{000} + \gamma_{001}(X_{..k} - \bar{X}_{..k}) + \gamma_{010}(X_{ijk} - \bar{X}_{..k}) + u_{00k} + r_{0jk} + r_{1jk}(X_{ijk} - \bar{X}_{..k}) + e_{ijk}
\]  

(180)

If we take the expectation within neighborhoods from the Level-1 equation, as above, we get:

\[
\mu_{Y_{jk}} = \pi_{0jk} - \pi_{1jk}(\bar{X}_{..k} - \bar{X}_{..k})
\]

\[
= \pi_{0jk}
\]

(181)

So \(\pi_{0jk}\) in the CWC case is an unadjusted mean, just like in previous models. Then \(\pi_{0jk}\) is interpreted as the unadjusted mean of Y for neighborhood j in city k. The Level-2 equation for the intercept in the CWC model is:

\[
\pi_{0jk} = \beta_{00k} + \beta_{01k}(\bar{X}_{..k} - \bar{X}_{..k}) + r_{0jk}
\]

(182)

Taking the expectation of \(\pi_{0jk}\) within cluster (city), we get:

\[
E(\pi_{0jk}) = \beta_{00k} + \beta_{01k}(\bar{X}_{..k} - \bar{X}_{..k})
\]

\[
= \beta_{00k}
\]

(183)
Thus, $\beta_{00k}$ is also an unadjusted mean, in this case the mean $\pi_{0jk}$ for city $k$. Next, we consider the Level-3 equation:

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(\bar{X}_{..k} - \bar{X}_{..}) + u_{00k}.$$ (184)

Taking the expectation of $\beta_{00k}$ we get:

$$E(\beta_{00k}) = \gamma_{000} + \gamma_{001}(\bar{X}_{..} - \bar{X}_{..})$$

$$= \gamma_{000}$$ (185)

The interpretation of $\gamma_{000}$ is then the unadjusted grand mean of $Y$. Note that if CGM is not used at Level-3, and the raw score is used, $\gamma_{000}$ has a slightly different interpretation. In this case, the Level-3 equation is:

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(\bar{X}_{..}) + u_{00k}$$ (186)

And the expectation of $\beta_{00k}$ is:

$$E(\beta_{00k}) = \gamma_{000} + \gamma_{001}(\bar{X}_{..})$$ (187)

Rearranging terms, we get:
\[ y_{000} = \mu_B - y_{001}(\bar{X}_.) \] \hspace{1cm} (188)

So in the CWC framework, when \( \bar{X}_{.,k} \) is not centered on the grand mean, the interpretation of \( y_{000} \) is an adjusted overall mean \( Y \) where the adjustment is \( y_{001} \) multiplied by the grand mean of \( X \).

**Model 3 Level-3 slope (\( y_{001} \)).** The upper-level slopes in Model 3 are the contextual effects discussed earlier. Under RAS and CGM, the interpretation of \( y_{001} \) is the difference in the effect of the city mean (\( \bar{X}_{.,k} \)) and the effect of the neighborhood mean (\( \bar{X}_{jk};(\beta_{city} - \beta_{neighborhood}) \)). Thus, \( y_{001} \) is a contextual effect. The resulting estimate is interpreted as the difference between the effect of the city average number of walking hours on health and the effect of the neighborhood average number of walking hours on health. A negative sign indicates that the neighborhood effect is larger than the city effect. Statistical significance in this coefficient indicates that the two effects (neighborhood, city) are different from one another, but doesn’t inform whether their relationship to health is statistically significant.

In the CWC model, the interpretation of \( y_{001} \) is the expected change in \( Y \) when \( (\bar{X}_{.,k} - \bar{X}_.) \) changes by one unit and when all other fixed effects equal zero; that is, the direct effect. Thus, \( y_{001} \) is the influence of the city mean (\( \bar{X}_{.,k} \)) on health for the average person in the average neighborhood.

**Model 3 Level-2 slope (\( y_{010} \)).** Under RAS and CGM, the interpretation of \( y_{010} \) is the difference in the effect of the neighborhood mean (\( \bar{X}_{jk} \)) and the effect of the person score, \( X_{ijk} \), on \( Y_{ijk} \) \((\beta_{neighborhood} - \beta_{person}) \). Thus, \( y_{010} \) is also a contextual effect. The resulting estimate is interpreted as the difference between the effect of the neighborhood
average number of walking hours and the effect of the individual person’s number of walking hours on health. A negative sign again indicates that the person effect is larger than the neighborhood effect, which is consistent with what might be expected (that is, my own walking habits have more influence on my health than the average walking hours of individuals in my neighborhood). Statistical significance in this coefficient indicates that the two effects (person, neighborhood) are different from one another, but again doesn’t inform whether their relationship to health is statistically significant.

In the CWC model, the interpretation of $\gamma_{010}$ is the expected change in Y when $(X_{jk} - \bar{X}_k)$ changes by one unit. Thus, $\gamma_{010}$ is the direct effect of the neighborhood walking average ($\bar{X}_{jk}$) on the health outcome.

**Model 3 Level-1 slope ($\gamma_{100}$).** In all three models, the interpretation of $\gamma_{100}$ is the effect of the person’s hours of walking when all other fixed effects equal zero. This interpretation is intuitive for the CWC model because CWC removes the between-neighborhood and between-city variation in the predictor variable so that all that remains is within-neighborhood variability. For CGM and RAS, as in Model 2, when the upper-level means are included in the model, the resulting regression coefficients are partial regression coefficients. As such, $\gamma_{100}$ is the influence of an individual resident’s walking habits controlling for neighborhood and city mean walking habits, while $\gamma_{010}$ and $\gamma_{001}$ estimate the influence of neighborhood and city mean walking habits on neighborhood and city mean health outcomes, respectively. This results in similar estimates of $\gamma_{100}$ under RAS, CGM and CWC.

Table 12 contains the gamma estimates from the artificial dataset using RAS, CWC and CGM in Model 3. Notice that the intercept estimate ($\gamma_{000}$) is identical to the
grand mean of Y when $\bar{X}_{.,k}$ is centered on the grand mean (62.942), and that the CGM estimate is an adjusted mean. The estimates of the Level-1 slope are identical across all three centering options ($\gamma_{100} = 5.008$). As in Model 2, the estimate of the Level-2 slope differs significantly between the centering options. The same is true of the estimate of the Level-3 slope. CGM and RAS estimates of the Level-2 slope are negative ($\gamma_{010} = -1.987$) while the CWC estimate is positive ($\gamma_{010} = 3.052$). Similarly, CGM and RAS estimates of the Level-3 slope are negative ($\gamma_{001} = -1.028$) while the CWC estimate is positive ($\gamma_{001} = 1.995$). As before, if the researcher does not recognize that the RAS and CGM estimates for $\gamma_{010}$ and $\gamma_{001}$ are contextual effects, he or she might erroneously interpret the relationship between the neighborhood walking mean, city walking mean and the health outcome as negative and conclude that increases in the average walking hours among neighborhoods and cities result in lower health outcomes.

Model 3 Variances and Covariances

The full hierarchical linear model utilizing RAS (shown above, copied here) contains fixed effects and random effects.

$$Y_{ijk} = \gamma_{000} + \gamma_{001}(\bar{X}_{.,k}) + \gamma_{010}(\bar{X}_{jk}) + \gamma_{100}(X_{ijk}) + u_{00k} + r_{0jk} + r_{1jk}(X_{ijk}) + e_{ijk}$$

(189)

The error term $e_{ijk}$ represents that which is unique to person i in neighborhood j and city k. In Model 3, $e_{ijk}$ is the error in $Y_{ijk}$ after adjusting for $X_{ijk}$. The assumption in HLM is that $e_{ijk}$ is normally distributed with homogenous variance across neighborhoods and cities. The variance-covariance matrix is $\Sigma$: 
\[ \Sigma = \sigma^2 \]  \hspace{1cm} (190)

The error terms \( r_{0jk} \) and \( r_{1jk} \) represent uniqueness resulting from residing in neighborhood \( j \) in city \( k \). The term \( r_{0jk} \) is the error remaining after adjusting for \( X_{jk} \). These terms arise from the random intercept \( (\pi_{0jk}) \) and random slope \( (\pi_{1jk}) \) in Level-1. The variance-covariance matrix for these terms is \( T_{\pi} \):

\[
T_{\pi} = \begin{bmatrix} \tau_{\pi_{00}} & \tau_{\pi_{01}} \\ \tau_{\pi_{10}} & \tau_{\pi_{11}} \end{bmatrix} \hspace{1cm} (191)
\]

The error term \( u_{00k} \) represents uniqueness resulting from residing in city \( k \). This term arises from the random intercept \( (\beta_{00k}) \) in Level-2 after adjusting for \( X_{..k} \). The variance-covariance matrix is \( T_{\beta} \):

\[
T_{\beta} = \tau_{\beta_{00}} \hspace{1cm} (192)
\]

**Model 3 variance of the Level-1 intercepts** \((\tau_{\pi_{00}})\). The interpretation of \( \tau_{\pi_{00}} \) under RAS is the variance in the neighborhood means of \( Y \) that remains after accounting for the variance explained by \( X_{jk} \) (a predictor in the model for the intercept for Model 3). Under CWC, \( \tau_{\pi_{00}} \) refers to the variance of the unadjusted neighborhood means of \( Y \) that remains after accounting for the variance explained by \((X_{jk} - \bar{X}_{..})\). When CGM is used, \( \tau_{\pi_{00}} \) is the variance of the adjusted neighborhood means after accounting for the variance.
explained by \((\bar{X}_{jk} - \bar{X})\). As in previous models, the random Level-1 slope can also affect the variance estimate of the intercept such that in the presence of random slopes, and when the neighborhood mean differs from the grand mean (and both differ from zero), the estimates for \(\tau_{\pi_{00}}\) will differ depending on whether RAS, CWC or CGM is used because of the differing location of the intercept in each model.

**Model 3 variance of the Level-1 slopes** \((\tau_{\pi_{11}})\). When CWC is employed, \(\tau_{\pi_{11}}\) refers to the variance of the Level-1 slopes \((\pi_{1jk})\). Because there is no between-group variation in the slopes when \(X_{ijk}\) is CWC-L2 \((X_{ijk} - \bar{X}_{jk})\), the CWC slopes represent the pooled regression of \(Y_{ijk}\) on \(X_{ijk}\) within each neighborhood and \(\tau_{\pi_{11}}\) represents the variation of these neighborhood slopes. The variance of the Level-1 slopes in the CGM model is significantly less straightforward because of the dependency between slopes and intercepts and because CGM does not eliminate between-group variation in the slope coefficient. As discussed in Chapter 2, when the intraclass correlation is greater than zero, as it is in this dataset, the interpretation of \(\pi_{1jk}\) is a blend of the within-cluster and between-cluster regression of \(Y\) on \(X\).

**Model 3 covariance of** \(r_{0jk}\) and \(r_{1jk}\) \((\tau_{\pi_{01}})\). Although assumed not to covary with residuals at other levels, the two residual terms at Level-2 are expected to covary with one another. \(\tau_{\pi_{01}}\) represents the covariance between these residuals. The influence of the random slope on the variance of the intercept described above also affects the estimates of the covariances between random intercepts and random slopes, \(\tau_{\pi_{01}}\).

