

The Pennsylvania State University

The Graduate School

**A MULTI-LEVEL MIXED-EFFECTS MODEL  
FOR INDIVIDUAL PARTICIPANT DATA META-ANALYSIS**

A Dissertation in

Biostatistics

by

Ying Zhang

© 2019 Ying Zhang

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

August 2019

The dissertation of Ying Zhang was reviewed and approved\* by the following:

Vernon M. Chinchilli  
Distinguished Professor and Chair, Department of Public Health Sciences  
Professor of Statistics  
Dissertation Advisor  
Chair of Committee

Lan Kong  
Professor, Department of Public Health Sciences

David T. Mauger  
Professor, Department of Public Health Sciences

Christopher N. Sciamanna  
Professor and Vice Chair for Research Affairs, Department of Medicine  
Professor, Department of Public Health Sciences

Arthur Berg  
Associate Professor of Public Health Sciences  
Head of Biostatistics PhD Program

\*Signatures are on file in the Graduate School

## ABSTRACT

Individual participant data (IPD) meta-analysis that combines and analyzes raw data from multiple studies is considered to be more powerful and flexible compared with meta-analysis based on summary statistics. A one-stage meta-analysis method based on IPD models all studies simultaneously, accounting for the clustering of participants within each study. We propose a statistical model that is a combination of a mixed-effect model and a multi-level model such that the new model (1) contains fixed and random effects at each level, such as participant and study, and (2) allows each study to have different lengths of follow-ups and different sets of covariates for adjustment. The model is firstly developed for data with continuous outcomes, and then extended to outcomes from an exponential family, such as binary, categorical, count outcomes, etc. We conducted simulation studies to compare the proposed model with other meta-analysis methods in 40 simulation scenarios with continuous and binary outcomes, respectively. We applied the proposed model to three randomized studies from the National Heart, Lung, and Blood Institute to evaluate the effect of reducing sodium intake on blood pressure control. The simulation studies indicate the proposed model properly estimates the variability of data and maintains around 95% coverage probability, while other meta-analytic methods tended to underestimate the variation and suffered insufficient coverage probability when heterogeneity increases and sample size decreases. The proposed one-stage model for IPD meta-analysis provides more flexibility than two-stage methods and fixed-effects methods. It can properly account for the variability of the data and provide reasonable pooled estimations, especially when a large amount of heterogeneity exists across studies.

Keywords: meta-analysis; individual participant data; longitudinal data; mixed-effects model; multi-level model; exponential family outcome

## TABLE OF CONTENTS

LIST OF FIGURES.....	vi
LIST OF TABLES.....	vii
ACKNOWLEDGEMENTS.....	viii
Chapter 1 Introduction.....	1
Chapter 2 Literature Review.....	3
2.1 Two-Stage IPD Meta-Analysis Methods.....	3
2.2 One-Stage IPD Meta-Analysis Methods.....	4
2.3 Methods for Different Outcomes and Longitudinal Data.....	4
Chapter 3 The One-Stage Multi-Level Mixed-Effects Model for IPD Meta-Analysis with Continuous Outcomes.....	6
3.1 Methods.....	6
3.1.1 Model Introduction.....	6
3.1.2 Maximum Likelihood (ML) Estimation.....	9
3.1.3 Restricted Maximum Likelihood (REML) Estimation.....	16
3.2 Statistical Inference.....	22
3.2.1 Inference for Fixed Effects $\beta$ .....	22
3.2.2 Prediction of Random Effects $\tilde{\gamma}_k$ .....	25
3.3 Simulation Study.....	25
3.3.1 Simulation Design.....	25
3.3.2 Simulation Results.....	30
3.3.3 Supplemental Simulation Study.....	35
3.4 Real Data Application.....	37
3.4.1 Example Data - Blood Pressure Studies.....	37
3.4.1 Results.....	39
Chapter 4 The One-Stage Multi-Level Mixed-Effects Model for IPD Meta-Analysis with Outcomes from An Exponential Family.....	43
4.1 Methods.....	43
4.1.1 Model Introduction.....	43
4.1.2 Quasi-likelihood Methods.....	46
4.1.3 Laplace's Method.....	52
4.1.3 Gaussian Quadrature.....	55
4.1.4 Simulated Maximum Likelihood.....	58
4.1.4 Expectation-Maximization (EM) algorithm.....	59
4.2 Inference for Fixed Effects $\beta$ .....	61
4.3 Simulation Study.....	62
4.3.1 Simulation Design.....	62

4.3.2 Simulation Results.....	66
4.4 Real Data Application.....	70
4.4.1 Example Data - Blood Pressure Studies .....	70
4.4.2 Results .....	72
Chapter 5 Summary and Future Work.....	77
Appendix A Chapter 3.....	81
A.1 Simulation Study.....	81
A.2 Real Data Application .....	99
Appendix B Chapter 4.....	102
B.1 Simulation Study.....	102
B.2 Real Data Application .....	121
Bibliography .....	127

## LIST OF FIGURES

Figure 3-1: 3-level and 4-level Simulated Data Structure .....	26
Figure 3-2: Simulation results for the 3-level data structure with common covariates.....	31
Figure 3-3: Simulation results for the 3-level data structure with distinct covariates.....	32
Figure 3-4: Simulation results for the 4-level data structure with common covariates.....	34
Figure 3-5: Simulation results for the 4-level data structure with distinct covariates.....	34
Figure 3-6: Individual and pooled estimates with outcome, the change of SBP.....	39
Figure 3-7: Meta-analysis methods comparison with outcomes, the change of SBP.....	40
Figure 3-8: Individual and pooled estimates with outcome, the change of DBP.....	41
Figure 3-9: Meta-analysis methods comparison with outcomes, the change of DBP.....	41
Figure 4-1: Simulation results for 3-level data structure with common covariates.....	67
Figure 4-2: Simulation results for 4-level data structure with common covariates.....	68
Figure 4-3: Simulation results for 3-level data structure with distinct covariates.....	69
Figure 4-4: Simulation results for 4-level data structure with distinct covariates.....	70
Figure 4-5: Individual estimates and pooled estimates with the binary SBP outcome.....	73
Figure 4-6: Meta-analysis methods comparison with the binary SBP outcome. ....	73
Figure 4-7: Individual estimates and pooled estimates with the binary DBP outcome.....	74
Figure 4-8: Meta-analysis methods comparison with the binary DBP outcome.....	74
Figure 4-9: Individual estimates and pooled estimates with the ordinal SBP outcome. ....	75
Figure 4-10: Meta-analysis methods comparison with the ordinal SBP outcome. ....	76
Figure 4-11: Individual estimates and pooled estimates with the ordinal DBP outcome. ....	76
Figure 4-12: Meta-analysis methods comparison with the ordinal DBP outcome.....	76

**LIST OF TABLES**

Table 3-1: Sample size design in the simulation study with continuous outcomes. ....	28
Table 3-2: Model specifications of the four test models. ....	30
Table 3-3: Model specifications of the supplemental simulation study. ....	35
Table 3-4: Results for the supplemental simulation study.....	36
Table 3-5: Study information summary of the NHLBI studies.....	37
Table 3-6: Model fitting with NHLBI studies. ....	38
Table 4-1: Sample size design in the simulation study with binary outcomes. ....	64
Table 4-2: Study information summary of the NHLBI studies.....	71
Table 4-3: Model fitting with NHLBI studies. ....	72

## ACKNOWLEDGEMENTS

I would like to sincerely express my appreciation to my dissertation advisor, Professor Vernon M. Chinchilli, for his persistent support to my dissertation research in the past four years. I am very grateful for his tremendous guidance, patience, inspiration, and knowledge. I would also like to thank my committee members Professor Lan Kong, Professor David T. Mauger and Professor Christopher N. Sciamanna for their valuable suggestions and comments all the way through my dissertation research. I would like to thank the National Heart, Lung, and Blood Institute for providing the data access to support this research work.

In addition, I would like to thank all the faculty members in the Department of Public Health Sciences at The Pennsylvania State University for their support during my coursework and dissertation research. I would like to thank my friends and classmates for their companionship and encouragement. I would also like to thank my advisors during my master program education, Dr. Xiaofei Wang at Duke University, and Dr. Herbert Pang at The University of Hong Kong.

Last but not the least, I would like to thank my parents for their unconditional love and support all the time in my life.



## **Chapter 1**

### **Introduction**

Meta-analysis aims to combine and analyze multiple related studies to address a common research question. Synthesizing information from related studies helps in improving the statistical power in estimating parameters of interest, i.e., effect size, and thus helps researchers to reach more accurate conclusions (Borenstein et al., 2011). Traditional meta-analysis usually aggregates study-level summary statistics, such as the estimated effect size and its standard error, from publications or study authors, and then estimates a weighted average value of these statistics based on a fixed-effects model or a random-effects model. A fixed-effects model assumes all studies in the meta-analysis share a common true effect, while a random-effects model assumes that the true effect of each study is sampled from a distribution of these true effects, and thus a random-effects model can accommodate between-study heterogeneity. With aggregate data, we are unable to investigate the interaction effect between effect size and study-level covariates when estimating the weighted average effect size, and we need to rely on another analysis technique, meta-regression, to explore the modification effects of study-level covariates.

Individual participant data (IPD) meta-analysis, an alternative method to aggregate data meta-analysis, was proposed and has become increasingly popular since 1990. Riley et al. (2010) state that the application of IPD meta-analysis has increased very quickly, with an average of 49 applications per year after 2005. IPD contains the data recorded for each participant in a study, which is a concept contrary to aggregate data. In an IPD meta-analysis, raw data from each study are collected, synthesized and analyzed directly, preserving the clustering of participants within a study. Many potential advantages of IPD meta-analysis over aggregate data meta-analysis have been demonstrated (Riley et al., 2008; Riley et al., 2010). Firstly, IPD meta-analysis provides

higher statistical power with expanded sample size, and can avoid ecological bias and publication bias caused by aggregate data meta-analysis. Aggregate data sometimes are poorly reported or derived differently across studies, so the across-study relationships may differ from within-study relationships. Also, aggregate data collected from published studies are likely to contain only statistically significant results, which may result in publication bias. With IPD, these issues can be avoided so that IPD meta-analysis can provide more consistent results. Secondly, more sophisticated statistical methods for meta-analysis can be applied using IPD. Various approaches have been developed for IPD meta-analysis, primarily including one-stage methods and two-stage methods. Moreover, a modification effect of patient-level covariates on effect size can be modeled and investigated simultaneously when estimating the synthesized effect size. Lastly, with raw data from all studies, IPD meta-analysis guarantees consistent data checking and data cleaning across studies, and provides more opportunities to address different research questions.

IPD meta-analysis has been shown to have several potential advantages over aggregate meta-analysis, but corresponding methods for repeated measurements with modelling on source-specific variances or for other types of outcomes are limited. Hierarchy is an inherent characteristic of the meta-analysis data structure. For example, a meta-analysis usually involves multiple studies, with participants within each study and sometimes multiple measurements within each participant. Therefore, an optimal utilization of participant-level data should integrate information from each hierarchical level, but it is often ignored in most meta-analysis research endeavors. In addition, meta-analysis methods for longitudinal IPD for other outcome types are also of interest, such as outcomes from an exponential family and time-to-event outcomes. All these issues stimulate the motivation to develop statistical models for longitudinal IPD that can incorporate different types of outcomes, and utilize the hierarchical feature of meta-analysis data.

## Chapter 2

### Literature Review

Most implementation methods for IPD meta-analysis can be categorized as two-stage or one-stage methods. Both types of methods will be described in the next two sections. In addition, the type of outcomes and the type of data may also vary in practice, and the methods that should be adopted in different situations are different. The current methods for different type of outcomes and longitudinal will be reviewed in section 2.3.

#### 2.1 Two-Stage IPD Meta-Analysis Methods

There is a general model framework for the two-stage methods (Riley et al., 2015). Let  $\hat{\theta}_k$  be the effect size estimate for the  $k^{th}$  study, such that we can specify the general model as

$$\begin{aligned}\hat{\theta}_k &\sim N(\theta_k, S_k^2) \\ \theta_k &\sim N(\mu, \sigma^2)\end{aligned}\tag{2.1}$$

where  $\hat{\theta}_k$  follows a normal distribution with mean  $\theta_k$  and variance  $S_k^2$  for study  $k$ , and  $\theta_k$  follows another normal distribution with mean  $\mu$  and variance  $\sigma^2$ .  $S_k^2$  represents within-study variance, while  $\sigma^2$  represents between-study variance. Model (1) is a random-effects model, and when  $\sigma^2$  is 0, model (2.1) reduces to a fixed-effects model. Based on IPD,  $\hat{\theta}_k$  and  $\hat{S}_k^2$  for each study are estimated in the first stage, and relevant covariates can be incorporated. The second stage is to obtain an estimate of  $\mu$  based on  $\hat{\theta}_k$  and  $\hat{S}_k^2$  through model (2.1).

## 2.2 One-Stage IPD Meta-Analysis Methods

For one-stage approaches, the IPD from all studies are modelled simultaneously, accounting for the clustering of participants within each study. Consider  $y_{kl}$  as the continuous outcome for the  $l^{th}$  participant within the  $k^{th}$  study. Then a general linear mixed-effects model framework is specified as

$$y_{kl} = \mathbf{x}_{kl}^T \boldsymbol{\beta} + \mathbf{z}_{kl}^T \boldsymbol{\gamma}_{kl} + \varepsilon_{kl} \quad (2.2)$$

where  $\mathbf{x}_{kl}$  is a  $p \times 1$  vector of the design effects for  $\boldsymbol{\beta}$ , a  $p \times 1$  vector of the parameters for fixed effects,  $\mathbf{z}_{kl}$  is a  $q \times 1$  vector of the design effects for the random effects,  $\boldsymbol{\gamma}_{kl}$  is a  $q \times 1$  vector of the parameters for random effects, and  $\varepsilon_{kl}$  is the random error term. The model is specified to accommodate fixed effects and/or random effects, and any relevant covariates. Some parameters in  $\boldsymbol{\beta}$  may be of study interest, such as the effect size parameter  $\mu$  in model (2.1). The one-stage and two-stage methods have been discussed further in Higgins et al. (2001), Riley et al. (2008), Mathew et al. (2010), and Riley et al. (2015).

## 2.3 Methods for Different Outcomes and Longitudinal Data

Besides continuous outcomes, researchers sometimes are interested in treatment effects on other outcomes, such as binary, ordinal, and time-to-event survival outcomes. For binary outcomes, Turner et al. (2000) introduced a multi-level model framework to synthesize aggregate data and IPD when IPD are not available in certain studies, and the authors discussed the differences of inference methods in IPD and aggregate settings. For ordinal outcomes, Whitehead et al. (2001) proposed a Bayesian proportional odds model framework. For a time-to-event outcome, there are several articles that discussed and compared one-stage and two-stage methods, or fixed-effects and

random-effects models (Smith et al., 2005; Simmonds et al., 2013; Bowden et al., 2011; Simonds et al., 2011; Rondeau et al., 2008).

In clinical studies, participants sometimes are followed for a period of time, and the measurements of outcome on one participant may be collected at multiple time points. The data with repeated measurements, or so-called longitudinal data, are common in clinical research studies. The key to analyzing repeated measurement data is to appropriately incorporate the correlation among repeated observations. With longitudinal IPD or aggregate data from multiple studies, multivariate meta-analysis should be adopted. Jones et al. (2009) discussed the implementation of multivariate meta-analysis based on longitudinal IPD. In the article, repeated measurement time points were treated as a factor or as a continuous variable. For either approach with a continuous outcome, one-stage or two-stage methods can be applied, and both types of methods were compared. The specification of the covariance structure of the residual error allows for the correlation between repeated observations on the same participant. However, the article focuses on fixed-effects models, and models the variances altogether, which has the drawback of being unable to distinguish the sources of variance. Another article from Trikalinos et al. (2012) applied multivariate meta-analysis to aggregate data, using both fixed-effects and random-effects models, treating time points as a factor.

## Chapter 3

### The One-Stage Multi-Level Mixed-Effects Model for IPD Meta-Analysis with Continuous Outcomes

This section will introduce the proposed model for IPD meta-analysis with continuous outcomes, which adopts the features of a mixed-effects model and a multi-level model. The proposed model can be applied to longitudinal or repeated measurement data, and will have the capability to incorporate both fixed and random effects from each hierarchical level. Section 3.1 will introduce the model for continuous outcomes; section 3.2 will present the statistical inference for fixed-effects and variance-covariance parameters; section 3.3 will describe the simulation study conducted to evaluate the propose model; and section 3.4 will introduce the real data application of the proposed model.

#### 3.1 Methods

##### 3.1.1 Model Introduction

For a continuous outcome variable, we propose a statistical model that is a combination of a linear mixed-effects model (LMM) and a multi-level model that contains (a) fixed effects and random effects for the longitudinal data from each participant within a study, and (b) fixed effects and random effects for each study. The studies may be (a) observational studies or (b) randomized interventional trials.

Let  $Y_{kl}(t)$  denote the continuous outcome variable measured at time  $t$  for the  $l^{th}$  participant within the  $k^{th}$  study,  $k = 1, 2, \dots, K$  and  $l = 1, 2, \dots, n_k$ . Then

$$\begin{aligned}
Y_{kl}(t) = & \mathbf{x}_{u,kl}^T(t)\boldsymbol{\beta}_{uk} + \mathbf{x}_{c,kl}^T(t)\boldsymbol{\beta}_c + \mathbf{z}_{kl}^T(t)\boldsymbol{\gamma}_{kl} + \varepsilon_{kl}(t) + \mathbf{x}_k^{*T}\boldsymbol{\beta}^* \\
& + \mathbf{z}_k^{*T}\boldsymbol{\gamma}_k^* + \varepsilon_k^*
\end{aligned} \tag{3.1}$$

where

- $\mathbf{x}_{u,kl}(t)$  is a participant-level, unique fixed-effects,  $r_{uk} \times 1$  vector of design effects and covariates at time  $t$  for the  $l^{th}$  participant within the  $k^{th}$  study
- $\boldsymbol{\beta}_{uk}$  is a participant-level, unique fixed-effects  $r_{uk} \times 1$  vector of parameters for the  $k^{th}$  study
- $\mathbf{x}_{c,kl}(t)$  is a participant-level, common fixed-effects,  $r_c \times 1$  vector of design effects and covariates at time  $t$  for the  $l^{th}$  participant within the  $k^{th}$  study
- $\boldsymbol{\beta}_c$  is a participant-level, common fixed-effects  $r_c \times 1$  vector of parameters
- $\mathbf{z}_{kl}(t)$  is a participant-level, random-effects  $s_k \times 1$  vector of design effects and covariates at time  $t$  for the  $l^{th}$  participant within the  $k^{th}$  study
- $\boldsymbol{\gamma}_{kl}$  is a participant-level, random-effects  $s_k \times 1$  vector of parameters for the  $l^{th}$  participant within the  $k^{th}$  study
- $\varepsilon_{kl}(t)$  is a participant-level, random error term at time  $t$  for the  $l^{th}$  participant within the  $k^{th}$  study
- $\mathbf{x}_k^*$  is a study-level, fixed-effects,  $r^* \times 1$  vector of design effects and covariates for the  $k^{th}$  study
- $\boldsymbol{\beta}^*$  is a study-level, fixed-effects,  $r^* \times 1$  vector of parameters
- $\mathbf{z}_k^*$  is a study-level, random-effects  $s^* \times 1$  vector of design effects and covariates for the  $k^{th}$  study
- $\boldsymbol{\gamma}_k^*$  is a study-level, random-effects  $s^* \times 1$  vector of parameters for the  $k^{th}$  study
- $\varepsilon_k^*$  is a study-level, random error term for the  $k^{th}$  study

With respect to distributional assumptions,

- the  $\boldsymbol{\gamma}_{kl}$ 's are independent with  $\boldsymbol{\gamma}_{kl} \sim N_{s_k}(\mathbf{0}, \boldsymbol{\Gamma}_k)$ , where  $\boldsymbol{\Gamma}_k$  is a positive definite matrix
- $\boldsymbol{\varepsilon}_{kl} = [\varepsilon_{kl}(t_{kl1}) \ \varepsilon_{kl}(t_{kl2}) \ \dots \ \varepsilon_{kl}(t_{klp_{kl}})]^T \sim N_{p_{kl}}(\mathbf{0}, \boldsymbol{\Sigma}_{kl})$ , and the  $\boldsymbol{\varepsilon}_{kl}$ 's are independent, where  $\boldsymbol{\Sigma}_{kl}$  is a positive definite matrix and a function of a parameter vector  $\boldsymbol{\xi}$
- the  $\boldsymbol{\gamma}_k^*$ 's are independent with  $\boldsymbol{\gamma}_k^* \sim N_{s^*}(\mathbf{0}, \boldsymbol{\Gamma}^*)$ , where  $\boldsymbol{\Gamma}^*$  is a positive definite matrix
- the  $\varepsilon_k^*$ 's are independent with  $\varepsilon_k^* \sim N(0, \sigma^{*2})$
- the  $\boldsymbol{\gamma}_{kl}$ 's, the  $\boldsymbol{\varepsilon}_{kl}$ 's, the  $\boldsymbol{\gamma}_k^*$ 's, and the  $\varepsilon_k^*$ 's are mutually independent

We can rewrite model (3.1) for the  $l^{th}$  participant within the  $k^{th}$  study as

$$\begin{aligned} \mathbf{Y}_{kl} = & \begin{bmatrix} \mathbf{X}_{u,kl} & \mathbf{X}_{c,kl} & \mathbf{1}_{p_{kl} \times 1} \mathbf{x}_k^{*T} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{uk} \\ \boldsymbol{\beta}_c \\ \boldsymbol{\beta}^* \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_{kl} & \mathbf{1}_{p_{kl} \times 1} \mathbf{z}_k^{*T} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{kl} \\ \boldsymbol{\gamma}_k^* \end{bmatrix} \\ & + (\boldsymbol{\varepsilon}_{kl} + \mathbf{1}_{p_{kl} \times 1} \varepsilon_k^*) \end{aligned} \quad (3.2)$$

or

$$\begin{aligned} \begin{bmatrix} Y_{kl}(t_1) \\ \vdots \\ Y_{kl}(t_{p_{kl}}) \end{bmatrix}_{p_{kl} \times 1} &= \begin{bmatrix} \mathbf{x}_{u,kl}^T(t_1) & \mathbf{x}_{c,kl}^T(t_1) & \mathbf{x}_k^{*T} \\ \vdots & \vdots & \vdots \\ \mathbf{x}_{u,kl}^T(t_{p_{kl}}) & \mathbf{x}_{c,kl}^T(t_{p_{kl}}) & \mathbf{x}_k^{*T} \end{bmatrix}_{p_{kl} \times (r_{uk} + r_c + r^*)} \begin{bmatrix} \boldsymbol{\beta}_{uk} \\ \boldsymbol{\beta}_c \\ \boldsymbol{\beta}^* \end{bmatrix}_{(r_{uk} + r_c + r^*) \times 1} \\ &+ \begin{bmatrix} \mathbf{z}_{kl}^T(t_1) & \mathbf{z}_k^{*T} \\ \vdots & \vdots \\ \mathbf{z}_{kl}^T(t_{p_{kl}}) & \mathbf{z}_k^{*T} \end{bmatrix}_{p_{kl} \times (s_k + s^*)} \begin{bmatrix} \mathbf{Y}_{kl} \\ \boldsymbol{\gamma}_k^* \end{bmatrix}_{(s_k + s^*) \times 1} \\ &+ \begin{bmatrix} \boldsymbol{\varepsilon}_{kl}(t_1) + \varepsilon_k^* \\ \vdots \\ \boldsymbol{\varepsilon}_{kl}(t_{p_{kl}}) + \varepsilon_k^* \end{bmatrix}_{p_{kl} \times 1} \end{aligned} \quad (3.3)$$

where  $\mathbf{Y}_{kl} = [Y_{kl}(t_1) \ Y_{kl}(t_2) \ \dots \ Y_{kl}(t_{p_{kl}})]^T$ ,  $\mathbf{X}_{u,kl} = [\mathbf{x}_{u,kl}^T(t_1) \ \mathbf{x}_{u,kl}^T(t_2) \ \dots \ \mathbf{x}_{u,kl}^T(t_{p_{kl}})]^T$ ,

$\mathbf{X}_{c,kl} = [\mathbf{x}_{c,kl}^T(t_1) \ \mathbf{x}_{c,kl}^T(t_2) \ \dots \ \mathbf{x}_{c,kl}^T(t_{p_{kl}})]^T$ ,  $\mathbf{Z}_{kl} = [\mathbf{z}_{kl}^T(t_1) \ \mathbf{z}_{kl}^T(t_2) \ \dots \ \mathbf{z}_{kl}^T(t_{p_{kl}})]^T$ , and  $\mathbf{1}_{u \times v}$  is a  $u \times v$  matrix of unit values.

