Within Study Dependence in Meta-Analysis: Comparison of GLS Method and Multilevel Approaches

Seungjin Lee
FLORIDA STATE UNIVERSITY
COLLEGE OF EDUCATION

WITHIN STUDY DEPENDENCE IN META-ANALYSIS:
COMPARISON OF GLS METHOD AND MULTILEVEL APPROACHES

By
SEUNGJIN LEE

A Dissertation submitted to the
Department of Educational Psychology and Learning Systems
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Fall Semester, 2014
Seungjin Lee defended this dissertation on November 6, 2014.
The members of the supervisory committee were:

Betsy Jane Becker  
Professor Directing Dissertation

Fred Huffer  
University Representative

Insu Paek  
Committee Member

Yanyun Yang  
Committee Member

The Graduate School has verified and approved the above named committee members, and certifies than the dissertation has been approved in accordance with university requirements.
I dedicate my dissertation to my mother, Sukja Jung and my father who is in heaven, Keehyuk Lee

I love you

Also, I express my sincere appreciation to my big sister, Yongman Lee
ACKNOWLEDGMENTS

My advisor, Dr. Betsy Jane Becker, I could not imagine my life in USA without her support and encouragement. She has been with me as my best academic advisor and she has been with me as my counsellor for my life. And it is my turn to be with her as her proud student and friend. First of all, I could not forget forever her warmth to show me in a meeting at her kitchen table on November 8, 2014. If I have a chance, I would like to draw the scene but I would not forget to add my little friend Tashi on there. Dr. Becker, I could not have completed my dissertation without your help and advice, I love you.

I would like to express my gratitude to my dissertation committee members, Dr Yanyun Yang, Dr. Fred Huffer, and Dr. Insu Paek. I am very thankful for their valuable advice, comments, and suggestions for my progress in my dissertation.

And my friend Fatih Orcan who is in Turkey now, I could not complete my dissertation without his contribution on my dissertation as well. He is my main partner to develop the R and SAS codes to simulate data in this dissertation. Also, he was my best friend who shared most of my daily life in Measurement and Statistics. We are planning to work together in the future. Thank you so much, Fatih.

I would like to extend my thanks to my family, my sister Yongha Lee, Bomoon Lee, and my brother Taehoon Lee, and my friends Jiyeo Yun, Kyunghwa Cho, Raesun Kim, Abdullah Alghamdi, Bernd Weiss, Christopher Thompson, and Jiyoun Kim. Also, I extend my love to my little lovely friend Tashi.
# TABLE OF CONTENTS

LIST OF TABLES ....................................................................................................................... viii

LIST OF FIGURES ....................................................................................................................... ix

ABSTRACT .................................................................................................................................... x

CHAPTER 1: INTRODUCTION ................................................................................................... 1

CHAPTER 2: LITERATURE REVIEW ........................................................................................ 5

Standardized-Mean-Difference Effect Size .................................................................................... 5

Models to Analyze Effect Sizes ...................................................................................................... 6

Univariate Meta-analysis ................................................................................................................ 8

    Traditional Meta-Analysis Model ............................................................................................. 8

    Univariate Two-Level Meta-Analysis Model ......................................................................... 11

Multivariate Meta-analysis ........................................................................................................... 12

    GLS Method ............................................................................................................................ 13

    Multivariate Two-Level Model ............................................................................................... 18

    Univariate Three-Level Model ............................................................................................... 21

Purpose of This Study ................................................................................................................... 31

    Overall Effects in GLS and Three-Level Approaches ............................................................ 32

    Outcome-Specific Effects in GLS and Three-level Approaches ............................................ 33

CHAPTER 3: METHODOLOGY ................................................................................................ 35

Simulation Conditions ................................................................................................................ 35
LIST OF TABLES

Table 2.1: Comparison of the Two-Level Multivariate Models. .................................................. 26

Table 2.2: Comparison of the Three-Level Models................................................................. 31

Table 2.3: Comparisons of Estimates Between Three-Level and GLS Method Approaches...... 34

Table 3.1: Simulation Conditions ................................................................................................. 37

Table 4.1: Partial Eta Squares for Difference in SE Bias Between Two Approaches in the Random-Effects Model....................................................................................................... 49
LIST OF FIGURES

Figure 4.1: Relative percentage bias of the SE in the overall effects for two outcomes in the random-effects model ........................................................................................................................................................................44

Figure 4.2: Relative percentage bias of the SE in the overall effects for five outcomes in the random-effects model ........................................................................................................................................................................45

Figure 4.3: Relative percentage bias of the SE of the outcome-specific effects for two outcomes in the random-effects model ........................................................................................................................................................................46

Figure 4.4: Relative percentage bias of the SE in the outcome-specific effects for five outcomes in the random-effects model ........................................................................................................................................................................47
ABSTRACT

Multivariate meta-analysis methods typically assume the dependence of effect sizes. One type of experimental-design study that generates dependent effect sizes is the multiple-endpoint study. While the generalized least squares (GLS) approach requires the sample covariance between outcomes within studies to deal with the dependence of the effect sizes, the univariate three-level approach does not require the sample covariance to analyze such multivariate effect-size data. Considering that it is rare that primary studies report the sample covariance, if the two approaches produce the same estimates and corresponding standard errors, the univariate three-level model approach could be an alternative to the GLS approach.

The main purpose of this dissertation was to compare these two approaches under the random-effects model for synthesizing standardized mean differences in multiple-endpoints experimental designs using a simulation study. Two data sets were generated under the random-effects model: one set with two outcomes and the other set with five outcomes. The simulation study in this dissertation found that the univariate three-level model yielded the appropriate parameter estimates and their standard errors corresponding to those in the multivariate meta-analysis using the GLS approach.
CHAPTER 1

INTRODUCTION

Meta-analysis is a statistical method to combine results from studies on the same or related topics, which is widely used in medicine and the social sciences. The results from studies are often integrated using statistical indicators or effect-size estimates (Hedges, 2007). Common effect-size estimates are correlation coefficients, standardized mean differences, and odds ratios.

Two types of meta-analysis have been developed to synthesize effect sizes across studies: univariate and multivariate meta-analysis. Univariate meta-analysis methods typically assume the independence of effect sizes and thus each estimated effect size does not affect the other estimated effect sizes in their directions and magnitudes. However, it is common that studies produce multiple effect sizes, and they might be correlated (Gleser & Olkin, 1994; Hedges & Olkin, 1985; Rosenthal & Rubin, 1986). For example, if a study produces multiple effect sizes for multiple outcome measures with common participants, the effect sizes in this study are not independent of each other. Multivariate meta-analysis methods, thus, account for the dependence of effect sizes within the same study.

Gleser and Olkin (1994) discussed two types of experimental-design studies that generate dependent effect sizes within the same study: multiple-treatment studies and multiple-endpoint studies. In multiple treatment studies, effect sizes are estimated with one common control group and multiple treatment groups. Even with only one outcome measure the effect sizes using the common control group are correlated (Kim & Becker, 2010). Multiple-endpoint studies may use
single treatment and control groups to estimate effect sizes for multiple outcome measures, and
effect sizes from the common participants are dependent on each other.

There are three approaches for dealing with the dependence of effect sizes within studies
in meta-analysis (Becker, 2000; Littell, Corcoran, & Pillai, 2008). First, a researcher might
ignore the dependence of effect sizes within studies and treat all effect sizes as independent.
Second, the dependence could be avoided. For example, if a study has three outcome measures
for the same participants, a researcher might perform three separate meta-analyses, one for each
individual outcome measure across studies. Third, a researcher could statistically model the
dependence. Three statistical multivariate meta-analysis approaches for multiple-endpoint
studies are explained in this dissertation: the multivariate meta-analysis model with generalized
least squares (GLS) estimation, the multivariate two-level model, and the univariate three-level
model estimated using hierarchical modeling methods.

Raudenbush, Becker, and Kalaian (1988) applied the GLS method to develop the
multivariate fixed-effects regression model and used it to examine the effect of the hours of
coaching on effect sizes for SAT-Math and SAT-Verbal scores. Gleser and Olkin (1994) also
described the GLS method for multivariate fixed-effects effect-size data. Both sets of authors
used the sample variance-covariance matrix of effect sizes within the same study to handle the
dependence of effect sizes.

The traditional meta-analysis model is a special case of a univariate two-level model. In
the two-level model, the dispersion of effect sizes is explained by two random variance
components: the participant-level (level 1) sampling error and the study-level (level-2) variation.
Kalaian and Raudenbush (1996) extended the multivariate two-level fixed-effects regression
model (Raudenbush, Becker, & Kalaian, 1988) to a multivariate two-level mixed-effects regression model to examine the variance and covariance of effect sizes between studies, based on the sample variances and covariances of effect sizes within studies. They also investigated whether hours of coaching would explain the differences among effect sizes from SAT coaching studies. The between-studies variances and covariances were estimated using restricted maximum likelihood (REML). The estimates of regression coefficients (i.e., the hours of coaching) and individual effect sizes were based on GLS methods and a Bayesian approach, respectively.

Finally, Van Den Noortgate, Lopez-Lopez, Marin-Martinez, and Sanchez-Meca (2012) proposed the univariate three-level model as an alternative approach to multivariate meta-analysis, based on inverse-variance weighting, the REML method, and an empirical Bayesian approach. The univariate three-level model was an extension of the traditional two-level model with the inclusion of an additional level, “the outcome level,” to account for the dependence of effect sizes. This three-level model, thus, accounts for three random variance components to explain the dispersion of effect sizes over three levels: the participant-level sampling error (level 1), the outcome-level variation (level 2), and the study-level variation (level 3). Van den Noortgate et al. (2012) investigated how the between-studies variance in the univariate three-level model reflected the dependence of effect sizes.

Unlike the multivariate fixed approach model based on the GLS method and the multivariate two-level model approach, the univariate three-level approach does not require the sample covariance between outcomes within studies to analyze the multivariate effect-size data. Considering that it is rare that primary studies report the sample covariance, the univariate three-level model is both applicable and relatively simple in comparison to the other two approaches.
If the three approaches produce the same estimates and corresponding standard errors, the univariate three-level model approach could be an alternative to the other two approaches.

Van den Noortgate et al. (2012) found that the parameter estimates and the corresponding standard errors in the univariate three-level model were not biased in comparison to those from a multivariate two-level model with a simulation study. Thus, the main purpose of this dissertation is to investigate whether the univariate three-level model would be a viable alternative to multivariate meta-analysis with GLS methods. To this end, I generated a multivariate data set based on a multiple-endpoints experimental design. The generated data was used to compute effect sizes for the corresponding outcome measures. Two multivariate meta-analysis models were examined with the effect sizes: the univariate three-level model approach and the multivariate meta-analysis model using GLS approach.
There are two types of meta-analysis: univariate meta-analysis and multivariate meta-analysis. Univariate meta-analysis analyzes studies which produce effect sizes to measure a single outcome variable and assumes the independence of effect sizes across studies. On the other hand, multivariate meta-analysis analyzes studies that may have multiple effect sizes for the multiple outcomes which are dependent on each other. Thus, the dependence of effect sizes within the same study is accounted for by the analysis in multivariate meta-analysis.

The main purpose of this chapter is to compare the multivariate meta-analysis model estimated using the GLS method and the multi-level meta-analysis model for the univariate three-level case. To this end, this chapter briefly begins with a description of the standardized-mean-difference effect size and two types of univariate meta-analysis models: traditional meta-analysis and meta-analysis with a univariate two-level multi-level model.

**Standardized-Mean-Difference Effect Size**

Meta-analysis involves combining effect sizes (e.g., correlation coefficients, standardized mean differences, odds ratios) from individual studies dealing with the same topic, and asking whether the effect sizes are homogenous across studies (Hedges, 2007).