**Model 3 variance of the Level-2 intercepts** \((\tau_{\beta_{00}})\). At Level-2, there is just one random effect \((\beta_{00k})\). \(\beta_{00k}\) represents the intercept of the model for the Level-1 intercept \((\pi_{0jk})\). When this term is modeled at Level-3, the equation includes one predictor \((\bar{X}_{.k})\)
and one error term \( (u_{00k}) \). The variance is \( \tau_{\beta_{00}} \). Since the city mean is a predictor in Model 3, \( \tau_{\beta_{00}} \) refers to the between-group (between-city) variation in the city intercepts after accounting for the variance explained by the city mean of \( X \).

Table 13 contains the equivalencies in the variances and covariances between the RAS, CWC and CGM estimates of Model 3.

**Model Misspecification**

When significant relationships exist between the upper-level means and the outcome, failure to model these relationships can influence the parameter estimates. Consider Model 1 where only \( X_{ijk} \) is present in the model: If the effects of \( \bar{X}_{jk} \) and \( \bar{X}_{k} \) are not different from the effect of \( X_{ijk} \), then failing to include these means as predictors in the model should have no influence on the estimate of the Level-1 slope (\( \pi_{1jk} \)). However, if Model 1 is estimated using data for which \( Y_{ijk} \) is significantly influenced by \( X_{ijk}, \bar{X}_{jk} \) and \( \bar{X}_{k} \), as was the case with Model 3 data, the resulting parameter estimates are affected. To demonstrate, Model 1 was estimated using the data from Model 3. The Model 3 data were created with significant regression effects at Level-1, Level-2 and Level-3, and these effects were shown in Table 9. The parameter estimates resulting from the analysis of Model 1 on Model 3 data are in Table 14.

**Intercept \( (\gamma_{000}) \).** The interpretation of the intercept is not different in the mismatched analysis. Under RAS, the interpretation of the intercept is the expected value of \( Y_{ijk} \) when \( X_{ijk} = 0 \), that is the expected health for a person who walks zero hours per week. Under CGM, the intercept, \( \gamma_{000} \), is an adjusted mean of \( Y \), or the adjusted mean health for all residents. Under CWC, the interpretation of \( \gamma_{000} \) is the unadjusted mean of
Y, or the mean health of all residents. Looking at the estimates in Table 14, notice that the use of CWC produces an estimate very close to the mean Y across the sample ($\gamma_{000} = 62.942$) while RAS and CGM produce estimates that differ from the estimates produced by Model 3 (the model that fits these data). This finding is consistent with our understanding of the CWC intercept as an unadjusted mean and the CGM as an adjusted mean. The unadjusted mean produced by CWC should not differ according to the presence or absence of contextual effects in the data, and inclusion or exclusion of these effects in the model. However, the adjustments that lead to the CGM and RAS intercepts are dependent on whether these effects are both present in the data and present in the model. So in the mismatched model, when the significant contextual effects are not modeled as parameters, it follows that the adjusted means produced by RAS and CGM estimates are different than the means produced by Model 3 where these contextual effects are included in the model.

**Level-1 slope ($\gamma_{100}$).** Recall from Chapter 2 the importance of centering when interpreting the Level-1 slope coefficient. Under CWC, there is no between-cluster variation in $X_{ijk}$ and the interpretation of $\gamma_{100}$ is the expected change in health when a person increases his or her walking by one hour per week. Notice in Table 14 that the CWC estimate for $\gamma_{100}$ recovers the estimate resulting from estimating Model 3 using Model 3 data ($\gamma_{100} = 5.009$). This is the within-neighborhood effect of walking on health.

Under CGM, however, there is still between-group variation in $X_{ijk}$, and because this is a three-level model, there are two sources of between-group variation: between-neighborhood and between-city. Unlike Model 3, these means are not re-introduced as
predictors at Level-2 and Level-3 but there are significant effects for each of these means in the data. Because the means are not in the model, \( \gamma_{100} \) is not a partial regression coefficient as it was in Model 3. Instead, \( \gamma_{100} \) is a complex blend of the Level-1, Level-2, and Level-3 regression coefficients.

**Variance and covariances.** When the model is misspecified, the variances and covariances are also affected. If Level-2 and Level-3 effects are present in the data but not included in the model, the Level-2 variance of the intercept \( \tau_{\pi_{00}} \) and Level-3 variance of the intercept \( \tau_{\rho_{00}} \) are both larger under RAS, CWC and CGM for the mismatched model/data pairing compared with the matched model/data pairing. This is because the aggregated X predictors at each of these levels explain variance in Y, but have not been included in the model.
CHAPTER 5: DISCUSSION

This study investigated the implications of centering in three-level models. The primary goal was to identify the interpretations of the coefficients and variance terms in three distinct three-level models when employing CWC, CGM and RAS. The secondary goal was to explore the impact on parameter estimates when the imposed model failed to include associations that existed in the data, a misspecification of the model. This study extends previous work by targeting three-level models. Unlike two-level models, the three-level model presents researchers with two sets of centering decisions: those for the Level-1 variables, and those for the Level-2 variables.

Model 1 included a single Level-1 predictor, random terms for the intercept at Level-2 and Level-3, and a random term for the slope at Level-2. Model 2 built on Model 1 by adding the cluster means of the Level-1 predictor variable at Level-2. Model 3 built on Model 2 by adding the cluster means at Level-3. In Model 2 and Model 3, the centering decision made at Level-1 was also made at Level-2 so that if the Level-1 variable was CWC, so too was the Level-2 variable.

Before addressing the interpretation of the parameter estimates, initial work was done to determine the theoretical equivalencies in the means and variances across the centering options in each model. This work was an extension of Kreft, de Leeuw and Aiken (1995) in which similar equivalencies were derived for the two-level case. Model 1 was equivalent in the means and variances for RAS and CGM, but neither RAS nor CGM was equivalent to CWC in either fixed or random terms. Like Model 1, Model 2 had equivalencies between RAS and CGM for the fixed effects and random effects, but neither RAS nor CGM were equivalent to CWC in the fixed or random effects. Model 3
demonstrated equivalencies in the means for RAS, CGM and CWC, but only equivalence in the variance terms between RAS and CGM.

Additional work was done to identify the algebraic relationships between the centering options in the two contextual models: Model 2 and Model 3. This work was an extension of Raudenbush (1989a) in which similar algebraic relationships were demonstrated for the two-level case. In Model 3, the coefficients associated with the Level-2 mean ($\bar{X}_{jk}$) and the Level-3 mean ($\bar{X}_{..k}$) in the CGM and RAS models were identified as contextual effects, while the coefficients associated with the Level-2 mean ($\bar{X}_{jk}$) and the Level-3 mean ($\bar{X}_{..k}$) in the CWC model were direct effects. Model 2 was significantly more complicated. If the Level-2 mean ($\bar{X}_{jk}$) was RAS or CGM, the results were similar to those in the two-level framework. That is, the Level-1 slope coefficient estimated the within-level relationship and the Level-2 slope coefficient estimated a contextual effect, specifically the difference between the relationship at Level-1 and the relationship at Level-2. However, if a significant effect of the Level-3 mean ($\bar{X}_{..k}$) is present in the data, then the coefficient associated with the Level-2 mean is a complex blend of the Level-2 and Level-3 effects. Alternately, if the Level-2 mean was CWC, the model is not equivalent in the fixed (or random) effects with the RAS or CGM models, and the coefficients are therefore not simple reparameterizations of one another. The findings here suggest that for estimating contextual effects in three-level models, the best approach is to include the mean at both Level-2 ($\bar{X}_{jk}$) and Level-3 ($\bar{X}_{..k}$), especially if a significant contextual effect at Level-3 is hypothesized to be present in the data.

The remaining results illustrated the interpretations of the terms in each of the models and used artificially generated datasets to provide examples. In the misspecified
model, it was shown that failing to model existing significant upper level effects can produce blended Level-1 slope coefficients that have complex interpretations. The details of these findings will be discussed in light of their relationship to research questions in the next section.

**Implications for Researchers**

Consistent with previous work in this area (in particular, Enders & Tofghi, 2007), what follows is a series of recommendations to researchers to assist with decision-making regarding centering options in three-level models similar to those presented. As Kreft, de Leeuw & Aiken (1995) point out, aside from the principle of parsimony there is no statistically “correct” decision when considering the centering options in a multilevel model. Centering does improve ability of the estimation process to converge and come to a stable solution, particularly in the presence of variables on dramatically different scales. However, either option of CGM or CWC would accomplish this goal. The choice of centering therefore rests heavily on substantive theory. In the three-level models, the choices of centering expand, as do the potential sources for misinterpretation. Thus, extreme caution needs to be exercised when making choices about centering, and interpreting the resulting parameter estimates.

To facilitate this discussion, I used a single dataset from the Model 3 datasets for demonstration. As with all of the Model 3 datasets, this dataset has a significant contextual effect at both Level-2 and Level-3. In the examples below, various models are used to demonstrate the recommendations based on the results in Chapter 4. I will continue with the example of residents nested in neighborhood, nested in cities, with number of walking hours per week as the predictor (X) and a measure of self-reported
health status as the outcome (Y). In the example dataset, the population value for the effect of an individual’s walking habits on the health status was 5.0; the effect of the neighborhood mean on the health status was 3.0; and the effect of the city mean on the health status was 2.0. This implies that the Level-2 contextual effect is expected to be -2.0 (that is $3.0 - 5.0 = -2.0$) and the Level-3 contextual effect is expected to be -1.0 (that is $2.0 - 3.0 = -1.0$). The negative contextual effects are the result of the lower level effect being larger than the upper level effect. In this example, the negative contextual effect suggests the effect of an individual’s walking habits on his/her health status is larger than the effect of the person’s neighborhood average walking habits.

**Interest in the effect of the Level-1 variable.** If the researcher’s primary interest is in the effect of a person’s walking habits on their health status, and neighborhood and city mean walking hours will not be included in the model (i.e. Model 1), CWC appears to be the best approach. This is consistent with the findings expressed by Enders & Tofighi (2007). The reason for this conclusion is that CWC produces an estimate of the within-group regression that is not confounded by the presence of between-group variance in the predictor variable. As shown in the misspecified model in Chapter 4, CGM and RAS do not remove this variation, and thus produce estimates of the regression slope that are a blend of the significant Level-1, Level-2 and Level-3 regression coefficients present in the data. Using the example dataset, we would expect the relationship between the person’s walking habits and health to be 5.0. The results were that the RAS and CGM estimates of the Level-1 regression were both 4.97, while the CWC estimate was closer to the population value at 5.0. A similar analysis done in
Chapter 4 using all ten datasets showed that CWC produced a closer estimate of the population value in the Model 1 analysis of Model 3 data (Table 14).

It is important to note again that the interpretation of the CWC coefficient is a within-cluster regression. This model suggests a substantive theory that involves an individual’s relative standing within his or her group, and that the effects are dependent on context. Klein, Dansereau, & Hall (1994) point out that in hypotheses of this type, “knowledge of the group context is not only informative but necessary to interpret an individual's placement or standing in the group” (p. 202). In our example, this means we would consider the walking “environment” of the neighborhood important. If an individual walks three hours per week in a neighborhood where few people walk at all, that individual might view himself/herself as more healthy than his/her neighbors and self-report a higher score on the health measure. On the other hand, an individual who walks three hours per week among neighbors who walk an hour each day might feel he/she is less healthy than his/her peers and self-report a lower score on the health measure. Thus, the same value on X results in different values of Y because of the context.