The distributional assumptions lead to the following expressions for the expectation vectors, the variance matrices, and the covariance matrices:



$$E(\mathbf{Y}_{kl}) = \begin{bmatrix} \mathbf{X}_{u,kl} & \mathbf{X}_{c,kl} & \mathbf{1}_{p_{kl} \times 1} \mathbf{x}_k^{*T} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{uk} \\ \boldsymbol{\beta}_c \\ \boldsymbol{\beta}^* \end{bmatrix} \quad (3.4)$$

$$\begin{aligned} \text{Var}(\mathbf{Y}_{kl}) &= \begin{bmatrix} \mathbf{Z}_{kl} & \mathbf{1}_{p_{kl} \times 1} \mathbf{z}_k^{*T} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Gamma}_k & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}^* \end{bmatrix} \begin{bmatrix} \mathbf{Z}_{kl}^T \\ \mathbf{z}_k^* \mathbf{1}_{1 \times p_{kl}} \end{bmatrix} \\ &\quad + (\boldsymbol{\Sigma}_{kl} + \sigma^{*2} \mathbf{1}_{p_{kl} \times p_{kl}}) \\ &= \mathbf{Z}_{kl} \boldsymbol{\Gamma}_k \mathbf{Z}_{kl}^T + \boldsymbol{\Sigma}_{kl} + (\mathbf{z}_k^{*T} \boldsymbol{\Gamma}^* \mathbf{z}_k^* + \sigma^{*2}) \mathbf{1}_{p_{kl} \times p_{kl}} \end{aligned} \quad (3.5)$$

$$\begin{aligned} \text{Cov}(\mathbf{Y}_{kl}, \mathbf{Y}_{km}) &= \text{Cov}_{\boldsymbol{\gamma}}[E(\mathbf{Y}_{kl}|\boldsymbol{\gamma}), E(\mathbf{Y}_{km}|\boldsymbol{\gamma})] + E_{\boldsymbol{\gamma}}[\text{Cov}(\mathbf{Y}_{kl}|\boldsymbol{\gamma}, \mathbf{Y}_{km}|\boldsymbol{\gamma})] \\ &= (\mathbf{z}_k^{*T} \boldsymbol{\Gamma}^* \mathbf{z}_k^*) \mathbf{1}_{p_{kl} \times p_{km}} + \sigma^{*2} \mathbf{1}_{p_{kl} \times p_{km}} \\ &= (\mathbf{z}_k^{*T} \boldsymbol{\Gamma}^* \mathbf{z}_k^* + \sigma^{*2}) \mathbf{1}_{p_{kl} \times p_{km}} \end{aligned} \quad (3.6)$$

$$\begin{aligned} \text{Cov}(\mathbf{Y}_{kl}, \mathbf{Y}_{k'm}) &= \text{Cov}_{\boldsymbol{\gamma}}[E(\mathbf{Y}_{kl}|\boldsymbol{\gamma}), E(\mathbf{Y}_{k'm}|\boldsymbol{\gamma})] \\ &\quad + E_{\boldsymbol{\gamma}}[\text{Cov}(\mathbf{Y}_{kl}|\boldsymbol{\gamma}, \mathbf{Y}_{k'm}|\boldsymbol{\gamma})] = 0 \end{aligned} \quad (3.7)$$

with  $k, k' = 1, 2, \dots, K$ ,  $l = 1, 2, \dots, n_k$  and  $m = 1, 2, \dots, n_{k'}$ .

The statistical model expressed in equations (3.2) and (3.3), along with its expectation vectors, variance matrices, and covariance matrices as expressed in equations (3.4)-(3.7), is in the form of a LMM (Laird et al., 1982) although it has the following extensions: (a) it is stratified according to study, which allows for unique variance-covariance parameters within each study; (b) it includes fixed effects and random effects at the study level. Because it is a LMM, we still can apply maximum likelihood estimation and restricted maximum likelihood estimation to derive parameter estimates and conduct statistical inference (Jennrich et al., 1986; Lindstrom et al., 1988; Vonesh et al., 1996) in the same manner as for the basic linear mixed-effects model.

### 3.1.2 Maximum Likelihood (ML) Estimation

We write the model for observations from the  $k^{\text{th}}$  study as

$$\mathbf{Y}_k = \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} + \mathbf{X}_k \boldsymbol{\beta} + \mathbf{Z}_k \boldsymbol{\gamma}_k + \boldsymbol{\varepsilon}_k \quad (3.8)$$

or equivalently,

$$\begin{aligned} \mathbf{Y}_k &= \begin{bmatrix} \mathbf{Y}_{k1} \\ \vdots \\ \mathbf{Y}_{kn_k} \end{bmatrix}_{(\sum_{l=1}^{n_k} p_{kl}) \times 1} = \begin{bmatrix} \mathbf{X}_{u,k1} \\ \vdots \\ \mathbf{X}_{u,kn_k} \end{bmatrix}_{(\sum_{l=1}^{n_k} p_{kl}) \times r_{uk}} \boldsymbol{\beta}_{uk} \, r_{uk} \times 1 \\ &\quad + \begin{bmatrix} \mathbf{X}_{c,k1} & \mathbf{1}_{p_{k1} \times 1} \mathbf{x}_k^{*T} \\ \vdots & \vdots \\ \mathbf{X}_{c,kn_k} & \mathbf{1}_{p_{kn_k} \times 1} \mathbf{x}_k^{*T} \end{bmatrix}_{(\sum_{l=1}^{n_k} p_{kl}) \times (r_c + r^*)} \begin{bmatrix} \boldsymbol{\beta}_c \\ \boldsymbol{\beta}^* \end{bmatrix}_{(r_c + r^*) \times 1} \\ &\quad + \begin{bmatrix} \mathbf{Z}_{k1} & \cdots & 0 & \mathbf{1}_{p_{k1} \times 1} \mathbf{z}_k^{*T} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \mathbf{Z}_{kn_k} & \mathbf{1}_{p_{kn_k} \times 1} \mathbf{z}_k^{*T} \end{bmatrix}_{(\sum_{l=1}^{n_k} p_{kl}) \times (s_k \times n_k + s^*)} \begin{bmatrix} \boldsymbol{\gamma}_{k1} \\ \vdots \\ \boldsymbol{\gamma}_{kn_k} \\ \boldsymbol{\gamma}_k^* \end{bmatrix}_{(s_k \times n_k + s^*) \times 1} \\ &\quad + \begin{bmatrix} \boldsymbol{\varepsilon}_{k1} + \mathbf{1}_{p_{k1} \times 1} \boldsymbol{\varepsilon}_k^* \\ \vdots \\ \boldsymbol{\varepsilon}_{kn_k} + \mathbf{1}_{p_{kn_k} \times 1} \boldsymbol{\varepsilon}_k^* \end{bmatrix}_{(\sum_{l=1}^{n_k} p_{kl}) \times 1} \end{aligned}$$

where  $\boldsymbol{\beta} = [\boldsymbol{\beta}_c \ \boldsymbol{\beta}^*]^T$  is the set of common fixed parameters of interest.

The distributional assumptions are given by

$$\boldsymbol{\gamma}_k = \begin{bmatrix} \boldsymbol{\gamma}_{k1} \\ \vdots \\ \boldsymbol{\gamma}_{kn_k} \\ \boldsymbol{\gamma}_k^* \end{bmatrix} \sim N_{s_k \times n_k + s^*}(\mathbf{0}, \mathbf{G}_k)$$

and

$$\boldsymbol{\varepsilon}_k = \begin{bmatrix} \boldsymbol{\varepsilon}_{k1} + \mathbf{1}_{p_{k1} \times 1} \boldsymbol{\varepsilon}_k^* \\ \vdots \\ \boldsymbol{\varepsilon}_{kn_k} + \mathbf{1}_{p_{kn_k} \times 1} \boldsymbol{\varepsilon}_k^* \end{bmatrix} \sim N_{\sum_{l=1}^{n_k} p_{kl}}(\mathbf{0}, \mathbf{R}_k)$$

where

$$\mathbf{G}_k = \begin{bmatrix} \boldsymbol{\Gamma}_k & 0 & \cdots & 0 \\ 0 & \ddots & \cdots & \vdots \\ \vdots & \vdots & \boldsymbol{\Gamma}_k & 0 \\ 0 & \cdots & 0 & \boldsymbol{\Gamma}^* \end{bmatrix}$$

and

$$\mathbf{R}_k = \begin{bmatrix} \boldsymbol{\Sigma}_{k1} + \sigma^{*2} \mathbf{1}_{p_{k1} \times p_{k1}} & \cdots & \sigma^{*2} \mathbf{1}_{p_{k1} \times p_{kn_k}} \\ \vdots & \ddots & \vdots \\ \sigma^{*2} \mathbf{1}_{p_{kn_k} \times p_{k1}} & \cdots & \boldsymbol{\Sigma}_{kn_k} + \sigma^{*2} \mathbf{1}_{p_{kn_k} \times p_{kn_k}} \end{bmatrix}$$

Therefore, the covariance matrix of  $\mathbf{Y}_k$  is

$$\text{Cov}(\mathbf{Y}_k) = \mathbf{Z}_k \mathbf{G}_k \mathbf{Z}_k^T + \mathbf{R}_k = \boldsymbol{\Sigma}_k(\boldsymbol{\xi})$$

The likelihood function  $L$  for the data,  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_K)$ , is constructed as

$$L(\boldsymbol{\beta}_{u1}, \dots, \boldsymbol{\beta}_{uK}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}) = \prod_{k=1}^K N(\mathbf{Y}_k; \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} + \mathbf{X}_k \boldsymbol{\beta}, \boldsymbol{\Sigma}_k(\boldsymbol{\xi})) \quad (3.9)$$

Here  $\boldsymbol{\beta}$  is a  $(r_c + r^*) \times 1$  vector of unknown regression parameters for common fixed effects across  $K$  studies and  $\boldsymbol{\xi}$  is a  $q \times 1$  vector of unknown covariance parameters for random effects and residual matrices. Taking the natural logarithm of  $L$ , the log-likelihood  $l$  is

$$\begin{aligned} l(\boldsymbol{\beta}_{u1}, \dots, \boldsymbol{\beta}_{uK}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}) &= \log \left\{ \prod_{k=1}^K N(\mathbf{Y}_k; \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} + \mathbf{X}_k \boldsymbol{\beta}, \boldsymbol{\Sigma}_k(\boldsymbol{\xi})) \right\} \\ &= \text{Constant} - \frac{1}{2} \sum_{k=1}^K \log |\boldsymbol{\Sigma}_k| - \frac{1}{2} \sum_{k=1}^K (\mathbf{Y}_k - \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} - \mathbf{X}_k \boldsymbol{\beta})^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} - \mathbf{X}_k \boldsymbol{\beta}) \end{aligned}$$

The score vector  $\mathbf{S}$  is defined as

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{\boldsymbol{\beta}_{u1}} \\ \vdots \\ \mathbf{S}_{\boldsymbol{\beta}_{uK}} \\ \mathbf{S}_{\boldsymbol{\beta}} \\ \mathbf{S}_{\boldsymbol{\xi}} \end{bmatrix} = \begin{bmatrix} \frac{\partial l}{\partial \boldsymbol{\beta}_{u1}} \\ \vdots \\ \frac{\partial l}{\partial \boldsymbol{\beta}_{uK}} \\ \frac{\partial l}{\partial \boldsymbol{\beta}} \\ \frac{\partial l}{\partial \boldsymbol{\xi}} \end{bmatrix}$$

where

$$\begin{aligned} \mathbf{S}_{\boldsymbol{\beta}_{uk}} &= \frac{\partial l}{\partial \boldsymbol{\beta}_{uk}} = \mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} - \mathbf{X}_k \boldsymbol{\beta}) = \mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_k \\ \mathbf{S}_{\boldsymbol{\beta}} &= \frac{\partial l}{\partial \boldsymbol{\beta}} = \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} - \mathbf{X}_k \boldsymbol{\beta}) = \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_k \end{aligned}$$

$$\begin{aligned}
[\mathcal{S}_\xi]_r &= \frac{\partial l}{\partial \xi_r} = -\frac{1}{2} \sum_{k=1}^K \text{tr}(\Sigma_k^{-1} \dot{\Sigma}_{k,r}) + \frac{1}{2} \sum_{k=1}^K \text{tr}(\mathbf{e}_k^T \Sigma_k^{-1} \dot{\Sigma}_{k,r} \Sigma_k^{-1} \mathbf{e}_k) \\
&= \frac{1}{2} \sum_{k=1}^K \text{tr}[\Sigma_k^{-1} (\mathbf{e}_k \mathbf{e}_k^T - \Sigma_k) \Sigma_k^{-1} \dot{\Sigma}_{k,r}]
\end{aligned}$$

with  $\mathbf{e}_k = \mathbf{Y}_k - \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} - \mathbf{X}_k \boldsymbol{\beta}$ ,  $\dot{\Sigma}_{k,r} = \frac{\partial \Sigma_k}{\partial \xi_r}$ ,  $k = 1, \dots, K$ , and  $r = 1, \dots, q$ .

The Hessian matrix  $\mathbf{H}$  is defined as

$$\begin{aligned}
\mathbf{H} &= \begin{bmatrix} \mathbf{H}_{\boldsymbol{\beta}_{u1}\boldsymbol{\beta}_{u1}} & \dots & \mathbf{H}_{\boldsymbol{\beta}_{u1}\boldsymbol{\beta}_{uK}} & \mathbf{H}_{\boldsymbol{\beta}_{u1}\boldsymbol{\beta}} & \mathbf{H}_{\boldsymbol{\beta}_{u1}\xi} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{H}_{\boldsymbol{\beta}_{uK}\boldsymbol{\beta}_{u1}} & \dots & \mathbf{H}_{\boldsymbol{\beta}_{uK}\boldsymbol{\beta}_{uK}} & \mathbf{H}_{\boldsymbol{\beta}_{uK}\boldsymbol{\beta}} & \mathbf{H}_{\boldsymbol{\beta}_{uK}\xi} \\ \mathbf{H}_{\boldsymbol{\beta}\boldsymbol{\beta}_{u1}} & \dots & \mathbf{H}_{\boldsymbol{\beta}\boldsymbol{\beta}_{uK}} & \mathbf{H}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{H}_{\boldsymbol{\beta}\xi} \\ \mathbf{H}_{\xi\boldsymbol{\beta}_{u1}} & \dots & \mathbf{H}_{\xi\boldsymbol{\beta}_{uK}} & \mathbf{H}_{\xi\boldsymbol{\beta}} & \mathbf{H}_{\xi\xi} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial^2 l}{\partial \boldsymbol{\beta}_{u1} \partial \boldsymbol{\beta}_{u1}} & \frac{\partial^2 l}{\partial \boldsymbol{\beta}_{u1} \partial \boldsymbol{\beta}_{uK}} & \frac{\partial^2 l}{\partial \boldsymbol{\beta}_{u1} \partial \boldsymbol{\beta}} & \frac{\partial^2 l}{\partial \boldsymbol{\beta}_{u1} \partial \xi} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 l}{\partial \boldsymbol{\beta}_{uK} \partial \boldsymbol{\beta}_{u1}} & \dots & \frac{\partial^2 l}{\partial \boldsymbol{\beta}_{uK} \partial \boldsymbol{\beta}_{uK}} & \frac{\partial^2 l}{\partial \boldsymbol{\beta}_{uK} \partial \boldsymbol{\beta}} & \frac{\partial^2 l}{\partial \boldsymbol{\beta}_{uK} \partial \xi} \\ \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}_{u1}} & \dots & \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}_{uK}} & \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}} & \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \xi} \\ \frac{\partial^2 l}{\partial \xi \partial \boldsymbol{\beta}_{u1}} & \frac{\partial^2 l}{\partial \xi \partial \boldsymbol{\beta}_{uK}} & \frac{\partial^2 l}{\partial \xi \partial \boldsymbol{\beta}} & \frac{\partial^2 l}{\partial \xi \partial \xi} \end{bmatrix}
\end{aligned}$$

where

$$\mathbf{H}_{\boldsymbol{\beta}_{uk}\boldsymbol{\beta}_{uk}} = \frac{\partial^2 l}{\partial \boldsymbol{\beta}_{uk} \partial \boldsymbol{\beta}_{uk}} = -\mathbf{X}_{u,k}^T \Sigma_k^{-1} \mathbf{X}_{u,k}$$

$$\mathbf{H}_{\boldsymbol{\beta}_{uk}\boldsymbol{\beta}_{uk'}} = \frac{\partial^2 l}{\partial \boldsymbol{\beta}_{uk} \partial \boldsymbol{\beta}_{uk'}} = \mathbf{0}$$

$$\mathbf{H}_{\boldsymbol{\beta}_{uk}\boldsymbol{\beta}} = \frac{\partial^2 l}{\partial \boldsymbol{\beta}_{uk} \partial \boldsymbol{\beta}} = -\mathbf{X}_{u,k}^T \Sigma_k^{-1} \mathbf{X}_k$$

$$[\mathbf{H}_{\boldsymbol{\beta}_{uk}\xi}]_r = \frac{\partial^2 l}{\partial \boldsymbol{\beta}_{uk} \partial \xi_r} = \frac{\partial \mathcal{S}_{\boldsymbol{\beta}_{uk}}}{\partial \xi_r} = -\mathbf{X}_{u,k}^T \Sigma_k^{-1} \dot{\Sigma}_{k,r} \Sigma_k^{-1} \mathbf{e}_k, r = 1, \dots, q$$

$$\mathbf{H}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}} = -\sum_{k=1}^K \mathbf{X}_k^T \Sigma_k^{-1} \mathbf{X}_k$$

$$\begin{aligned}
[\mathbf{H}_{\beta\xi}]_r &= \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \xi_r} = \frac{\partial \mathbf{S}_\beta}{\partial \xi_r} = - \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_k \\
[\mathbf{H}_{\xi\xi}]_{rs} &= -\frac{1}{2} \sum_{k=1}^K \text{tr} [-\boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,s} + \boldsymbol{\Sigma}_k^{-1} \ddot{\boldsymbol{\Sigma}}_{k,rs}] \\
&\quad -\frac{1}{2} \sum_{k=1}^K \text{tr} [-\mathbf{e}_k^T (-\boldsymbol{\Sigma}_k^{-1} (\dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,s} - \ddot{\boldsymbol{\Sigma}}_{k,rs} \\
&\quad\quad\quad + \dot{\boldsymbol{\Sigma}}_{k,s} \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r}) \boldsymbol{\Sigma}_k^{-1}) \mathbf{e}_k] \\
&= \frac{1}{2} \sum_{k=1}^K \text{tr} [\boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,s} - \boldsymbol{\Sigma}_k^{-1} \ddot{\boldsymbol{\Sigma}}_{k,rs} \\
&\quad\quad\quad - \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,s} \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_k \mathbf{e}_k^T + \boldsymbol{\Sigma}_k^{-1} \ddot{\boldsymbol{\Sigma}}_{k,rs} \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_k \mathbf{e}_k^T \\
&\quad\quad\quad - \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,s} \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \mathbf{e}_k \mathbf{e}_k^T] \\
&= \frac{1}{2} \sum_{k=1}^K \text{tr} [\boldsymbol{\Sigma}_k^{-1} (\mathbf{e}_k \mathbf{e}_k^T - \boldsymbol{\Sigma}_k) \boldsymbol{\Sigma}_k^{-1} \ddot{\boldsymbol{\Sigma}}_{k,rs}] \\
&\quad\quad\quad - \frac{1}{2} \sum_{k=1}^K \text{tr} [\boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} (2\mathbf{e}_k \mathbf{e}_k^T - \boldsymbol{\Sigma}_k) \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,s}]
\end{aligned}$$

where  $\ddot{\boldsymbol{\Sigma}}_{k,rs} = \frac{\partial^2 \boldsymbol{\Sigma}_k}{\partial \xi_r \partial \xi_s}$ ,  $k' = 1, \dots, K$ ,  $k \neq k'$ , and  $r, s = 1, \dots, q$ .