The estimated standardized mean difference (effect size) between the treatment group and the control group for one outcome measure is computed as
where $\bar{Y}_C$ and $\bar{Y}_T$ are the means of the outcome variable in the control group and treatment group in study $i$, respectively, and $S_i$ is the pooled standard deviation for the effect size in study $i$, which is calculated as follows

$$S_i = \sqrt{\frac{(n_i^T - 1)S_{T_i}^2 + (n_i^C - 1)S_{C_i}^2}{(n_i^T - 1) + (n_i^C - 1)}}, \quad (2)$$

where $n_i^T$ and $n_i^C$ are group sample sizes, and $S_{T_i}^2$ and $S_{C_i}^2$ are the corresponding group variances.

Hedges (1981) showed that effect-size estimates from small samples are biased and this small sample bias is corrected by $E[g_i] = \delta_i / c(m_i)$ where $c(m_i) = 1 - 3/(4m_i - 1)$, and $m_i = n_i^T + n_i^C - 2$. Effect-size estimates $d_i = g_i c(m_i)$ are approximately normally distributed with a mean of the true effect size $\delta_i = (\mu^T_i - \mu^C_i) / \sigma_i$ where $\sigma_i$ is the population pooled standard deviation for the effect size in study $i$ and a variance of

$$\sigma^2_{d_i} = \frac{1}{n_i^T} + \frac{1}{n_i^C} + \frac{\delta_i^2}{2(n_i^T + n_i^C)}, \quad (3)$$

Typically, $\sigma^2_{d_i}$ is estimated by substituting $d_i$ for $\delta_i$ in Equation 3.

**Models to Analyze Effect Sizes**

Meta-analysis studies use statistical methods (i.e., homogeneity tests) to examine whether effect sizes estimate the same population effect. Two types of models can be used to analyze effect sizes: the fixed-effects model and the random-effects model. In fixed-effects models, the
dispersion of effect sizes around the population mean is viewed as resulting from the participant-level sampling error alone, and each effect size estimates the common population value. If each effect size does not estimate a common population mean, the variability of effect sizes is determined to contain true differences between studies in addition to the participant-level sampling error, which leads to the random-effects model.

Effect sizes are estimated from studies of different sizes, thus the sampling variances for each effect size are not equal. Thus, in analyses each effect size is weighted by the inverse variance. In the fixed-effects model, the source of variation for effect sizes is sampling error and each effect size is weighted by the inverse sampling variance. However, under the random-effects model, the variability of the effect sizes comes from sampling error and true differences in effects across studies. Thus, each effect is weighted by the inverse of the sum of the sampling variance and the between-studies variance in the random-effects model.

In the random-effects model, the true differences between studies can be characterized by three models: fixed-effects ANOVA or regression models or mixed-effects models (Hedges, 1982; Hedges & Olkin, 1985). In the fixed-effects ANOVA or regression models, the observed distributions of effect sizes are more variable than for the simplest fixed-effects models. The added variability is not random, but has a systematic portion which can be explained by moderators (study characteristics) to differentiate studies. Finally, the mixed-effects model for ANOVA or regression adds another random component to the fixed-effects ANOVA or regression models for the further differences in effect sizes across studies; this reflects a “leftover” portion which is not explained by moderators.
The main purpose of this dissertation is to compare the multivariate meta-analysis model based on the GLS approach and the multivariate meta-analysis model for the univariate three-level case under the random-effects models without moderators.

**Univariate Meta-analysis**

Univariate meta-analysis assumes the independence of effect sizes for a single outcome measure across studies. Two types of univariate meta-analysis approaches are described: traditional meta-analysis models and meta-analysis with a univariate two-level model.

**Traditional Meta-Analysis Model**

*Homogeneity test.* The $Q$ statistic test (i.e., homogeneity test) examines whether the variability among the effect sizes is greater than the variance expected from the participant-level sampling error. The null hypothesis for the $Q$ statistic is that all effect-size estimates arise from a common population effect (i.e., $H_0: \hat{\delta}_1 = \ldots = \hat{\delta}_k = \delta$). $Q$ is distributed as a chi-square with $k-1$ degrees of freedom where $k$ is the number of effect sizes. The formula is

$$Q = \sum_{i=1}^{k} \frac{(d_i - d_*)^2}{\sigma_{d_i}^2},$$

(4)

where $d_i$ is the effect size estimate in study $i$ for $i=1,\ldots,k$, and $\sigma_{d_i}^2$ is the variance of the effect size in study $i$. Also, $d_*$ is the fixed-effects weighted mean effect-size estimate over $k$ effect sizes (Hedges & Olkin, 1985), which is expressed as

$$d_* = \frac{\sum [d_i / \sigma_{d_i}^2]}{\sum [1 / \sigma_{d_i}^2]},$$

(5)
where each effect size $d_i$ is weighted with an inverse variance $\frac{1}{\sigma_i^2}$. The variance of the common population mean effect is $\sum \frac{1}{1/\sigma_i^2}$ and the standard deviation of the mean effect is the square root of the variance. While meta-analysts may choose to adopt the random-effects model as a matter of principle, in other cases, the $Q$ test can be used to decide whether the fixed or random-effects model applies.

**Fixed-effects model.** If $Q$ is smaller than the critical value of $\chi^2$ with $k-1$ degrees of freedom, we can conclude that all effect sizes are from one common population effect, which leads to the model (Lipsey & Wilson, 2001)

$$d_i = \delta + \varepsilon_i,$$

where $d_i$ is an observed effect size in study $i$, $\delta$ is the corresponding true or population effect size, and $\varepsilon_i$ is an error term in study $i$. The random effect $\varepsilon_i$ is approximately normally distributed with a mean zero and a variance $\sigma_i^2$. Thus, the estimated effect size $d_i$ is normally distributed about the true effect size $\delta$ (i.e., with the mean of $\delta$) with a variance of $\sigma_i^2$. The variance $\sigma_i^2$ is treated as a known variance and is typically estimated using Equation 3. In the fixed-effects model, all variation arises from sampling error within studies alone, and all effect sizes estimate the common population effect size $\delta$.

**Random-effects model.** Under the random-effects model, effect sizes do not estimate a common population effect. The random-effects model assumes that true effect sizes $\delta_i$ vary and another random component, the between-studies variance $\tau$, is included to explain the true
differences in effect sizes. The variation in effect sizes comes from true differences in effect sizes across studies in addition to sampling error within studies, specifically

\[ d_i = \delta_i + \varepsilon_i \quad \text{or} \quad d_i = \delta^* + \varepsilon_i + u_i \]  

(7)

where \( d_i \) is an observed effect size in study \( i \), \( \delta_i \) is the corresponding population effect size, \( \delta^* \) is an overall effect (i.e., the average of the population effect sizes) across studies and \( u_i \) is a random effect showing between-studies variation around the overall effect size. The random effects \( u_i \) are normally distributed with a mean zero and a variance \( \tau \). The overall effect size is fixed, and thus the variance of the random effects (\( u_i \)) is also the variance of the true effect sizes (\( \delta_i \)). Thus, the true effect sizes \( \delta_i \) can be viewed as being normally distributed around the overall effect size \( \delta^* \) with a variance \( \tau \). The overall population mean effect \( \delta^* \) (the average effect size) is estimated as

\[ \delta^* = \frac{\sum d_i / [\sigma_{\delta_i}^2 + \hat{\tau}]}{\sum 1 / [\sigma_{\delta_i}^2 + \hat{\tau}]} \]  

(8)

Two method-of-moments estimators of the between-studies variance (\( \tau \)) are available (Raudenbush, 1994). While one approach is not weighted, the other approach is weighted. That is, the between-studies variance is computed using a “typical” sample variance of the observed effect sizes in the first approach. In the second approach, the between-studies variance is estimated as

\[ \hat{\tau} = \frac{Q - (k - 1)}{\sum w_j - (\sum w_i^2 / \sum w_j)} \]  

(9)
where $Q$ is the value of the homogeneity test in Equation 4, $k$ is the number of effect sizes and $w_i$ is the inverse sampling variance for study $i$.

**Univariate Two-Level Meta-Analysis Model**

Effect sizes are nested within studies in the meta-analysis data structure. Thus, multilevel modeling is a useful framework for conducting meta-analysis, accounting for variation in all levels.

*Fixed- and random-effects models.* The traditional simplest random-effects model in Equation 7 typically follows a two-level structure using the one-way ANOVA model (Hox, 2002; Konstantopoulos, 2011). The two-level random-effects model accounts for both within-study and between-studies variances in level 1 and level 2, respectively.

The first level model (within-study level) is

$$d_i = \delta_i + \epsilon_i,$$

(10)

the second-level model (between-studies level) is

$$\delta_i = \delta^* + u_i,$$

(11)

and the combined model with within-study and between-studies levels leads to

$$d_i = \delta^* + u_i + \epsilon_i.$$  

(12)

All terms in Equations 10, 11, and 12 were defined extensively for Equation 7. The true effect sizes $\delta_i$ are estimated using a Bayesian approach based on the estimates given in Equations 1 and 3.
If the effect sizes are not significantly different across studies and the between-studies variance $\tau$ is not significantly different from zero, the effect sizes across studies are considered homogeneous and the fixed-effects model ($d_i = \delta + \varepsilon_i$) can be adopted. If the between-studies variance $\tau$ is significantly different from zero, effect sizes are treated as heterogeneous, and the random-effects model ($d_i = \delta^* + u_i + \varepsilon_i$) would be adopted. The overall population effect across studies $\delta^*$ is then estimated using a weight that includes both the sampling variance and between-studies variance.

**Homogeneity test.** The between-studies variance $\tau$ is estimated using REML and is tested (i.e., $H_0: \tau^2 = 0$) to examine whether effect sizes estimate the common population effect size using a chi-square test, which corresponds to the $Q$ test in the traditional meta-analysis approach.

**Multivariate Meta-analysis**

Several types of studies with multivariate data structures can lead to dependence of effect sizes in meta-analysis: multiple-treatment studies, multiple-endpoint studies, and multiple time-point studies. Gleser and Olkin (1994) provided such covariance matrices for two types of experimental designs that may generate dependent effect sizes in the same study. First, when a study has multiple treatments and one common control group given a dependent variable, the effect sizes from this study are dependent on each other; such studies called multiple-treatment studies. Second, in an experimental study that measures multiple outcomes with single treatment and control groups, the effect sizes are not independent; this is called a multiple-endpoint study.

This dissertation focuses on multiple-endpoint studies but similar results are expected to hold for other dependence structures. It is desirable that each study has the exact same set of
outcome measures in multivariate meta-analysis. However, it is rare that all studies have the
same set of effect sizes. Thus, the multivariate meta-analysis methods discussed in this
dissertation allow different studies to have different subsets of effect sizes or outcome variables.

Three types of multivariate meta-analysis approaches to estimate mean effects across
studies are described: First, the multivariate fixed-effects meta-analysis approach with GLS
methods, next the multivariate two-level approach, and third, the univariate three-level approach.
The first approach estimates mean effects using GLS methods. In the second approach, the mean
effects are estimated using GLS methods, given the REML estimates of the variances and
covariances across studies. The mean effects are estimated using the inverse variances based on
REML in the univariate three-level approach.

GLS Method

The GLS estimation method accounts for the dependence of effect sizes with a sample
variance-covariance matrix (Berkey, Anderson, & Hoaglin, 1996; Berkey, Hoaglin, Antczak-
Bouckoms, Mosteller, & Colditz, 1998; Gleser & Olkin, 1994; Raudenbush, Becker, & Kalaian,
1988). In a multiple-endpoints experimental design (Gleser & Olkin, 1994; Raudenbush et al.,
1988), the population variances ($\sigma_j^2$) and covariances ($\sigma_{ij}$) of effect sizes in the same study $i$
are estimated, respectively, as

$$\hat{\psi}_j = \hat{\sigma}_j^2 \approx \frac{n_j^T + n_j^C}{n_j^T n_j^C} + \frac{d_{ij}^2}{2(n_j^T + n_j^C)}; j = 1, ..., p; i = 1, ..., k,$$ (13)

and

$$\hat{\psi}_{ij} = \hat{\sigma}_{ij} \approx \frac{n_j^T + n_j^C}{n_j^T n_j^C} r_{ij} + \frac{d_{ij}^2 r_{ij}^2}{2(n_j^T + n_j^C)},$$ (14)
where \( n_i^T \) and \( n_i^C \) are the sample sizes of the treatment group and control group in study \( i \), respectively, \( d_{ij} \) and \( d_{ij'} \) are effect-size estimates for outcomes \( j \) and \( j' \) in study \( i \), and \( r_{jj'} \) is the sample correlation between the two outcome variables \( Y_j \) and \( Y_{j'} \).