**Interest in the effect of the Level-2 mean.** Consider a situation where the researcher might be interested in the effect of the neighborhood mean walking hours, controlling for an individual’s walking hours. This research question involves a Level-2 variable. In the three-level model, the Level-2 variables are unique. On the one hand, they represent an upper-level cluster (neighborhoods are the clusters of individual residents) and on the other hand, they are also lower-level units (neighborhoods are clustered within
cities). In Model 2 and Model 3, the Level-2 predictor is the neighborhood mean hours of walking among residents ($\bar{X}_{jk}$).

There are two centering decisions to be made due to the three-level nature of the data. The first decision is about the Level-1 variable, $X_{ijk}$. By centering $X_{ijk}$ on the grand mean (CGM) or using the raw score (RAS), the correlation between $X_{ijk}$ and $\bar{X}_{jk}$ is preserved. This is important if the researcher is interested in the effect of the Level-2 mean controlling for the Level-1 effect. When CGM or RAS are used at Level-1, the resulting Level-2 coefficient associated with $\bar{X}_{jk}$ is the contextual effect, or the difference between the effect of $X_{ijk}$ and the effect of $\bar{X}_{jk}$. This is the effect of $\bar{X}_{jk}$ controlling for the effect of $X_{ijk}$. If the researcher were to instead use CWC at Level-1, the resulting Level-2 coefficient is the effect of $\bar{X}_{jk}$ independent of the effect of $X_{ijk}$. This is because applying CWC to the Level-1 variable removes the correlation between $X_{ijk}$ and $\bar{X}_{jk}$. As a result, testing the statistical significance of this parameter estimate when CWC is employed at Level-1 would not be addressing a hypothesis about the Level-2 mean controlling for the Level-1 effect, but instead determining whether the relationship between $\bar{X}_{jk}$ and the outcome differed from zero independent of the effect of $X_{ijk}$.

The second centering decision to be made is for the Level-2 mean ($\bar{X}_{jk}$). If the Level-3 mean ($\bar{X}_{.k}$) is not included in the model, the Level-2 mean should be CWC in order to remove any between-city variation from $\bar{X}_{jk}$. If this centering is not done, the resulting coefficient will be a blend of the Level-2 and the Level-3 effect because the Level-2 mean ($\bar{X}_{jk}$) would still contain Level-3 variation. In the event the Level-3 mean
and the Level-2 mean have the same relationship with the outcome variable, which is not the case in this example dataset, this is not an issue.

Using the example dataset, we would expect the relationship between the person’s walking habits and health to be approximately 5.0 and the Level-2 contextual effect to be approximately -2.0. I estimated Model 2 on the sample data using two centering combinations. The first model used CGM at Level-1 and CWC at Level-2 and produced an estimate of the Level-2 coefficient closer to the population value of -2.0 ($\gamma_{010} = -2.03$) than the estimate produced when the both the Level-1 and Level-2 variables are CGM (Model B; $\gamma_{010} = -2.16$). Although a significant Level-3 effect was present in the data, it was not included in this model. Thus, when CGM was used at Level-2, the coefficient associated with $X_{jk}$ was a blend of the Level-2 coefficient and the Level-3 coefficient and further from the population value for the effect of the Level-2 mean.

**Interest in the effect of the Level-3 mean.** Consider a situation where the researcher might be interested in the effect of the city mean walking hours, controlling for an individual’s walking hours and the average walking hours in a person’s neighborhood. Model 3 included predictors at all three levels of the model ($X_{ijk}$, $X_{jk}$, and $X_{k}$).

Again there are two centering decisions to make: a decision at Level-1 and a decision at Level-2. As before, centering $X_{ijk}$ on the grand mean (CGM) or using the raw score (RAS), preserves the correlation between $X_{ijk}$, $X_{jk}$ and $X_{k}$. This is important if the researcher is interested in the context effect. As we saw in Table 5, employing CWC at Level-1 results in zero correlations between $X_{ijk}$, and both $X_{jk}$ and $X_{k}$. 
The second centering decision to be made is for the Level-2 mean ($\bar{X}_{jk}$). The same logic that applies for the Level-1 variable in Model 2 also applies here. By centering $\bar{X}_{jk}$ on the grand mean (CGM) or using the raw score (RAS), the correlation between $\bar{X}_{jk}$ and $\bar{X}_{..k}$ is preserved. This is important if the researcher is interested in the effect of the Level-3 mean controlling for the effect of the Level-2 mean. When CGM or RAS are used at Level-2, the resulting Level-3 coefficient associated with $\bar{X}_{..k}$ is the effect of $\bar{X}_{..k}$ controlling for the effect of $\bar{X}_{jk}$. If the researcher were to instead use CWC at Level-2, the resulting Level-3 coefficient is the direct effect of $\bar{X}_{..k}$ independent of the effect of $\bar{X}_{jk}$. This is because applying CWC to the lower-level variable removes the correlation between $\bar{X}_{jk}$ and $\bar{X}_{..k}$.

Unlike Model 2, I showed in Chapter 4 that the fixed effects in Model 3 are equivalent across RAS, CWC and CGM. This means that algebraic manipulation of one set of estimates can produce another set of estimates. Thus, through the use of user-defined constraints, a researcher could test the effect of the city mean controlling for the neighborhood mean and the individual’s walking hours using CWC instead of CGM or RAS at the lower levels. In this way, making a particular centering decision in this specific three-level model does not preclude making tests associated with another centering decision. However, interpretation of the resulting coefficients does require caution on the part of the researcher.

I estimated Model 3 on the sample data using CGM at all levels (identical to the CGM version of Model 3 in Chapter 4). The resulting estimate for the Level-3 contextual effect was very close to the population value of -1.0 ($\gamma_{001} = -0.92$).
Interest in the effect of a variable at all three levels. What we learned about the equivalence between CWC and CGM in Model 3 also has direct implications for a researcher interested in the effects of a variable at all three levels of the model. Model 3 produces equivalent fixed effects under CWC, CGM and RAS, so the centering option chosen when the interest is in the effect of a variable at all three levels does not matter. What is critically important is that the researcher understands the appropriate interpretation of the parameter estimate produced by the model she has chosen. This means keeping in mind that the CGM and RAS models produce Level-2 and Level-3 estimates that quantify the contextual effects and that the CWC approach results in quantification of the direct effects.

An issue that researchers may encounter when examining effects at all three levels is a lack of statistical significance for one or both of the contextual effects under the CGM or RAS centering option. When the estimate for a contextual effect is non-significant, assuming there is enough power to detect an effect, the conclusion is that the upper-level effect is equal to the lower-level effect. If the Level-3 contextual effect is non-significant, the conclusion would be that the effect of the city mean is equal to the effect of the neighborhood mean. If the Level-2 contextual effect is non-significant, the conclusion would be that the effect of the neighborhood mean is equal to the effect of the person score. In either case, the most parsimonious model is one that does not include the non-significant contextual effect. If both the Level-2 and Level-3 contextual effects are non-significant, the researcher might consider dropping both from the model. The result is a more efficient Level-1 parameter estimate (Raudenbush & Bryk, 2002).
Variances and Covariances

In this dissertation, all three models had the same variance and covariance structure. That is, a random intercept at Level-1 and Level-2, and a random slope at Level-1. What is known about the influence of centering on the variances and covariances in the two-level model directly apply to the three-level model, and are summarized here. As noted in Chapter 2 and throughout the models in Chapter 4, centering influences the estimates of variance and covariance in a number of ways. Under CWC, the intercept at Level-1 and Level-2 are unadjusted means and therefore the interpretation of the variances of these intercepts ($\tau_{\pi_{00}}$ and $\tau_{\beta_{00}}$) are between-neighborhood variation in the Level-1 intercept and and between-city variation in the Level-2 intercepts, respectively. CGM and RAS, on the other hand, produce estimates of the intercepts that are adjusted means and therefore the interpretation of the variances of the intercepts reflects this. Additionally, changing the location of the Level-1 intercept through the use of centering in the presence of a random Level-1 slope (which exists in Model 1, Model 2 and Model 3), can also affect the variance estimate for the intercept.

The variance estimate of the slope can also be affected by centering. In this dissertation there was just one slope variance estimate for the random slope at Level-1 ($\tau_{\pi_{11}}$; present in Model 1, Model 2 and Model 3). Under CWC, $\tau_{\pi_{11}}$ is the variance of the pooled within-group regression of $Y$ on $X$. Under CGM, there is a strong relationship between the intercept and slope which can influence and potentially minimize the estimate of $\tau_{\pi_{11}}$. For this reason, Raudenbush and Bryk (2002) recommend CWC for the best estimate of $\tau_{\pi_{11}}$. 
The estimate of the covariance of the random Level-1 intercept and random Level-1 slope ($\tau_{\pi_{01}}$; present in Model 1, Model 2 and Model 3) is affected by centering insofar as the estimates of $\tau_{\pi_{00}}$ and $\tau_{\pi_{11}}$ are influenced by centering decisions.

**Contextual Variables**

The inclusion or exclusion of aggregated contextual variables, such as the upper level means in Model 2 and Model 3 also deserve more attention. One issue is that these contextual variables may be reliable proxies for characteristics of the environment that were unmeasured in the study, as in the walking conditions example discussed early. They may also be constructs that are difficult to measure as a static group variable. An example of this might be workplace stress. Since workplaces are not people, but instead a collection of people, aggregation of stress measures across individual employees is a reasonable way to operationalize this construct. In either case, inclusion of these upper level means may explain additional variability in the outcome measure of interest.

Failing to include these means, when they do explain variation in the outcome over and above the lower level variable can dramatically bias the resulting Level-1 slope coefficient under RAS or CGM. Kreft, de Leeuw and Aiken (1995) point out that “It is more common practice that researchers specify why they *include* certain effects, than to explain why they *exclude* certain effects” (p. 20). As multilevel models become increasingly complex, researchers may need to do more to explain why upper level means of lower level variables are not included in their models. That is, the decision to exclude an upper-level mean should be accompanied by theoretical support for why the researcher expects the effect of the upper-level mean to be the same as that of the lower-level
variable. In the case of more than two levels, the researcher would need to provide theoretical support for identical aggregate effects across all levels of the model for that variable.

**Limitations**

As discussed in the Introduction, this work was not without limitations. Not every combination of centering options was considered in this dissertation, but instead a subset based on previous work in this area (Kreft, de Leeuw, & Aiken, 1995). For simplicity, the models were also limited to a single Level-1 variable that was aggregated to the upper levels. Finally, the limitations inherent in the use of artificially generated data also affect this work.