### ***Solve $\boldsymbol{\beta}_{uk}$ , $\boldsymbol{\beta}$ and $\xi$ Using Newton-Raphson and Fisher Scoring Algorithms***

The Newton-Raphson algorithm is an iterative procedure that computes new parameter values  $\tilde{\boldsymbol{\beta}}_{uk}$ ,  $\tilde{\boldsymbol{\beta}}$ , and  $\tilde{\xi}$  from current values  $\boldsymbol{\beta}_{uk}$ ,  $\boldsymbol{\beta}$ , and  $\xi$  using

$$\begin{bmatrix} \tilde{\boldsymbol{\beta}}_{u1} \\ \vdots \\ \tilde{\boldsymbol{\beta}} \\ \tilde{\xi} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_{u1} \\ \vdots \\ \boldsymbol{\beta} \\ \xi \end{bmatrix} - \begin{bmatrix} \mathbf{H}_{\boldsymbol{\beta}_{u1}\boldsymbol{\beta}_{u1}} & \cdots & \mathbf{H}_{\boldsymbol{\beta}_{u1}\xi} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{\xi\boldsymbol{\beta}_{u1}} & \cdots & \mathbf{H}_{\xi\xi} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{S}_{\boldsymbol{\beta}_{u1}} \\ \vdots \\ \mathbf{S}_\beta \\ \mathbf{S}_\xi \end{bmatrix}$$

The Fisher scoring algorithm replaces the Hessian matrix by its expectation.

$$\mathbf{E}(\mathbf{H}_{\boldsymbol{\beta}_{uk}\boldsymbol{\beta}_{uk}}) = \mathbf{H}_{\boldsymbol{\beta}_{uk}\boldsymbol{\beta}_{uk}} = -\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k}$$

$$E(\mathbf{H}_{\beta_{uk}\beta_{uk'}}) = \mathbf{0}$$

$$E(\mathbf{H}_{\beta_{uk}\beta}) = \mathbf{H}_{\beta_{uk}\beta} = -\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k$$

$$E(\mathbf{H}_{\beta_{uk}\xi}) = \mathbf{0}$$

$$E(\mathbf{H}_{\beta\beta}) = \mathbf{H}_{\beta\beta} = -\sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k$$

$$E(\mathbf{H}_{\xi\beta}) = E(\mathbf{H}_{\beta\xi}) = \mathbf{0}$$

$$E\left([\mathbf{H}_{\xi\xi}]_{rs}\right) = \frac{1}{2} \sum_{k=1}^K \text{tr}[\boldsymbol{\Sigma}_k^{-1} (\boldsymbol{\Sigma}_k - \boldsymbol{\Sigma}_k) \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,rs}] -$$

$$\frac{1}{2} \sum_{k=1}^K \text{tr}[\boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} (2\boldsymbol{\Sigma}_k - \boldsymbol{\Sigma}_k) \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,s}] = -\frac{1}{2} \sum_{k=1}^K \text{tr}[\boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,s}]$$

Because  $E(\mathbf{H}_{\beta_{uk}\xi}) = \mathbf{0}$  and  $E(\mathbf{H}_{\beta\xi}) = \mathbf{0}$ , the updates of  $(\tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}})$  and  $\tilde{\xi}$  based on the Fisher scoring algorithm can be separated. The new values  $(\tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}})$  are obtained through:

$$\begin{bmatrix} \tilde{\boldsymbol{\beta}}_{u1} \\ \vdots \\ \tilde{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_{u1} \\ \vdots \\ \boldsymbol{\beta} \end{bmatrix} - \begin{bmatrix} E(\mathbf{H}_{\beta_{u1}\beta_{u1}}) & \cdots & E(\mathbf{H}_{\beta_{u1}\beta}) \\ \vdots & \ddots & \vdots \\ E(\mathbf{H}_{\beta\beta_{u1}}) & \cdots & E(\mathbf{H}_{\beta\beta}) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{S}_{\beta_{u1}} \\ \vdots \\ \mathbf{S}_{\beta} \end{bmatrix} \quad (3.10)$$

The new values  $\tilde{\xi}$  are then obtained through:

$$\tilde{\xi} = \xi - E(\mathbf{H}_{\xi\xi})^{-1} \mathbf{S}_{\xi}(\tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}})$$

### ***Solve $\beta_{uk}$ , $\beta$ and $\xi$ Using Generalized EM Scoring Algorithm for the Balanced, Incomplete Data Model***

The generalized EM Scoring Algorithm is suitable for the balanced, but incomplete, data model. If the variance-covariance matrix for a complete set of measurements on any study participant is large and some participants display missing data in the balanced design, then this algorithm has advantages over the Newton-Raphson and Fisher scoring methods.

The steps of this algorithm are as follows:

- (i)  $(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta})$  is updated via  $(\tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}})$  given in equation (3.10).

(ii) Let

$$\mathbf{e}_k^* = \begin{bmatrix} \mathbf{e}_k \\ \mathbf{e}_k^+ \end{bmatrix} \sim N(\mathbf{0}, \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{k11} & \mathbf{\Sigma}_{k12} \\ \mathbf{\Sigma}_{k21} & \mathbf{\Sigma}_{k22} \end{bmatrix})$$

where  $\mathbf{e}_k^*$  represents complete data for the  $k$ th study using sub-vectors  $\mathbf{e}_k = \mathbf{Y}_k - \mathbf{X}_{u,k}\boldsymbol{\beta}_{uk} - \mathbf{X}_k\boldsymbol{\beta}$  as observed data and  $\mathbf{e}_k^+$  as unobserved data.  $\mathbf{\Sigma}$  is partitioned correspondingly.

Let  $E(\mathbf{e}_k^* | \mathbf{e}_k) = \hat{\mathbf{e}}_k^*$ ,  $Cov(\mathbf{e}_k^* | \mathbf{e}_k) = \mathbf{P}_k$ . Then they can be calculated as

$$\hat{\mathbf{e}}_k^* = \begin{bmatrix} E(\mathbf{e}_k) \\ E(\mathbf{e}_k^+ | \mathbf{e}_k) \end{bmatrix} = \begin{bmatrix} \mathbf{e}_k \\ \mathbf{\Sigma}_{k21} \mathbf{\Sigma}_{k11}^{-1} \mathbf{e}_k \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{\Sigma}_{k21} \mathbf{\Sigma}_{k11}^{-1} \end{bmatrix} \mathbf{e}_k = \mathbf{M}_k \mathbf{e}_k \quad (3.11)$$

$$\mathbf{P}_k = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{k22} - \mathbf{\Sigma}_{k21} \mathbf{\Sigma}_{k11}^{-1} \mathbf{\Sigma}_{k12} \end{bmatrix}$$

(iii) If  $\mathbf{\Sigma}$  is unstructured, then the new values  $\tilde{\mathbf{\Sigma}}$  are estimated as

$$\tilde{\mathbf{\Sigma}} = \frac{1}{K} \sum_{k=1}^K (\hat{\mathbf{e}}_k^* \hat{\mathbf{e}}_k^{*T} + \mathbf{P}_k)$$

If  $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\xi})$  is structured, then  $\mathbf{\Sigma}$  is updated through a ‘‘scoring step’’

$$\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi} - E(\mathbf{H}_{\tilde{\boldsymbol{\xi}}})^{-1} \mathbf{s}$$

where

$$[\mathbf{s}]_r = \frac{1}{2} \text{tr}[\mathbf{\Sigma}^{-1}(\tilde{\mathbf{\Sigma}} - \mathbf{\Sigma})\mathbf{\Sigma}^{-1}\dot{\mathbf{\Sigma}}_r]$$

and

$$[E(\mathbf{H}_{\tilde{\boldsymbol{\xi}}})]_{rs} = -\frac{1}{2} \text{tr}[\mathbf{\Sigma}^{-1}\dot{\mathbf{\Sigma}}_r\mathbf{\Sigma}^{-1}\dot{\mathbf{\Sigma}}_s]$$

$r, s = 1, \dots, q$ .

(iv) Let  $h(\boldsymbol{\xi}) = -\log|\mathbf{\Sigma}(\boldsymbol{\xi})| - \text{tr}[\mathbf{\Sigma}^{-1}(\boldsymbol{\xi})\tilde{\mathbf{\Sigma}}]$ . To guarantee that the likelihood is increasing at each step, check to see if  $h(\tilde{\boldsymbol{\xi}}) > h(\boldsymbol{\xi})$ . If  $h(\boldsymbol{\xi})$  is not increased, then use partial stepping to increase it (i.e., replace  $E(\mathbf{H}_{\tilde{\boldsymbol{\xi}}})^{-1} \mathbf{s}$  by  $(E(\mathbf{H}_{\tilde{\boldsymbol{\xi}}})^{-1} \mathbf{s})/2$  until  $h(\tilde{\boldsymbol{\xi}}) > h(\boldsymbol{\xi})$ ).

### 3.1.3 Restricted Maximum Likelihood (REML) Estimation

In ML estimation, the estimate of  $\xi$  depends on the form of design matrices  $\mathbf{X}_{u,k}$  and  $\mathbf{X}_k$ , and incorrect specifications of  $\mathbf{X}_{u,k}$  and  $\mathbf{X}_k$  may result in an inconsistent estimate of  $\xi$ . Restricted maximum likelihood (REML) estimation can be adopted to address this issue, which is based on the likelihood function of transformed outcome data  $\mathbf{Y}^*$ . If we stack observations from all  $K$  studies together, we can have

$$\begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_K \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{u,1} & \cdots & 0 & \mathbf{X}_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \mathbf{X}_{u,K} & \mathbf{X}_K \end{bmatrix} \begin{bmatrix} \beta_{u1} \\ \vdots \\ \beta_{uK} \\ \beta \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{Z}_K \end{bmatrix} \begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_K \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_K \end{bmatrix}$$

or equivalently,

$$\mathbf{Y} = \mathbf{X}\bar{\boldsymbol{\beta}} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

The transformed outcome data is constructed as  $\mathbf{Y}^* = \mathbf{B}^T \mathbf{Y}$  such that the distribution of  $\mathbf{Y}^*$  does not depend on the form of  $\mathbf{X}$ . One choice of  $\mathbf{B}$  is based on the decomposition of ordinary least square residuals, such that

$$\mathbf{B}\mathbf{B}^T = \mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

$$\mathbf{B}^T \mathbf{B} = \mathbf{I}^*$$

$$\text{cov}(\mathbf{B}^T \mathbf{Y}, \bar{\boldsymbol{\beta}}) = \mathbf{0}$$

where  $\mathbf{B}$  is a  $(\sum_{k=1}^K \sum_{l=1}^{n_k} p_{kl}) \times q$  matrix,  $\mathbf{I}$  is the  $(\sum_{k=1}^K \sum_{l=1}^{n_k} p_{kl}) \times (\sum_{k=1}^K \sum_{l=1}^{n_k} p_{kl})$  identity matrix,  $\mathbf{I}^*$  is the  $q \times q$  identity matrix, and  $\bar{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_{u1}, \dots, \hat{\boldsymbol{\beta}}_{uK}, \hat{\boldsymbol{\beta}})$  is the generalized least square estimator of  $\bar{\boldsymbol{\beta}}$ .

REML estimation leads to a consistent estimate of  $\xi$  and it has the advantage that its estimate of  $\xi$  has less statistical bias than the ML estimate. The restricted log-likelihood function  $l_R$  is constructed as the log-likelihood function of  $\mathbf{Y}^*$ , and since



$$l_R(\widehat{\boldsymbol{\beta}}_{u1}, \dots, \widehat{\boldsymbol{\beta}}_{uK}, \widehat{\boldsymbol{\beta}}, \boldsymbol{\xi} | \mathbf{Y}) = l(\boldsymbol{\beta}_{u1}, \dots, \boldsymbol{\beta}_{uK}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}^*) \propto \frac{l(\boldsymbol{\beta}_{u1}, \dots, \boldsymbol{\beta}_{uK}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y})}{l(\boldsymbol{\beta}_{u1}, \dots, \boldsymbol{\beta}_{uK}, \boldsymbol{\beta}, \boldsymbol{\xi} | \widehat{\boldsymbol{\beta}}_{u1}, \dots, \widehat{\boldsymbol{\beta}}_{uK}, \widehat{\boldsymbol{\beta}})},$$

we can obtain

$$\begin{aligned} & l_R(\widehat{\boldsymbol{\beta}}_{u1}, \dots, \widehat{\boldsymbol{\beta}}_{uK}, \widehat{\boldsymbol{\beta}}, \boldsymbol{\xi} | \mathbf{Y}) \\ &= l(\widehat{\boldsymbol{\beta}}_{u1}, \dots, \widehat{\boldsymbol{\beta}}_{uK}, \widehat{\boldsymbol{\beta}}, \boldsymbol{\xi} | \mathbf{Y}) - \frac{1}{2} \log \left| \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right| \\ &\quad - \frac{1}{2} \sum_{k=1}^K \log |\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k}| \\ &= \text{Constant} - \frac{1}{2} \sum_{k=1}^K \log |\boldsymbol{\Sigma}_k| \\ &\quad - \frac{1}{2} \sum_{k=1}^K (\mathbf{Y}_k - \mathbf{X}_{u,k} \widehat{\boldsymbol{\beta}}_{uk}(\boldsymbol{\xi}) - \mathbf{X}_k \widehat{\boldsymbol{\beta}}(\boldsymbol{\xi}))^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y}_k \\ &\quad - \mathbf{X}_{u,k} \widehat{\boldsymbol{\beta}}_{uk}(\boldsymbol{\xi}) - \mathbf{X}_k \widehat{\boldsymbol{\beta}}(\boldsymbol{\xi})) - \frac{1}{2} \log \left| \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right| \\ &\quad - \frac{1}{2} \sum_{k=1}^K \log |\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k}| \end{aligned} \tag{3.12}$$

The REML estimates of  $\boldsymbol{\xi}$ ,  $\widehat{\boldsymbol{\beta}}_{uk}(\boldsymbol{\xi})$ , and  $\widehat{\boldsymbol{\beta}}(\boldsymbol{\xi})$  may also be calculated by redefining  $l_R$  as a function of  $\boldsymbol{\xi}$ ,  $\boldsymbol{\beta}_{uk}$ , and  $\boldsymbol{\beta}$ . Therefore, the redefined  $l_R$  will be

$$\begin{aligned} & l_R(\boldsymbol{\beta}_{u1}, \dots, \boldsymbol{\beta}_{uK}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}) \\ &= \text{Constant} - \frac{1}{2} \sum_{k=1}^K \log |\boldsymbol{\Sigma}_k| \\ &\quad - \frac{1}{2} \sum_{k=1}^K (\mathbf{Y}_k - \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} - \mathbf{X}_k \boldsymbol{\beta})^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} - \mathbf{X}_k \boldsymbol{\beta}) \\ &\quad - \frac{1}{2} \log \left| \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right| - \frac{1}{2} \sum_{k=1}^K \log |\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k}| \end{aligned}$$

The score vector  $\mathbf{S}_R$  is defined as

$$\mathbf{S}_R = \begin{bmatrix} \mathbf{S}_{R,\beta_{u1}} \\ \vdots \\ \mathbf{S}_{R,\beta_{uK}} \\ \mathbf{S}_{R,\beta} \\ \mathbf{S}_{R,\xi} \end{bmatrix} = \begin{bmatrix} \frac{\partial l_R}{\partial \beta_{u1}} \\ \vdots \\ \frac{\partial l_R}{\partial \beta_{uK}} \\ \frac{\partial l_R}{\partial \beta} \\ \frac{\partial l_R}{\partial \xi} \end{bmatrix}$$

where

$$\begin{aligned} \mathbf{S}_{R,\beta_{uk}} &= \frac{\partial l_R}{\partial \beta_{uk}} = \mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} - \mathbf{X}_k \boldsymbol{\beta}) = \mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_k = \mathbf{S}_{\beta_{uk}} \\ \mathbf{S}_{R,\beta} &= \frac{\partial l_R}{\partial \beta} = \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y}_k - \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} - \mathbf{X}_k \boldsymbol{\beta}) = \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_k = \mathbf{S}_\beta \\ [\mathbf{S}_{R,\xi}]_r &= \frac{\partial l_R}{\partial \xi_r} = -\frac{1}{2} \sum_{k=1}^K \text{tr}(\boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r}) + \frac{1}{2} \sum_{k=1}^K \text{tr}(\mathbf{e}_k^T \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_k) \\ &\quad + \frac{1}{2} \sum_{k=1}^K \text{tr} \left[ \left( \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right)^{-1} \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k^T \right] \\ &\quad + \frac{1}{2} \sum_{k=1}^K \text{tr} \left[ (\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k})^{-1} \mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k}^T \right] \\ &= \frac{1}{2} \sum_{k=1}^K \text{tr} \left[ \boldsymbol{\Sigma}_k^{-1} \left( \mathbf{e}_k \mathbf{e}_k^T - \boldsymbol{\Sigma}_k + \mathbf{X}_k \left( \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right)^{-1} \mathbf{X}_k^T \right. \right. \\ &\quad \left. \left. + \mathbf{X}_{u,k} (\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k})^{-1} \mathbf{X}_{u,k}^T \right) \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \right], \end{aligned}$$

$r = 1, \dots, q$ .

The Hessian matrix  $\mathbf{H}_R$  is defined as

$$\mathbf{H}_R = \begin{bmatrix} \mathbf{H}_{R,\beta_{u1}\beta_{u1}} & \dots & \mathbf{H}_{R,\beta_{u1}\beta_{uK}} & \mathbf{H}_{R,\beta_{u1}\beta} & \mathbf{H}_{R,\beta_{u1}\xi} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{H}_{R,\beta_{uK}\beta_{u1}} & \dots & \mathbf{H}_{R,\beta_{uK}\beta_{uK}} & \mathbf{H}_{R,\beta_{uK}\beta} & \mathbf{H}_{R,\beta_{uK}\xi} \\ \mathbf{H}_{R,\beta\beta_{u1}} & \dots & \mathbf{H}_{R,\beta\beta_{uK}} & \mathbf{H}_{R,\beta\beta} & \mathbf{H}_{R,\beta\xi} \\ \mathbf{H}_{R,\xi\beta_{u1}} & \dots & \mathbf{H}_{R,\xi\beta_{uK}} & \mathbf{H}_{R,\xi\beta} & \mathbf{H}_{R,\xi\xi} \end{bmatrix}$$

where

$$\mathbf{H}_{\mathbf{R}, \beta_{uk} \beta_{uk}} = \frac{\partial^2 l_R}{\partial \beta_{uk} \partial \beta_{uk}} = -\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k}$$

$$\mathbf{H}_{\mathbf{R}, \beta_{uk} \beta_{uk'}} = \frac{\partial^2 l_R}{\partial \beta_{uk} \partial \beta_{uk'}} = \mathbf{0}$$

$$\mathbf{H}_{\mathbf{R}, \beta_{uk} \beta} = \frac{\partial^2 l_R}{\partial \beta_{uk} \partial \beta} = -\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k$$

$$[\mathbf{H}_{\mathbf{R}, \beta_{uk} \xi}]_r = \frac{\partial^2 l}{\partial \beta_{uk} \partial \xi_r} = \frac{\partial \mathbf{S}_{\beta_{uk}}}{\partial \xi_r} = -\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_k, r = 1, \dots, q$$

$$\mathbf{H}_{\mathbf{R}, \beta \beta} = \frac{\partial^2 l}{\partial \beta \partial \beta} = -\sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k$$

$$[\mathbf{H}_{\mathbf{R}, \beta \xi}]_r = \frac{\partial^2 l}{\partial \beta \partial \xi_r} = \frac{\partial \mathbf{S}_{\beta}}{\partial \xi_r} = -\sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_k$$

$$\begin{aligned} & [\mathbf{H}_{\mathbf{R}, \xi \xi}]_{rs} \\ &= -\frac{1}{2} \sum_{k=1}^K \text{tr} [-\boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,s} + \boldsymbol{\Sigma}_k^{-1} \ddot{\boldsymbol{\Sigma}}_{k,rs}] \\ & -\frac{1}{2} \sum_{k=1}^K \text{tr} [-\mathbf{e}_k^T (-\boldsymbol{\Sigma}_k^{-1} (\dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,s} - \ddot{\boldsymbol{\Sigma}}_{k,rs} + \dot{\boldsymbol{\Sigma}}_{k,s} \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r}) \boldsymbol{\Sigma}_k^{-1}) \mathbf{e}_k] \\ & -\frac{1}{2} \left\{ -\text{tr} \left[ \left( \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right)^{-1} \sum_{k=1}^K (\mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k) \left( \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right)^{-1} \sum_{k=1}^K (\mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k) \right] \right. \\ & \left. + \text{tr} \left[ \left( \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right)^{-1} \sum_{k=1}^K (\mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} (\dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,s} - \ddot{\boldsymbol{\Sigma}}_{k,rs} + \dot{\boldsymbol{\Sigma}}_{k,s} \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r}) \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k) \right] \right\} \\ & -\frac{1}{2} \left\{ -\text{tr} \left[ (\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k})^{-1} (\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k}) (\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k})^{-1} (\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k}) \right] \right. \\ & \left. + \text{tr} \left[ (\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k})^{-1} (\mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} (\dot{\boldsymbol{\Sigma}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,s} - \ddot{\boldsymbol{\Sigma}}_{k,rs} + \dot{\boldsymbol{\Sigma}}_{k,s} \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\Sigma}}_{k,r}) \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k}) \right] \right\} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2} \sum_{k=1}^K \text{tr} \left[ \Sigma_k^{-1} \left( \mathbf{e}_k \mathbf{e}_k^T - \Sigma_k + \mathbf{X}_k \left( \sum_{k=1}^K \mathbf{X}_k^T \Sigma_k^{-1} \mathbf{X}_k \right)^{-1} \mathbf{X}_k^T + \mathbf{X}_{u,k} \left( \mathbf{X}_{u,k}^T \Sigma_k^{-1} \mathbf{X}_{u,k} \right)^{-1} \mathbf{X}_{u,k}^T \right) \Sigma_k^{-1} \dot{\Sigma}_{k,rs} \right] \\
&- \frac{1}{2} \sum_{k=1}^K \text{tr} \left[ \Sigma_k^{-1} \dot{\Sigma}_{k,r} \Sigma_k^{-1} \left( 2 \mathbf{e}_k \mathbf{e}_k^T - \Sigma_k + 2 \mathbf{X}_k \left( \sum_{k=1}^K \mathbf{X}_k^T \Sigma_k^{-1} \mathbf{X}_k \right)^{-1} \mathbf{X}_k^T \right. \right. \\
&\quad \left. \left. + 2 \mathbf{X}_{u,k} \left( \mathbf{X}_{u,k}^T \Sigma_k^{-1} \mathbf{X}_{u,k} \right)^{-1} \mathbf{X}_{u,k}^T \right) \Sigma_k^{-1} \dot{\Sigma}_{k,s} \right] \\
&+ \frac{1}{2} \text{tr} \left[ \left( \sum_{k=1}^K \mathbf{X}_k^T \Sigma_k^{-1} \mathbf{X}_k \right)^{-1} \sum_{k=1}^K \left( \mathbf{X}_k^T \Sigma_k^{-1} \dot{\Sigma}_{k,r} \Sigma_k^{-1} \mathbf{X}_k \right) \left( \sum_{k=1}^K \mathbf{X}_k^T \Sigma_k^{-1} \mathbf{X}_k \right)^{-1} \sum_{k=1}^K \left( \mathbf{X}_k^T \Sigma_k^{-1} \dot{\Sigma}_{k,r} \Sigma_k^{-1} \mathbf{X}_k \right) \right] \\
&+ \frac{1}{2} \text{tr} \left[ \left( \mathbf{X}_{u,k}^T \Sigma_k^{-1} \mathbf{X}_{u,k} \right)^{-1} \left( \mathbf{X}_{u,k}^T \Sigma_k^{-1} \dot{\Sigma}_{k,r} \Sigma_k^{-1} \mathbf{X}_{u,k} \right) \left( \mathbf{X}_{u,k}^T \Sigma_k^{-1} \mathbf{X}_{u,k} \right)^{-1} \left( \mathbf{X}_{u,k}^T \Sigma_k^{-1} \dot{\Sigma}_{k,r} \Sigma_k^{-1} \mathbf{X}_{u,k} \right) \right], \\
&r, s = 1, \dots, q
\end{aligned}$$