Each study has a complete vector of true effect sizes, \( \delta = (\delta_1, \ldots, \delta_p) \) for the corresponding outcome measures. Different studies could report different subsets of the complete effect sizes but I consider the case where each study \( i \) has its own vector of all effect-size estimates, \( \mathbf{d}_i = (d_{i1}, \ldots, d_{ip}) \). Each study \( i \) produces its own estimated variance-covariance matrix \( \hat{\Psi}(i) \) with dimension \( p \times p \) which accounts for the dependence of effect sizes in the study. Variances and covariances in the matrix are computed using Equations 13 and 14, respectively.

The combined estimated column vector of effect sizes for \( k \) primary studies

\[
(\mathbf{d} = (\mathbf{d}_1, \ldots, \mathbf{d}_k))
\]

has dimension \( p^* \times 1 \) where

\[
p^* = \sum_{i=1}^{k} p_i
\]

The corresponding combined estimated variance-covariance matrix will be a block diagonal matrix \( \hat{\Psi} \) of dimensionality \( p^* \times p^* \). In the matrix \( \hat{\Psi} \), the main diagonal blocks are the variance and covariance matrices \( \hat{\Psi}(i) \) from the individual primary studies, and off-diagonal blocks are zero which indicates the independence of studies. This is represented as

\[
\hat{\Psi} = \begin{bmatrix}
\hat{\Psi}(1) & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & \hat{\Psi}(k)
\end{bmatrix}.
\]

The multivariate fixed-effects model. The multivariate fixed-effects model (with homogeneous effect sizes over studies) is
Where $\mathbf{d}$ is the combined estimated column vector of effect sizes for $k$ primary studies with dimension $p^* \times 1$. The design matrix $\mathbf{X}$ has dimension $p^* \times p$. The values of the design matrix are dummy variables. If an effect size is estimated for a corresponding outcome measure in study $i$, its $\mathbf{X}$ matrix element is equal to 1, and equal to 0 otherwise. $\mathbf{\delta}$ is a column vector which contains the common effect sizes $(\delta_1, ..., \delta_p)'$ for the corresponding outcome measures across studies. The common effects $(\hat{\delta}_j)$ for the $p$ endpoints across studies are estimated by

$$
\hat{\mathbf{\delta}} = (\hat{\delta}_1, ..., \hat{\delta}_p)' = (\mathbf{X}' \hat{\psi}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\psi}^{-1} \mathbf{d},
$$

where the matrix $(\mathbf{X}' \hat{\psi}^{-1} \mathbf{X})^{-1}$ is the estimated variance-covariance matrix of common effect sizes.

The standard errors of the $p$ common effects are the square roots of the diagonal elements in the variance-covariance matrix $(\mathbf{X}' \hat{\psi}^{-1} \mathbf{X})^{-1}$. Finally, the vector of errors $\mathbf{\varepsilon}$ has dimension $p^* \times 1$.

The errors $\mathbf{\varepsilon}_i$ in study $i$ are assumed to have a multivariate normal distribution, which is expressed as $\mathbf{\varepsilon}_i \sim N(0, \psi(i))$. For example, when study $i$ reports two effect sizes for outcomes 1 and 2, the variance and covariance matrix $\psi(i)$ of the two effect sizes is

$$
\begin{bmatrix}
\sigma^2_{d_1} & \sigma_{d_1d_2} \\
\sigma_{d_1d_2} & \sigma^2_{d_2}
\end{bmatrix},
$$

where $\sigma^2_{d_1}$ and $\sigma^2_{d_2}$ are the sampling variances of effect sizes for outcomes 1 and 2, respectively, and $\sigma_{d_1d_2}$ is the covariance of the two effect sizes in the study.
When each study could have two effect sizes \((p = 2)\) for two outcome measures over \(k\) studies, the estimated multivariate regression GLS model in Equation 16 is expressed in matrix notation as 

\[
\begin{bmatrix}
  d_{11} \\
  d_{12} \\
  d_{21} \\
  d_{31} \\
  \vdots \\
  d_{k1} \\
  d_{k2}
\end{bmatrix}
= 
\begin{bmatrix}
  1 & 0 \\
  0 & 1 \\
  1 & 0 \\
  1 & 0 \\
  \vdots & \vdots \\
  1 & 0 \\
  0 & 1
\end{bmatrix}
\begin{bmatrix}
  \delta_1 \\
  \delta_2
\end{bmatrix}
+ 
\begin{bmatrix}
  \varepsilon_{11} \\
  \varepsilon_{12} \\
  \varepsilon_{21} \\
  \varepsilon_{31} \\
  \vdots \\
  \varepsilon_{k1} \\
  \varepsilon_{k2}
\end{bmatrix}. 
\]  

(19)

The design matrix \(X\) has dummy variables in the first and second columns. The first column has a dummy variable that is equal to 1 when the effect size is estimated for the first outcome measure in study \(i\), and equal to 0, otherwise. In the second column, the dummy variable takes value 1 when the second effect size (for the second outcome measure) is reported, 0, otherwise. This matrix shows that the first study reported two effect sizes for two outcomes whereas the second study reported only the effect size for the first outcome.

The multivariate random-effects model. The multivariate random-effects model is 

\[
d = X\delta^* + \varepsilon. 
\]  

(20)

In the random-effects model, the variation of effect sizes is composed of the true differences in studies (between-studies variance) and the within-studies variance, and each true effect size is dispersed from the averaged population mean effect size \(\delta^*\). \(\delta^*\) is a column vector which contains the averaged effect sizes \((\delta^*_1, \ldots, \delta^*_p)\) for the corresponding outcome measures across studies. The averaged effects \((\delta^*_1)\) for the \(p\) endpoints across studies are estimated by
\[ \hat{\delta}^* = (\hat{\delta}_1^*, \ldots, \hat{\delta}_p^*)' = (X' \hat{\psi}^{-1} X)^{-1} X' \hat{\psi}^{-1} d, \quad (21) \]

and the matrix \((X' \hat{\psi}^{-1} X)^{-1}\) is the estimated variance and covariance matrix of the average effect sizes. In addition to the sampling variances and covariances among effect sizes \(\psi(i)\) within study \(i\), the variance-covariance among effect sizes across studies, \(\tau\), is included to capture all variation in effect sizes in the multivariate random-effects model. The total variance-covariance among effect sizes in a study, \(\psi(i)^*\), is thus obtained by the sum of within-study-variance-covariance \(\psi(i)\) and between-studies-variance-covariance \(\tau\),

\[ \psi(i)^* = \psi(i) + \tau = \begin{bmatrix} \sigma_{d_1}^2 & \sigma_{d_d_2} \\ \sigma_{d_d_2} & \sigma_{d_2}^2 \end{bmatrix} + \begin{bmatrix} \tau_{11} \\ \tau_{12} \end{bmatrix}. \quad (22) \]

Thus, the standard errors for the corresponding overall effects are computed by taking the square roots of the diagonal elements of the variance-covariance matrix among the overall effects \(v_\delta^* = (X' \hat{\psi}^{-1} X)^{-1}\). All other terms were defined for Equations 20 and 21, or earlier.

The multivariate random-effects model for two effect sizes \((p=2)\) across \(k\) studies is

\[ d = X\hat{\delta}^* + \varepsilon \quad \text{or} \quad \begin{bmatrix} d_{11} \\ d_{12} \\ \vdots \\ d_{k1} \\ d_{k2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \end{bmatrix} \begin{bmatrix} \delta_1^* \\ \delta_2^* \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{k1} \\ \varepsilon_{k2} \end{bmatrix}. \quad (23) \]

**Homogeneity test.** The \(Q\) statistic for the homogeneity test is computed as

\[ Q = d' \hat{\psi}^{-1} d - \hat{\delta} X' \hat{\psi}^{-1} X \hat{\delta}, \quad (24) \]
which follows a chi-square distribution with degrees of freedom equal to the dimension $p^*$ minus $p$. The $Q$ statistic tests the null hypothesis that all effect sizes across all outcomes and all studies arise from the same population, which is denoted as

$$H_0: \delta_{11} = \ldots = \delta_{kp} = \delta,$$

where $\delta$ is the common population effect size mean across all outcomes and studies. The matrix notation for the homogeneity of effect size across $k$ studies and two outcomes is represented as

$$d = X\delta + \varepsilon \text{ or } \begin{bmatrix} d_{11} \\ d_{12} \\ \vdots \\ d_{k1} \\ d_{k2} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{k1} \\ \varepsilon_{k2} \end{bmatrix},$$

where all terms were defined in Equations 21 and 22, except for the design matrix $X$ which is a column vector of 1s. The common population effect size $\delta$ is estimated as

$$\hat{\delta} = (X'\hat{\psi}^{-1}X)^{-1}X'\hat{\psi}^{-1}d.$$  

The standard errors of the overall common effects are the square roots of the diagonal elements of $(X'\hat{\psi}^{-1}X)^{-1}$ in Equation 27. Thus, the overall population effect in the random-effects model $\delta$ is estimated by $\hat{\delta} = (X'\hat{\psi}^{*-1}X)^{-1}X'\hat{\psi}^{*-1}d$.

**Multivariate Two-Level Model**

Kalaian and Raudenbush (1996) developed a two-level multivariate model to conduct multivariate mixed-effects analyses in multiple-endpoints studies. The first level model (for within the study) represents the relationship between the true effect sizes and the corresponding
effect-size estimates in each study. The second level (between-studies) is for the distribution of these true effect sizes around their population mean.

**Multivariate two-level random-effects model.** The level 1 (within study) model is expressed as

\[ d_{ij} = \sum_{j=1}^{p} \delta_{ij} X_{ij} + e_{ij} : i=1,...,k; X_{ij} = 0 \text{ or } 1 \]  

(28)

where \( d_{ij} \) and \( \delta_{ij} \) are the effect-size estimate and the corresponding true effect size for outcome \( j \) in study \( i \). The variable \( X_{ij} \) indicates the presence of each effect size for the corresponding outcome \( j \) in study \( i \). Each study is assumed to have a complete vector of true effect sizes, \( \delta = (\delta_1, ..., \delta_p) \) for the \( p \) outcome measures. Each study could report a different subset of effect sizes and study \( i \) has own vector of sample effect sizes, \( d_i = (d_{i1}, ..., d_{ip}) \). For example, in two-endpoint studies, if study \( i \) reports two effect-size estimates, the vector \( d_i \) is \( (d_{i1}, d_{i2}) \). If study \( i \) estimates the effect size for outcome 1, the value \( X_{i1} \) is equal to 1, and equal to 0, otherwise.