**Summary**

This dissertation highlights the importance of thoughtfully articulating three-level hierarchical models to best understand the interpretation of the parameter estimates resulting from the analyst’s chosen method of centering. As in all analytic decisions, sound theoretical framework must be a keystone. This dissertation incorporated an example (walking habits as a predictor of health) that demonstrated how cross-sectional, multilevel, contextual designs might be applied to the field of epidemiology.
REFERENCES


<table>
<thead>
<tr>
<th>Model 1</th>
<th>Expectations</th>
<th>Dispersions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{ij}$ at Level-1</td>
<td>RAS = CGM ≠ CWC</td>
<td>RAS = CGM ≠ CWC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 2</th>
<th>Expectations</th>
<th>Dispersions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{ij}$ at Level-1</td>
<td>RAS = CGM = CWC</td>
<td>RAS = CGM ≠ CWC</td>
</tr>
<tr>
<td>$\bar{X}_j$ at Level-2</td>
<td>RAS = CGM = CWC</td>
<td>RAS = CGM ≠ CWC</td>
</tr>
</tbody>
</table>
Table 2: Model 1, Model 2 and Model 3 raw score population parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{000}$</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>$\gamma_{100}$</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>$\gamma_{010}$</td>
<td>NA</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>$\gamma_{001}$</td>
<td>NA</td>
<td>NA</td>
<td>-1</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>$\tau_{\pi00}$</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>$\tau_{\pi01}$</td>
<td>$\rho = -0.50$</td>
<td>$\rho = -0.50$</td>
<td>$\rho = -0.50$</td>
</tr>
<tr>
<td>$\tau_{\pi11}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\tau_{\beta00}$</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
### Table 3: Intraclass Correlations for Model 1, Model 2 and Model 3

<table>
<thead>
<tr>
<th>Equation</th>
<th>ICC for Y</th>
<th></th>
<th>ICC for X</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of variance within Level-2 units</td>
<td></td>
<td>Model 1 Data</td>
<td>Model 2 Data</td>
<td>Model 3 Data</td>
</tr>
<tr>
<td>( \frac{\sigma^2}{(\sigma^2 + \tau_\pi + \tau_\beta)} )</td>
<td>0.27</td>
<td>0.23</td>
<td>0.23</td>
<td>0.49</td>
</tr>
<tr>
<td>Proportion of variance between Level-2 units</td>
<td></td>
<td>Model 1 Data</td>
<td>Model 2 Data</td>
<td>Model 3 Data</td>
</tr>
<tr>
<td>(within Level-3 units)</td>
<td></td>
<td>0.63</td>
<td>0.66</td>
<td>0.72</td>
</tr>
<tr>
<td>( \frac{\tau_\pi}{(\sigma^2 + \tau_\pi + \tau_\beta)} )</td>
<td></td>
<td>0.10</td>
<td>0.11</td>
<td>0.05</td>
</tr>
<tr>
<td>Proportion of variance between Level-3 units</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \frac{\tau_\beta}{(\sigma^2 + \tau_\pi + \tau_\beta)} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4: Model 3 data neighborhood and city means (number1model3full.dta in stata file)

<table>
<thead>
<tr>
<th></th>
<th>$Y_{ijk}$</th>
<th>$X_{ijk}$</th>
<th>$X_{ijk} - \bar{Y}_{jk}$</th>
<th>$X_{ijk} - \bar{X}_{jk}$</th>
<th>$\bar{Y}<em>{jk} - \bar{X}</em>{jk}$</th>
<th>$\bar{Y}<em>{jk} - \bar{X}</em>{jk}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>neighborhood 1</td>
<td>46.84</td>
<td>18.02</td>
<td>0</td>
<td>-4.93</td>
<td>-5.57</td>
<td>-4.93</td>
</tr>
<tr>
<td>neighborhood 2</td>
<td>65.63</td>
<td>26.49</td>
<td>0</td>
<td>3.55</td>
<td>2.90</td>
<td>3.55</td>
</tr>
<tr>
<td>neighborhood 3</td>
<td>87.88</td>
<td>25.23</td>
<td>0</td>
<td>2.29</td>
<td>1.64</td>
<td>2.29</td>
</tr>
<tr>
<td>neighborhood 4</td>
<td>53.90</td>
<td>25.49</td>
<td>0</td>
<td>2.54</td>
<td>1.90</td>
<td>2.54</td>
</tr>
<tr>
<td>city 1</td>
<td>59.47</td>
<td>23.59</td>
<td>0</td>
<td>0.65</td>
<td>0</td>
<td>0.65</td>
</tr>
<tr>
<td>city 2</td>
<td>69.84</td>
<td>27.24</td>
<td>0</td>
<td>4.30</td>
<td>0</td>
<td>4.30</td>
</tr>
</tbody>
</table>
Table 5: Model 3 Data Bivariate Correlations (number1model3full.dta in stata file)

<table>
<thead>
<tr>
<th>Level-1 Variables</th>
<th>Level-2 Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>$Y_{ijk}$</td>
<td>$X_{ijk}$</td>
</tr>
<tr>
<td>1.00</td>
<td>0.62</td>
</tr>
<tr>
<td>B</td>
<td>$X_{ijk}$</td>
</tr>
<tr>
<td>0.62</td>
<td>1.00</td>
</tr>
<tr>
<td>C</td>
<td>$(X_{ijk} - \bar{X}_{jk})$</td>
</tr>
<tr>
<td>0.45</td>
<td>0.46</td>
</tr>
<tr>
<td>D</td>
<td>$(X_{ijk} - \bar{X}_{..})$</td>
</tr>
<tr>
<td>0.62</td>
<td>1.00</td>
</tr>
<tr>
<td>E</td>
<td>$\bar{X}_{jk}$</td>
</tr>
<tr>
<td>0.47</td>
<td>0.89</td>
</tr>
<tr>
<td>F</td>
<td>$(\bar{X}<em>{jk} - \bar{X}</em>{..})$</td>
</tr>
<tr>
<td>0.35</td>
<td>0.64</td>
</tr>
<tr>
<td>G</td>
<td>$(\bar{X}<em>{jk} - \bar{X}</em>{..})$</td>
</tr>
<tr>
<td>0.47</td>
<td>0.89</td>
</tr>
<tr>
<td>H</td>
<td>$\bar{X}_{..}$</td>
</tr>
<tr>
<td>0.31</td>
<td>0.61</td>
</tr>
</tbody>
</table>
Table 6: Variance-Covariance Matrix for Three-Level Model with Random Intercept and Random Level-1 Slope

<table>
<thead>
<tr>
<th>C*</th>
<th>N*</th>
<th>P*</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>(\tau_\beta + \tau_{a00})</td>
<td>+ 2(\tau_{a01})X</td>
<td>+ (\tau_{a11}X^2)</td>
<td>+ (\sigma^2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>(\tau_\beta + \tau_{a00})</td>
<td>+ (\tau_{a01}(X_{111} + X_{112}))</td>
<td>+ (\tau_{a11}(X_{111})) (X_{112})</td>
<td>+ (\tau_{a11}X^2)</td>
<td>+ (\sigma^2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>(\tau_\beta + \tau_{a00})</td>
<td>+ (\tau_{a01}(X_{111} + X_{112} + X_{113}))</td>
<td>+ (\tau_{a11}(X_{111} + X_{112})) (X_{113})</td>
<td>+ (\tau_{a11}X^2)</td>
<td>+ (\sigma^2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>(\tau_\beta)</td>
<td>+ (\tau_{a00})</td>
<td>+ 2(\tau_{a01})X</td>
<td>+ (\tau_{a11}X^2)</td>
<td>+ (\sigma^2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>(\tau_\beta)</td>
<td>+ (\tau_{a00})</td>
<td>+ (\tau_{a01}(X_{121} + X_{122})) (X_{123})</td>
<td>+ (\tau_{a11}X^2)</td>
<td>+ (\sigma^2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>$\tau_\beta$</td>
<td>$\tau_\beta$</td>
<td>$\tau_\beta$</td>
<td>$\tau_\beta + \tau_{a00}$</td>
<td>$\tau_{a01}(X_{121} + X_{123}) + \tau_{a11}(X_{121})$</td>
<td>$\tau_\beta + \tau_{a00}$</td>
<td>$\tau_{a01}(X_{122} + X_{123}) + \tau_{a11}(X_{122})$</td>
<td>$\tau_\beta + \tau_{a00}$</td>
<td>$\tau_{a01}X + \tau_{a11}X^2 + \sigma^2$</td>
<td>$\tau_\beta$</td>
<td>$\tau_\beta$</td>
<td>$\tau_\beta$</td>
<td>$\tau_\beta + \tau_{a00}$</td>
<td>$\tau_{a01}(X_{231} + X_{232}) + \tau_{a11}(X_{231})$</td>
<td>$\tau_\beta + \tau_{a00}$</td>
<td>$\tau_{a01}(X_{232} + X_{233}) + \tau_{a11}(X_{232})$</td>
<td>$\tau_\beta + \tau_{a00}$ + $2\tau_{a01}X + \tau_{a11}X^2 + \sigma^2$</td>
</tr>
</tbody>
</table>

*C indicates city, N indicates neighborhood, P indicates person."
<table>
<thead>
<tr>
<th>Model</th>
<th>Expectations</th>
<th>Dispersions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>$X_{ijk}$ at Level-1</td>
<td>RAS = CGM ≠ CWC</td>
</tr>
<tr>
<td>Model 2</td>
<td>$X_{ijk}$ at Level-1</td>
<td>RAS = CGM ≠ CWC</td>
</tr>
<tr>
<td>Model 3</td>
<td>$X_{ijk}$ at Level-1</td>
<td>RAS = CGM = CWC</td>
</tr>
</tbody>
</table>

*Table 7: Summary of the Equivalencies*
Table 8: Model 1 parameter estimates averaged across 10 datasets.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Population Value</th>
<th>RAS</th>
<th>CGM</th>
<th>CWC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{000}$</td>
<td>15</td>
<td>15.071</td>
<td>134.675</td>
<td>134.642</td>
</tr>
<tr>
<td>$\gamma_{100}(X_{ijk})$</td>
<td>5</td>
<td>5.000</td>
<td>5.000</td>
<td>5.000</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>12</td>
<td>12.102</td>
<td>12.101</td>
<td>12.100</td>
</tr>
<tr>
<td>$\tau_{\pi_{00}}$</td>
<td>6</td>
<td>5.146</td>
<td>537.820</td>
<td>765.275</td>
</tr>
<tr>
<td>$\tau_{\pi_{11}}$</td>
<td>1</td>
<td>1.018</td>
<td>1.018</td>
<td>1.025</td>
</tr>
<tr>
<td>$\tau_{\pi_{01}}$</td>
<td>-1.23**</td>
<td>-1.049</td>
<td>23.308</td>
<td>23.468</td>
</tr>
<tr>
<td>$\tau_{\rho_{00}}$</td>
<td>3</td>
<td>2.702</td>
<td>2.703</td>
<td>124.240</td>
</tr>
</tbody>
</table>

$\bar{X}_{..} = 23.924$; $\bar{Y}_{..} = 134.642$; ** expected covariance = $\rho * SQRT(\tau_{\pi_{00}}) * SQRT(\tau_{\pi_{11}}) = (-0.50)(2.45)(1) = -1.23$; All coefficients are statistically significant at $p < .02$. 
Table 9: Model 1 RAS and CGM equivalencies using estimates averaged across all ten datasets (Table 8)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimated RAS</th>
<th>Estimated CGM*</th>
<th>RAS in terms of CGM</th>
<th>Calculated RAS from CGM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{000}$</td>
<td>15.071</td>
<td>134.675</td>
<td>$\gamma_{000} = \gamma_{000}^* - \gamma_{100}^* (\bar{X}_{..})$</td>
<td>15.055</td>
</tr>
<tr>
<td>$\gamma_{100}$</td>
<td>5.000</td>
<td>5.000</td>
<td>$\gamma_{100} = \gamma_{100}^*$</td>
<td>5.000</td>
</tr>
<tr>
<td>$(X_{ijk})$</td>
<td>5.000</td>
<td>5.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>12.102</td>
<td>12.101</td>
<td>$\sigma^2 = \sigma_{2*}^2$</td>
<td>12.101</td>
</tr>
<tr>
<td>$\tau_{\pi_{00}}$</td>
<td>5.146</td>
<td>537.820</td>
<td>$\tau_{\pi_{00}} = \tau_{\pi_{00}}^* - 2\tau_{\pi_{01}}^<em>(\bar{X}<em>{..}) + \tau</em>{\pi_{11}}^</em>(\bar{X}_{..})^2$</td>
<td>5.239</td>
</tr>
<tr>
<td>$\tau_{\pi_{11}}$</td>
<td>1.018</td>
<td>1.018</td>
<td>$\tau_{\pi_{11}} = \tau_{\pi_{11}}^*$</td>
<td>1.018</td>
</tr>
<tr>
<td>$\tau_{\pi_{01}}$</td>
<td>-1.049</td>
<td>23.308</td>
<td>$\tau_{\pi_{01}} = \tau_{\pi_{01}}^* - \tau_{\pi_{11}}^*(\bar{X}_{..})$</td>
<td>-1.047</td>
</tr>
<tr>
<td>$\tau_{\beta_{00}}$</td>
<td>2.702</td>
<td>2.703</td>
<td>$\tau_{\beta_{00}} = \tau_{\beta_{00}}^*$</td>
<td>2.703</td>
</tr>
</tbody>
</table>