### *Solve $\beta_{uk}$ , $\beta$ and $\xi$ Using Newton-Raphson and Fisher Scoring Algorithms*

Based on the Newton-Raphson algorithm, new parameter values  $\tilde{\beta}_{uk}$ ,  $\tilde{\beta}$ , and  $\tilde{\xi}$  are calculated as

$$\begin{bmatrix} \tilde{\beta}_{u1} \\ \vdots \\ \tilde{\beta} \\ \tilde{\xi} \end{bmatrix} = \begin{bmatrix} \beta_{u1} \\ \vdots \\ \beta \\ \xi \end{bmatrix} - \begin{bmatrix} \mathbf{H}_{R,\beta_{u1}\beta_{u1}} & \cdots & \mathbf{H}_{R,\beta_{u1}\xi} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{R,\xi\beta_{u1}} & \cdots & \mathbf{H}_{R,\xi\xi} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{S}_{R,\beta_{u1}} \\ \vdots \\ \mathbf{S}_{R,\beta} \\ \mathbf{S}_{R,\xi} \end{bmatrix}$$

The expectations of the elements of  $\mathbf{H}_R$  are

$$\mathbf{E}(\mathbf{H}_{R,\beta_{uk}\beta_{uk}}) = \mathbf{H}_{R,\beta_{uk}\beta_{uk}} = -\mathbf{X}_{u,k}^T \Sigma_k^{-1} \mathbf{X}_{u,k}$$

$$\mathbf{E}(\mathbf{H}_{R,\beta_{uk}\beta_{uk'}}) = \mathbf{0}$$

$$\mathbf{E}(\mathbf{H}_{R,\beta_{uk}\beta}) = \mathbf{H}_{R,\beta_{uk}\beta} = -\mathbf{X}_{u,k}^T \Sigma_k^{-1} \mathbf{X}_k$$

$$\mathbf{E}(\mathbf{H}_{R,\beta_{uk}\xi}) = \mathbf{0}$$

$$\mathbf{E}(\mathbf{H}_{R,\beta\beta}) = \mathbf{H}_{R,\beta\beta} = -\sum_{k=1}^K \mathbf{X}_k^T \Sigma_k^{-1} \mathbf{X}_k$$

$$\mathbf{E}(\mathbf{H}_{R,\xi\beta}) = \mathbf{E}(\mathbf{H}_{R,\beta\xi}) = \mathbf{0}$$

$$\begin{aligned}
& \mathbf{E}\left([\mathbf{H}_{\mathbf{R},\xi\xi}]_{rs}\right) \\
&= \frac{1}{2} \sum_{k=1}^K \text{tr} \left[ \boldsymbol{\Sigma}_k^{-1} \left( \mathbf{X}_k \left( \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right)^{-1} \mathbf{X}_k^T + \mathbf{X}_{u,k} \left( \mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k} \right)^{-1} \mathbf{X}_{u,k}^T \right) \boldsymbol{\Sigma}_k^{-1} \ddot{\boldsymbol{\xi}}_{k,rs} \right] \\
&- \frac{1}{2} \sum_{k=1}^K \text{tr} \left[ \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\xi}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \left( \boldsymbol{\Sigma}_k + 2\mathbf{X}_k \left( \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right)^{-1} \mathbf{X}_k^T + 2\mathbf{X}_{u,k} \left( \mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k} \right)^{-1} \mathbf{X}_{u,k}^T \right) \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\xi}}_{k,s} \right] \\
&+ \frac{1}{2} \text{tr} \left[ \left( \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right)^{-1} \sum_{k=1}^K \left( \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\xi}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right) \left( \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right)^{-1} \sum_{k=1}^K \left( \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\xi}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right) \right] \\
&+ \frac{1}{2} \text{tr} \left[ \left( \mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k} \right)^{-1} \left( \mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\xi}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k} \right) \left( \mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k} \right)^{-1} \left( \mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \dot{\boldsymbol{\xi}}_{k,r} \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k} \right) \right]
\end{aligned}$$

Similar to ML estimation, the updates of  $(\tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}})$  and  $\tilde{\boldsymbol{\xi}}$  based on the Fisher scoring algorithm can be separated. The new values  $(\tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}})$  are still obtained through:

$$\begin{bmatrix} \tilde{\boldsymbol{\beta}}_{u1} \\ \vdots \\ \tilde{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_{u1} \\ \vdots \\ \boldsymbol{\beta} \end{bmatrix} - \begin{bmatrix} \mathbf{E}(\mathbf{H}_{\mathbf{R},\boldsymbol{\beta}_{u1}\boldsymbol{\beta}_{u1}}) & \cdots & \mathbf{E}(\mathbf{H}_{\mathbf{R},\boldsymbol{\beta}_{u1}\boldsymbol{\beta}}) \\ \vdots & \ddots & \vdots \\ \mathbf{E}(\mathbf{H}_{\mathbf{R},\boldsymbol{\beta}\boldsymbol{\beta}_{u1}}) & \cdots & \mathbf{E}(\mathbf{H}_{\mathbf{R},\boldsymbol{\beta}\boldsymbol{\beta}}) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{S}_{\mathbf{R},\boldsymbol{\beta}_{u1}} \\ \vdots \\ \mathbf{S}_{\mathbf{R},\boldsymbol{\beta}} \end{bmatrix}$$

The new values  $\tilde{\boldsymbol{\xi}}$  are then obtained through:

$$\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi} - \mathbf{E}(\mathbf{H}_{\mathbf{R},\xi\xi})^{-1} \mathbf{S}_{\mathbf{R},\xi}(\tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}})$$

### ***Solve $\boldsymbol{\beta}_{uk}$ , $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$ Using Generalized EM Scoring Algorithm for the Balanced, Incomplete Data Model***

Steps (i)-(ii) are the same as described for ML estimations, and  $(\tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}})$ ,  $\hat{\boldsymbol{e}}_k^*$ ,  $\mathbf{R}_k$ , and  $\mathbf{S}$  are computed. At step (iii), we define and compute

$$\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_2 &= \tilde{\boldsymbol{\Sigma}} + \frac{1}{K} \sum_{k=1}^K \mathbf{M}_k \left\{ \mathbf{X}_{u,k} \left( \mathbf{X}_{u,k}^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_{u,k} \right)^{-1} \mathbf{X}_{u,k}^T \right. \\
&\quad \left. + \mathbf{X}_k \left( \sum_{k=1}^K \mathbf{X}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{X}_k \right)^{-1} \mathbf{X}_k^T \right\} \mathbf{M}_k
\end{aligned}$$

where  $\mathbf{M}_k$  is defined in equation (9). If  $\boldsymbol{\Sigma}$  is unstructured, then the new values of  $\boldsymbol{\Sigma}$  can be estimated by  $\tilde{\boldsymbol{\Sigma}}_2$ . If  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\xi})$  is structured, then  $\boldsymbol{\Sigma}$  is updated through a ‘‘scoring step’’

$$\tilde{\xi} = \xi - \mathbf{E}(\mathbf{H}_{\xi\xi})^{-1} \mathbf{s}'$$

where

$$[\mathbf{s}']_r = \frac{1}{2} \text{tr}[\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{\Sigma}}_2 - \boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}_r],$$

$r, s = 1, \dots, q$ .

Let  $h'(\xi) = -\log|\boldsymbol{\Sigma}(\xi)| - \text{tr}[\boldsymbol{\Sigma}^{-1}(\xi)\tilde{\boldsymbol{\Sigma}}_2]$ .  $h'(\tilde{\xi}) > h'(\xi)$  is required to guarantee the increase in the restricted likelihood. If  $h'(\xi)$  is not increasing, then use partial stepping to increase it (i.e., replace  $\mathbf{E}(\mathbf{H}_{\xi\xi})^{-1} \mathbf{s}'$  by  $(\mathbf{E}(\mathbf{H}_{\xi\xi})^{-1} \mathbf{s}')/2$  until  $h'(\tilde{\xi}) > h'(\xi)$ ).

### 3.2 Statistical Inference

#### 3.2.1 Inference for Fixed Effects $\boldsymbol{\beta}$

$\boldsymbol{\beta}$  is the set of common fixed parameters across all studies, which is usually of research interest. For the inference of  $\boldsymbol{\beta}$ , large sample tests can be performed with related confidence intervals estimated, by using either the Wald chi-square test, adjusted Wald test, likelihood ratio test, or score test. For a contrast matrix  $\mathbf{L}_{M \times (r_c + r^*)}$ , we want to test the null hypothesis  $H_0: \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$  versus  $H_1: \mathbf{L}\boldsymbol{\beta} \neq \mathbf{0}$ .

##### 1) Wald Chi-square Test

Based on the final estimate  $\hat{\boldsymbol{\beta}}$ , its estimated covariance matrix is

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \left( \sum_{k=1}^K \mathbf{X}_k^T \hat{\boldsymbol{\Sigma}}_k^{-1} \mathbf{X}_k \right)^{-1}$$

where  $\widehat{\Sigma}_k$  is the ML or the REML estimate of  $\Sigma_k$ , and  $\sum_{k=1}^K \mathbf{X}_k^T \widehat{\Sigma}_k^{-1} \mathbf{X}_k$  is the empirical Fisher information matrix. To test the null hypothesis, the Wald statistic is

$$C^2 = (\mathbf{L}\widehat{\boldsymbol{\beta}})^T \left\{ \mathbf{L} \left( \sum_{k=1}^K \mathbf{X}_k^T \widehat{\Sigma}_k^{-1} \mathbf{X}_k \right)^{-1} \mathbf{L}^T \right\}^{-1} (\mathbf{L}\widehat{\boldsymbol{\beta}}),$$

and  $C^2$  is compared to percentiles from a  $\chi_M^2$  distribution. If the number of rows of  $\mathbf{L}$ ,  $M$ , equals 1, then an approximate  $(1 - \alpha) \times 100\%$  confidence interval for  $\mathbf{L}\boldsymbol{\beta}$  is given by  $\mathbf{L}\widehat{\boldsymbol{\beta}} \pm z_{1-\alpha/2} \sqrt{\mathbf{L} \text{Cov}(\widehat{\boldsymbol{\beta}}) \mathbf{L}^T}$ .

## 2) Adjusted Wald Test (Satterthwaite et al., 1941)

The Wald test relies on a large sample normal approximation to the sampling distribution of  $\widehat{\boldsymbol{\beta}}$ . When the sample size is not large enough, the Wald test statistic tends to be anti-conservative, since additional variability is introduced through the estimate of  $\widehat{\Sigma}_k$ . For a small sample, an adjusted Wald statistic can be used, and it follows an  $F$  distribution. The adjusted Wald statistic is defined as

$$F = \frac{(\mathbf{L}\widehat{\boldsymbol{\beta}})^T \left\{ \mathbf{L} \left( \sum_{k=1}^K \mathbf{X}_k^T \widehat{\Sigma}_k^{-1} \mathbf{X}_k \right)^{-1} \mathbf{L}^T \right\}^{-1} (\mathbf{L}\widehat{\boldsymbol{\beta}})}{\text{rank}(\mathbf{L})}$$

which follows an approximate  $F$  distribution with numerator degrees of freedom as  $\text{rank}(\mathbf{L})$  and denominator degrees of freedom estimated from the data. A typical choice for the denominator degrees of freedom is the total sample size minus  $(r_c + r^*)$ , the dimension of  $\boldsymbol{\beta}$ .

## 3) Likelihood Ratio Test

An alternative to the Wald test is the likelihood ratio test (LRT). Denote the maximized ML log-likelihoods under null and alternative hypotheses as  $\hat{l}_{reduced}(\widetilde{\boldsymbol{\beta}})$  and  $\hat{l}_{full}(\widehat{\boldsymbol{\beta}})$ . The LRT for two nested models can be constructed by comparing  $\hat{l}_{reduced}(\widetilde{\boldsymbol{\beta}})$  and  $\hat{l}_{full}(\widehat{\boldsymbol{\beta}})$ . The larger the

difference between  $\hat{l}_{reduced}(\tilde{\boldsymbol{\beta}})$  and  $\hat{l}_{full}(\hat{\boldsymbol{\beta}})$ , the stronger the evidence that the reduced model is inadequate. The LRT statistic is

$$G^2 = -2(\hat{l}_{reduced}(\tilde{\boldsymbol{\beta}}) - \hat{l}_{full}(\hat{\boldsymbol{\beta}}))$$

and  $G^2$  is compared to percentiles from a  $\chi_M^2$  distribution. If  $M$ , the number of rows of  $\mathbf{L}$ , equals 1, then an approximate  $(1 - \alpha) \times 100\%$  confidence interval for  $\mathbf{L}\boldsymbol{\beta}$  can also be obtained by inverting the LRT.

However, the REML log-likelihood cannot be used to compare nested models for  $\boldsymbol{\beta}$ , because the extra term in the REML log-likelihood depends on the model specification and the two nested models for the mean are based on two entirely different sets of transformed responses.

#### 4) Score Test

A score test is based on the null hypothesis, so the score function is obtained at  $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$ :

$$\mathbf{u}(\tilde{\boldsymbol{\beta}}) = \frac{\partial l(\boldsymbol{\beta}, \tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\xi}} | \mathbf{Y})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}$$

The information matrix  $\mathbf{I}_{\tilde{\boldsymbol{\beta}}}$  can be either the observed or expected information matrix evaluated at  $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$ .

$$\mathbf{I}_{\tilde{\boldsymbol{\beta}}} = -\frac{\partial^2 l(\boldsymbol{\beta}, \tilde{\boldsymbol{\xi}} | \mathbf{Y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \Big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}} \quad \text{or} \quad \mathbf{I}_{\tilde{\boldsymbol{\beta}}} = E\left(-\frac{\partial^2 l(\boldsymbol{\beta}, \tilde{\boldsymbol{\xi}} | \mathbf{Y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \Big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}\right)$$

The score test statistic is defined as

$$s(\tilde{\boldsymbol{\beta}}) = \mathbf{u}(\tilde{\boldsymbol{\beta}})^T \mathbf{I}_{\tilde{\boldsymbol{\beta}}}^{-1} \mathbf{u}(\tilde{\boldsymbol{\beta}})$$

$s(\tilde{\boldsymbol{\beta}})$  is compared to percentiles from a  $\chi_M^2$  distribution. If  $M$  equals 1, then an approximate  $(1 - \alpha) \times 100\%$  confidence interval for  $\mathbf{L}\boldsymbol{\beta}$  can also be obtained by inverting the score test.



### 3.2.2 Prediction of Random Effects $\tilde{\boldsymbol{\gamma}}_k$

Consider the joint distribution of study-specific observations  $\mathbf{Y}_k$  and random effects  $\boldsymbol{\gamma}_k$  within the  $k^{th}$  study,  $k = 1, 2, \dots, K$ .

$$\begin{bmatrix} \mathbf{Y}_k \\ \boldsymbol{\gamma}_k \end{bmatrix} \sim N \left( \begin{bmatrix} \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} + \mathbf{X}_k \boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_k(\boldsymbol{\xi}) & \mathbf{Z}_k \mathbf{G}_k \\ \mathbf{G}_k \mathbf{Z}_k^T & \mathbf{G}_k \end{bmatrix} \right) \quad (3.13)$$

where  $\mathbf{G}_k$  is defined in equation (3.8). Equation (3.13) implies that the conditional distribution of  $\boldsymbol{\gamma}_k | \mathbf{Y}_k$  is

$$\boldsymbol{\gamma}_k | \mathbf{Y}_k \sim N(\mathbf{G}_k \mathbf{Z}_k^T \boldsymbol{\Sigma}_k(\boldsymbol{\xi})^{-1} (\mathbf{Y}_k - \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} - \mathbf{X}_k \boldsymbol{\beta}), \boldsymbol{\Sigma}_k^*(\boldsymbol{\xi}))$$

where

$$\boldsymbol{\Sigma}_k^*(\boldsymbol{\xi}) = \mathbf{G}_k - \mathbf{G}_k \mathbf{Z}_k^T \boldsymbol{\Sigma}_k(\boldsymbol{\xi})^{-1} \mathbf{Z}_k \mathbf{G}_k = \mathbf{G}_k - \mathbf{G}_k \mathbf{Z}_k^T (\mathbf{Z}_k \mathbf{G}_k \mathbf{Z}_k^T + \mathbf{R}_k)^{-1} \mathbf{Z}_k \mathbf{G}_k$$

Therefore, the empirical Bayes estimator for  $\boldsymbol{\gamma}_k | \mathbf{Y}_k$  is given by

$$\tilde{\boldsymbol{\gamma}}_k = E[\boldsymbol{\gamma}_k | \mathbf{Y}_k, \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi}] = \mathbf{G}_k \mathbf{Z}_k^T \boldsymbol{\Sigma}_k(\boldsymbol{\xi})^{-1} (\mathbf{Y}_k - \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} - \mathbf{X}_k \boldsymbol{\beta})$$

which also is called the best linear unbiased predictor (BLUP).  $\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi}$  can be replaced by their ML or REML estimates.

## 3.3 Simulation Study

### 3.3.1 Simulation Design

We conducted a simulation study to evaluate the proposed one-stage model for IPD meta-analysis. The simulated data mimicked multi-center clinical trial data, with a two-arm, placebo-controlled, parallel design, which evaluated the effect of an active treatment on reducing a patient's total cholesterol level.

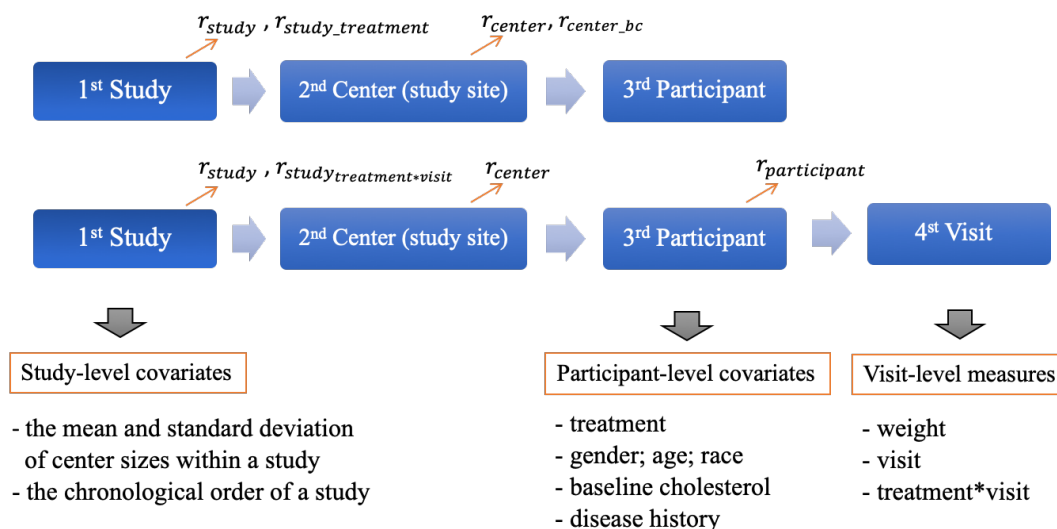


Figure 3-1: 3-level and 4-level Simulated Data Structure

We designed the simulated data to have a 3-level or a 4-level structure (Figure 3-1). For the 3-level data, we assume the clinical data were collected from multiple studies, and each study had different numbers of centers (study sites), and each center had different numbers of participants. For the 4-level data, we assumed each participant within a center had different numbers of visits, which was the visit level and resulted in longitudinal data. For both 3-level and 4-level multi-center data, we simulated several participant-level variables, including treatment assignment (TRT; placebo or active treatment), baseline cholesterol level (base\_cho; mg/dL), age at enrollment, gender (male or female), race (Caucasian (Cau), African American (AA), Hispanic (His) and other), diabetes disease status (yes or no), and cardiovascular disease status (CVD; yes or no). Both base\_cho and age were centered by their population means, respectively. We also generated three study-level covariates to reflect the characteristics of each study, including the mean of cohort sizes of centers within each study ( $SS_{mean}$ ), the standard deviation of cohort sizes of centers within each study ( $SS_{std}$ ), and the chronological order of each study (order). We generated two sets of study-level random effects, including study-level random intercept ( $r_{study}$ ) and random effect for treatment effects across studies ( $r_{study\_treatment}$ ). We imposed another two sets of

random effects to the center level, including center-level random intercept ( $r_{center}$ ) and random effect for baseline cholesterol level ( $r_{center\_bc}$ ). With the 4-level data structure, we generated each participant's weight (lbs) as a repeated measurement variable, and each participant might have 3 to 5 visits. Weight was centered by its population mean before it was used. We removed the random effect  $r_{center\_bc}$  and included an additional set of random intercepts for participant level ( $r_{participant}$ ). The study-level treatment random effect ( $r_{study\_treatment}$ ) is replaced by the study-level random effect for treatment×visit ( $r_{study\_treatment*visit}$ ). The primary parameter of interest is the difference of effect between treatment groups in the 3-level data structure, and the parameter of interest is the difference in slopes between treatment groups in the 4-level data structure.