Thus, the equation becomes for \( j = 1 \)

\[ d_{i1} = \delta_{i1} X_{i1} + \delta_{i2} X_{i2} + e_{i1} = \delta_{i1}(1) + \delta_{i2}(0) + e_{i1}. \]  

(29)

For \( j = 2 \),

\[ d_{i2} = \delta_{i1} X_{i1} + \delta_{i2} X_{i2} + e_{i2} = \delta_{i1}(0) + \delta_{i2}(1) + e_{i2}. \]  

(30)

If we express this in matrix notation,

\[ d_i = X_i \delta_i + e_i, \text{ or } \begin{bmatrix} d_{i1} \\ d_{i2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \delta_{i1} \\ \delta_{i2} \end{bmatrix} + \begin{bmatrix} e_{i1} \\ e_{i2} \end{bmatrix}. \]  

(31)
The errors $e_{ij}$ are assumed to have a multivariate normal distribution $e_i \sim N(0, \psi(i))$, where $\psi(i)$ is a $p \times p$ variance-covariance matrix in study $i$. In this matrix $\psi(i)$, the variances ($\sigma_{ij}^2$) and covariances ($\sigma_{ij}^d$) of effect sizes are estimated, using Equations 13 and 14, respectively, so that

$$
\begin{bmatrix}
e_{i1} \\
e_{i2}
\end{bmatrix} \sim N\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\sigma_{d1}^2 & \sigma_{d1}^d \\
\sigma_{d1}^d & \sigma_{d2}^2
\end{bmatrix}\right),
$$

(32)

The between-studies model (level 2) is

$$
\delta_{ij} = \delta_j^* + u_{ij},
$$

(33)

where $\delta_{ij}$ is the true effect size in study $i$, $\delta_j^*$ is the effect size mean across studies for the corresponding outcome measure. The random effects $u_{ij}$ are deviations of the true effect sizes from the population means $\delta_j^*$, and are assumed to have a multivariate normal distribution with a mean of zero and a $p \times p$ variance and covariance matrix $\tau$, which is expressed as $u_i \sim N(0, \tau)$,

$$
\begin{bmatrix}
u_{i1} \\
u_{i2}
\end{bmatrix} \sim N\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\tau_{11}^2 & \tau_{12} \\
\tau_{12} & \tau_{22}^2
\end{bmatrix}\right).
$$

(34)

The combined model including the within-study model in Equation 28 and the between-studies model in Equation 33 is $d_y = \sum_{j=1}^n (\delta_j^* + u_{ij})X_{ij} + e_{ij}$. The random components in the variance and covariance matrix $\tau$ are estimated using REML and the estimates of the overall effects $\hat{\delta}_j^*$ are estimated on the GLS method, given the estimated variance and covariance in $\tau$.

**Multivariate two-level fixed-effects model.** If the effect sizes are not significantly different across studies, and there are no true differences among effect sizes in studies, the combined
model is \( d_j = \sum_{i=1}^{p} (\delta_j) X_{ij} + e_j \), which leads to the multivariate fixed-effects model. This fixed-effects model is comparable to the model in the GLS method in Equation 20. The common-effects estimates for each outcome \( \hat{\delta}_j \) in \( d_j = \sum_{i=1}^{p} (\delta_j) X_{ij} + e_j \) are thus comparable to the corresponding elements in the effect-size column vector \( \hat{\delta} = (\hat{\delta}_{i_1}, \ldots, \hat{\delta}_{i_p}) = (X' \psi^{-1} X)^{-1} X' \psi^{-1} d \), in Equation 21.

**Homogeneity test.** The between-studies variances and covariances in \( \tau \) are tested using a chi-square distribution to examine whether effect sizes estimate the common population effect size \( (H_0; \tau_{ij}^2 = 0; \tau_{jj'} = 0) \), which corresponds to the \( Q \) test in the multivariate fixed-effects meta-analysis approach with GLS methods (Equation 24).

**Univariate Three-Level Model**

A major limitation of the multivariate meta-analysis approach for both the GLS method and the two-level model is that they require the sample covariance matrix of the effect sizes within a study. However, recently Van den Noortgate et al. (2013) have proposed a univariate three-level model which does not require a sample covariance matrix for the multivariate effect-size data. The univariate three-level model was extended from the traditional two-level model with the inclusion of an additional level (i.e., the outcome level) to account for the dependence of effect sizes. This model contains three types of random variance components over three levels: the participant-level sampling error (level 1), the outcome-level variation (level 2), and the study-level variation (level 3). Thus, the effect sizes vary over participants, outcomes, and studies. Van den Noortgate et al. (2012) simulated a multivariate two-level data set to investigate how the between-studies variance in the univariate three-level model reflects the dependence of effect sizes in the multivariate two-level model. They found that the parameter estimates and the
corresponding standard errors in the univariate three-level model were not biased in comparison
to those from a multivariate two-level model with a simulation study.

*The multivariate two-level model to generate data.* Van den Noortegate et al. (2013)
simulated multivariate two-level data for two outcome measures in multiple-endpoint studies for
300 studies with a sample size of 50 per group (the treatment group and the control group). The
raw data were used to examine four models (summarized in Tables 2.1 and 2.2). All parameter
values (i.e., true values) to generate data are described in Table 2.1.

In the multivariate two-level model for two outcomes, the participant-level (level-1)
models for two outcomes are

\[
Y_{is1} = \delta_{i01} + \delta_{i11}T_{is} + e_{is1} \quad \text{and} \quad Y_{is2} = \delta_{i02} + \delta_{i12}T_{is} + e_{is2},
\]

\[
i = 1, \ldots, p; s = 1, \ldots, n; j = 1, \ldots, p
\]

where \(Y_{isj}\) is an outcome value for participant \(s\) for outcome \(j\) in study \(i\) and \(T\) is a dummy
treatment factor variable. If a participant \(s\) is assigned into the treatment group, his or her \(T_{is}\)
value is 1, and the value \(T_{is}\) is 0 for control-group members. The coefficient \(\delta_{i0j}\) is the expected
value (i.e., intercept) for the control group and \(\delta_{i1j}\) is the difference in the expected values
between the treatment group and the control group (i.e., slope for the treatment effect or effect
size) for outcome \(j\) in study \(i\). In a study, each participant produces two scores (for two
outcomes), and the errors \(e_{isj}\) in the two outcomes are multivariate normally distributed as

\[
\begin{bmatrix}
e_{is1} \\
e_{is2}
\end{bmatrix} \sim N \left( 0, \begin{bmatrix} \sigma_{e1}^2 & \sigma_{e1e2} \\ \sigma_{e2e1} & \sigma_{e2}^2 \end{bmatrix} \right).
\]

(36)
The study-level (level-2) models for outcome 1 from Equation 35 are

\[ \delta_{i01} = \delta_{001}^* + u_{i01} \quad \text{and} \]
\[ \delta_{i11} = \delta_{011}^* + u_{i11}, \quad (37) \]

For outcome 2,

\[ \delta_{i02} = \delta_{002}^* + u_{i02} \quad \text{and} \]
\[ \delta_{i12} = \delta_{012}^* + u_{i12}, \quad (38) \]

where \( \delta_{001}^* \) and \( \delta_{002}^* \) are the means of expected values (intercepts) across studies for outcomes 1 and 2, respectively. The parameters \( \delta_{011}^* \) and \( \delta_{012}^* \) are the means of the treatment effects across studies for each outcome. Level 2 has four random components: two random effects from the expected values in the control groups, \( u_{i01} \) and \( u_{i02} \), and the two random effects from the treatment effects in the treatment group (\( u_{i11} \) and \( u_{i12} \)). The error \( u_{igj} \) where \( g \) is the group (0=control group and 1=treatment group) is multivariate normally distributed as \( \mathbf{u}_i \sim N(0, \tau) \). \( \tau \) is a variance-covariance matrix for the residuals of the two treatment effects and the two expected values for the corresponding outcomes, specifically,

\[
\begin{bmatrix}
  u_{i01} \\
  u_{i02} \\
  u_{i11} \\
  u_{i12}
\end{bmatrix}
\sim N(0, 
\begin{bmatrix}
  \tau_{01}^2 & \tau_{0102} & \tau_{0111} & \tau_{0112} \\
  \tau_{0102} & \tau_{02}^2 & \tau_{1102} & \tau_{1202} \\
  \tau_{0111} & \tau_{1102} & \tau_{11}^2 & \tau_{1211} \\
  \tau_{0112} & \tau_{1202} & \tau_{1211} & \tau_{12}^2
\end{bmatrix}
), \quad (39)\text{or}
\]

The generated data were used to examine four models.
Model 1. In model 1, the two separate expected values ($\delta_{001}^*$ and $\delta_{002}^*$) for each outcome were aggregated and the overall expected value ($\delta_{000}^*$) was estimated. In the same way, one overall treatment effect ($\delta_{010}^*$) across studies was estimated, which was an aggregation of the two separate treatment effects ($\delta_{011}^*$ and $\delta_{012}^*$). The models for the expected values for the study level (Equations 36 and 37) were modified into

$$\delta_{01} = \delta_{000} + u_{01} \quad \text{and}$$

$$\delta_{02} = \delta_{000} + u_{02},$$

(40)

where $\delta_{000}^*$ is the overall expected value and $u_{01}$ and $u_{02}$ are the deviations from the overall expected value $\delta_{000}^*$ (i.e., not from the separate expected values $\delta_{001}^*$ and $\delta_{002}^*$) for the corresponding outcome measures in Equations 37 and 38.

For treatment effects,

$$\delta_{11} = \delta_{010}^* + u_{11} \quad \text{and}$$

$$\delta_{12} = \delta_{010}^* + u_{12},$$

(41)

where $\delta_{010}^*$ is the overall treatment effect in the study level and $u_{11}$ and $u_{12}$ are the study-specific deviations from the overall treatment effect $\delta_{010}^*$, not from the separate treatment-effect means ($\delta_{011}^*$ and $\delta_{012}^*$). Thus, when the treatment effects truly differ by outcome, the variation of the treatment effects around the overall treatment effect ($\delta_{010}^*$) is likely to be greater than the variation about the separate treatment effect means ($\delta_{011}^*$ and $\delta_{012}^*$). This difference would increase the between-studies variances ($\sigma_{u11}^2$ and $\sigma_{u12}^2$) for each treatment effect, and decrease the
covariance between the two effects \( (\sigma_{u_{ij}u_{ij2}}) \) across studies. For example, the estimated overall treatment effect \( \delta_{010}^* = 0.209 \) was in the middle of the two separate treatment effect means (0.109 and 0.310) from model 1, as expected.

In model 1, in the control group with \( n \) participants, the expected values for outcome \( j \) have two sources of errors, the sampling variation \( \frac{\sigma_{e_j}^2}{n} \) (i.e., the error for the expected values) in level 1 and the between-studies variance \( \sigma_{u_{0j}}^2 \) in level 2. The multilevel method assumes the independence of residuals across levels (Raudenbush & Bryk, 2002), and the total variance in the expected value is the sum of the sampling variation and the between-studies variance \( \frac{\sigma_{e_j}^2}{n} + \sigma_{u_{0j}}^2 \).

Similarly, the total covariance between the expected values over two levels is the sum of the sample covariance \( \frac{\sigma_{e,e'}_{ij}}{n} \) in level 1 and the covariance between expected values \( \sigma_{u_{0j}u_{0j'}} \) between the two expected values in level 2 as \( \sigma_{u_{0j}u_{0j'}} + \frac{\sigma_{e,e'}_{ij}}{n} \).

The treatment effect (i.e., effect size) is the mean difference between the control and treatment groups. The expected mean differences between two groups vary within studies and across studies because of the sampling variation and the between-studies variance, respectively. Specifically,

\[
E(\bar{Y}_{yj} - \bar{Y}_{yjC}) = (\delta_{00j} + \delta_{11j} + \bar{e}_{yj} - \bar{e}_{yjC}) - (\delta_{0ij} + \bar{e}_{yj} - \bar{e}_{yjC}) = \delta_{01j} + u_{1j} + \bar{e}_{yj} - \bar{e}_{yjC}.
\]
Unlike the expected value for the control group, for the expected mean differences between two groups, two sources of sampling variation are involved for the effects: sampling variation in the control group $e_{ijc}$ (i.e., the error for the expected values in the control group) and sampling variation in the treatment group $e_{ijt}$ (i.e., the error for the expected value in the treatment group) in Equation 42. If we assume that the sampling variance is the same in both groups, the total covariance of the expected treatment effects is 

\[ \sigma_{u_{ij}, u_{ij}} + \frac{2\sigma_{e_{ij}, e_{ij}}}{n}. \]

**Table 2.1: Comparison of the Two-Level Multivariate Models.**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Fixed effects</th>
<th>Random effects</th>
<th>Random effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intercept</td>
<td>.000</td>
<td>.100</td>
</tr>
<tr>
<td></td>
<td>Outcome 1($\delta_{o1}$)</td>
<td>.000</td>
<td>.100</td>
</tr>
<tr>
<td></td>
<td>Outcome 2($\delta_{o2}$)</td>
<td>.000</td>
<td>.100</td>
</tr>
<tr>
<td></td>
<td>Treatment effect</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Outcome 1($\delta_{o11}$)</td>
<td>.100</td>
<td>.100</td>
</tr>
<tr>
<td></td>
<td>Outcome 2($\delta_{o12}$)</td>
<td>.300</td>
<td>.300</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2</td>
<td>Intercept</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.100 .050]</td>
<td>[.080 .040]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.050 .100]</td>
<td>[.040 .11]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.055 .111]</td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td></td>
<td>[.100 .02]</td>
<td>[.098 .006]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.02 .100]</td>
<td>[.006 .100]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.038 .100]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[1 0.8]</td>
<td>[.990 .790]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.8 1]</td>
<td>.992</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.790 .994]</td>
<td></td>
</tr>
</tbody>
</table>

**Model 2.** Model 2 was developed to investigate how the covariance of the expected treatment effects for two outcomes in the study level reflects the total covariance of the expected
treatment effects over two levels; the participant level and the study level. For this, model 2 assumed that the sample covariance was zero, specifically \( \left[ e_{ij} \right] \sim N(0, \sigma^2_e) \) or \( N\left( 0, \begin{bmatrix} \sigma^2_e & 0 \\ 0 & \sigma^2_e \end{bmatrix} \right) \).