$\bar{X}_{..} = 23.924$; Deviance for RAS Model 1 run on one dataset with random intercepts only: 43424.11; Deviance for CGM Model 1 run on one dataset with random intercepts only: 43424.11.
Table 10: Model 2 parameter estimates averaged across 10 datasets

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Population Value</th>
<th>RAS</th>
<th>CGM</th>
<th>CWC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{000}$</td>
<td>15</td>
<td>15.325</td>
<td>87.012</td>
<td>86.959</td>
</tr>
<tr>
<td>$\gamma_{010}$ (X_{jk})</td>
<td>-2</td>
<td>-2.016</td>
<td>-2.016</td>
<td>2.969</td>
</tr>
<tr>
<td>$\gamma_{100}$ (X_{ijk})</td>
<td>5</td>
<td>5.006</td>
<td>5.006</td>
<td>5.006</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>12</td>
<td>11.962</td>
<td>11.961</td>
<td>11.974</td>
</tr>
<tr>
<td>$\tau_{\pi_{00}}$</td>
<td>6</td>
<td>8.254</td>
<td>470.404</td>
<td>501.094</td>
</tr>
<tr>
<td>$\tau_{\pi_{11}}$</td>
<td>1</td>
<td>1.006</td>
<td>1.007</td>
<td>1.003</td>
</tr>
<tr>
<td>$\tau_{\pi_{01}}$</td>
<td>-1.23**</td>
<td>-2.446</td>
<td>21.700</td>
<td>21.850</td>
</tr>
<tr>
<td>$\tau_{\beta_{00}}$</td>
<td>3</td>
<td>2.525</td>
<td>2.519</td>
<td>120.169</td>
</tr>
</tbody>
</table>

$\bar{X}_{..} = 23.979$; $\bar{Y}_{..} = 86.959$; ** expected covariance = $\rho * SQRT(\tau_{\pi_{00}}) * SQRT(\tau_{\pi_{11}}) = (-0.50)(2.45)(1) = -1.23$; All coefficients are statistically significant at $p < .01$. 


Table 11: Model 2 RAS and CGM equivalencies using estimates from Table 10

<table>
<thead>
<tr>
<th>Parameter</th>
<th>RAS</th>
<th>CGM*</th>
<th>RAS in terms of CGM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{000}$</td>
<td>15.325</td>
<td>87.012</td>
<td>$\gamma_{000} = \gamma_{000}^{<em>} - \gamma_{010}^{</em>}(\bar{X}<em>{\ldots}) - \gamma</em>{100}^{*}(\bar{X}_{\ldots}) = 15.315$</td>
</tr>
<tr>
<td>$\gamma_{010}$ (\bar{X}_{jk})</td>
<td>-2.016</td>
<td>-2.016</td>
<td>$\gamma_{010} = \gamma_{010}^{*} = -2.016$</td>
</tr>
<tr>
<td>$\gamma_{100}$ (X_{ijk})</td>
<td>5.006</td>
<td>5.006</td>
<td>$\gamma_{100} = \gamma_{100}^{*} = 5.006$</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>11.962</td>
<td>11.961</td>
<td>$\sigma^2 = \sigma^2^{*} = 11.961$</td>
</tr>
<tr>
<td>$\tau_{\pi_{00}}$</td>
<td>8.254</td>
<td>470.404</td>
<td>$\tau_{\pi_{00}} = \tau_{\pi_{00}}^{<em>} - 2\tau_{\pi_{01}}^{</em>}(\bar{X}<em>{\ldots}) + \tau</em>{\pi_{11}}^{*}(\bar{X}_{\ldots}^{2}) = 8.733$</td>
</tr>
<tr>
<td>$\tau_{\pi_{11}}$</td>
<td>1.006</td>
<td>1.007</td>
<td>$\tau_{\pi_{11}} = \tau_{\pi_{11}}^{*} = 1.007$</td>
</tr>
<tr>
<td>$\tau_{\pi_{01}}$</td>
<td>-2.446</td>
<td>21.700</td>
<td>$\tau_{\pi_{01}} = \tau_{\pi_{01}}^{<em>} - \tau_{\pi_{11}}^{</em>}(\bar{X}_{\ldots}) = -2.447$</td>
</tr>
<tr>
<td>$\tau_{\beta_{00}}$</td>
<td>2.525</td>
<td>2.519</td>
<td>$\tau_{\beta_{00}} = \tau_{\beta_{00}}^{*} = 2.519$</td>
</tr>
</tbody>
</table>

$\bar{X}_{\ldots} = 23.979$; Deviance for RAS Model 2 run on one dataset with random intercepts only: 42130.25; Deviance for CGM Model 2 run on one dataset with random intercepts only: 42130.25.
Table 12: Model 3 parameter estimates averaged across 10 datasets

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Population Value</th>
<th>RAS</th>
<th>CGM</th>
<th>CWC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{000}$</td>
<td>15</td>
<td>15.321</td>
<td>62.824</td>
<td>15.379</td>
</tr>
<tr>
<td>$\gamma_{001}$ ($\bar{X}_{..k}$)</td>
<td>-1</td>
<td>-1.028</td>
<td>-1.028</td>
<td>1.995</td>
</tr>
<tr>
<td>$\gamma_{010}$ ($\bar{X}_{..j\bar{k}}$)</td>
<td>-2</td>
<td>-1.987</td>
<td>-1.987</td>
<td>3.052</td>
</tr>
<tr>
<td>$\gamma_{100}$ ($\bar{X}_{ij\bar{k}}$)</td>
<td>5</td>
<td>5.008</td>
<td>5.008</td>
<td>5.008</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>12</td>
<td>11.984</td>
<td>11.984</td>
<td>11.977</td>
</tr>
<tr>
<td>$\tau_{\pi_{00}}$</td>
<td>6</td>
<td>7.115</td>
<td>544.996</td>
<td>563.172</td>
</tr>
<tr>
<td>$\tau_{\pi_{11}}$</td>
<td>1</td>
<td>1.049</td>
<td>1.049</td>
<td>1.049</td>
</tr>
<tr>
<td>$\tau_{\pi_{01}}$</td>
<td>-1.23**</td>
<td>-1.248</td>
<td>23.776</td>
<td>23.624</td>
</tr>
<tr>
<td>$\tau_{\beta_{00}}$</td>
<td>3</td>
<td>1.886</td>
<td>1.885</td>
<td>1.388</td>
</tr>
</tbody>
</table>

$\bar{X}_{..} = 23.839; \bar{Y}_{..} = 62.942; *$ indicates estimate of $\gamma_{000}$ if $\bar{X}_{..k}$ CGM instead of RAS;

** expected covariance = $\rho \times \text{SQRT}(\tau_{\pi_{00}}) \times \text{SQRT}(\tau_{\pi_{11}}) = (-0.50)(2.45)(1) = -1.23$; With the exception of $\tau_{\beta_{00}}$ under CWC ($p = .122$), all coefficients are statistically significant at $p < .02$. 
Table 13: Model 3 RAS, CGM and CWC equivalencies using estimates from Table 12

<table>
<thead>
<tr>
<th>Parameter</th>
<th>RAS</th>
<th>CGM*</th>
<th>CWC**</th>
<th>RAS in terms of CGM</th>
<th>RAS in terms of CWC</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_{000} )</td>
<td>15.32</td>
<td>62.82</td>
<td>15.38</td>
<td>( \gamma_{000} = \gamma_{000}^* - \gamma_{010}^<em>(\bar{X}<em>{...}) - \gamma</em>{001}^</em>(\bar{X}_{...}) = 15.31 )</td>
<td>( \gamma_{000} = \gamma_{000}^{**} = 15.38 )</td>
</tr>
<tr>
<td>( \gamma_{001} (\bar{X}_{...}) )</td>
<td>-1.03</td>
<td>-1.03</td>
<td>2.00</td>
<td>( \gamma_{001} = \gamma_{001}^* = -1.03 )</td>
<td>( \gamma_{001} = \gamma_{001}^{<strong>} - \gamma_{010}^{</strong>} = -1.06 )</td>
</tr>
<tr>
<td>( \gamma_{010} (\bar{X}_{...}) )</td>
<td>-1.99</td>
<td>-1.99</td>
<td>3.05</td>
<td>( \gamma_{010} = \gamma_{010}^* = -1.99 )</td>
<td>( \gamma_{010} = \gamma_{010}^{<strong>} - \gamma_{100}^{</strong>} = -1.96 )</td>
</tr>
<tr>
<td>( \gamma_{100} (X_{ijk}) )</td>
<td>5.01</td>
<td>5.01</td>
<td>5.01</td>
<td>( \gamma_{100} = \gamma_{100}^* = 5.01 )</td>
<td>( \gamma_{100} = \gamma_{100}^{**} = 5.01 )</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>11.98</td>
<td>11.98</td>
<td>11.98</td>
<td>( \sigma^2 = \sigma^2^{**} = 11.98 )</td>
<td></td>
</tr>
<tr>
<td>( \tau_{\pi_{00}} )</td>
<td>7.12</td>
<td>545.00</td>
<td>563.17</td>
<td>( \tau_{\pi_{00}} = \tau_{\pi_{00}}^* - 2\tau_{\pi_{01}}^<em>(\bar{X}<em>{...}) + \tau</em>{\pi_{11}}^</em>(\bar{X}_{...}^2) = 7.55 )</td>
<td></td>
</tr>
<tr>
<td>( \tau_{\pi_{11}} )</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
<td>( \tau_{\pi_{11}} = \tau_{\pi_{11}}^* = 1.05 )</td>
<td></td>
</tr>
<tr>
<td>( \tau_{\pi_{01}} )</td>
<td>-1.25</td>
<td>23.78</td>
<td>23.62</td>
<td>( \tau_{\pi_{01}} = \tau_{\pi_{01}}^* - \tau_{\pi_{11}}^*(\bar{X}_{...}) = -1.23 )</td>
<td></td>
</tr>
<tr>
<td>( \tau_{\beta_{00}} )</td>
<td>1.89</td>
<td>1.88</td>
<td>1.39</td>
<td>( \tau_{\beta_{00}} = \tau_{\beta_{00}}^* = 1.88 )</td>
<td></td>
</tr>
</tbody>
</table>