Based on 3-level data, the mean function conditional on random effects used to generate the outcome is

$$\begin{aligned}
E(Y_{kcl} | r_{study}, r_{study\_treatment}, r_{center}, r_{center\_bc}) \\
&= 220 - 8 \times TRT + 0.2 \times base\_cho + 0.2 \times age \\
&\quad - 4 \times gender + 4 \times race_{AA} + 6 \times race_{His} \\
&\quad - 0.8 \times race_{other} + 6 \times diabetes + 6 \times CVD \\
&\quad - 0.3 \times SS_{mean} + 1.5 \times SS_{std} - 0.5 \times order + r_{study} \\
&\quad + r_{study\_treatment} \times TRT + r_{center} \\
&\quad + r_{center\_bc} \times base\_cho
\end{aligned} \tag{3.14}$$

where  $Y_{kcl}$  is the outcome for the  $l^{\text{th}}$  participant from the  $c^{\text{th}}$  center in the  $k^{\text{th}}$  study.

Based on 4-level data, the mean function conditional on random effects used to generate the outcome is

$$\begin{aligned}
& E(Y_{kcl}(t) | r_{study}, r_{study\_treatment*visit}, r_{center}, r_{participant}) \\
& = 220 + 0.2 \times base\_cho + 0.2 \times age - 4 \times gender \\
& + 4 \times race_{AA} + 6 \times race_{His} - 0.8 \times race_{other} \\
& + 6 \times diabetes + 6 \times CVD - 0.3 \times SS_{mean} \\
& + 1.5 \times SS_{std} - 0.5 \times order - t + 0.4 \times weight \times t \\
& - 4 \times TRT \times t + r_{study} \\
& + r_{study\_treatment*visit} \times TRT \times t + r_{center} \\
& + r_{participant}
\end{aligned} \tag{3.15}$$

where  $Y_{kcl}(t)$  is the outcome at visit  $t$  for the  $l^{\text{th}}$  participant from the  $c^{\text{th}}$  center in the  $k^{\text{th}}$  study. The outcome  $Y_{kcl}$  or  $Y_{kcl}(t)$  is then generated from a normal distribution  $N(0, \sigma^2)$ , with  $\sigma^2$  as the variance for the random error.

We developed 40 simulation scenarios by considering 3 factors, namely, the sample size of the meta-analysis data, the intra-cluster correlation coefficient (ICC), and common variables across studies. Regarding sample size, the meta-analysis data could contain 12 or 8 studies for the 3-level data structure and 10 or 6 studies for the 4-level data structure, and balance or imbalance sample size within each study. With balanced sample size, every study has a similar or a comparable number of centers and participants, while with imbalanced sample sizes, studies might have larger differences in the number of centers and participants (Table 3-1).

Table 3-1: Sample size design in the simulation study with continuous outcomes.

3-level data				
Balance design	balance	imbalance	balance	imbalance
No. of studies	12	8	12	8
No. of centers within a study	8-16	8-12	study 1-4: 6; study 5-8: 9; study 9-12: 12	study 1-2: 6; study 3-6: 9; study 7-8: 12
No. of participants within a center	24-70	20-30	study 1-4: 16; study 5-8: 20; study 9-12: 26	study 1-2: 16; study 3-6: 20; study 7-8: 26

4-level data				
Balance design	balance	imbalance	balance	imbalance
No. of studies	10	6	10	6
No. of centers within a study	6-10	4-8	study 1-3: 4; study 4-6: 8; study 7-10: 12	study 1-2: 3; study 3-4: 6; study 5-6: 9
No. of participants within a center	24-70	20-30	study 1-3: 16; study 4-6: 20; study 7-10: 26	study 1-2: 16; study 3-4: 20; study 5-6: 26
No. of visits of a participant			3-5	

The ICC is usually defined as the between-study variance  $\sigma_{between}^2$  divided by the sum of the between-study variance and the within-study variance ( $\sigma_{between}^2 + \sigma_{within}^2$ ) (Donner, 1986). As the value of the ICC may be associated with the participant's characteristics, we specified the ICC values for the reference group, and altered the ICC as 0.1, 0.2, 0.3, 0.4, and 0.5. The reference group is referred to subjects who are Caucasian males assigned to the control treatment group, without diabetes or CVD and with mean base\_cho and age. For example, suppose with 3-level data, we have two subjects in the reference group from the same center of a study, then the ICC is equal to

$$ICC = \frac{\sigma_{r\_study}^2 + \sigma_{r\_center}^2}{\sigma_{r\_study}^2 + \sigma_{r\_center}^2 + \sigma^2}$$

where  $r_{study} \sim N(0, \sigma_{r\_study}^2)$ ,  $r_{center} \sim N(0, \sigma_{r\_center}^2)$ . We specified  $\sigma^2$  as 225 in the simulation study and  $\sigma_{r\_study}^2 : \sigma_{r\_center}^2 = 1 : 2$ . As ICC ranged from 0.1, 0.2 to 0.5, we could decide the values of  $\sigma_{r\_study}^2$  and  $\sigma_{r\_center}^2$ . More details about value specification could be found in the appendix A.

The covariates collected from each study could be the same or different. We considered either common covariates from each study, or with distinct/unique variables. In scenarios with distinct/unique variables, we chose gender, race, CVD, and diabetes information as study-specific variables. In particular, gender was only available in study no. 3 and 5, race was available in study

1, diabetes was available in study 7, and CVD was available in study 8. We examined each scenario in 1000 simulation runs.

With each simulated dataset, we applied 4 test models to the data. The models were a two-stage fixed-effects model, a two-stage random-effects model, a one-stage fixed-effects model, and a one-stage model (our proposed model). Both two-stage methods still include center-level and participant-level random effects in the first stage estimates, while the one-stage fixed-effects model ignores all random effects, and the one-stage proposed model captures all random effects (Table 3-2). To assess the performance of the 4 test models, we compare the average of mean estimates, average of standard errors from simulation, and the coverage probability of containing the true treatment effect among 1000 simulations.

Table 3-2: Model specifications of the four test models.

Test model	Model fitting
two-stage fixed-effects	1st stage: study-level covariates, participant-level covariates, visit-level covariates*, $r_{center}$ , $r_{center\_bc}^{\Delta}$ , $r_{participant}^*$ ; 2nd stage: a fixed-effects model to pool treatment effect
two-stage random-effects	1st stage: study-level covariates, participant-level covariates, visit-level covariates*, $r_{center}$ , $r_{center\_bc}^{\Delta}$ , $r_{participant}^*$ ; 2nd stage: a random-effects model to pool treatment effect
one-stage fixed-effects	participant-level covariates, visit-level covariates*
one-stage proposed	study-level covariates, participant-level covariates, visit-level covariates*, $r_{study}$ , $r_{study\_treatment}$ , $r_{center}$ , $r_{center\_bc}^{\Delta}$ , $r_{participant}^*$

\*: only used in 4-level data;  $\Delta$ : only used in 3-level data

### 3.3.2 Simulation Results

The average of mean estimates, the average of standard errors from each estimate, and the coverage probability from 1000 simulations for the scenarios with common and distinct/unique

covariates in the 3-level data structure settings are shown in Figures 3-2 and 3-3, with different combinations of sample size and ICC value. Results for the 4-level data structure are shown in Figures 3-4 and 3-5.

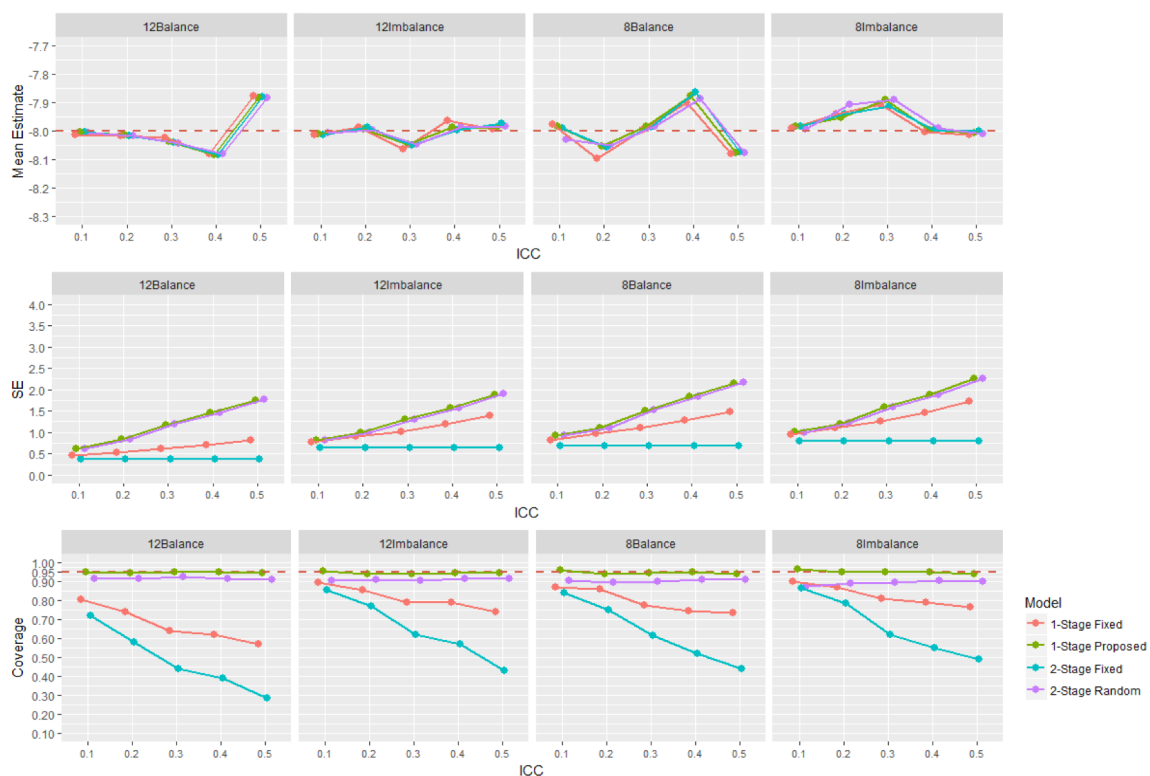


Figure 3-2: Simulation results for the 3-level data structure with common covariates.

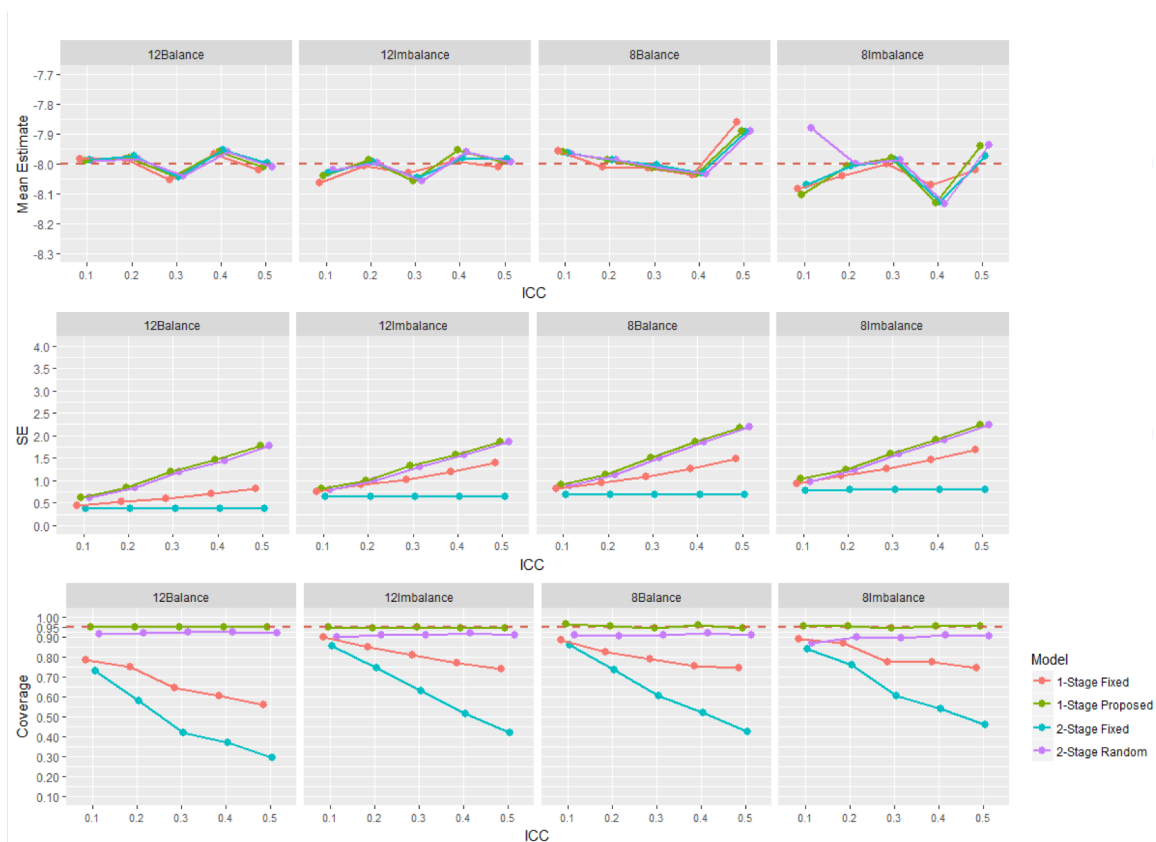


Figure 3-3: Simulation results for the 3-level data structure with distinct covariates.

Each figure for the simulation results is presented as a 3 by 4 plot matrix. The three rows in the plot matrix show the mean estimate, standard error (SE), and coverage probability from the top to bottom, and the four columns in the plot matrix show the 4 combinations of the number of studies and balanced/imbalanced design, yielding ‘12Balance’, ‘12Imbalance’, ‘8Balance’, and ‘8Imbalance’ for 3-level data structure and ‘10Balance’, ‘10Imbalance’, ‘6Balance’, and ‘6Imbalance’ for the 4-level data structure. Each frame of the plot matrix displays the estimates as the ICC value changes from 0.1 to 0.5. Four colors of lines represent four different fitted models, which are one-stage fixed-effects model (1-Stage Fixed), one-stage proposed model (1-Stage Proposed), two-stage fixed-effects model (2-Stage Fixed), and two-stage random-effects model (2-Stage Random).



For simulated data with the 3-level data structure, the true value of the difference of treatment effect was specified as -8. As seen in Figures 3-2 and 3-3, simulations with common or distinct covariates show similar results. The mean estimates are accurate in most scenarios, which lie between -8.15 and -7.86, though with increases of the ICC, the sample size decreases, or imbalanced design, the mean estimates become less accurate and more fluctuated. Both fixed-effects models tend to underestimate the SE when the ICC increases, and as a result, both fixed-effects models yield poor coverage probabilities. When the ICC is greater than or equal to 0.20, the coverage probabilities for the two-stage fixed-effects and one-stage fixed-effects models are equal or less than 0.87. The two-stage random-effects and one-stage proposed models behave similarly regarding the estimation of SE and coverage in most scenarios, but when the sample size decreases or becomes imbalanced, the two-stage random-effects model shows insufficient coverage probability (ranging from 0.868 to 0.927). The one-stage proposed model is able to appropriately estimate the variation between studies and to maintain about 0.95 coverage probability in all scenarios (ranging from 0.939 to 0.966).

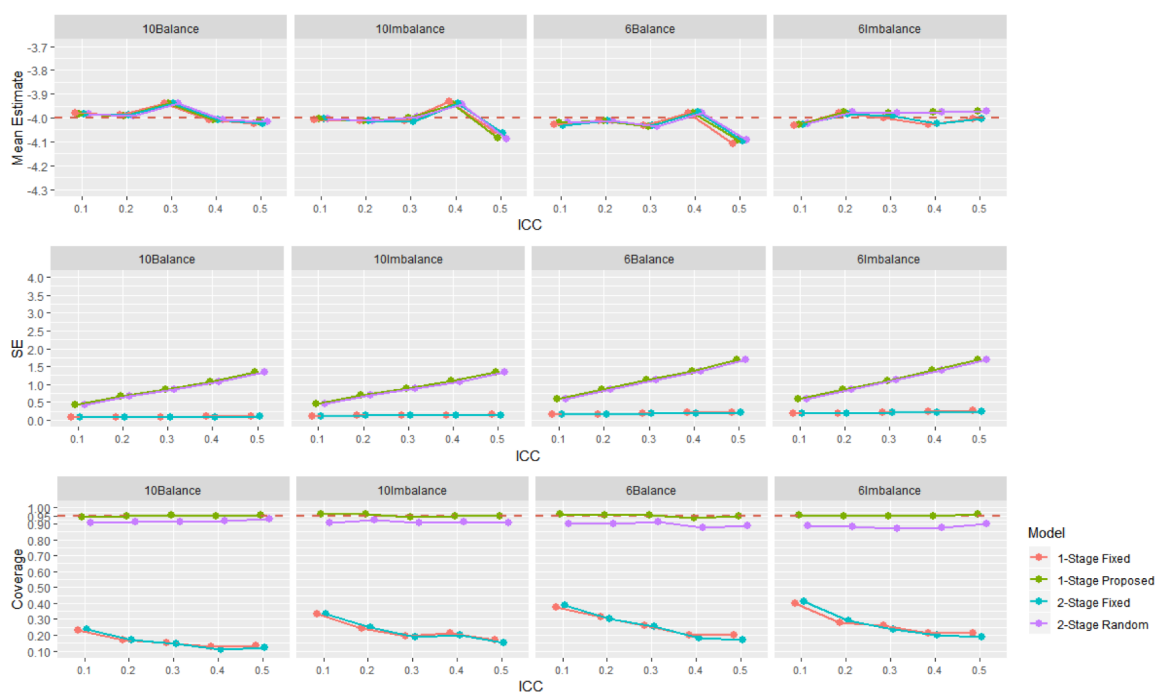


Figure 3-4: Simulation results for the 4-level data structure with common covariates.

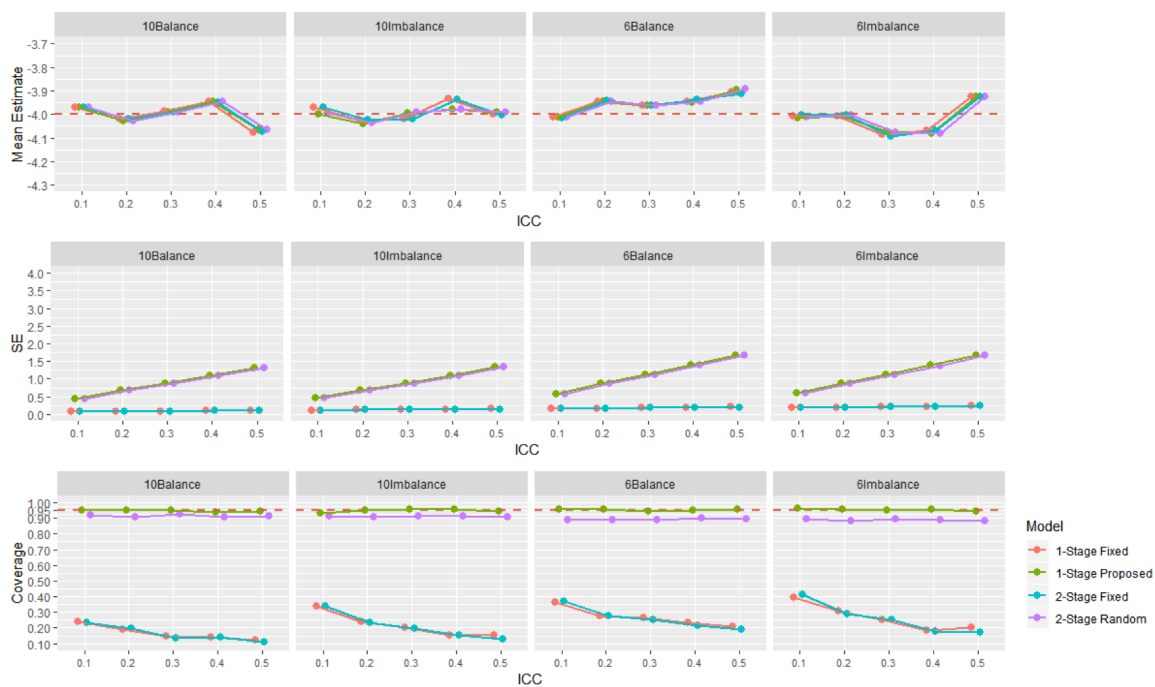


Figure 3-5: Simulation results for the 4-level data structure with distinct covariates.

For simulations with longitudinal data (Figures 3-4 and 3-5), the true value specified for the difference of slopes between groups is  $-4$ . Similarly, the impact of common or distinct covariates seems to be negligible on the results. Similar observations can be made for the mean estimate and standard error compared with the 3-level structure data. The mean estimates are accurate in most scenarios, and become less accurate and more fluctuated when the ICC increases, the sample size decreases, or with imbalanced design. The performance of the two fixed-effects models is poor compared with the other two models with the longitudinal data. The coverage probabilities for both fixed-effects models are equal or lower than 0.45 in all scenarios. When comparing the two-stage random-effects model and the one-stage proposed model, we still can see insufficient coverage probabilities shown by the two-stage random-effects model, and the discrepancy regarding coverage increases when the sample size decreases or becomes imbalanced.