Van den Noortegate et al. (2012) found that the covariance between the treatment effects in level 2 in model 2 was equal to the total covariance of the expected treatment effects over two levels,

\[
\sigma_{n_{pj}n_{pj}} + \frac{2\sigma_{e_i}}{n} \quad \text{(which was} \quad 0.006 + \frac{2 \times 0.790}{50} = 0.038 \text{)} \quad \text{in model 1. Therefore they argued that the ignored sampling covariance between outcomes in level 1 appears as part of the overestimated covariances between the expected treatment effects for the two outcomes in level 2.}
\]

**Model 3.** Model 3 incorporates an additional level (the outcome level as a within-studies level) into model 2 to account for the dependence of outcomes. Levels 1, 2, and 3 are the participant level, the outcome level within studies, and the study level, respectively. Each level accounts for a different source of errors: participant-level sampling error, between-outcomes variance, and between-studies variance, respectively (see Table 2).

The participant-level (level-1) model is

\[
Y_{ij} = \delta_{0j} + \delta_{1j}T + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2_e) \quad : i = 1, \ldots, k; j = 1, \ldots, p; s = 1, \ldots, n. \quad (43)
\]

\( Y_{ij} \) is a value of outcome \( j \) of participant \( s \) in study \( i \). \( T \) is a dummy treatment factor (1= treatment group and 0= control group). \( \delta_{0j} \) and \( \delta_{1j} \) are the expected values in the control group and the treatment effects for outcome \( j \) in study \( i \), respectively.

The outcome level (level 2) model is

\[
\delta_{0j} = \delta^*_{0j} + v_{0j}
\]
\[ \delta_{ij} = \delta_{i0}^* + v_{ij}, \text{ with} \]
\[ \begin{bmatrix} v_{i0j} \\ v_{i1j} \end{bmatrix} \sim N \left( 0, \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix} \right), \quad (44) \]

For the study level (level 3),

\[ \delta_{i0}^* = \delta_{000}^* + u_{i00} \]
\[ \delta_{i10}^* = \delta_{010}^* + u_{i10}, \text{ with} \]
\[ \begin{bmatrix} u_{i00} \\ u_{i10} \end{bmatrix} \sim N \left( 0, \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix} \right), \quad (45) \]

where \( \delta_{000}^* \) and \( \delta_{010}^* \) are the expected value of the mean in the control group, and the treatment effect mean over outcomes within studies, and \( \delta_{000}^* \) and \( \delta_{010}^* \) refer to the overall expected value and the overall treatment effect across studies and outcomes. The errors \( v_{ij} \) and \( u_i \) are assumed to have multivariate normal distributions with mean vectors of zeros and \( 2 \times 2 \) variance-covariance matrices between the expected values and the treatment effects. These are \( \mathbf{V} \) and \( \mathbf{\tau} \) in the outcome level and the study level, respectively. Thus \( v_i \sim N(0, \mathbf{V}) \) and \( u_i \sim N(0, \mathbf{\tau}) \). The expected value and the treatment effect vary over outcomes and studies in this model.

The variances of the expected values for the two outcomes in model 2 are redistributed into the variances over the outcome level (level 2) and the study level (level 3) in model 3, the between-outcomes variance and the between-studies variance in the study level. Thus, the total variance of the expected values was 0.095 (0.04 + 0.055) from levels 2 and 3 in model 3 was midway between the two expected value variances (0.08 and 0.11) for outcome 1 and outcome 2 in the study level in model 2 in Table 2.2. Similarly, the total variance of the treatment effects
over the two levels was 0.099 = 0.061 + 0.038 in model 3. This value was between the two study-level treatment effect variances (0.098 and 0.100) for outcomes 1 and 2 in model 2. The between-studies variances for the expected values and the expected treatment effects in the study level (level 3), 0.055 and 0.038 in model 3 were equal to the covariances among the expected values and the expected treatment effect for two outcomes (level 2) in model 2. Also, considering that the covariances at the study level were the total covariance of the expected treatment effects and the expected values over two levels – the participant level and the study level. The overall expected value, the overall treatment effect across studies, and the corresponding standard errors were not changed from those in models 1 and 2.

**Model 4.** While Model 3 was the three-level model for raw data, Model 4 was the univariate three-level model developed for effect sizes. That is, the generated data were used to compute effect sizes per study. The effect sizes were used to analyze the univariate three-level model. Thus, the expected value for the control group was excluded in the univariate three-level model,

\[
d_{ij} = \delta_{ij} + e_{ij} \quad \text{with } e_{ij} \sim N(0, \sigma_{e_i}^2),
\]

\[
\delta_{ij} = \delta_{ij0} + v_{ij} \quad \text{with } v_{ij} \sim N(0, \sigma_{v_i}^2),
\]

\[
\delta_{ij0} = \delta_{j0} + u_{ij0} \quad \text{with } u_{ij0} \sim N(0, \sigma_{u_i}^2); i = 1, \ldots, k; j = 1, \ldots, p.
\]

Equation 46 states that the dispersion of effect sizes consists of the sampling error, between-outcomes variance, and between-studies variance. Here, \( d_{ij} \) is the observed effect size for outcome \( j \) in study \( i \), estimated using Equations 1 and 2, \( \delta_{ij} \) is the true effect size for the corresponding observed effect size \( d_{ij} \), and \( e_{ij} \) is a residual which is asymptotically normally distributed with a mean of zero and a variance \( \sigma_{e_i}^2 \). The variance \( \sigma_{e_i}^2 \) is replaced by the sampling
variance $\hat{\sigma}_v^2$, estimated using Equation 3. The estimated effect size, $d_{ij}$, is thus normally distributed with mean $\delta_{ij}$ and variance $\sigma_{eij}^2$. Also $\delta_{i0}$ is the effect-size mean over outcomes within studies. The random effect $v_{ij}$ is the deviation of $\delta_{ij}$ from the mean effect $\delta_{i0}$ within study $i$. The true effect sizes $\delta_{ij}$ are normally distributed with mean $\delta_{i0}$ and variance $\sigma_v^2$. Finally, $\delta_{00}$ is the overall mean effect size across studies. The random effects $u_{i0}$ are residuals from the overall effect $\delta_{00}$ for each study. The effect-size means $\delta_{i0}$ over outcomes are normally distributed with a mean of $\delta_{00}$ and the error variance of the random effects $u_{i0}$ equal to $\sigma_u^2$. In model 4 the estimated overall treatment effect $\delta_{00}$ across studies, the corresponding standard error, between-outcomes variance $\sigma_v^2$, and the between-studies variance $\sigma_u^2$ which reflected the total covariance of the expected treatment effects and the expected values over two levels were almost identical to the estimates in model 3.

If the effect sizes are not significantly different across outcomes and studies, and the between-outcomes variance (level 2) and the between-studies variance (level 3) are zero ($H_0: \sigma_v^2 = 0; \sigma_u^2 = 0$), the effect sizes across studies are considered homogeneous and the fixed-effects model ($d_{ij} = \delta_{ij} + e_{ij}$) can be adopted. If the variances are significantly different from zero the random-effects model ($d_i = \delta^* + v_{ij} + u_i + e_i$) would be adopted. The overall population effects are estimated with weights that incorporate both the sampling variance and between-studies variance.

*Homogeneity test in the univariate three-level model.* The between-outcomes and between-studies variances are estimated using REML and are tested with a chi-square test to examine whether effect sizes estimate a common population effect size.
Table 2.2: *Comparison of the Three-Level Models.*

<table>
<thead>
<tr>
<th></th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.003(.021)</td>
<td>0.210(0.021)</td>
</tr>
<tr>
<td>Treatment Effect</td>
<td>0.210(0.021)</td>
<td>0.210(0.021)</td>
</tr>
<tr>
<td>Random effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 3</td>
<td>[ 0.055  -0.035 ]</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>[-0.035  0.038]</td>
<td></td>
</tr>
<tr>
<td>Level 2</td>
<td>[0.040  0.012]</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>[-0.012  0.061]</td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td>0.0992</td>
<td>*</td>
</tr>
</tbody>
</table>

* The sampling variances were computed before analyzing model 5.

**Purpose of This Study**

A major limitation of multivariate meta-analysis conducted using the GLS approach and the multivariate two-level model is that they require the sample covariance matrix of effect sizes within a study. The covariance of the two effect sizes involves the sample correlation between two outcomes. However, it is rare that all primary studies report the sample covariance or correlation between outcomes. Van den Noortegate et al. (2012) developed the univariate three-level approach which does not require the sample covariance matrix of effect sizes to account for the dependence of effect sizes. They found with a simulation study that the univariate three-level model yielded appropriate parameter estimates and standard errors corresponding to those from the multivariate two-level approach.

Thus, the main purpose of this dissertation is to compare the corresponding parameter estimates and the standard errors in the multivariate meta-analysis using the GLS method and the univariate three-level model approach with a simulation study. For this, two multivariate data sets were generated under the random-effects model: one set with two outcomes and the other set with five outcomes. Each generated data set was used to estimate the overall effect over
outcomes, and the specific-outcome effects associated with the overall effect. For example, three parameters were estimated from the data set with two outcomes: the overall effect across two outcomes and the two specific outcome effects. The description of the parameter estimates and their standard errors which I intended to compare in both approaches follows.

**Overall Effects in GLS and Three-Level Approaches**

*GLS approach.* The dispersion of effect sizes is explained by two random variance components in the multivariate random-effects model: the sampling variance and covariance among effect sizes \( \psi(i) \) within study \( i \) and the variance-covariance among effect sizes across studies \( \tau \). Thus, the total variance-covariance among effect sizes is obtained by the sum of within-study-variance-covariance and between-studies-variance-covariance

\[
\psi(i) = \psi(i) + \tau = \begin{bmatrix}
\sigma_{d_1}^2 & \sigma_{d_2}^2 \\
\sigma_{d_1d_2} & \sigma_{d_2}^2
\end{bmatrix} + \begin{bmatrix}
\tau_1^2 & \tau_{12} \\
\tau_{12} & \tau_2^2
\end{bmatrix}.
\]

(48)

The overall mean effect-size estimate (\( \hat{\delta} \)) across outcomes, across all studies is estimated as

\[
\hat{\delta} = (X' \hat{\psi}^{-1} X)^{-1} X' \hat{\psi}^{-1} d,
\]

described in Chapter 2. The corresponding standard error is computed as the square root of \( \psi_\hat{\delta} = (X' \hat{\psi}^{-1} X)^{-1} \). All terms were defined in section 2.4.2.

*Three-level approach.* The univariate three-level model to estimate the overall population effect follows

\[
\begin{align*}
d_{ij} &= \delta_{ij} + e_{ij} \quad \text{with} \quad e_{ij} \sim N(0, \sigma_{e_{ij}}^2) , \\
\delta_{ij} &= \delta_{i0}^* + v_{ij} \quad \text{with} \quad v_{ij} \sim N(0, \sigma_{v_{ij}}^2) , \\
\delta_{i0}^* &= \delta_{i0} + u_{i0} \quad \text{with} \quad u_{i0} \sim N(0, \sigma_{u_{i0}}^2) : i = 1, ..., k; j = 1, ..., p.
\end{align*}
\]

(47)
The coefficient $\delta_{00}^*$ is the overall effect across outcomes and studies. All other terms were defined for Equation 46 or above.