\( \bar{X}_{...} = 23.84; \) Deviance for RAS Model 3 run on one dataset with random intercepts only: 41940.73; Deviance for CGM Model 3 run on one dataset with random intercepts only: 41940.74; Deviance for CWC Model 3 run on one dataset with random intercepts only: 41940.73.
Table 14: Model 1 estimates resulting from analysis on Model 3 data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Population Value</th>
<th>RAS</th>
<th>CGM</th>
<th>CWC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{000}$</td>
<td>15</td>
<td>-55.411</td>
<td>63.345</td>
<td>62.942</td>
</tr>
<tr>
<td>$\gamma_{100}$ \ ($X_{ijk}$)</td>
<td>5</td>
<td>4.981</td>
<td>4.981</td>
<td>5.009</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>12</td>
<td>11.970</td>
<td>11.970</td>
<td>11.976</td>
</tr>
<tr>
<td>$\tau_{\pi_{00}}$</td>
<td>6</td>
<td>63.979</td>
<td>594.489</td>
<td>707.641</td>
</tr>
<tr>
<td>$\tau_{\pi_{11}}$</td>
<td>1</td>
<td>1.048</td>
<td>1.048</td>
<td>1.051</td>
</tr>
<tr>
<td>$\tau_{\pi_{01}}$</td>
<td>-1.23</td>
<td>-1.391</td>
<td>23.621</td>
<td>23.968</td>
</tr>
<tr>
<td>$\tau_{\beta_{00}}$</td>
<td>3</td>
<td>121.157</td>
<td>121.157</td>
<td>45.403</td>
</tr>
</tbody>
</table>

$\bar{X}_{\text{...}} = 23.84$. 
Figure 1: Two-level hierarchical data structure.
Figure 2: Example of longitudinal data represented in two-level model.
Figure 3: Three-level hierarchical data structure.
Figure 4: Intercepts under random slope and fixed slope models.
Appendix A

Model 1

data number1modellla;
/* CREATE Xijk AND RESIDUAL TERMS */

seed1 = 2168;
seed2 = 81016;
seed3 = 36172;
seed4 = 2149;
seed5 = 83422;
seed6 = 20047;
seed7 = 21034;

taubeta00 = 3;
taupi00 = 6;
sigmasq = 12;
xlevel3var = 5;
xlevel2var = 8;
xlevel1var = 12;
tauill = 1;
meanx = 24;

ncity = 30;

/* 30 cities of 15 hoods each = 450 neighborhoods */
nhood = 15;

/* 450 neighborhoods of 15 persons = 6750 people */

nperson = 15;
do city=1 to ncity;
    XU00k = rannor(seed1)*sqrt(xlevel3var); /*create L3 X residual*/
    U00k = rannor(seed2)*sqrt(taubeta00);  /*create L3 Y residual*/

do hood=1 to nhood;
    XR0jk = rannor(seed3)*sqrt(xlevel2var); /*create L2 X residual*/
    tempR0jk = rannor(seed4);  /*create temp L2 residual for Y*/
    tempR1jk = rannor(seed5); /*create temp L2 residual for slope*/
    temp2R1jk = tempR0jk*-0.50 + tempR1jk*(sqrt(1-.25));
    R0jk = tempR0jk*sqrt(taupi00); /*rescale for desired variances*/
    R1jk = temp2R1jk*sqrt(tauill) ;

do person=1 to nperson;
    XEijk = rannor(seed6)*sqrt(xlevel1var); /* create L1 X residual*/
    YEijk = rannor(seed7)*sqrt(sigmasq);  /* create L1 Y residual*/
    Xijk = meanx + XEijk + XR0jk + XU00k ;  /* create x from mean(x) + person, hood, city errors */

cityid = city;
hoodid = hood + (city-1)*nhood; /* creates unique hood id */
personid = person + (hood-1)*nperson + (city-1)*nhood*nperson;
/* creates unique person id */

output;
end;
end;
end;
keep cityid hoodid personid xijk xeijk xr0jk r1jk xu00k yeijk r0jk r1jk u00k;
run;

proc corr cov data=number1modella;
var r0jk r1jk u00k;
run;

/* SORT BY NEIGHBORHOOD */
proc sort data=number1modella;
by hoodid;
run;

/* CREATE Xjk */
data number1modellb;
  nhood = 30;
  nperson = 15;
  do until(last.hoodid);
    set number1modella;
    by hoodid;
    sum_Xijk=sum(sum_Xijk,Xijk,0); /* creates sum(Xijk) for each hood */
  end;
  Xjk=sum Xijk/nperson; /* creates mean(Xijk) for each hood */
  do until(last.hoodid);
    set number1modella;
    by hoodid;
    output;
  end;
keep cityid hoodid personid xijk xjk yeijk r0jk r1jk u00k;
run;

/* SORT BY CITY */
proc sort data=number1modellb;
by cityid;
run;

/* CREATE Xk */
data number1modellc;
  nhood = 30;
  nperson = 15;
  do until(last.cityid);
    set number1modellb;
    by cityid;
    sum_Xijk=sum(sum_Xijk,Xijk,0); /* creates sum(Xijk) for each city */
  end;
  Xk=sum Xijk/(nperson*nhood); /* creates mean(Xijk) for each city */
  do until(last.cityid);
    set number1modellb;
    by cityid;
    output;
  end;
keep cityid hoodid personid xijk xjk xk yeijk r0jk r1jk u00k;
run;

/* CREATE FINAL DATASET */
data number1modell;
set number1model1c;

gamma000 = 15; /* grand intercept */
gamma100 = 5; /* (Xijk) slope coefficient */

Yijk = gamma000 + gamma100*Xijk + yEijk + R0jk + R1jk*Xijk + U00k;

output;
keep cityid hoodid personid yijk xijk xjk xk yeijk r0jk r1jk u00k;
run;
Model 2

data number1model2a;
/* CREATE Xijk AND RESIDUAL TERMS */

seed1 = 22556;
seed2 = 3693;
seed3 = 17181;
seed4 = 29039;
seed5 = 62838;
seed6 = 89538;
seed7 = 48546;

taubeta00 = 3;
taupi00 = 6;
sigmasq = 12;
xlevel3var = 12;
xlevel2var = 14;
xlevel1var = 8;
taupill = 1;
meanx = 24;

ncity = 30;
nhood = 15;
nperson = 15;

do city=1 to ncity;
   XU00k = rannor(seed1)*sqrt(xlevel3var);
   U00k = rannor(seed2)*sqrt(taubeta00);
   do hood=1 to nhood;
      XR0jk = rannor(seed3)*sqrt(xlevel2var);
      tempR0jk = rannor(seed4);
      tempR1jk = rannor(seed5);
      temp2R1jk = tempR0jk*-.93 + tempR1jk*(sqrt(1-.86));
      R0jk = tempR0jk*sqrt(taupi00);
      R1jk = temp2R1jk*sqrt(taupill);
   end;
   do person=1 to nperson;
      XEijk = rannor(seed6)*sqrt(xlevel1var);
      YEijk = rannor(seed7)*sqrt(sigmasq);
      Xijk = meanx + XEijk + XR0jk + XU00k;
   end;
   cityid = city;
   hoodid = hood + (city-1)*nhood;
   personid = person + (hood-1)*nperson + (city-1)*nhood*nperson;
   output;
end;
end;

keep cityid hoodid personid xijk xeijk xr0jk r1jk xu00k yeijk r0jk r1jk u00k;
run;
**proc corr cov data=number1model2a;**

**var r0jk r1jk u00k;**

**run;**

/* SORT BY NEIGHBORHOOD */
**proc sort data=number1model2a;**
**by hoodid;**

**run;**

/* CREATE Xjk */
**data number1model2b ;**
*nhood = 15;**
nperson = 15;
**do until(last.hoodid);**
**set number1model2a;**
**by hoodid;**
sum_Xijk=sum(sum_Xijk,Xijk,0);
**end;**
Xjk=sum_Xijk/nperson;
**do until(last.hoodid);**
**set number1model2a;**
**by hoodid;**
**output;**
end;

keep cityid hoodid personid xijk xjk xeijk xr0jk xu00k yeijk r0jk r1jk u00k;
**run;**

/* SORT BY CITY */
**proc sort data=number1model2b;**
**by cityid;**

**run;**

/* CREATE Xk */
**data number1model2c ;**
*nhood = 15;**
nperson = 15;
**do until(last.cityid);**
**set number1model2b;**
**by cityid;**
sum_Xijk=sum(sum_Xijk,Xijk,0);
**end;**
Xk=sum_Xijk/(nperson*nhood);
**do until(last.cityid);**
**set number1model2b;**
**by cityid;**
**output;**
end;

keep cityid hoodid personid xijk xjk xk xeijk xr0jk xu00k yeijk r0jk r1jk u00k;
**run;**

/* CREATE FINAL DATASET */
**data number1model2;**
**set number1model2c;**
\[
\begin{align*}
\gamma_{000} &= 15; \quad /* \text{grand \"intercept\" */} \\
\gamma_{100} &= 5; \quad /* \text{(Xijk) slope coefficient */} \\
\gamma_{010} &= -2; \quad /* \text{(Xjk) slope coefficient */} \\

Y_{ijk} &= \gamma_{000} + \gamma_{100}X_{ijk} + \gamma_{010}X_{jk} + \eta_{Eijk} + R_{0jk} + \\
& \quad R_{1jk}X_{ijk} + U_{00k}; \\
\end{align*}
\]

\textbf{output;} \\
\textbf{keep} cityid hoodid personid yijk xijk xjk xeijk xr0jk xu00k yeijk \\
\textbf{r0jk r1jk u00k ; \\
\textbf{run;}} \\

\textbf{Model 3}

data number1model3a; \\
/* \text{CREATE Xijk AND RESIDUAL TERMS */} \\
seed1 = 47665; \\
seed2 = 83892; \\
seed3 = 1227; \\
seed4 = 61395; \\
seed5 = 47793; \\
seed6 = 57434; \\
seed7 = 69101; \\
\[\begin{align*}
\tau_{\beta00} &= 3; \\
\tau_{p00} &= 6; \\
\sigma_{\text{sq}} &= 12; \\
\text{xlevel3var} &= 12; \\
\text{xlevel2var} &= 14; \\
\text{xlevel1var} &= 8; \\
\tau_{p11} &= 1; \\
\text{meanx} &= 24; \\
\end{align*}\]

ncity = 30; \\
nhood = 15; \\
nperson = 15; \\
do city=1 to ncity; \\
XU00k = rannor(seed1)*sqrt(xlevel3var); \\
U00k = rannor(seed2)*sqrt(taubeta00); \\
do hoo0d=1 to nhood; \\
XR0jk = rannor(seed3)*sqrt(xlevel2var); \\
tempR0jk = rannor(seed4); \\
tempR1jk = rannor(seed5); \\
temp2R1jk = tempR0jk*-.50 + tempR1jk*(sqrt(1-.25)); \\
R0jk = tempR0jk*sqrt(tau00p); \\
R1jk = temp2R1jk*sqrt(tau011); \\
do person=1 to nperson; \\
XEijk = rannor(seed6)*sqrt(xlevel1var); \\
YEijk = rannor(seed7)*sqrt(sigmasq);
Xijk = meanx + XEijk + XR0jk + XU00k ;

cityid = city;
hoodid = hood + (city-1)*nhood;
personid = person + (hood-1)*nperson + (city-1)*nhood*nperson;

output;
end;
end;
end;

keep cityid hoodid personid xijk xeijk xr0jk xu00k yeijk r0jk r1jk u00k;
run;

/* SORT BY NEIGHBORHOOD */
proc sort data=number1model3a;
by hoodid;
run;

/* CREATE Xjk */
data number1model3b;
  nhood = 15;
  nperson = 15;
  do until(last.hoodid);
    set number1model3a;
    by hoodid;
    sum_Xijk=sum(sum_Xijk,Xijk,0);
  end;
  Xjk=sum_Xijk/nperson;
  do until(last.hoodid);
    set number1model3a;
    by hoodid;
    output;
  end;

keep cityid hoodid personid xijk xjk xeijk xr0jk xu00k yeijk r0jk r1jk u00k;
run;