### 3.3.3 Supplemental Simulation Study

In the simulation study described above, we assume the random effects used to generate simulated data are the same to that included in the proposed model. In this section, we conducted an additional simulation study to evaluate situations that random effects used to generate the data and the random effects used to fit the proposed model are different. We generated the simulated data with 12 studies, balance sample size design, and common covariates. Only two sets of random effects were included,  $r_{study}$  and  $r_{center}$ . The mean function conditional on random effects for the outcome variable is

$$\begin{aligned}
 E(Y_{kcl} | r_{study}, r_{center}) &= 220 - 8 \times TRT + 0.2 \times base\_cho + 0.2 \times age - 4 \times gender + 4 \times race_{AA} \\
 &+ 6 \times race_{His} - 0.8 \times race_{other} + 6 \times diabetes + 6 \times CVD - 0.3 \times SS_{mean} \\
 &+ 1.5 \times SS_{std} - 0.5 \times order + r_{study} + r_{center}
 \end{aligned}$$

Based on the simulated data, we fitted five test models as described in Table 3-3. Both two-stage methods ignored the center-level random effect in the first stage. The one-stage proposed model included three sets of random effects, with one redundant random effect  $r_{study\_treatment}$ .

Table 3-3: Model specifications of the supplemental simulation study.

Test model	Model fitting
two-stage fixed-effects	1st stage: study-level covariates, participant-level covariates; 2nd stage: a fixed-effects model to pool treatment effect
two-stage random-effects	1st stage: study-level covariates, participant-level covariates; 2nd stage: a random-effects model to pool treatment effect
one-stage fixed-effects	participant-level covariates
one-stage proposed	study-level covariates, participant-level covariates, $r_{study}$ , $r_{study\_treatment}$ , $r_{center}$

We altered the ICC from 0.1, 0.2, to 0.5, and conducted 1000 simulations for each ICC value. The simulation results are summarized in Table 3-4. The one-stage proposed model had a higher chance in running into the convergence problem as it included a redundant random effect term, and its variance was not able to be decided. In this case, it also indicates the need to remove certain random effects. All the four test models provided similar coverage probabilities and mean estimates, though the two-stage random-effects model still showed relatively insufficient coverage probabilities compared with other methods.

Table 3-4: Results for the supplemental simulation study.

ICC	Model	N	Mean estimate	SE	Coverage
0.1	two-stage fixed-effects	1000	-7.998	0.383	0.950
	one-stage proposed	565	-7.988	0.363	0.970
	one-stage fixed-effects	1000	-7.996	0.392	0.965
	two-stage random-effects	1000	-8.098	0.363	0.907
0.2	two-stage fixed-effects	1000	-8.024	0.398	0.954
	one-stage proposed	553	-8.017	0.363	0.966
	one-stage fixed-effects	1000	-8.021	0.414	0.968
	two-stage random-effects	1000	-8.021	0.358	0.887
0.3	two-stage fixed-effects	1000	-8.025	0.412	0.970
	one-stage proposed	1000	-8.027	0.391	0.977
	one-stage fixed-effects	1000	-8.026	0.441	0.981
	two-stage random-effects	1000	-7.955	0.355	0.897
0.4	two-stage fixed-effects	1000	-7.993	0.433	0.982
	one-stage proposed	540	-7.989	0.364	0.978
	one-stage fixed-effects	1000	-7.993	0.473	0.992
	two-stage random-effects	1000	-8.018	0.357	0.882
0.5	two-stage fixed-effects	1000	-7.985	0.461	0.982
	one-stage proposed	547	-7.980	0.364	0.973
	one-stage fixed-effects	1000	-7.981	0.517	0.993

---

two-stage random-effects      1000      -11.848      0.366      0.855

---

### 3.4 Real Data Application

#### 3.4.1 Example Data - Blood Pressure Studies

To assess the performance of the one-stage proposed model in a real setting, we selected 3 studies sponsored by the National Heart, Lung, and Blood Institute (NHLBI) to determine the effect of reducing sodium intake on lowering blood pressure. The 3 studies identified are DASH-Sodium (Sacks et al., 2001), PREMIER (Appel et al., 2003), and TOHP (phase II) (Trials of Hypertension Prevention Collaborative Research Group, 1997), and the study information are summarized in Table 3-5. IPD from each study were collected from the NHLBI website (<https://biolincc.nhlbi.nih.gov/home/>).

Table 3-5: Study information summary of the NHLBI studies.

Study	DASH-Sodium	PREMIER	TOHP (phase II)
Study period	1997-2002	1998-2004	1986-1998
Inclusion	SBP 120-159 mmHg; DBP 80-95 mmHg; free of anti-hypertensive medications	SBP 120-159 mmHg; DBP 80-95 mmHg; free of anti-hypertensive medications	SBP <140 mmHg; DBP 83-89 mmHg; free of anti-hypertensive medications
Age range	≥22	≥25	30-54
Study design	crossover RCT; DASH diet + sodium reduction	parallel RCT; DASH diet + sodium reduction	parallel RCT; weight loss + sodium reduction
Study arms used	low vs high (control) sodium level	low vs control sodium level	low vs control sodium level
Subjects used	204	541	1191
Number of centers	5	4	9
Outcome	change of SBP after 1 month change of DBP after 1 month	change of SBP after 3 months change of DBP after 3 months	change of SBP after 6 months change of DBP after 6 months

SBP: systolic blood pressure; DBP: diastolic blood pressure; RCT: randomized clinical trial.

The 3 studies were designed as randomized clinical trials, comparing the effect on lowering the blood pressure by combining treatment strategies of sodium intake reduction and weight loss or DASH diet, targeting on the high-risk population of hypertension (Moore et al., 2011). The normal blood pressure range is currently defined as systolic blood pressure (SBP) <120 mmHg and diastolic blood pressure (DBP) <80 mmHg (World Health Organization, 2015). For the purpose of this real data application, we did not include all treatment arms for analysis, and only focus on the comparison of sodium intake reduction group versus control, advice only, or usual care. DASH-Sodium had a crossover randomization design, and participants from DASH-Sodium were used as self-comparison. The measurement times of the 3 studies were not the same, and we use the closest measurement times between studies, which were the end of 1 month, 3 months, and 6 months for DASH-Sodium, PREMIER, and TOHP, respectively. For the outcome variables, we focused on the change of blood pressure from baseline to the selected end time point of each study, including the change of systolic blood pressure (SBP) and the change of diastolic blood pressure (DBP).

Table 3-6: Model fitting with NHLBI studies.

Covariates	DASH-sodium	PREMIER	TOHP (phase II)
Participant-level			
Sodium level (low or control)	X	X	X
Age	X	X	X
Female	X	X	X
Race	X	X	X
Baseline SBP/DBP	X	X	X
Baseline weight	X	X	X
Study level			
SS_mean	X	X	X
SS_std	X	X	X
Random effect			
Study_treatment	X	X	X
Center	X	X	X

SBP: systolic blood pressure; DBP: diastolic blood pressure.

We included participant-level variables, study-level variables, and random effects in the one-stage proposed model (Table 3-6). For comparison, we calculated the individual treatment effects estimated from each study. We also fitted the two-stage fixed-effects model, the two-stage random-effects model, and the one-stage fixed-effects model to compare with the one-stage proposed model. We conducted all analyses in SAS 9.4.

### 3.4.1 Results

The individual treatment effect estimates along with the overall estimates from the proposed model and estimates from other meta-analysis models are presented in Figures 3-6 and 3-7 for the change of SBP. For the change of DBP, the results are presented in Figures 3-8 and 3-9.

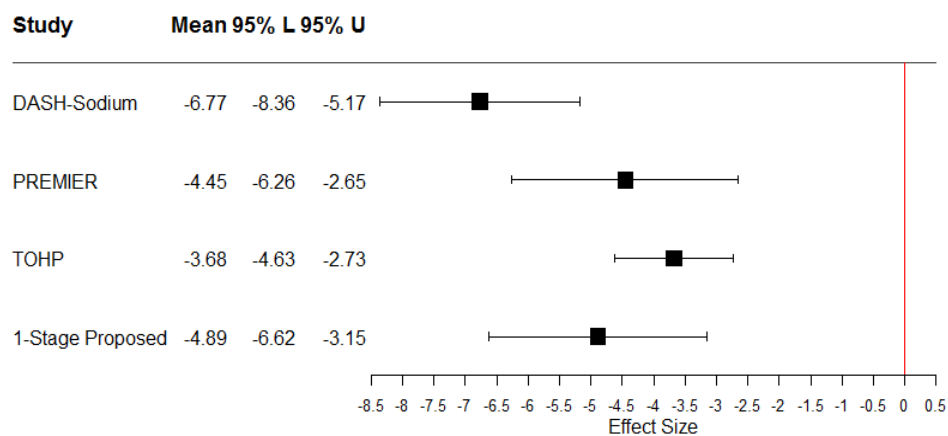


Figure 3-6: Individual and pooled estimates with outcome, the change of SBP.

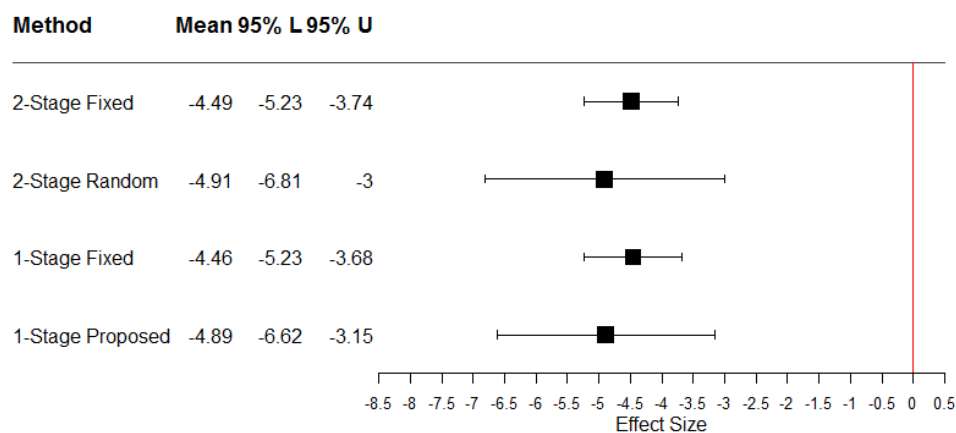


Figure 3-7: Meta-analysis methods comparison with outcomes, the change of SBP.

For the outcome variable, the change of SBP, all estimates yield statistically significant results. As seen from the individual estimates from each study (Figure 3-6), DASH-Sodium (-6.77, 95% CI (-8.36, -5.17)) shows the largest effect size estimate, followed by PREMIER (-4.45, 95% CI (-6.26, -2.65)) and TOHP (-3.68, 95% CI (-4.63, -2.73)). The overall treatment effect estimate obtained from the one-stage proposed model is (-4.89, 95% CI (-6.62, -3.15)). A certain amount of heterogeneity of treatment effect exists across studies, and the one-stage proposed model gives an appropriate overall estimate. Comparing the estimate from the one-stage proposed model with other tested models (Figure 3-7), both two-stage and one-stage fixed-effects models yielded estimates closer to the TOHP's individual estimate, and with narrower 95% confidence intervals (two-stage fixed-effects: -4.49, 95% CI (-5.23, -3.74); one-stage fixed-effects: -4.46, 95% CI (-5.23, -3.68)). Since the TOHP study has the largest sample size among all three studies, and both fixed-effects models could not appropriately incorporate the heterogeneity across studies, their estimates are greatly dominated by the TOHP study. The two-stage random-effects model yields similar results as the one-stage proposed model. But as seen in the simulation study, the two-stage random effects model may result in insufficient coverage probability when sample size becomes smaller or the sample sizes are not equivalent between studies.



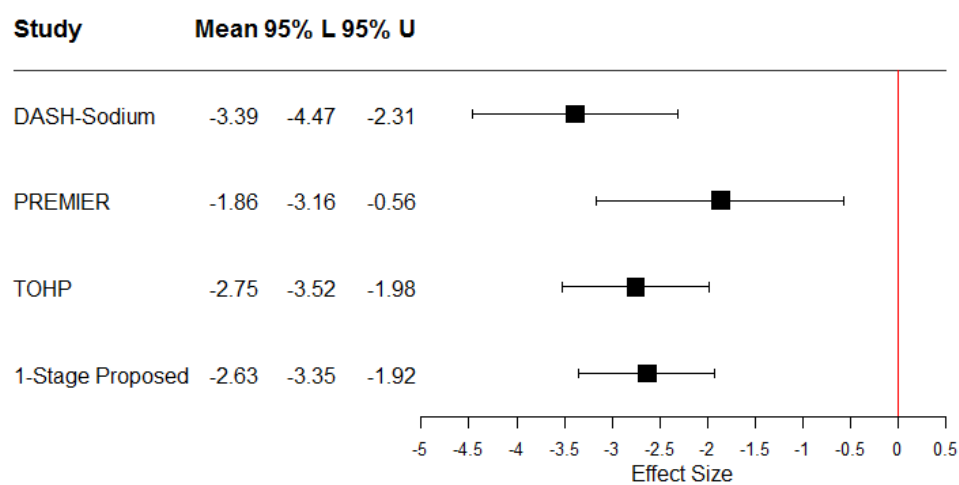


Figure 3-8: Individual and pooled estimates with outcome, the change of DBP.

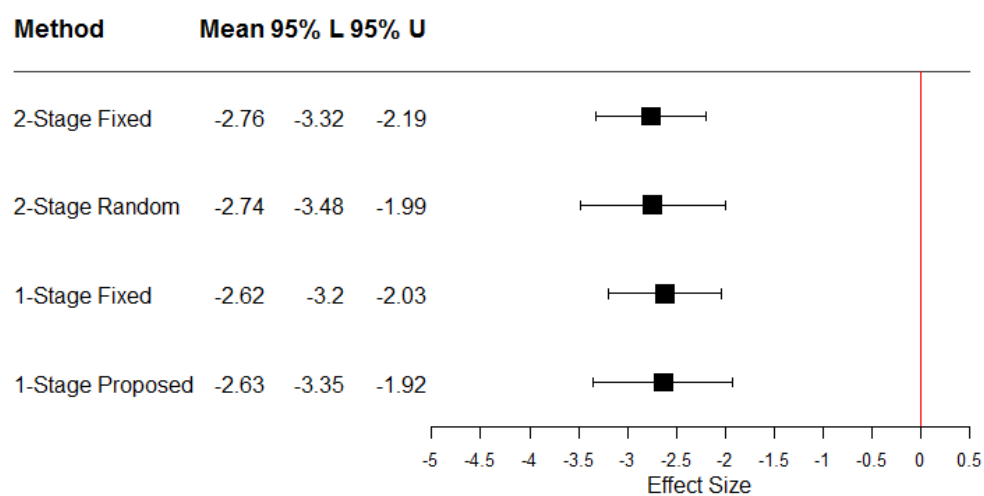


Figure 3-9: Meta-analysis methods comparison with outcomes, the change of DBP.

For the outcome variable, the change of DBP, all estimates still yield statistically significant results though with a smaller effect size scale. As seen from the individual estimates from each study (Figure 3-8), DASH-Sodium (-3.39, 95% CI (-4.47, -2.31)) shows the largest effect size estimate, followed by TOHP (-2.75, 95% CI (-3.52, -1.98)) and PREMIER (-1.86, 95% CI (-

3.16, -0.56)). The overall treatment effect estimate obtained from the one-stage proposed model is (-2.63, 95% CI (-3.35, -1.92)). A certain amount of heterogeneity of treatment effect exists across studies, but the four tested models present small differences regarding the mean estimate (Figure 3-7). This may be due to the study with the largest sample size, TOHP, having an intermediate estimate among the three individual estimates, so both fixed-effects models seem to yield reasonable overall estimates. The one-stage proposed model gives an appropriate overall estimate. When comparing the estimate from the one-stage proposed model with other tested models, both two-stage and one-stage fixed-effects models still show narrower 95% confidence intervals (two-stage fixed-effects: -2.76, 95% CI (-3.32, -2.19); one-stage fixed-effects: -2.74, 95% CI (-3.48, -1.99)). Little difference is observed between the results of the two-stage random-effects model and the one-stage proposed model (two-stage random-effects: -2.74, 95% CI (-3.48, -1.99)).

## Chapter 4

### The One-Stage Multi-Level Mixed-Effects Model for IPD Meta-Analysis with Outcomes from An Exponential Family

In the previous chapter, we proposed an IPD meta-analysis method for data with a continuous outcome. But in practice, it also is very common to have data with other types of outcomes, such as binary and count outcomes. In this chapter, we extend the previous work to accommodate outcomes from an exponential family, which can greatly expand the application range of the method. If the primary outcome variable has a distribution from the exponential family (e.g., binary, binominal, Poisson, negative binominal with a fixed scale parameter, exponential, gamma, beta, central t, etc.), then we can invoke a generalized linear mixed-effects model (GLMM) (Breslow et al., 1993; Diggle et al., 2002; Karim et al., 1992; McCulloch et al., 2001).

## 4.1 Methods

### 4.1.1 Model Introduction

The GLMM is comprised of a generalized linear model with normal random effects. Let  $Y_{kl}(t)$  denote the outcome variable measured at time  $t$  for the  $l^{th}$  participant within the  $k^{th}$  study,  $k = 1, 2, \dots, K$  and  $l = 1, 2, \dots, n_k$ . We assume that a monotonic link function,  $g$ , of the expected value of  $Y_{kl}(t)$ , conditional on the fixed-effect parameters and the random-effect parameters, can be expressed as a linear predictor, i.e.,

$$\begin{aligned} g[E\{Y_{kl}(t) \mid \mathbf{x}_{u,kl}^T(t), \boldsymbol{\beta}_{uk}, \mathbf{x}_{c,kl}^T(t), \boldsymbol{\beta}_c, \mathbf{z}_{kl}(t), \boldsymbol{\gamma}_{kl}, \mathbf{x}_k^*, \boldsymbol{\beta}^*, \mathbf{z}_k^*, \boldsymbol{\gamma}_k^*\}] \\ = \mathbf{x}_{u,kl}^T(t) \boldsymbol{\beta}_{uk} + \mathbf{x}_{c,kl}^T(t) \boldsymbol{\beta}_c + \mathbf{z}_{kl}^T(t) \boldsymbol{\gamma}_{kl} + \mathbf{x}_k^{*T} \boldsymbol{\beta}^* + \mathbf{z}_k^{*T} \boldsymbol{\gamma}_k^* \end{aligned}$$

where

- $\mathbf{x}_{u,kl}(t)$  is a participant-level, unique fixed-effects,  $r_{uk} \times 1$  vector of design effects and covariates at time  $t$  for the  $l^{th}$  participant within the  $k^{th}$  study
- $\boldsymbol{\beta}_{uk}$  is a participant-level, unique fixed-effects  $r_{uk} \times 1$  vector of parameters for the  $k^{th}$  study
- $\mathbf{x}_{c,kl}(t)$  is a participant-level, common fixed-effects,  $r_c \times 1$  vector of design effects and covariates at time  $t$  for the  $l^{th}$  participant within the  $k^{th}$  study
- $\boldsymbol{\beta}_c$  is a participant-level, common fixed-effects  $r_c \times 1$  vector of parameters
- $\mathbf{z}_{kl}(t)$  is a participant-level, random-effects  $s_k \times 1$  vector of design effects and covariates at time  $t$  for the  $l^{th}$  participant within the  $k^{th}$  cohort study
- $\boldsymbol{\gamma}_{kl}$  is a participant-level, random-effects  $s_k \times 1$  vector of parameters for the  $l^{th}$  participant within the  $k^{th}$  cohort study
- $\mathbf{x}_k^*$  is a cohort-level, fixed-effects,  $r^* \times 1$  vector of design effects and covariates for the  $k^{th}$  cohort study
- $\boldsymbol{\beta}^*$  is a cohort-level, fixed-effects  $r^* \times 1$  vector of parameters for the  $k^{th}$  cohort study
- $\mathbf{z}_k^*$  is a cohort-level, random-effects  $s^* \times 1$  vector of design effects and covariates for the  $k^{th}$  cohort study
- $\boldsymbol{\gamma}_k^*$  is a cohort-level, random-effects  $s^* \times 1$  vector of parameters for the  $k^{th}$  cohort study
- the  $\boldsymbol{\gamma}_{kl}$ 's are independent with  $\boldsymbol{\gamma}_{kl} \sim N_{s_k}(\mathbf{0}, \boldsymbol{\Gamma}_k)$ , where  $\boldsymbol{\Gamma}_k$  is a positive definite matrix
- the  $\boldsymbol{\gamma}_k^*$ 's are independent with  $\boldsymbol{\gamma}_k^* \sim N_{s^*}(\mathbf{0}, \boldsymbol{\Gamma}^*)$ , where  $\boldsymbol{\Gamma}^*$  is a positive definite matrix
- the  $\boldsymbol{\gamma}_{kl}$ 's and the  $\boldsymbol{\gamma}_k^*$ 's are mutually independent

Conditional on random effects  $\boldsymbol{\gamma}_{kl}$  and  $\boldsymbol{\gamma}_k^*$ , the  $Y_{kl}(t)$ 's are independent, with a density from the exponential family

$$f(Y_{kl}(t)|\boldsymbol{\gamma}_{kl}, \boldsymbol{\gamma}_k^*) = \exp \left\{ \frac{Y_{kl}(t)\theta_{klt} - b(\theta_{klt})}{a_{klt}(\phi)} + c(Y_{kl}(t), \phi) \right\}$$

where  $\theta_{klt}$  is known as the canonical parameter,  $\phi$  is a fixed dispersion parameter,  $a_{klt}(\cdot)$  is some specific function of  $\phi$ , and  $b(\cdot)$  is some specific function of  $\theta_{klt}$ . The forms of  $a_{klt}(\cdot)$  and  $b(\cdot)$  depend on the distribution of the outcome. There are two properties of outcomes from the exponential family, described as follows:

$$E(Y_{kl}(t)|\boldsymbol{\gamma}_{kl}, \boldsymbol{\gamma}_k^*) = \mu_{klt} = b'(\theta_{klt})$$

$$\text{Var}(Y_{kl}(t)|\boldsymbol{\gamma}_{kl}, \boldsymbol{\gamma}_k^*) = b''(\theta_{klt})a_{klt}(\phi)$$

Parameter estimation proceeds by maximizing the marginal likelihood  $L(\boldsymbol{\beta}_{u1}, \dots, \boldsymbol{\beta}_{uK}, \boldsymbol{\beta}_c, \boldsymbol{\beta}^*, \boldsymbol{\xi}|\mathbf{Y})$ , which is constructed by integrating the full likelihood function with respect to the distribution of the random effects

$$\begin{aligned} L(\boldsymbol{\beta}_{u1}, \dots, \boldsymbol{\beta}_{uK}, \boldsymbol{\beta}_c, \boldsymbol{\beta}^*, \boldsymbol{\xi}|\mathbf{Y}) &= \prod_{k=1}^K \prod_{l=1}^{n_k} f(\mathbf{Y}_{kl}; \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}_c, \boldsymbol{\beta}^*, \boldsymbol{\xi}) \\ &= \prod_{k=1}^K \left\{ \int \left( \prod_{l=1}^{n_k} \int f(\mathbf{Y}_{kl} | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}_c, \boldsymbol{\beta}^*, \boldsymbol{\xi}, \boldsymbol{\gamma}_{kl}, \boldsymbol{\gamma}_k^*) f(\boldsymbol{\gamma}_{kl} | \boldsymbol{\xi}) d\boldsymbol{\gamma}_{kl} \right) f(\boldsymbol{\gamma}_k^* | \boldsymbol{\xi}) d\boldsymbol{\gamma}_k^* \right\} \end{aligned}$$

where  $\boldsymbol{\xi}$  is the vector of variance-covariance parameters. However, if the link function is not linear, the marginal likelihood can become intractable, and approximation of the integral, such as linearization methods (quasi-likelihood methods), Laplace's method, and numerical approximation, are necessary.