**Outcome-Specific Effects in GLS and Three-level Approaches**

**GLS approach.** The overall effect for each outcome is estimated by

$$\hat{\delta}^* = (\hat{\delta}_1^*, \hat{\delta}_2^*) = (X^* \hat{\psi}^{-1} X^*)^{-1} X^* \hat{\psi}^{-1} d.$$  

The standard errors for the corresponding overall effects are computed as the square roots of the diagonal elements of the variance-covariance matrix among the overall effects,  

$$v_{\hat{\delta}}^* = (X^* \hat{\psi}^{-1} X^*)^{-1}.$$  

**Three-level model.** The univariate three-level model to estimate the overall mean effect for separate outcomes is

\[ d_{ij} = \delta_{ij} + e_{ij} \quad \text{with } e_{ij} \sim N(0, \sigma_{e_{ij}}^2), \]

\[ \delta_{ij} = \delta_{1i}^* X_1 + \delta_{2i}^* X_2 + v_{ij}^* \quad \text{with } v_{ij}^* \sim N(0, \sigma_{v_{ij}}^2), \]

\[ \delta_{1i}^* = \delta_{01}^* + u_{(1)i1} \quad \text{with } u_{(1)i1} \sim N(0, \sigma_{u_{(1)i1}}^2), \]

\[ \delta_{2i}^* = \delta_{02}^* + u_{(2)i2} \quad \text{with } u_{(2)i2} \sim N(0, \sigma_{u_{(2)i2}}^2). \]  

(49)

where $d_{ij}$ is the observed effect size for outcome $j$ ($j = 1, \ldots, p$) in study $i$ ($i = 1, \ldots, k$) and is normally distributed with a mean of the true effect size $\delta_{ij}$ and variance $\sigma_{e_{ij}}^2$. Effects $\delta_{1i}^*$ and $\delta_{2i}^*$ are the true mean effects for outcomes 1 and 2, respectively. The dummy variable $X_j$ shows the presence of an effect size for the $j^{th}$ outcome. For example, if study $i$ estimates the effect size for outcome 1 ($d_{1i}$), the corresponding $X$ value for the slope $\delta_{1i}$ is equal to 1, and equal to 0, otherwise. The random effect $v_{ij}^*$ is the residual within study $i$. The coefficients $\delta_{01}^*$ and $\delta_{02}^*$ are the overall mean effects across studies for outcomes 1 and 2, respectively. The random effects $u_{(1)i1}$ and $u_{(2)i2}$ are the deviations from each overall effect $\delta_{01}^*$ and $\delta_{02}^*$. Thus, the true treatment effect
for each outcome (δ_{i1} or δ_{i2}) is normally distributed with mean δ_{01} or δ_{02} and variance σ^2_{i1} or σ^2_{i2}, respectively.

In summary, the overall population effect δ_{00} and the specific-outcome effects associated with the overall effect, δ_{01} and δ_{02}, and their standard errors in the univariate three-level model are comparable to the \( \hat{\delta} = (X'\psi^{-1}X)^{-1}X'\psi^{-1}d \) and the corresponding elements in the common-effects vector \( \hat{\delta} = (\hat{\delta}_1, \hat{\delta}_2)' = (X'^*\psi'^{-1}X'^*)^{-1}X'^*\psi'^{-1}d \), respectively (Table 2.3).

Table 2.3: Comparisons of Estimates Between Three-Level and GLS Method Approaches.

<table>
<thead>
<tr>
<th></th>
<th>Three-level estimates</th>
<th>GLS method estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>The mean effect size in the random-effects model</td>
<td>( \hat{\delta}_{00} )</td>
<td>( \hat{\delta} = (X'\psi^{-1}X)^{-1}X'\psi^{-1}d )</td>
</tr>
<tr>
<td>The two mean effect sizes in the random-effects model</td>
<td>( \hat{\delta}<em>{01} ) and ( \hat{\delta}</em>{02} )</td>
<td>( \hat{\delta} = (\hat{\delta}_1, \hat{\delta}_2)' = (X'^<em>\psi'^{-1}X'^</em>)^{-1}X'^*\psi'^{-1}d )</td>
</tr>
</tbody>
</table>
CHAPTER 3

METHODOLOGY

This chapter describes the conditions for the simulation study design based on the multivariate two-level model, and the data-generation procedure. In addition, data analysis and assessment criteria needed to evaluate the results are provided. This dissertation applied the same simulation conditions as Van den Noortegate et al. (2012), summarized in Table 4.1. The description of each condition in Table 4.1 follows.

Simulation Conditions

*Number of studies, sample sizes, and outcomes.* The number of studies is set to 25 and 50. The number of units in the higher level is smaller than the number of units in the lower level in the multilevel model. If the number of units is not sufficient, the parameter estimates of the variance components and their standard errors are underestimated (Van Der Leeden & Busing, 1994). Den NorBrowne and Draper (2000) argued that six to twelve units in the highest level are needed to estimate reasonable variance components in multilevel analysis. Kreft and De Leeuw (1998) suggested that at least 30 units are required to produce unbiased parameter estimates and their standard errors.

*The overall effect size.* The overall effects over \( p \) outcomes across studies \( \delta_{100} \) are set to 0.2 and 0.4 which are moderate effects in the social and behavior sciences (Cohen, 1988).

*The deviations of the separate effects from the overall effect size.* The differences of the mean effects for corresponding outcomes around the overall effect sizes are -0.20 and +0.20 for
two outcomes and -0.2, -0.1, 0, 0.1, and 0.2 for five outcomes, relatively. For example, for the effect-size means associated with the overall effect of 0.4 are 0.2 and 0.6 for two outcomes and 0.2, 0.3, 0.4, 0.5, and 0.6 for five outcomes. Most of these are in the small to medium range.

*The variances and covariances of outcomes in level 1.* The variance is 1 for all outcomes. Three covariance values between outcomes were selected, which correspond with the correlation coefficients 0, 0.4, and 0.8 as small, moderate and large correlations, respectively.

*The between-studies variance for each effect in level 2.* The between-studies variance is either 0.1 (i.e., a standard deviation of 0.33).

*The covariances between treatment effects in level 2.* The covariance between two outcome in the participant level is more likely to be higher than that in the study level because the covariance in the participant level arises from data from one study, whereas the covariance at the study level arises from different studies (Riley, 2009). Thus, relatively small covariances were examined, corresponding to the correlation coefficients 0, 0.2, and 0.4.

*The control group in data generation.* In the two-level multivariate model used to generate data, the participant level model was \( Y_{ij} = \delta_{0j} + \delta_{ij} X_i + e_{ij} \). The errors \( e_{ij} \) were assumed to have mean of zero and a variance of 1. In the models, the slope \( \delta_{ij} \) is thus the standardized mean difference (i.e., effect size). Considering that meta-analysis examines effect sizes, the expected value (the intercept) \( \delta_{0j} \) in the control groups was not needed in the model. Thus, the expected term \( \delta_{0j} \) is not needed to generate data, essentially, \( \delta_{0j} = 0 \). The participant level model thus is \( Y_{ij} = \delta_{ij} X_i + e_{ij} \).
Table 3.1: *Simulation Conditions*

<table>
<thead>
<tr>
<th>Description</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of outcomes ($p$)</td>
<td>2</td>
</tr>
<tr>
<td>The number of studies ($k$)</td>
<td>30 and 60</td>
</tr>
<tr>
<td>The group sizes ($n$)</td>
<td>25 and 50</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
</tr>
<tr>
<td>Overall effect ($\delta_{010}$ or $\delta_{010}$)</td>
<td>0.2 and 0.4</td>
</tr>
<tr>
<td>Outcome 1 ($\delta_{011}$)</td>
<td>0 0.2</td>
</tr>
<tr>
<td>Outcome 2 ($\delta_{012}$)</td>
<td>0.4 0.6</td>
</tr>
<tr>
<td>Outcome 3 ($\delta_{013}$)</td>
<td>0.2 0.4</td>
</tr>
<tr>
<td>Outcome 4 ($\delta_{014}$)</td>
<td>0.3 0.5</td>
</tr>
<tr>
<td>Outcome 5 ($\delta_{015}$)</td>
<td>0.4 0.6</td>
</tr>
<tr>
<td>Random effects</td>
<td></td>
</tr>
<tr>
<td>Study level ($\tau$)</td>
<td>$\begin{bmatrix} 0.1 &amp; \tau_{12} \tau_{12} \ \tau_{12} &amp; 0.1 \end{bmatrix}$</td>
</tr>
<tr>
<td>Participant level ($\sigma_{e}^2$ and $\sigma_{y}$)</td>
<td>$\begin{bmatrix} 1 &amp; 0.4, 8 \ 0.4, 8 &amp; 1 \end{bmatrix}$</td>
</tr>
</tbody>
</table>

*Covariances corresponding to the correlation coefficients 0, 0.2, and 0.4.*

**Data Generation Procedure**

Data were generated in R (version 2.13.2) based on the conditions in Table 4.1. The study-level (level 2) data were generated first and they were combined with the level 1 information to generate outcome scores for each participant.

**Level 2 (Study Level)**

First, the independent random effects were obtained for each treatment for the corresponding outcome measures. The random effects (i.e., errors) were normally distributed with a mean of zero and a variance 1. Second, the uncorrelated random components were converted into correlated values (the multivariate data) $u_y$ using the Cholesky decomposition of the population values of the variance-covariance matrix among the treatment effects in Table 4.
Third, the generated random-effects values $u_{ij}$ were combined with the corresponding overall effect-parameter values $\delta_{01j}$ to estimate the true treatment effects (effect sizes) $\delta_{11j}$ for each outcome across studies, specifically, by computing $\delta_{11j} = \delta_{01j} + u_{ij}.$

**Level 1 (Participant Level)**

First, the independent random effects for each outcome were generated; these were normally distributed with a mean of zero and a variance 1. Second, the uncorrelated random components were converted into correlated values (the multivariate data) $e_{ij}$ using the Cholesky decomposition of the population values of the variance-covariance matrix among outcomes in Table 4. Third, one constant variable $X = 1$ was created. Finally, the generated random-effect values $e_{ij}$ were combined with the corresponding treatment effects $\delta_{11j}$ which were multiplied by $X.$ Each pseudo participant then had outcome scores for the corresponding outcomes $Y_{ij}$ in

$$Y_{ij} = \delta_{11j}X + e_{ij}.$$  

Based on the simulation conditions (summarized in Table 4) the number of outcomes, sample sizes, number of studies, the magnitude of overall effect size, and the magnitudes of covariances in level 1 and 2, $2\times2\times2\times2\times3\times3=144$ combinations were created. For each combination I generated 1000 data sets, thus 144,000 datasets were generated in total. Each data set was used to compute effect sizes, their variances, and covariances. The multivariate meta-analysis was then conducted with the GLS method and the univariate three-level model approach.

**Data Analysis**

The effect sizes were analyzed using the univariate three-level model and the multivariate GLS method to estimate the overall effect over outcomes across studies and the specific-outcome
effects associated with the overall effect. SAS/PROC MIXED (SAS Institute, 1995) and R were used to perform the univariate three-level model analysis with Restricted Maximum Likelihood (RML) estimation and the multivariate meta-analysis using GLS methods, respectively. Across 1000 replications, the averages of the parameter estimates and the SEs of the parameter estimates, as well as the SDs of the estimates (the empirical SEs of the estimate) were computed. These indices were used to compute the performance measures: the bias of the parameter estimates and the bias of the SEs of the parameter estimates. The two performance measures were compared based on the sample size, the number of studies, the magnitudes of the overall effects, the size of covariance in levels 1 and 2. Finally, ANOVA tests were conducted to further understand the impact of the five factors on the biases of the SEs of the parameter estimates.