/* SORT BY CITY */
proc sort data=number1model3b;
by cityid;
run;

/* CREATE Xk */
data number1model3c;
  nhood = 15;
  nperson = 15;
  do until(last.cityid);
    set number1model3b;
    by cityid;
    sum_Xijk=sum(sum_Xijk,Xijk,0);
  end;
  Xk=sum_Xijk/(nperson*nhood);
  do until(last.cityid);
    set number1model3b;
  end;
by cityid;
output;
end;
keep cityid hoodid personid xijk xjk xx xrijk xu00k yeijk r0jk r1jk u00k;
run;

/* CREATE FINAL DATASET */
data number1model3;
set number1model3c;

gamma000 = 15; /* grand "intercept" */
gamma100 = 5; /* (Xijk) slope coefficient */
gamma010 = -2; /* (Xjk) slope coefficient */
gamma001 = -1; /* (Xk) slope coefficient */

Yijk = gamma000 + gamma100*Xijk + gamma010*Xjk + gamma001*Xk + yeijk + R0jk + R1jk*Xijk + U00k;

output;
keep cityid hoodid personid yijk xijk xjk xx xrijk xu00k yeijk r0jk r1jk u00k;
run;
Appendix B

Equivalencies for Model 1

RAS Model 1:

\[ Y_{ijk} = \gamma_{000} + \gamma_{100}X_{ijk} + e_{ijk} + r_{0jk} + r_{1jk}X_{ijk} + u_{00k} \]

CGM Model 1:

\[ Y_{ijk} = \gamma_{000}^* + \gamma_{100}^*(X_{ijk} - \bar{X}_{..}) + e_{ijk}^* + r_{0jk}^* + r_{1jk}^*(X_{ijk} - \bar{X}_{..}) + u_{00k}^* \]

CWC Model 1:

\[ Y_{ijk} = \gamma_{000}^{**} + \gamma_{100}^{**}(X_{ijk} - \bar{X}_{..}) + e_{ijk}^{**} + r_{0jk}^{**} + r_{1jk}^{**}(X_{ijk} - \bar{X}_{..}) + u_{00k}^{**} \]

Fixed Effects (RAS = CGM)

\[ \gamma_{000} + \gamma_{100}(X_{ijk}) = \gamma_{000}^* + \gamma_{100}^*(X_{ijk} - \bar{X}_{..}) \]
\[ \gamma_{000} + \gamma_{100}(X_{ijk}) = \gamma_{000}^* + \gamma_{100}^*(X_{ijk}) - \gamma_{100}^*(\bar{X}_{..}) \]
\[ \gamma_{000} + \gamma_{100}(X_{ijk}) = [\gamma_{000}^* - \gamma_{100}^*(\bar{X}_{..})] + \gamma_{100}^*(X_{ijk}) \]
\[ \gamma_{000} = \gamma_{000}^* - \gamma_{100}^*(\bar{X}_{..}) \]
\[ \gamma_{100} = \gamma_{100}^* \]

Fixed Effects (RAS = CWC)

\[ \gamma_{000} + \gamma_{100}(X_{ijk}) = \gamma_{000}^{**} + \gamma_{100}^{**}(X_{ijk} - \bar{X}_{..}) \]
\[ \gamma_{000} + \gamma_{100}(X_{ijk}) = \gamma_{000}^{**} + \gamma_{100}^{**}(X_{ijk}) - \gamma_{100}^{**}(\bar{X}_{..}) \]

The problem here is the \( \bar{X}_{..} \) term. There is no counterpart in the RAS or CGM model, thus equivalency cannot be determined.
Random Effects (RAS=CGM)

\[ \text{Var}(Y_{ijk}) = E[Y - E(Y)]^2 = E[e_{ijk} + r_{0jk} + r_{1jk}X_{ijk} + u_{00k}]^2 \]
\[ = E[e_{ijk}^2 + r_{0jk}^2 + 2r_{0jk}r_{1jk}X_{ijk} + r_{1jk}^2(X_{ijk})^2 + u_{00k}^2] \]
\[ = \sigma^2 + \tau_{\pi00} + 2\tau_{\pi01}(X_{ijk}) + \tau_{\pi11}(X_{ijk})^2 + \tau_{\beta00} \]

\[ \text{Var}(Y_{ijk}) \]

RAS:

\[ \sigma^2 + \tau_{\pi00} + 2\tau_{\pi01}(X_{ijk}) + \tau_{\pi11}(X_{ijk})^2 + \tau_{\beta00} \]

CGM:

\[ \sigma^{2*} + \tau_{\pi00}^{*} + 2\tau_{\pi01}^{*}(X_{ijk} - \bar{X}_{...}) + \tau_{\pi11}^{*}(X_{ijk} - \bar{X}_{...})^2 + \tau_{\beta00}^{*} \]
\[ = \sigma^{2*} + \tau_{\pi00}^{*} + 2\tau_{\pi01}^{*}X_{ijk} - 2\tau_{\pi01}^{*}\bar{X}_{...} + \tau_{\pi11}^{*}(X_{ijk})^2 - 2\tau_{\pi11}^{*}X_{ijk}\bar{X}_{...} \]
\[ + \tau_{\pi11}^{*}(\bar{X}_{...})^2 + \tau_{\beta00}^{*} \]

CWC:

\[ \sigma^{2**} + \tau_{\pi00}^{**} + 2\tau_{\pi01}^{**}(X_{ijk} - \bar{X}_{jk}) + \tau_{\pi11}^{**}(X_{ijk} - \bar{X}_{jk})^2 + \tau_{\beta00}^{**} \]
\[ = \sigma^{2**} + \tau_{\pi00}^{**} + 2\tau_{\pi01}^{**}X_{ijk} - 2\tau_{\pi01}^{**}\bar{X}_{jk} + \tau_{\pi11}^{**}(X_{ijk})^2 - 2\tau_{\pi11}^{**}X_{ijk}\bar{X}_{jk} \]
\[ + \tau_{\pi11}^{**}(\bar{X}_{jk})^2 + \tau_{\beta00}^{**} \]
RAS = CGM

\[ \sigma^2 + \tau_{\pi_{00}} + 2\tau_{\pi_{01}}(X_{ijk}) + \tau_{\pi_{11}}(X_{ijk})^2 + \tau_{\beta_{00}} \]

\[ = \sigma^{2*} + \left[ \tau^{*}_{\pi_{00}} - 2\tau^{*}_{\pi_{01}}\bar{X}_{...} + \tau^{*}_{\pi_{11}}(\bar{X}_{...}^2) \right] + \left[ 2\tau^{*}_{\pi_{01}}X_{ijk} - 2\tau^{*}_{\pi_{11}}X_{ijk}\bar{X}_{...} \right] \]

\[ + \tau^{*}_{\pi_{11}}(X_{ijk})^2 + \tau^{*}_{\beta_{00}} \]

\[ \sigma^2 = \sigma^{2*} \]

\[ \tau_{\pi_{00}} = \tau^{*}_{\pi_{00}} - 2\tau^{*}_{\pi_{01}}(\bar{X}_{...}) + \tau^{*}_{\pi_{11}}(\bar{X}_{...}^2) \]

\[ \tau_{\pi_{01}} = \tau^{*}_{\pi_{01}} - \tau^{*}_{\pi_{11}}(\bar{X}_{...}) \]

\[ \tau_{\pi_{11}} = \tau^{*}_{\pi_{11}} \]

\[ \tau_{\beta_{00}} = \tau^{*}_{\beta_{00}} \]

RAS = CWC

\[ \sigma^2 + \tau_{\pi_{00}} + 2\tau_{\pi_{01}}(X_{ijk}) + \tau_{\pi_{11}}(X_{ijk})^2 + \tau_{\beta_{00}} = \sigma^{2**} + \tau^{**}_{\pi_{00}} + 2\tau^{**}_{\pi_{01}}X_{ijk} - 2\tau^{**}_{\pi_{01}}\bar{X}_{...} \]

\[ + \tau^{**}_{\pi_{11}}(X_{ijk})^2 - 2\tau^{**}_{\pi_{11}}X_{ijk}\bar{X}_{...} + \tau^{**}_{\pi_{11}}(\bar{X}_{...}^2) + \tau^{**}_{\beta_{00}} \]

Because the CWC variance equation includes \( \bar{X}_{...} \) and \( (\bar{X}_{...})^2 \) terms, but the RAS (nor CGM) equation does not, these two variances cannot be equivalent.
Appendix C

Equivalencies for Model 2

RAS Model 2:

\[
Y_{ijk} = \gamma_{000} + \gamma_{010}(\bar{X}_{jk}) + \gamma_{100}(X_{ijk}) + e_{ijk} + \tau_{0jk} + \tau_{1jk}(X_{ijk}) + u_{00k}
\]

CGM Model 2:

\[
Y_{ijk} = \gamma^*_{000} + \gamma^*_{010}(\bar{X}_{jk} - \bar{X}_-) + \gamma^*_{100}(X_{ijk} - \bar{X}_-) + e^*_{ijk} + \tau^*_{0jk} + \tau^*_{1jk}(X_{ijk} - \bar{X}_-) + u^*_{00k}
\]

CWC Model 2:

\[
Y_{ijk} = \gamma^{**}_{000} + \gamma^{**}_{010}(\bar{X}_{jk} - \bar{X}_-) + \gamma^{**}_{100}(X_{ijk} - \bar{X}_-) + e^{**}_{ijk} + \tau^{**}_{0jk} + \tau^{**}_{1jk}(X_{ijk} - \bar{X}_-) + u^{**}_{00k}
\]

Fixed Effects (RAS = CGM)

\[
\gamma_{000} + \gamma_{010}(\bar{X}_{jk}) + \gamma_{100}(X_{ijk}) \\
= \gamma^*_{000} + \gamma^*_{010}(\bar{X}_{jk} - \bar{X}_-) + \gamma^*_{100}(X_{ijk} - \bar{X}_-) \\
= \gamma^*_{000} + \gamma^*_{010}(\bar{X}_{jk}) - \gamma^*_{010}(\bar{X}_-) + \gamma^*_{100}(X_{ijk}) - \gamma^*_{100}(\bar{X}_-) \\
= [\gamma^*_{000} - \gamma^*_{010}(\bar{X}_-) - \gamma^*_{100}(\bar{X}_-)] + + \gamma^*_{010}(\bar{X}_{jk}) + \gamma^*_{100}(X_{ijk}) \\
\gamma_{000} = \gamma^*_{000} - \gamma^*_{010}(\bar{X}_-) - \gamma^*_{100}(\bar{X}_-) \\
\gamma_{010} = \gamma^*_{010} \\
\gamma_{100} = \gamma^*_{100}
\]

Fixed Effects (RAS = CWC)

\[
\gamma_{000} + \gamma_{010}(\bar{X}_{jk}) + \gamma_{100}(X_{ijk}) 
\]
\[ \gamma_{000} + \gamma_{010}^\star (\bar{X}_{jk} - \bar{X}_{.k}) + \gamma_{100}^\star (X_{ijk} - \bar{X}_{jk}) \]

\[ = \gamma_{000} + \gamma_{010}^\star (\bar{X}_{jk}) - \gamma_{010}^\star (\bar{X}_{..k}) + \gamma_{100}^\star (X_{ijk}) - \gamma_{100}^\star (\bar{X}_{jk}) \]

The problem here is the \( X_{.k} \) term in the CWC model. There is no counterpart in the RAS or CGM model, thus equivalency cannot be determined.