There are two general types of integral approximation for GLMMs with normally distributed random effects. The first type is to approximate the integrand, so that the integral of the approximation has a closed form, which is also called analytical integration. The two typical approaches are quasi-likelihood methods (Breslow et al., 1993; Diggle et al., 2002; Karim et al., 1992; McCulloch et al., 2001; Wolfinger et al., 1993) and Laplace's method (Tuerlinckx et al., 2006). The quasi-likelihood methods include penalized quasi-likelihood (PQL) and marginal quasi-

likelihood (MQL). The second type of integral approximation for GLMMs is to approximate the integral numerically. The typical approaches are Gauss quadrature (McCulloch et al., 1993; Tuerlinckx et al, 2006), simulated maximum likelihood method (McCulloch et al., 1993; Tuerlinckx et al, 2006; McCulloch et al., 1997), and the expectation-maximization (EM) algorithm (McCulloch et al., 1997; Tuerlinckx et al, 2006; McCulloch et al., 1997). The following sections introduce the methods separately. In the analysis of the simulation study and the real data application, we will invoke PQL to obtain the estimates of fixed effects and variance-covariance parameters.

#### 4.1.2 Quasi-likelihood Methods

##### 1) Penalized Quasi-likelihood (PQL)

Instead of defining a specific distribution, only conditional means and variances are defined in PQL. Given random effects  $\boldsymbol{\gamma}_{kl}$  and  $\boldsymbol{\gamma}_k^*$ , the  $Y_{kl}(t)$  are conditionally independent with means

$$E[Y_{kl}(t)|\boldsymbol{\gamma}_{kl}, \boldsymbol{\gamma}_k^*] = \mu_{klt}$$

and variances

$$Var[Y_{kl}(t)|\boldsymbol{\gamma}_{kl}, \boldsymbol{\gamma}_k^*] = a_{klt}(\phi)v(\mu_{klt}) = a_{klt}(\phi)v_{klt},$$

where  $v(\cdot)$  is a specific function of  $\mu_{klt}$  depending on the distribution of the outcome. The conditional mean is related to a linear predictor  $\eta_{klt}$  through a link function  $g(\cdot)$ :

$$g(\mu_{klt}) = \eta_{klt} = \mathbf{x}_{u,kl}^T(t)\boldsymbol{\beta}_{uk} + \mathbf{x}_{c,kl}^T(t)\boldsymbol{\beta}_c + \mathbf{z}_{kl}^T(t)\boldsymbol{\gamma}_{kl} + \mathbf{x}_k^{*T}\boldsymbol{\beta}^* + \mathbf{z}_k^{*T}\boldsymbol{\gamma}_k^*$$

By stacking  $g(\mu_{klt})$ , we construct

$$g(\boldsymbol{\mu}_{kl}) = \begin{bmatrix} \mathbf{X}_{u,kl} & \mathbf{X}_{c,kl} & \mathbf{1}_{p_{kl} \times 1} \mathbf{x}_k^{*T} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{uk} \\ \boldsymbol{\beta}_c \\ \boldsymbol{\beta}^* \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_{kl} & \mathbf{1}_{p_{kl} \times 1} \mathbf{z}_k^{*T} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma}_{kl} \\ \boldsymbol{\gamma}_k^* \end{bmatrix}$$

and

$$\begin{aligned}
g(\boldsymbol{\mu}_k) &= \begin{bmatrix} g(\boldsymbol{\mu}_{k1}) \\ \vdots \\ g(\boldsymbol{\mu}_{kn_k}) \end{bmatrix}_{(\sum_{l=1}^{n_k} p_{kl}) \times 1} = \begin{bmatrix} \mathbf{X}_{u,k1} \\ \vdots \\ \mathbf{X}_{u,kn_k} \end{bmatrix}_{(\sum_{l=1}^{n_k} p_{kl}) \times r_{uk}} \boldsymbol{\beta}_{uk} \quad r_{uk} \times 1 \\
&+ \begin{bmatrix} \mathbf{X}_{c,k1} & \mathbf{1}_{p_{k1} \times 1} \mathbf{x}_k^{*T} \\ \vdots & \vdots \\ \mathbf{X}_{c,kn_k} & \mathbf{1}_{p_{kn_k} \times 1} \mathbf{x}_k^{*T} \end{bmatrix}_{(\sum_{l=1}^{n_k} p_{kl}) \times (r_c + r^*)} \begin{bmatrix} \boldsymbol{\beta}_c \\ \boldsymbol{\beta}^* \end{bmatrix}_{(r_c + r^*) \times 1} \\
&+ \begin{bmatrix} \mathbf{Z}_{k1} & \cdots & 0 & \mathbf{1}_{p_{k1} \times 1} \mathbf{z}_k^{*T} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \mathbf{Z}_{kn_k} & \mathbf{1}_{p_{kn_k} \times 1} \mathbf{z}_k^{*T} \end{bmatrix}_{(\sum_{l=1}^{n_k} p_{kl}) \times (s_k \times n_k + s^*)} \begin{bmatrix} \boldsymbol{\gamma}_{k1} \\ \vdots \\ \boldsymbol{\gamma}_{kn_k} \\ \boldsymbol{\gamma}_k^* \end{bmatrix}_{(s_k \times n_k + s^*) \times 1} \\
&= \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} + \mathbf{X}_k \boldsymbol{\beta} + \mathbf{Z}_k \boldsymbol{\gamma}_k
\end{aligned}$$

where

$$\boldsymbol{\gamma}_k \sim N_{s_k \times n_k + s^*}(\mathbf{0}, \mathbf{G}_k(\boldsymbol{\xi})) = \begin{bmatrix} \boldsymbol{\Gamma}_k & \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \vdots & \cdots & \boldsymbol{\Gamma}_k & \vdots \\ 0 & \cdots & \cdots & \boldsymbol{\Gamma}^* \end{bmatrix}.$$

Then, the marginal likelihood is approximated by the quasi-likelihood (QL), which is used to estimate  $\boldsymbol{\beta}_{uk}$ ,  $\boldsymbol{\beta}$ , and  $\boldsymbol{\xi}$ :

$$\begin{aligned}
L(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}) &= \prod_{k=1}^K \int f(\mathbf{Y}_k | \boldsymbol{\gamma}_k, \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}) f(\boldsymbol{\gamma}_k; \boldsymbol{\xi}) d\boldsymbol{\gamma}_k \\
&= \prod_{k=1}^K \int \prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\gamma}_k, \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}) f(\boldsymbol{\gamma}_k; \boldsymbol{\xi}) d\boldsymbol{\gamma}_k \\
&\approx \prod_{k=1}^K \int \exp \left[ \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} Q(Y_{kl}(t); \mu_{klt}) \right] |\mathbf{G}_k|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \boldsymbol{\gamma}_k^T \mathbf{G}_k^{-1} \boldsymbol{\gamma}_k \right] d\boldsymbol{\gamma}_k \\
&= \prod_{k=1}^K |\mathbf{G}_k|^{-\frac{1}{2}} \int \exp \left[ \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} Q(Y_{kl}(t); \mu_{klt}) \right] \exp \left[ -\frac{1}{2} \boldsymbol{\gamma}_k^T \mathbf{G}_k^{-1} \boldsymbol{\gamma}_k \right] d\boldsymbol{\gamma}_k
\end{aligned}$$

where  $Q(Y_{kl}(t); \mu_{klt}) = \int_{Y_{kl}(t)}^{\mu_{klt}} \frac{Y_{kl}(t) - u}{a_{klt}(\phi) v_{klt}} du$  is the quasi-likelihood for  $Y_{kl}(t)$ . In this case,

$\sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} Q(Y_{kl}(t); \mu_{klt})$  represents the true log-likelihood of the data.

Now let

$$\begin{aligned} k(\boldsymbol{\gamma}_k) &= - \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} Q(Y_{kl}(t); \mu_{klt}) + \frac{1}{2} \boldsymbol{\gamma}_k^T \mathbf{G}_k^{-1} \boldsymbol{\gamma}_k \\ &= - \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \int_{Y_{kl}(t)}^{\mu_{klt}} \frac{Y_{kl}(t) - u}{a_{klt}(\phi) v_{klt}} du + \frac{1}{2} \boldsymbol{\gamma}_k^T \mathbf{G}_k^{-1} \boldsymbol{\gamma}_k \end{aligned}$$

Then based on Laplace's approximation, the marginal log-likelihood can be approximated by

$$l(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}) \approx \sum_{k=1}^K \left[ -\frac{1}{2} \log |\mathbf{G}_k| - \frac{1}{2} \log |k''(\tilde{\boldsymbol{\gamma}}_k)| - k(\tilde{\boldsymbol{\gamma}}_k) \right] \quad (4.1)$$

where  $\tilde{\boldsymbol{\gamma}}_k = \tilde{\boldsymbol{\gamma}}_k(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi})$  is the solution to  $k'(\boldsymbol{\gamma}_k) = 0$  that minimizes  $k(\boldsymbol{\gamma}_k)$ .  $k'(\boldsymbol{\gamma}_k)$  is derived as

$$\begin{aligned} k'(\boldsymbol{\gamma}_k) &= - \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \frac{\partial Q(Y_{kl}(t); \mu_{klt})}{\partial \mu_{klt}} \frac{\partial \mu_{klt}}{\partial \eta_{klt}} \frac{\partial \eta_{klt}}{\partial \boldsymbol{\gamma}_k} + \mathbf{G}_k^{-1} \boldsymbol{\gamma}_k \\ &= - \left[ \begin{array}{c} \sum_{t=1}^{p_{k1}} \frac{Y_{k1}(t) - \mu_{k1t_1}}{a_{k1t}(\phi) v_{k1t} g'(\mu_{k1t})} \mathbf{z}_{k1}(t) \\ \vdots \\ \sum_{t=1}^{p_{kn_k}} \frac{Y_{kn_k}(t) - \mu_{kn_k t}}{a_{kn_k t}(\phi) v_{kn_k t} g'(\mu_{kn_k t})} \mathbf{z}_{kn_k}(t) \\ \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \frac{Y_{kl}(t) - \mu_{klt}}{a_{klt}(\phi) v_{klt} g'(\mu_{klt})} \mathbf{z}_k^* \end{array} \right]_{(S_k \times n_k + S^*) \times 1} \\ &\quad + \mathbf{G}_k^{-1} \boldsymbol{\gamma}_k \end{aligned}$$

Differentiating with respect to  $\boldsymbol{\gamma}_k$ ,  $k''(\boldsymbol{\gamma}_k)$  is derived as

$$k''(\boldsymbol{\gamma}_k) = \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \frac{\partial Q'(Y_{kl}(t); \mu_{klt})}{\partial \mu_{klt}} \frac{\partial \mu_{klt}}{\partial \eta_{klt}} \frac{\partial \eta_{klt}}{\partial \boldsymbol{\gamma}_k} + \mathbf{G}_k^{-1}$$



$$\begin{aligned}
&= \begin{bmatrix} \sum_{t=1}^{p_{k1}} \frac{\mathbf{z}_{k1}(t)\mathbf{z}_{k1}(t)^T}{a_{k1t}(\phi)v_{k1t}[g'(\mu_{k1t})]^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \frac{\mathbf{z}_k^* \mathbf{z}_k^{*T}}{a_{klt}(\phi)v_{klt}[g'(\mu_{klt})]^2} \end{bmatrix} \\
&+ \begin{bmatrix} \sum_{t=1}^{p_{k1}} (Y_{k1}(t) - \mu_{k1t}) \mathbf{z}_{k1}(t) \frac{\partial}{\partial \boldsymbol{\gamma}_k} \left( \frac{1}{a_{k1t}(\phi)v_{k1t}g'(\mu_{k1t})} \right) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} (Y_{kl}(t) - \mu_{klt}) \mathbf{z}_k^* \frac{\partial}{\partial \boldsymbol{\gamma}_k} \left( \frac{1}{a_{klt}(\phi)v_{klt}g'(\mu_{klt})} \right) \end{bmatrix} \\
&+ \mathbf{G}_k^{-1} \\
&= \mathbf{Z}_k^T \mathbf{W}_k \mathbf{Z}_k + \mathbf{R}_k + \mathbf{G}_k^{-1} \\
&\approx \mathbf{Z}_k^T \mathbf{W}_k \mathbf{Z}_k + \mathbf{G}_k^{-1}
\end{aligned}$$

where  $\mathbf{W}_k = \text{diag} \left( \frac{1}{a_{klt}(\phi)v_{klt}[g'(\mu_{klt})]^2} \right)$ ,  $l = 1, \dots, n_k$  and  $t = 1, \dots, p_{kl}$ , is the  $(\sum_{l=1}^{n_k} p_{kl}) \times (\sum_{l=1}^{n_k} p_{kl})$  diagonal matrix that can be recognized as generalized linear model (GLM) iterated weights.  $\mathbf{R}_k$  is the reminder term that has expectation 0 and has lower order than the terms  $\mathbf{Z}_k^T \mathbf{W}_k \mathbf{Z}_k$  and  $\mathbf{G}_k^{-1}$ . By ignoring  $\mathbf{R}_k$  and replacing  $k(\tilde{\boldsymbol{\gamma}}_k)$  and  $k''(\tilde{\boldsymbol{\gamma}}_k)$  in  $l(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y})$ ,

$$\begin{aligned}
l(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}) &\approx \sum_{k=1}^K \left[ -\frac{1}{2} \log |\mathbf{G}_k| - \frac{1}{2} \log |k''(\tilde{\boldsymbol{\gamma}}_k)| - k(\tilde{\boldsymbol{\gamma}}_k) \right] \\
&= \sum_{k=1}^n \left[ -\frac{1}{2} \log |\mathbf{G}_k| - \frac{1}{2} \log |\mathbf{Z}_k^T \mathbf{W}_k \mathbf{Z}_k + \mathbf{G}_k^{-1}| + \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} Q(Y_{kl}(t); \mu_{klt}) \right. \\
&\quad \left. - \frac{1}{2} \tilde{\boldsymbol{\gamma}}_k^T \mathbf{G}_k^{-1} \tilde{\boldsymbol{\gamma}}_k \right] \\
&= \sum_{k=1}^n \left[ -\frac{1}{2} \log |\mathbf{Z}_k^T \mathbf{W}_k \mathbf{Z}_k \mathbf{G}_k + \mathbf{I}| + \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} Q(Y_{kl}(t); \mu_{klt}) - \frac{1}{2} \tilde{\boldsymbol{\gamma}}_k^T \mathbf{G}_k^{-1} \tilde{\boldsymbol{\gamma}}_k \right] \quad (4.2)
\end{aligned}$$

Assuming  $(\mathbf{Z}_k^T \mathbf{W}_k \mathbf{Z}_k \mathbf{G}_k + \mathbf{I})$  varies slowly since the iterative weights vary slowly as a function of the mean, we ignore the first term in (4.2). Therefore, the PQL can be defined as

$$PQL = \sum_{k=1}^K \left[ \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} Q(Y_{kl}(t); \mu_{klt}) - \frac{1}{2} \boldsymbol{\gamma}_k^T \mathbf{G}_k^{-1} \boldsymbol{\gamma}_k \right] \quad (4.3)$$

The estimates  $(\widehat{\boldsymbol{\beta}}_{uk}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}) = (\widehat{\boldsymbol{\beta}}_{uk}(\boldsymbol{\xi}), \widehat{\boldsymbol{\beta}}(\boldsymbol{\xi}), \widehat{\boldsymbol{\gamma}}(\boldsymbol{\xi}))$  can be obtained by jointly maximizing the PQL in (4.3). Differentiation of PQL with respect to  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}_k$  leads to the score equations:

$$\frac{\partial PQL}{\partial \boldsymbol{\beta}_{uk}} = \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \frac{(Y_{kl}(t) - \mu_{klt}) \mathbf{x}_{u,klt}}{a_{klt}(\phi) v_{klt} g'(\mu_{klt})} = 0 \quad (4.4)$$

$$\frac{\partial PQL}{\partial \boldsymbol{\beta}} = \sum_{k=1}^K \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \frac{(Y_{kl}(t) - \mu_{klt}) \mathbf{x}_{klt}}{a_{klt}(\phi) v_{klt} g'(\mu_{klt})} = 0 \quad (4.5)$$

$$\frac{\partial PQL}{\partial \boldsymbol{\gamma}_k} = \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \frac{(Y_{kl}(t) - \mu_{klt}) \mathbf{z}_{klt}}{a_{klt}(\phi) v_{klt} g'(\mu_{klt})} - \mathbf{G}_k^{-1} \boldsymbol{\gamma}_k = 0 \quad k = 1, \dots, K \quad (4.6)$$

(4.4) - (4.6) can be solved through an iterated weighted least squares approach using a linearization technique. Let

$$E(\mathbf{Y}_k | \boldsymbol{\gamma}_k) = \boldsymbol{\mu}_k = g^{-1}(\mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} + \mathbf{X}_k \boldsymbol{\beta} + \mathbf{Z}_k \boldsymbol{\gamma}_k) = g^{-1}(\boldsymbol{\eta}_k)$$

$$Var(\mathbf{Y}_k | \boldsymbol{\gamma}_k) = \mathbf{E}_k = \text{diag}(a_{klt}(\phi) v_{klt})$$

Based on a first-order Taylor series expansion at the current estimates  $\boldsymbol{\beta}_{uk} = \widetilde{\boldsymbol{\beta}}_{uk}$ ,  $\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}$ , and  $\boldsymbol{\gamma}_k = \widetilde{\boldsymbol{\gamma}}_k$ ,

$$\begin{aligned} \mathbf{Y}_k &= g^{-1}(\widetilde{\boldsymbol{\eta}}_k) + \widetilde{\Delta}^{-1} \mathbf{X}_{u,k} (\boldsymbol{\beta}_{uk} - \widetilde{\boldsymbol{\beta}}_{uk}) + \widetilde{\Delta}^{-1} \mathbf{X}_k (\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}) + \widetilde{\Delta}^{-1} \mathbf{Z}_k (\boldsymbol{\gamma}_k - \widetilde{\boldsymbol{\gamma}}_k) \\ &\quad + \boldsymbol{\varepsilon}_k \end{aligned}$$

where  $\widetilde{\Delta}^{-1} = \text{diag} \left( \frac{\partial g^{-1}(\boldsymbol{\eta}_k)}{\partial \boldsymbol{\eta}_k} \Big|_{\boldsymbol{\eta}_k = \widetilde{\boldsymbol{\eta}}_k} \right) = \text{diag}(g'(\widetilde{\boldsymbol{\mu}}_k))^{-1}$ , and  $Var(\boldsymbol{\varepsilon}_k) = \mathbf{E}_k$ .

For observations within the  $k^{th}$  study, let

$$\begin{aligned} \mathbf{Y}_k^* &= \widetilde{\Delta} [\mathbf{Y}_k - g^{-1}(\widetilde{\boldsymbol{\eta}}_k)] + \mathbf{X}_{u,k} \widetilde{\boldsymbol{\beta}}_{uk} + \mathbf{X}_k \widetilde{\boldsymbol{\beta}} + \mathbf{Z}_k \widetilde{\boldsymbol{\gamma}}_k \\ &= \mathbf{X}_{u,k} \boldsymbol{\beta}_{uk} + \mathbf{X}_k \boldsymbol{\beta} + \mathbf{Z}_k \boldsymbol{\gamma}_k + \boldsymbol{\varepsilon}_k \end{aligned}$$

$$Var(\mathbf{Y}_k^*) = \widetilde{\Delta} Var(\mathbf{Y}_k) \widetilde{\Delta}^T = \mathbf{Z}_k \mathbf{G}_k \mathbf{Z}_k^T + \widetilde{\Delta} \mathbf{E}_k \widetilde{\Delta}^T = \mathbf{Z}_k \mathbf{G}_k \mathbf{Z}_k^T + \mathbf{W}_k^{-1}$$

Thus, we can construct  $\mathbf{Y}^*$  for all studies as

$$\begin{aligned}\mathbf{Y}^* &= \begin{bmatrix} \mathbf{Y}_1^* \\ \vdots \\ \mathbf{Y}_K^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{u,1} & \cdots & 0 & \mathbf{X}_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \mathbf{X}_{u,K} & \mathbf{X}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{u1} \\ \vdots \\ \boldsymbol{\beta}_{uK} \\ \boldsymbol{\beta} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{Z}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \vdots \\ \boldsymbol{\gamma}_K \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_K \end{bmatrix} \\ &= \mathbf{X}\bar{\boldsymbol{\beta}} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\end{aligned}$$

with  $E(\mathbf{Y}^*) = \mathbf{X}\bar{\boldsymbol{\beta}}$  and  $Var(\mathbf{Y}^*) = \mathbf{ZGZ}^T + \mathbf{W}^{-1}$ , where  $\mathbf{G} = diag(\mathbf{G}_k)$  and  $\mathbf{W}^{-1} = diag(\mathbf{W}_k^{-1})$ . Based on  $\mathbf{Y}^*$ , estimation can proceed and be solved as a linear mixed-effects model (see chapter 3).