**Relative Percentage Bias**

In the simulation study, I evaluated the amount of bias and variability of the parameter estimates to compare the two methods or approaches. The method which yields the least biased parameter estimates with smallest variability is considered the better method (Burton, Altman, Roystom, & Holder, 2006).

Bias indicates the difference between the parameter estimate and the corresponding population value. Two types of biases were calculated: relative percentage bias for the parameter estimates, and relative percentage bias for the standard error of the parameter estimate. While less than 5% bias was considered as good for the parameter estimates, less than 10% was treated as acceptable for the standard errors (Hoogland & Boomsma, 1998).

Relative percentage bias for parameter estimate is computed as

\[ \text{Relative Bias}(\hat{\beta}) = \frac{\bar{\hat{\beta}} - \beta}{\beta} \times 100\%, \]  

(51)
where $\beta$ is the population parameter, $\hat{\beta}$ is the mean of the parameter estimates across all replications.

If the true value of a parameter estimate is zero, its relative percentage bias would be infinite in Equation 51. For such cases the raw biases $\bar{\beta} - \beta$ were calculated. Raw bias beyond the range from $\frac{1}{2}SD(\hat{\beta})$ to $2SD(\hat{\beta})$ is considered troublesome (Sinharay, Stern, & Russell, 2001).

When a parameter estimate does not show bias, the accuracy of the estimate is evaluated by the standard error of the estimates, because the SE is used to test the parameter values and to construct confidence intervals. Relative percentage bias for the standard error of the parameter estimates was calculated as

$$\text{Relative Bias}(SE(\hat{\beta})) = \frac{SE(\hat{\beta}) - SD(\hat{\beta})}{SD(\hat{\beta})} \times 100\%,$$

where $SE(\hat{\beta})$ was the standard-error estimate for the corresponding parameter from each replication. $SE(\hat{\beta})$ was the mean of the standard-error estimates across all replications. Finally, $SD(\hat{\beta})$ was the empirical standard deviation of the parameter estimates over all replications in each condition.

**ANOVA Tests**

In order to examine the effects of the five design factors on the relative biases of the standard errors of the parameter estimates, analysis of variance (ANOVA) tests were conducted. Three assumptions are needed for ANOVA tests: normality of errors, equality of variance across populations, and independence of errors. In this dissertation, each observation was the mean of
parameter estimates across 1000 replication (i.e., 1000 data sets). The 1000 data sets were randomly drawn for each condition and the error scores were independent. Thus, the errors of the averaged parameter estimates from different conditions were independent. The sample sizes were the same across groups, thus the homogeneity of error variances was not a concern.

The difference scores of relative biases between the GLS and the univariate three-level model were computed to examine impact of the five design factors on the difference scores in a 5-way ANOVA (without interaction effects). Factors were the sample size, the number of studies, the size of overall effects, and the magnitude of covariance in levels 1 and 2. Two types of outcome measures (i.e., difference scores) were created: the overall effect across outcomes and the individual specific-outcome effects associated with the overall effect. In order to further understand the effects of factors on the bias values, Partial Eta square values were examined. Partial Eta Square is the proportion of variation in the outcome variable explained by a main factor when other factors in the model are controlled.
CHAPTER 4

RESULTS

In each condition, the two performance measures were evaluated for the univariate three-level-model and GLS approaches: the bias of parameter estimates and the bias of the standard error of the parameter estimates. Finally, ANOVA tests were conducted to further understand the impact of the five factors (the size of the overall effect, the sample size, the number of studies, the size of overall effects, and the size of covariances in level 1 and 2) on the difference of the biases of the SE of the parameter estimates between the two approaches.

Convergence

All replications converged in the multivariate meta-analyses using the GLS method. No replication failed to converge for the overall effects in the univariate three-level model. However, the frequency of replications that converged ranged from 953 to 1000 (among 1000 replications) for specific-outcome effects. The range was from 957 to 1000 for five outcomes, and from 962 to 1000 for two outcomes. The average number of replications that converged was 982.40 across two and five outcomes.

The replications that did not converge were omitted from this dissertation, and that led to some degree of non-comparability with the GLS data set where all replications were estimated properly. Thus, I reran several conditions with the relatively large number of replications that did not converge (i.e., the sensitivity analysis) to check whether the missing values made a difference. For example, I reran the condition with 969 converged replications for two
outcomes. This was associated with the overall effect of 0.2, sample size 30, 50 studies, and covariances of 0 and 0.04 in levels 1 and 2, respectively. I also chose this case because the biases seemed to increase with increasing of covariance in level 2 in this condition while the biases seemed to decrease in other conditions. For this, I removed the 29 biases in the GLS approach for data sets that did not converge, and obtained the biases only for combinations that converged in both approaches (Appendix D). The results were rarely different from datasets that excluded the replications that did not converge.

**Bias**

**Parameter Estimates**

Both the univariate three-level model and the GLS approaches did not yield any important relative percentage bias values for the parameter estimates. No relative percentage bias values exceeded 0.1%. The raw biases $\beta - \hat{\beta}$ were calculated when the true value of a parameter estimate was zero because its relative percentage bias would be infinite in Equation 51. Raw bias beyond the range from $\frac{1}{2}SD(\hat{\beta})$ to $2SD(\hat{\beta})$ was considered troublesome (Sinharay et al., 2001). The outcome-specific effects based on true effects with value zero did not show a raw bias considered troublesome.

**SE of Parameter Estimate**

Only 15 (i.e., about 1%) relative percentage biases of the standard errors of the parameter estimates were larger than 10%; those ranged from 10.05% to 13.54%. 14 of these biases were from the univariate three-level approach with a covariance of 0.02 in level 2, associated with two
outcomes. Even though no important relative percent biases were considered troublesome, I examined the biases to investigate potential patterns across conditions.

*Relative Bias of the SE of Overall Effects.* For the overall effects associated with two outcomes in Figure 4.1, the relative biases from the three-level model approach were relatively larger and more varied than those from the GLS-approach estimates.

![Figure 4.1](image.png)

*Figure 4.1.* Relative percentage bias of the SE in the overall effects for two outcomes in the random-effects model.

The bias decreased as the covariance increased in levels 1 and 2 in the three-level model approach. The difference in biases between the two approaches decreased with an increasing number of studies, and with larger covariances in level 1 and 2. Specifically, almost no difference was seen between the two approaches when the covariance was 0.04 in level 2, and the difference was largest when the covariance was zero in level 2. The difference in biases
between the two approaches was relatively large when the covariance was zero in levels 1 and 2. The sample sizes did not seem to relate to the differences in bias values between the two approaches.

In the overall effects for five outcomes in Figure 4.2, the biases of the univariate three-level-model approach were relatively small in comparison to the bias for the overall effects for two outcomes.

*Figure 4.2. Relative percentage bias of the SE in the overall effects for five outcomes in the random-effects model*

The difference in bias between the two approaches was relatively large when the covariance was zero in levels 1 and 2, which was also seen for the overall effects for two outcomes.
outcomes. The sample size and the number of studies did not seem to affect the difference in biases between the two approaches.

_Bias of the SEs of Specific-Outcome Effects._ For the two specific-outcome effects in Figure 4.3, the biases decreased with a decrease in the number of studies, associated with the overall effect of 0.2, while the number of studies did not seem to affect the amount of bias for the overall effect of 0.4. The difference in biases between the two approaches seemed to decrease with increasing numbers of studies.

Figure 4.3. Relative percentage bias of the SE of the outcome-specific effects for two outcomes in the random-effects model
For the five specific-outcome effects associated with the corresponding overall effect in Figure 4.4, the biases in the GLS approach were relatively larger than those in the three-level model. In both approaches, the biases reduced as the sample size increased. The difference between biases of the two approaches seemed to decrease with an increasing number of studies and an increase in the sample size.

Figure 4.4. Relative percentage bias of the SE in the outcome-specific effects for five outcomes in the random-effects model

ANOVA Tests

Table 4.1 shows the partial eta squares for the main effects of the five factors on the outcome variables. Most main-effects factors significantly affected the difference in the bias of
the standard errors of the parameter estimates at the alpha level of 0.05. However, their partial
etta squares were not large, ranging from 0 to .64. The average of the partial eta squares was .18
across conditions.

At the alpha level of .05, except for the size of overall effects, all main effects
significantly affected the differences in the biases between the two approaches for the overall
effects associate with two outcomes. All five factors were significant for the five specific-
outcome effects. For overall effects associated with two outcomes, the partial eta squares of
covariance in level 1 and 2, the number of studies and the sample size are .64, .59, .25, and .10,
respectively, holding all other factors constant, which indicates that 64%, 59%, 25%, and 10% of
variation in the difference of bias between two approaches were explained by covariance in
levels 1 and 2, the number of studies, and the sample size, respectively. This result was
consistent with the results reported in Figure 4.1 for the plot of the bias of SEs. Figure 4.1
showed that the difference of biases between two approaches decreased with an increasing
number of studies, and with larger covariances in level 1 and 2. However, the sample sizes did
not seem to affect the difference in bias values between the two approaches, which was
consistent with the fact that the partial eta square of the sample size was not large (.1) even
though the sample size had a significant effect on the difference in bias between two approaches
in Table 4.

For the five outcome-specific effects, relatively large partial eta squares were found for
number of studies and sample sizes, .33 and .29, respectively. Only about 10% of the variation
was explained by covariance values in levels 1 and 2. This result was consistent with the finding
in Figure 4.4. Figure 4.4 showed that the difference in biases between the two approaches
seemed to decrease with an increasing number of studies and an increase in the sample size.
However, the covariances in levels 1 and 2 did not seem to affect the difference in biases between the two approaches.

The sample size and the number of studies significantly affected the difference of the biases of the two specific-outcome effects, but their eta-squares were small (.06 and .07, respectively). 17% and 18% of the variation of the difference of biases was explained by the covariances in level 1 and 2 for the overall effects associated with five outcomes. These results were consistent with the findings in Figures 4.2 and 4.3 that the difference in biases between the two approaches seemed to be small over all simulation conditions.

Table 4.1: Partial Eta Squares for Difference in SE Bias Between Two Approaches in the Random-Effects Model

<table>
<thead>
<tr>
<th>Outcome Number</th>
<th>The Difference in Biases of SEs</th>
<th>Size of Overall Effect Size</th>
<th>Sample Size</th>
<th>Number of Studies</th>
<th>Level 1 Covariance</th>
<th>Level 2 Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two Overall Effect</td>
<td>.012</td>
<td>.100*</td>
<td>.254*</td>
<td>.589*</td>
<td>.641*</td>
<td></td>
</tr>
<tr>
<td>Effect-Size Means</td>
<td>.001</td>
<td>.063*</td>
<td>.070*</td>
<td>.000</td>
<td>.018</td>
<td></td>
</tr>
<tr>
<td>Five Overall Effect</td>
<td>.009</td>
<td>.029</td>
<td>.014</td>
<td>.172*</td>
<td>.188*</td>
<td></td>
</tr>
<tr>
<td>Effect-Size Means</td>
<td>.023</td>
<td>.290*</td>
<td>.327*</td>
<td>.095*</td>
<td>.107*</td>
<td></td>
</tr>
</tbody>
</table>

*Significant at the alpha level of 0.05
CHAPTER 5

CONCLUSIONS AND DISCUSSION

Multivariate meta-analysis methods typically assume dependence of effect sizes. Gleser and Olkin (1994) applied the GLS approach to develop the multivariate meta-analysis model. In the GLS approach, sampling covariances among effect sizes within a study are computed to handle the dependence of effect sizes. Van Den Noortegate et al. (2012) developed the univariate three-level model approach, which does not require the sampling covariances, as an alternative for the multivariate meta-analysis.

The main purpose of this dissertation was to compare the corresponding parameter estimates and their standard errors in the two approaches in multiple-endpoints experimental-design studies. I did this in two ways: examining the relative percentage bias for the parameter estimates and their standard errors, and conducting a five-way ANOVA test. Both the univariate three-level model and GLS approaches did not yield any important relative percentage bias for the parameter estimates or their standard errors. In the ANOVA, the partial eta squares of the five main factors were not large. They ranged from 0 to .64, and the average of the partial eta squares was .18. This indicates that the variation of the difference of the bias between two approaches was not much explained by the five factors. This result was consistent with my expectation. I expected that both the univariate three-level model approach and the GLS approach would produce similar parameter estimates and standard error, thus they are reasonable alternatives. Thus, it is concluded that the univariate three-level model yielded the appropriate
parameter estimates and standard errors corresponding to those from the multivariate meta-analysis using the GLS approach.