**Random Effects**

\[
\text{Var}(Y_{ijk}) = E[Y - E(Y)]^2 = E[e_{ijk} + r_{0jk} + r_{1jk}(X_{ijk}) + u_{00k}]^2
\]

\[
= E\left[e_{ijk}^2 + r_{0jk}^2 + 2r_{0jk}r_{1jk}(X_{ijk}) + r_{1jk}^2(X_{ijk})^2 + u_{00k}^2\right]
\]

\[
= \sigma^2 + \tau_{\pi_{00}} + 2\tau_{\pi_{01}}(X_{ijk}) + \tau_{\pi_{11}}(X_{ijk})^2 + \tau_{\beta_{00}}
\]

**RAS:**

\[
\sigma^2 + \tau_{\pi_{00}} + 2\tau_{\pi_{01}}(X_{ijk}) + \tau_{\pi_{11}}(X_{ijk})^2 + \tau_{\beta_{00}}
\]

**CGM:**

\[
\sigma^{*2} + \tau_{\pi_{00}}^* + 2\tau_{\pi_{01}}^*(X_{ijk} - \bar{X}_{..}) + \tau_{\pi_{11}}^*(X_{ijk} - \bar{X}_{..})^2 + \tau_{\beta_{00}}^*
\]

\[= \sigma^{*2} + \tau_{\pi_{00}}^* + 2\tau_{\pi_{01}}^*X_{ijk} - 2\tau_{\pi_{01}}^*X_{ijk} - \bar{X}_{..} + \tau_{\pi_{11}}^*(X_{ijk})^2 - 2\tau_{\pi_{11}}^*X_{ijk}\bar{X}_{..}
\]

\[+ \tau_{\pi_{11}}^*(\bar{X}_{..})^2 + \tau_{\beta_{00}}^*
\]

**CWC:**

\[
\sigma^{2*2} + \tau_{\pi_{00}}^{*2} + 2\tau_{\pi_{01}}^{*2}(X_{ijk} - \bar{X}_{.j.k}) + \tau_{\pi_{11}}^{*2}(X_{ijk} - \bar{X}_{.j.k})^2 + \tau_{\beta_{00}}^{*2}
\]

\[= \sigma^{2*2} + \tau_{\pi_{00}}^{*2} + 2\tau_{\pi_{01}}^{*2}X_{ijk} - 2\tau_{\pi_{01}}^{*2}X_{ijk} - \bar{X}_{.j.k} + \tau_{\pi_{11}}^{*2}(X_{ijk})^2 - 2\tau_{\pi_{11}}^{*2}X_{ijk}\bar{X}_{.j.k}
\]

\[+ \tau_{\pi_{11}}^{*2}X_{.j.k}^2 + \tau_{\beta_{00}}^{*2}
\]
Note that the variance terms in Model 2 are identical to Model 1. Thus, the random effect equivalencies between RAS and CGM are the same, and neither RAS nor CGM are equivalent to CWC in random effects.

\[
\begin{align*}
\text{RAS} &= \text{CGM} \\
\sigma^2 &= \sigma^{2*} \\
\tau_{\pi_{00}} &= \tau^*_{\pi_{00}} - 2\tau^*_{\pi_{01}} (\bar{X}_\ldots) + \tau^*_{\pi_{11}} (\bar{X}^2) \\
\tau_{\pi_{01}} &= \tau^*_{\pi_{01}} - \tau^*_{\pi_{11}} (\bar{X}_\ldots) \\
\tau_{\pi_{11}} &= \tau^*_{\pi_{11}} \\
\tau_{\beta_{00}} &= \tau^*_{\beta_{00}}
\end{align*}
\]
Appendix D

Equivalencies for Model 3

RAS Model 3:

\[
Y_{ijk} = \gamma_{000} + \gamma_{001}(\bar{X}_{..k}) + \gamma_{010}(\bar{X}_{..jk}) + \gamma_{100}(X_{ijk}) + e_{ijk} + r_{0jk} + r_{1jk}(X_{ijk}) + u_{00k}
\]

CGM Model 3:

\[
Y_{ijk} = \gamma_{000}^* + \gamma_{001}^*(\bar{X}_{..k} - \bar{X}_{...}) + \gamma_{010}^*(\bar{X}_{..jk} - \bar{X}_{...}) + \gamma_{100}^*(X_{ijk} - \bar{X}_{...}) + e_{ijk}^* + r_{0jk}^*
\]

+ \gamma_{100}^*(X_{ijk} - \bar{X}_{...}) + u_{00k}^*

CWC Model 3:

\[
Y_{ijk} = \gamma_{000}^{**} + \gamma_{001}^{**}(\bar{X}_{..k}) + \gamma_{010}^{**}(\bar{X}_{..jk}) + \gamma_{100}^{**}(X_{ijk}) + e_{ijk}^{**} + r_{0jk}^{**}
\]

+ \gamma_{100}^{**}(X_{ijk} - \bar{X}_{...}) + u_{00k}^{**}

Fixed Effects (RAS = CGM)

\[
\gamma_{000} + \gamma_{001}(\bar{X}_{..k}) + \gamma_{010}(\bar{X}_{..jk}) + \gamma_{100}(X_{ijk})
\]

\[
= \gamma_{000} + \gamma_{001}^*(\bar{X}_{..k} - \bar{X}_{...}) + \gamma_{010}^*(\bar{X}_{..jk} - \bar{X}_{...}) + \gamma_{100}^*(X_{ijk} - \bar{X}_{...})
\]

\[
= \gamma_{000} + \gamma_{001}^*(\bar{X}_{..k}) + \gamma_{010}^*(\bar{X}_{..jk}) - \gamma_{010}^*(\bar{X}_{...}) + \gamma_{100}^*(X_{ijk}) - \gamma_{100}^*(\bar{X}_{...})
\]

\[
= [\gamma_{000} - \gamma_{001}^*(\bar{X}_{...}) - \gamma_{010}^*(\bar{X}_{...}) - \gamma_{100}^*(\bar{X}_{...})] + \gamma_{001}^*(\bar{X}_{..k}) + \gamma_{010}^*(\bar{X}_{..jk}) + \gamma_{100}^*(X_{ijk})
\]

\[
\gamma_{000} = \gamma_{000}^* - \gamma_{001}^*(\bar{X}_{...}) - \gamma_{010}^*(\bar{X}_{...}) - \gamma_{100}^*(\bar{X}_{...})
\]

\[
\gamma_{010} = \gamma_{010}^*
\]

\[
\gamma_{100} = \gamma_{100}^*
\]
\[ \gamma_{001} = \gamma^{*}_{001} \]

**Fixed Effects (RAS = CWC)**

\[
\gamma_{000} + \gamma_{001}(\bar{x}_{..k}) + \gamma_{010}(\bar{x}_{jk}) + \gamma_{100}(x_{ijk})
\]

\[
= \gamma_{000}^{**} + \gamma_{001}^{**}(\bar{x}_{..k}) + \gamma_{010}^{**}(\bar{x}_{..k}) - \gamma_{100}^{**}(\bar{x}_{..k}) + \gamma_{100}^{**}(x_{ijk})
\]

\[
= \gamma_{000}^{**} + [\gamma_{001}^{**}(\bar{x}_{..k}) - \gamma_{010}^{**}(\bar{x}_{..k})] + [\gamma_{010}^{**}(\bar{x}_{..k}) - \gamma_{100}^{**}(\bar{x}_{..k})] + \gamma_{100}^{**}(x_{ijk})
\]

In Model 3, the presence of the cluster means in the centered model is not a problem because the cluster means also appear in the RAS and CGM models. Thus,

\[
\gamma_{000} = \gamma_{000}^{**}
\]

\[
\gamma_{001} = \gamma_{001}^{**} - \gamma_{010}^{**}
\]

\[
\gamma_{010} = \gamma_{010}^{**} - \gamma_{100}^{**}
\]

\[
\gamma_{100} = \gamma_{100}^{**}
\]

**Random Effects**

\[
Var(Y_{ijk}) = E[Y - E(Y)]^2 = E[e_{ijk} + r_{0jk} + r_{1jk}(x_{ijk}) + u_{00k}]^2
\]

\[
= E\left[e_{ijk}^2 + r_{0jk}^2 + 2r_{0jk}r_{1jk}(x_{ijk}) + r_{1jk}^2(x_{ijk})^2 + u_{00k}^2\right]
\]

\[
= \sigma^2 + \tau_{\pi_{00}} + 2\tau_{\pi_{01}}(x_{ijk}) + \tau_{\pi_{11}}(x_{ijk})^2 + \tau_{\beta_{00}}
\]

\[
Var(Y_{ijk})
\]

**RAS:**

\[
\sigma^2 + \tau_{\pi_{00}} + 2\tau_{\pi_{01}}(x_{ijk}) + \tau_{\pi_{11}}(x_{ijk})^2 + \tau_{\beta_{00}}
\]
CGM:

\[ \sigma^2 + \tau_{\pi_{00}}^* + 2\tau_{\pi_{01}}^* (X_{ijk} - \bar{X}_{..}) + \tau_{\pi_{11}}^* (X_{ijk} - \bar{X}_{..})^2 + \tau_{\beta_{00}}^* = \sigma^2 + \tau_{\pi_{00}}^* + 2\tau_{\pi_{01}}^* X_{ijk} - 2\tau_{\pi_{01}}^* \bar{X}_{..} + \tau_{\pi_{11}}^* (X_{ijk})^2 - 2\tau_{\pi_{11}}^* X_{ijk} \bar{X}_{..} + \tau_{\pi_{11}}^* (\bar{X}_{..})^2 + \tau_{\beta_{00}}^* \]

CWC:

\[ \sigma^{2*} + \tau_{\pi_{00}}^{2*} + 2\tau_{\pi_{01}}^{2*} (X_{ijk} - \bar{X}_{jk}) + \tau_{\pi_{11}}^{2*} (X_{ijk} - \bar{X}_{jk})^2 + \tau_{\beta_{00}}^{2*} = \sigma^{2*} + \tau_{\pi_{00}}^{2*} + 2\tau_{\pi_{01}}^{2*} X_{ijk} - 2\tau_{\pi_{01}}^{2*} \bar{X}_{jk} + \tau_{\pi_{11}}^{2*} (X_{ijk})^2 - 2\tau_{\pi_{11}}^{2*} X_{ijk} \bar{X}_{jk} + \tau_{\pi_{11}}^{2*} (\bar{X}_{jk})^2 + \tau_{\beta_{00}}^{2*} \]

Note that the variance terms in Model 2 are identical to Model 1. Thus, the random effect equivalencies between RAS and CGM are the same, and neither RAS nor CGM are equivalent to CWC in random effects.

RAS = CGM

\[ \sigma^2 = \sigma^{2*} \]

\[ \tau_{\pi_{00}} = \tau_{\pi_{00}}^* - 2\tau_{\pi_{01}}^* (\bar{X}_{..}) + \tau_{\pi_{11}}^* (\bar{X}_{..})^2 \]

\[ \tau_{\pi_{01}} = \tau_{\pi_{01}}^* - \tau_{\pi_{11}}^* (\bar{X}_{..}) \]

\[ \tau_{\pi_{11}} = \tau_{\pi_{11}}^* \]

\[ \tau_{\beta_{00}} = \tau_{\beta_{00}}^* \]