The study has several limitations. First, I simulated only balanced data sets with the same sample size, the same number of studies, and the same number of outcomes. However, it rarely occurs that all studies have the same number of effect sizes with a balanced data set. Thus, it would be useful to simulate data with unbalanced data sets in the future. Second, only the same value of the between-studies variance over conditions was applied. Third, I used two different values of the between-studies covariance, but the covariance was same for all pairs of outcomes. For example, in studies with four outcomes where two outcomes are math scores and the other two outcomes are reading scores, the two outcomes within each content area are likely to be correlated strongly each other. In contrast, math scores may be more weakly correlated with reading scores. Therefore, further studies could consider different values of the covariances between pairs of outcomes.

Fourth, the relative percentage bias of the parameter estimates was measured for two overall effects, 0.2 and 0.4. Less than 5% bias was considered as acceptable for the parameter estimates. However, the size of the overall effect relate to the real magnitude of the bias. In other words, 5% of an overall effect of 0.2 is 0.01, while 5% of an effect with value 0.4 is 0.02. That is, a greater degree of variation is considered acceptable for larger overall effects, in comparison to that for the small overall effect. So it is not quite comparable to apply the 5% criteria to parameters whose magnitudes are different. However, considering that the sampling variance of the effect size is larger as the effect size increases (see Equation 3), less than 5% bias could be considered as acceptable for the parameter estimates regardless of the magnitude of the parameters.
Finally, while all replications converged for the GLS approach and for the overall effect in the univariate three-level model, up to 2% of replications did not converge for some conditions for the specific-outcome effect means in univariate three-level model approach. Nonconvergence can result from a number of causes, thus a future study could consider why nonconvergence occurred, as well as how to reduce the number of replications that do not converge in the univariate three-level model.

This dissertation found that the univariate three-level model yielded appropriate parameter estimates and standard errors corresponding to those in the multivariate meta-analysis using the GLS approach. Considering it is rare that the primary studies do not report the sampling covariances or correlations between outcomes, the univariate three-level model is both applicable and relatively simple in comparison to the multivariate meta-analysis using GLS method. Also, there is no “canned program” for the GLS approach; that makes GLS harder to run for meta-analysts who do not program themselves. Thus, the univariate three-level model is more convenient for the researcher who is not familiar with programming.
APPENDIX A

R-CODES FOR DATA GENERATION

rm(list=ls())

num.rep<-1000

RO<-RO_se<-matrix(NA, num.rep, 1)

R12<-R12_se<-matrix(NA, num.rep, 2)

eff.size.all<-matrix(NA,1,5+1)

ss<-60
K<-25

g110<-.2

g120<-.6

L1_cov<-.4

tau12<-.02

g010<0

g020<0

tau1<tau2<-.1

for (rep in 1:num.rep){
    level2<-matrix(NA,K,8)
    B01<-B02<-B11<-B12<-matrix(NA,K,1)
    cov.out<-matrix(NA,K,2)
    dev<-matrix(0, 2*K,2*K)
    design1<-matrix(1,2*K,1)
    design2<-matrix(rep(c(1,0,0,1),K),2*K,2, byrow=TRUE)
    Outcome1<Outcome2<-matrix(NA,K,ss)
    colnames(level2)<-c("u01", "u02", "u11", "u12","B01","B02","B11","B12")
    eff.size<-matrix(NA, 2*K, 5)
    colnames(eff.size)<-c("Study", "Outcome", "ES", "Var", "W")

    C<-matrix(c(.1,.05,-.025, -.025,.05,.1,-.025,-.025,.05,-.025,.1,.02, -.025,-.025,.02,.1, -.025,.02, -.025,.1,.02, -.025,.02, .1,.02, -.025,.02, -.025,.02, .1,.02),4)
    C<-matrix(-.025, 2*2, 2*2)
    diag(C)<-tau1
    C[1,2]<-.05
    C[2,1]<-.05
ran.u01<-rnorm(K,0,1)
ran.u02<-rnorm(K,0,1)
ran.u11<-rnorm(K,0,1)
ran.u12<-rnorm(K,0,1)
R<-as.matrix(cbind(ran.u01,ran.u02,ran.u11,ran.u12))
U<-chol(C)
corr.U<-R%*%U

for(i in 1:K){
  B01[i]<-g010+corr.U[i,1]
  B02[i]<-g020+corr.U[i,2]
  B11[i]<-g110+corr.U[i,3]
  B12[i]<-g120+corr.U[i,4]
}

level2<-cbind(corr.U, B01,B02,B11,B12)
colnames(level2)<-c("u01", "u02", "u11", "u12","B01","B02","B11","B12")

for (l2 in 1:K){
  ran.e1<-rnorm(ss/2,0,1)
  ran.e2<-rnorm(ss/2,0,1)
  C0<-matrix(L1_cov,2,2)
  diag(C0)<-1
  U0<-chol(C0)
  R0<-as.matrix(cbind(ran.e1,ran.e2))
  corr.E0<-R0%*%U0
  data.E0<-cbind(0,corr.E0)

  ran.e1<-rnorm(ss/2,0,1)
  ran.e2<-rnorm(ss/2,0,1)
  C1<-matrix(L1_cov,2,2)
  diag(C1)<-1
  U1<-chol(C1)
  R1<-as.matrix(cbind(ran.e1,ran.e2))
  corr.E1<-R1%*%U1
  data.E1<-cbind(1,corr.E1)

corr.E<-rbind(data.E0,data.E1)
colnames(corr.E)<-c("X", "E1", "E2")

for(l1 in 1:ss)
{
  Outcome1[l2,l1]<-level2[l2,5]+level2[l2,7]*corr.E[l1,1]+corr.E[l1,2]
  Outcome2[l2,l1]<-level2[l2,6]+level2[l2,8]*corr.E[l1,1]+corr.E[l1,3]
}

link1<-paste("C:/dis/Data/Cov2_02/Cov1_4/Over_M_4/2/K_25/SS_30")
link2<-paste(link1, "/rep.",rep, sep="")
link3<-paste(link2,"/study.", l2,".txt", sep="")

54
data<-cbind(l2, corr.E[,1], Outcome1[l2,], Outcome2[l2,])
APPENDIX B

R-CODES FOR PARAMETER ESTIMATE (GLS)

######################################################################## Effect size and Variance ###############################

group.M0<-apply(data[1:ss/2,], 2,mean)
group.M1<-apply(data[(ss/2+1):ss,], 2,mean)
group.V0<-apply(data[1:ss/2,], 2,var)
group.V1<-apply(data[(ss/2+1):ss,], 2,var)
eff.size[2*l2-1,1]<-eff.size[2*l2-1]<-l2
eff.size[2*l2-1,2]<-(2*l2-1)
eff.size[2*l2,2]<-2*l2
s_out2<-sqrt((group.V0[4]+group.V1[4])/2)
d_out1<-(group.M1[3]-group.M0[3])/s_out1
d_out2<-(group.M1[4]-group.M0[4])/s_out2
m=ss-2
cm=1-3/(4*m-1)
d_out1<-d_out1/cm
d_out2<-d_out2/cm
d_var_out1 <--(8+(d_out1*d_out1))/(2*ss)
W_out1<-(1/d_var_out1
d_var_out2 <--(8+(d_out2*d_out2))/(2*ss)
W_out2<-(1/d_var_out2
eff.size[2*l2-1,3]<-d_out1
eff.size[2*l2,3]<-d_out2
eff.size[2*l2-1,4]<-d_var_out1
eff.size[2*l2,4]<-d_var_out2
eff.size[2*l2-1,5]<-W_out1
eff.size[2*l2,5]<-W_out2

########################################################################Covariance########################################################################

cov.out[l2,1]<-cor(data[,3],data[,4])
cov.out[l2,2]<-((4/ss)*cov.out[l2,1]+(d_out1*d_out2*cov.out[l2,1]*cov.out[l2,1])/(2*ss)
dev[(2*l2-1),(2*l2-1)]<-(d_var_out1+tau1
dev[(2*l2),(2*l2)]<-(d_var_out2+tau2
dev[(2*l2),(2*l2-1)]<cov.out[l2,2]+tau12
dev[(2*l2-1),(2*l2)]<-(dev[(2*l2),(2*l2-1)]

}

or<-solve(t(design1)%*%solve(dev)*%design1)%*%t(design1)%*%solve(design1))*%as.matrix(eff.size[,3])
nor_se<-sqrt(solve(t(design1)%*%solve(dev)*%as.matrix(eff.size[,3]))

56
r12<-solve(t(design2)%*%solve(dev)%*%design2)%*%t(design2)%*%solve(dev)%*%as.matrix(eff.size[,3])

r12_se<-(solve(t(design2)%*%solve(dev)%*%design2))

r12_se2<-sqrt(diag(r12_se))

SE_GLS<-c(ro_se,r12_se2)

eff.size2<-cbind(rep,eff.size)

#write.table(eff.size2, paste(link1, "/data/Effect_Size_",rep,".txt", sep=""),quote=FALSE, col.names=TRUE, row.names=FALSE, sep="\t")

print(rep)
RO[rep,1]<-ro
RO_se[rep,1]<-ro_se
R12[rep,]<-r12
R12_se[rep,]<-r12_se2

Raw_data<-cbind(RO,R12)

Raw_mean<-apply(Raw_data,2,mean)

Raw_SD<-apply(Raw_data,2,sd)

Bias_ParEst<-((Raw_mean-c(.4,.2,.6))/c(.4,.2,.6))*100

Bias_SE<-((SE_GLS-Raw_SD)/Raw_SD)*100

asd<-cbind(Raw_mean,Raw_SD, SE_GLS, Bias_ParEst, Bias_SE)

link4<-paste(link1,"/raw_data", sep="")

dir.create(link4)

write.table(Raw_data, paste(link4, name_1, sep=""),quote=FALSE, col.names=FALSE, row.names=FALSE, sep="\t")
APPENDIX C

SAS-CODES FOR PARAMETER ESTIMATE (THREE-LEVEL)

```
libname Lib_sjl "C:\dis\Data\Cov2_0\Cov1_4\Over_M_4\5\K_50\SS_60\raw_data";

/* Overall Outcome Model*/
ods output SolutionF(persist=proc)=prms;
ods output LRT = lrt;
ods listing close;

proc mixed data=data_i method=reml;
class STUDY OUTCOME;
model ES= /solution ddfm=satterthwaite;
weight W;
random intercept/sub=STUDY;
random intercept/sub=OUTCOME;
parms 1 1 1 /hold= 3;
run;
ods output close;
ods listing;

data sj0; set prms; Out_0=Estimate; SE0=StdErr; probt0=probt; Bias0=((Out_0-.4)/.4);
    keep Out_0 SE0 probt0 Bias0; run;
******;
```
APPENDIX D

WITHOUT AND WITH REPLICATIONS THAT DID NOT CONVERGE

Two Effect-Size Means for the Overall Effect of 0.4 (n=30) with Replications That Did Not Converge

Two Effect-Size Means for the Overall Effect of 0.4 (n=30) without Replications That Did Not Converge
REFERENCES


BIIOGRAPHICAL SKETCH

Seungjin Lee was born in Seoul, South Korea. Seungjin received his BA degrees in Economics from Hanshin University, Gyeonggdo, Korea in 1997. In 2007, he started the Measurement and Statistics program in the College of Education at Florida State University. She earned an MS degree in 2008, and continued to work on her PhD in the same program. During his stay at Florida State University, he worked as a teaching assistant for numerous courses and research assistant for numerous projects. Currently, she is working with the Division of Accountability, Research and Measurement at Florida Department of Education as a psychometrician. In her dissertation, she focused on the use of multilevel and GLS approaches in multivariate meta-analysis.