The iterations for solving a GLMM based on the PQL algorithm is as follows:

- 1) Obtain initial estimates  $\tilde{\boldsymbol{\beta}}_{uk}$ ,  $\tilde{\boldsymbol{\beta}}$ , and  $\tilde{\boldsymbol{\gamma}}$ ;
- 2) Construct the pseudo outcome  $\mathbf{Y}^*$  based on current estimates  $\tilde{\boldsymbol{\beta}}_{uk}$ ,  $\tilde{\boldsymbol{\beta}}$ , and  $\tilde{\boldsymbol{\gamma}}$ ;
- 3) Update  $\tilde{\boldsymbol{\beta}}_{uk}$ ,  $\tilde{\boldsymbol{\beta}}$ , and  $\tilde{\boldsymbol{\gamma}}$ , and obtain the estimate  $\tilde{\boldsymbol{\xi}}$  by solving the GLM of  $\mathbf{Y}^*$  (ML or REML);
- 4) Repeat 2) to 3) until convergence.

## 2) Marginal Quasi-likelihood (MQL)

The MQL method is similar to PQL method. It also starts with a first-order Taylor series expansion, but the expansion is now at the current estimate for  $\boldsymbol{\beta}_{uk} = \tilde{\boldsymbol{\beta}}_{uk}$ ,  $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$ , and  $\boldsymbol{\gamma}_k = \mathbf{0}$ .

Let

$$E(\mathbf{Y}_k) = \boldsymbol{\mu}_k = g^{-1}(\mathbf{X}_{u,k}\boldsymbol{\beta}_{uk} + \mathbf{X}_k\boldsymbol{\beta}) = g^{-1}(\boldsymbol{\eta}'_k)$$

Then, based on a first-order Taylor series expansion at the current estimate  $\boldsymbol{\beta}_{uk} = \tilde{\boldsymbol{\beta}}_{uk}$ ,  $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$ , and  $\boldsymbol{\gamma}_k = \mathbf{0}$ ,

$$\mathbf{Y}_k = g^{-1}(\boldsymbol{\eta}'_k) + \tilde{\Delta}^{-1}\mathbf{X}_{u,k}(\boldsymbol{\beta}_{uk} - \tilde{\boldsymbol{\beta}}_{uk}) + \tilde{\Delta}^{-1}\mathbf{X}_k(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \tilde{\Delta}^{-1}\mathbf{Z}_k\boldsymbol{\gamma}_k + \boldsymbol{\varepsilon}_k$$

where  $\tilde{\Delta}^{-1} = \text{diag}\left(\frac{\partial g^{-1}(\eta'_k)}{\partial \eta'_k} \Big|_{\eta'_k = \tilde{\eta}'_k}\right) = \text{diag}(g'(\tilde{\boldsymbol{\mu}}_k))^{-1}$ , and  $\text{Var}(\boldsymbol{\varepsilon}_k) = \mathbf{E}_k$ . For observations within the  $k^{\text{th}}$  study, let

$$\begin{aligned} \mathbf{Y}_k^* &= \tilde{\Delta}[\mathbf{Y}_k - g^{-1}(\tilde{\boldsymbol{\eta}}'_k)] + \mathbf{X}_{u,k}\tilde{\boldsymbol{\beta}}_{uk} + \mathbf{X}_k\tilde{\boldsymbol{\beta}} = +\mathbf{X}_{u,k}\boldsymbol{\beta}_{uk} + \mathbf{X}_k\boldsymbol{\beta} + \mathbf{Z}_k\boldsymbol{\gamma}_k + \boldsymbol{\varepsilon}_k \\ \text{Var}(\mathbf{Y}_k^*) &= \tilde{\Delta}\text{Var}(\mathbf{Y}_k)\tilde{\Delta}^T = \mathbf{Z}_k\mathbf{G}_k\mathbf{Z}_k^T + \tilde{\Delta}\mathbf{E}_k\tilde{\Delta}^T = \mathbf{Z}_k\mathbf{G}_k\mathbf{Z}_k^T + \mathbf{W}_k^{-1} \end{aligned}$$

Thus,

$$\begin{aligned} \mathbf{Y}^* &= \begin{bmatrix} \mathbf{Y}_1^* \\ \vdots \\ \mathbf{Y}_K^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{u,1} & \cdots & 0 & \mathbf{X}_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \mathbf{X}_{u,K} & \mathbf{X}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{u1} \\ \vdots \\ \boldsymbol{\beta}_{uK} \\ \boldsymbol{\beta} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{Z}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \vdots \\ \boldsymbol{\gamma}_K \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_K \end{bmatrix} \\ &= \mathbf{X}\bar{\boldsymbol{\beta}} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \end{aligned}$$

with  $E(\mathbf{Y}^*) = \mathbf{X}\bar{\boldsymbol{\beta}}$  and  $\text{Var}(\mathbf{Y}^*) = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{W}^{-1}$ , where  $\mathbf{G} = \text{diag}(\mathbf{G}_k)$  and  $\mathbf{W}^{-1} = \text{diag}(\mathbf{W}_k^{-1})$ .

Again, it can be considered and solved as a linear mixed-effects model (see chapter 3). The iterations for solving a GLMM based on the MQL algorithm is slightly different from that for PQL:  $\tilde{\boldsymbol{\gamma}}$  is not estimated at each iteration using the MQL, and instead,  $\tilde{\boldsymbol{\gamma}}$  will be estimated using BLUP after convergence.

### 4.1.3 Laplace's Method

Laplace's method is usually used to approximate an integral of the form  $\int e^{l(\mathbf{t})} d\mathbf{t}$ , which requires  $l(\mathbf{t})$  to be a smooth, bounded, and unimodal function. Starting from the marginal likelihood function of a GLMM,

$$\begin{aligned}
L(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}) &= \prod_{k=1}^K f(\mathbf{Y}_k; \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi}) = \prod_{k=1}^K \int f(\mathbf{Y}_k, \boldsymbol{\gamma}_k; \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi}) d\boldsymbol{\gamma}_k \\
&= \prod_{k=1}^K \int f(\mathbf{Y}_k | \boldsymbol{\gamma}_k, \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}) f(\boldsymbol{\gamma}_k; \boldsymbol{\xi}) d\boldsymbol{\gamma}_k \\
&= \prod_{k=1}^K \int \prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\gamma}_k, \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}) f(\boldsymbol{\gamma}_k; \boldsymbol{\xi}) d\boldsymbol{\gamma}_k \\
&\propto \prod_{k=1}^K \int |\mathbf{G}_k|^{-\frac{1}{2}} \exp \left\{ \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \log[f(Y_{kl}(t) | \boldsymbol{\gamma}_k, \boldsymbol{\beta}_{uk}, \boldsymbol{\beta})] \right. \\
&\quad \left. - \frac{1}{2} \boldsymbol{\gamma}_k^T \mathbf{G}_k^{-1} \boldsymbol{\gamma}_k \right\} d\boldsymbol{\gamma}_k
\end{aligned}$$

Let

$$d(\boldsymbol{\gamma}_k) = - \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \log[f(Y_{kl}(t) | \boldsymbol{\gamma}_k, \boldsymbol{\beta}_{uk}, \boldsymbol{\beta})] + \frac{1}{2} \boldsymbol{\gamma}_k^T \mathbf{G}_k^{-1} \boldsymbol{\gamma}_k$$

Then based on a second-order Taylor series expansion about  $\hat{\boldsymbol{\gamma}}_k = \hat{\boldsymbol{\gamma}}_k(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi})$  at the maximum values of  $d(\boldsymbol{\gamma}_k)$  such that  $d'(\hat{\boldsymbol{\gamma}}_k) = 0$ , we have

$$d(\boldsymbol{\gamma}_k) \approx d(\hat{\boldsymbol{\gamma}}_k) + \frac{1}{2} (\boldsymbol{\gamma}_k - \hat{\boldsymbol{\gamma}}_k)^T d''(\hat{\boldsymbol{\gamma}}_k) (\boldsymbol{\gamma}_k - \hat{\boldsymbol{\gamma}}_k)$$

where  $d''(\boldsymbol{\gamma}_k)$  is the Hessian matrix which contains the second-order partial derivatives of  $d(\boldsymbol{\gamma}_k)$ .

Then,  $\int \exp[d(\boldsymbol{\gamma}_k)] d\boldsymbol{\gamma}_k$  can be approximated as

$$\begin{aligned}
\int \exp[d(\boldsymbol{\gamma}_k)] d\boldsymbol{\gamma}_k &\approx \int \exp \left[ d(\hat{\boldsymbol{\gamma}}_k) + \frac{1}{2} (\boldsymbol{\gamma}_k - \hat{\boldsymbol{\gamma}}_k)^T d''(\hat{\boldsymbol{\gamma}}_k) (\boldsymbol{\gamma}_k - \hat{\boldsymbol{\gamma}}_k) \right] d\boldsymbol{\gamma}_k \\
&\approx \exp[d(\hat{\boldsymbol{\gamma}}_k)] \int \exp \left[ -\frac{1}{2} (\boldsymbol{\gamma}_k - \hat{\boldsymbol{\gamma}}_k)^T (-d''(\hat{\boldsymbol{\gamma}}_k)) (\boldsymbol{\gamma}_k - \hat{\boldsymbol{\gamma}}_k) \right] d\boldsymbol{\gamma}_k \\
&\approx \exp[d(\hat{\boldsymbol{\gamma}}_k)] (2\pi)^{(s_k \times n_k + s^*)/2} |-d''(\hat{\boldsymbol{\gamma}}_k)|^{-\frac{1}{2}}
\end{aligned}$$

and the marginal likelihood can be approximated by

$$l(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}) \propto \sum_{k=1}^n \left[ -\frac{1}{2} \log |\mathbf{G}_k| + d(\hat{\boldsymbol{\gamma}}_k) + \frac{1}{2} \log |d''(\hat{\boldsymbol{\gamma}}_k)| \right]$$

Since  $\theta_{klt}$  is the canonical parameter, we have  $\theta_{klt} = \eta_{klt} = g(\mu_{klt})$ . Therefore,  $\hat{\boldsymbol{\gamma}}_k$  can be solved via

$$\begin{aligned} d'(\boldsymbol{\gamma}_k) &= - \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \frac{\partial \log[f(Y_{kl}(t) | \boldsymbol{\gamma}_k, \boldsymbol{\beta}_{uk}, \boldsymbol{\beta})]}{\partial \theta_{klt}} \frac{\partial \eta_{klt}}{\partial \boldsymbol{\gamma}_k} + \mathbf{G}_k^{-1} \boldsymbol{\gamma}_k = \mathbf{0} \\ &= - \begin{bmatrix} \sum_{t=1}^{p_{k1}} \frac{Y_{k1}(t) - b'(\theta_{k1t_1})}{a_{k1t}(\phi)} \mathbf{z}_{k1}(t) \\ \vdots \\ \sum_{t=1}^{p_{kn_k}} \frac{Y_{kn_k}(t) - b'(\theta_{kn_k t})}{a_{kn_k t}(\phi)} \mathbf{z}_{kn_k}(t) \\ \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \frac{Y_{kl}(t) - b'(\theta_{klt})}{a_{klt}(\phi)} \mathbf{z}_k^* \end{bmatrix}_{(s_k \times n_k + s^*) \times 1} + \mathbf{G}_k^{-1} \boldsymbol{\gamma}_k = \mathbf{0}. \end{aligned}$$

Meanwhile, the Hessian matrix  $d''(\boldsymbol{\gamma}_k)$  can be estimated as

$$\begin{aligned} d''(\boldsymbol{\gamma}_k) &= \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \frac{\partial \{\log[f(Y_{kl}(t) | \boldsymbol{\gamma}_k, \boldsymbol{\beta}_{uk}, \boldsymbol{\beta})]\}' \}}{\partial \theta_{klt}} \frac{\partial \eta_{klt}}{\partial \boldsymbol{\gamma}_k} + \mathbf{G}_k^{-1} \\ &= \begin{bmatrix} \sum_{t=1}^{p_{k1}} \frac{b''(\theta_{k1t_1}) \mathbf{z}_{k1}(t) \mathbf{z}_{k1}(t)^T}{a_{k1t}(\phi)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \frac{b''(\theta_{klt}) \mathbf{z}_k^* \mathbf{z}_k^{*T}}{a_{klt}(\phi)} \end{bmatrix} + \mathbf{G}_k^{-1} \\ &= \mathbf{Z}_k^T \mathbf{Q}_k \mathbf{Z}_k + \mathbf{G}_k^{-1} \end{aligned}$$

where  $\mathbf{Q}_k = \text{diag} \left( \frac{b''(\theta_{klt})}{a_{klt}(\phi)} \right)$ ,  $l = 1, \dots, n_k$  and  $t = 1, \dots, p_{kl}$ , is the  $(\sum_{l=1}^{n_k} p_{kl}) \times (\sum_{l=1}^{n_k} p_{kl})$  diagonal matrix. The estimates of  $\boldsymbol{\beta}_{uk}$ ,  $\boldsymbol{\beta}$ , and  $\boldsymbol{\xi}$  can be obtained through a Newton-Raphson or a Fisher scoring algorithm. The estimates of  $\boldsymbol{\gamma}_k$  can be obtained through BLUP after the convergence of estimating  $\boldsymbol{\beta}_{uk}$ ,  $\boldsymbol{\beta}$ , and  $\boldsymbol{\xi}$ .

### 4.1.3 Gaussian Quadrature

Another category of methods to approximate the integral of a function is to approximate the integral numerically. This technique is useful when the function cannot be analytically integrated or the analytical form of an integral is very complicated. Gaussian quadrature is one typical approach. Let's consider a general weighted integral  $\int_a^b f(x)w(x)dx$ , where  $w(x)$  is a weight function and is a continuous, non-negative function. Gaussian quadrature can approximate the integral by a finite weighted sum based on  $m$  selected nodes within the given domain  $(a, b)$ :

$$\int_a^b f(x)w(x)dx \approx \sum_{i=1}^m f(x_i)A_i(x_i)$$

where  $x_i, i = 1 \dots m$ , are the nodes and  $A_i(x_i)$  are their corresponding weights. The theory on Gaussian quadrature states that the  $m$  nodes and their corresponding weights are optimally spaced and calculated, such that the quadrature, the finite weighted sum, will exactly integrate any polynomial with degree of  $2m - 1$ .

#### 1) Gauss-Hermite Quadrature

For GLMM, random effects are assumed to be normally distributed, and the numerical integration is usually done through Gauss-Hermite quadrature. Gauss-Hermite quadrature can approximate integrals of the form  $\int f(x)e^{-x^2} dx$  as follows:

$$\int f(x)e^{-x^2} dx \approx \sum_{i=1}^m f(x_i)A_i(x_i)$$

where  $x_i$  and  $A_i(x_i)$  are the Gauss-Hermite quadrature nodes and weights that can be found in the literature (Abramowitz et al., 1966; Golub et al.). The  $w(x)$  here has a standard normal kernel  $e^{-x^2}$ . With some transformations, the approximation can be extended to the multivariate version

with random effects  $\mathbf{x}_{Q \times 1}$  following a multivariate normal distribution with mean  $\boldsymbol{\mu}$  and variance-covariance matrix  $\boldsymbol{\Sigma}$ :

$$\int f(y|\mathbf{x})f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})d\mathbf{x} \approx \left(\frac{1}{\sqrt{\pi}}\right)^Q \sum_{b_1=1}^m \dots \sum_{b_Q=1}^m f\left(y|\mathbf{x}_b^* = \boldsymbol{\mu} + \sqrt{2}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{x}_b\right) A_{b_1}(x_{b_1}) \dots A_{b_Q}(x_{b_Q})$$

where  $m$  is the number of nodes selected for each element of  $\mathbf{x}$ ,  $\mathbf{x}_b = (x_{b_1}, \dots, x_{b_Q})^T$ ,  $x_{b_i}$  and  $A_{b_i}(x_{b_i})$  are the node and the weight for the  $b_i^{th}$  node of the  $i^{th}$  random effect, respectively, based on the Gauss-Hermite quadrature nodes and weights,  $i = 1, \dots, Q$ .

Under the GLMM framework with random effects  $\boldsymbol{\gamma}_k \sim N_{s_k \times n_k + s^*}(\mathbf{0}, \mathbf{G}_k)$ , the marginal likelihood function can be approximated by

$$L(\boldsymbol{\beta}, \boldsymbol{\xi}|\mathbf{Y}) = \prod_{k=1}^K \int \prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k) f(\boldsymbol{\gamma}_k; \boldsymbol{\xi}) d\boldsymbol{\gamma}_k$$

$$\approx \prod_{k=1}^K \left\{ \left(\frac{1}{\sqrt{\pi}}\right)^Q \sum_{b_1=1}^m \dots \sum_{b_Q=1}^m \left[ \prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f\left(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k^* = \sqrt{2}\mathbf{G}_k^{\frac{1}{2}}\mathbf{x}_b\right) A_{b_1}(x_{b_1}) \dots A_{b_Q}(x_{b_Q}) \right] \right\}$$

where  $Q = s_k \times n_k + s^*$ ,  $m$ ,  $\mathbf{x}_b$ , and  $A_{b_i}(x_{b_i})$  are defined similarly as above.

## 2) Adaptive Gauss-Hermite Quadrature

An improved method of Gauss-Hermite quadrature is adaptive Gauss-Hermite quadrature, which selects Gauss-Hermite quadrature nodes differently for each study  $k$  according to its most likely value of random effects conditional on current estimates of  $\tilde{\boldsymbol{\beta}}_{uk}$ ,  $\tilde{\boldsymbol{\beta}}$ , and  $\tilde{\boldsymbol{\xi}}$ . In other words, the random effects of the studies will be centered and scaled differently. When the between-study variance is large, more nodes ( $m$ ) will be needed in Gauss-Hermite quadrature to achieve an ideal approximation. The advantage of adaptive Gauss-Hermite quadrature is that fewer nodes are needed to achieve equivalent accuracy as for Gauss-Hermite quadrature. With a smaller  $m$ , the placements of the  $m$  nodes for each study will depend on its specific distribution of random effects.



Let

$$a(\boldsymbol{\gamma}_k | Y_{kl}(t), \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi}) = \prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k) f(\boldsymbol{\gamma}_k; \boldsymbol{\xi})$$

and  $a(\boldsymbol{\gamma}_k | Y_{kl}(t), \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi})$  can be treated as the posterior distribution of random effects for the  $k^{th}$  study. Therefore, for the  $k^{th}$  study, the estimate of  $\boldsymbol{\gamma}_k = \hat{\boldsymbol{\gamma}}_k = \hat{\boldsymbol{\gamma}}_k(\tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\xi}})$  can be obtained via

$$\frac{\partial \log a(\boldsymbol{\gamma}_k | Y_{kl}(t), \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi})}{\partial \boldsymbol{\gamma}_k} \Big|_{\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\xi}}} = 0$$

where  $\tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\xi}}$  are the current estimates of  $\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi}$ .  $\mathbf{G}_k$  can also be estimated as  $\hat{\mathbf{G}}_k$ , using the observed or expected information matrix

$$\hat{\mathbf{G}}_k = - \frac{\partial^2 \log a(\boldsymbol{\gamma}_k | Y_{kl}(t), \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi})}{\partial \boldsymbol{\gamma}_k \partial \boldsymbol{\gamma}_k} \Big|_{\tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\xi}}}$$

or

$$\hat{\mathbf{G}}_k = E \left[ - \frac{\partial^2 \log a(\boldsymbol{\gamma}_k | Y_{kl}(t), \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi})}{\partial \boldsymbol{\gamma}_k \partial \boldsymbol{\gamma}_k} \Big|_{\tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\xi}}} \right]$$

Then, the marginal likelihood can be approximated by

$$\begin{aligned} L(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}) &= \prod_{k=1}^K \int \prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k) f(\boldsymbol{\gamma}_k; \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \mathbf{G}_k) d\boldsymbol{\gamma}_k \\ &= \prod_{k=1}^K \int \frac{\prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k) f(\boldsymbol{\gamma}_k; \mathbf{0}, \mathbf{G}_k)}{f(\boldsymbol{\gamma}_k; \hat{\boldsymbol{\gamma}}_k, \hat{\mathbf{G}}_k)} f(\boldsymbol{\gamma}_k; \hat{\boldsymbol{\gamma}}_k, \hat{\mathbf{G}}_k) d\boldsymbol{\gamma}_k \\ &\approx \prod_{k=1}^K \left\{ \left( \frac{1}{\sqrt{\pi}} \right)^Q \sum_{b_1=1}^m \dots \sum_{b_Q=1}^m \left[ \prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k^*) f(\boldsymbol{\gamma}_k^*; \mathbf{0}, \mathbf{G}_k) (2\pi)^{\frac{Q}{2}} |\hat{\mathbf{G}}_k|^{\frac{1}{2}} \exp \left( \frac{1}{2} (\boldsymbol{\gamma}_k^* \right. \right. \right. \\ &\quad \left. \left. \left. - \hat{\boldsymbol{\gamma}}_k)^T \hat{\mathbf{G}}_k^{-1} (\boldsymbol{\gamma}_k^* - \hat{\boldsymbol{\gamma}}_k) \right) A_{b_1}(x_{b_1}) \dots A_{b_Q}(x_{b_Q}) \right] \right\} \end{aligned}$$

Replacing nodes  $\boldsymbol{\gamma}_k^*$  with  $\hat{\boldsymbol{\gamma}}_k + \sqrt{2} \hat{\mathbf{G}}_k^{\frac{1}{2}} \mathbf{x}_b$ ,

$$L(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}) \approx \prod_{k=1}^K \left\{ (2)^{\frac{Q}{2}} |\widehat{\mathbf{G}}_k|^{\frac{1}{2}} \sum_{b_1=1}^m \dots \sum_{b_Q=1}^m \left[ \prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f\left(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}_k + \sqrt{2\widehat{\mathbf{G}}_k} \frac{1}{2} \mathbf{x}_b\right) f\left(\widehat{\boldsymbol{\gamma}}_k + \sqrt{2\widehat{\mathbf{G}}_k} \frac{1}{2} \mathbf{x}_b; \mathbf{0}, \mathbf{G}_k\right) \exp(\mathbf{x}_b^T \mathbf{x}_b) A_{b_1}(x_{b_1}) \dots A_{b_Q}(x_{b_Q}) \right] \right\}$$

#### 4.1.4 Simulated Maximum Likelihood

Simulated maximum likelihood estimation also is called Monte Carlo integration, which is a stochastic numerical approximation to the integral. Starting with the marginal likelihood, it can be written as

$$\begin{aligned} L(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}) &= \prod_{k=1}^K \int \prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k) f(\boldsymbol{\gamma}_k; \boldsymbol{\xi}) d\boldsymbol{\gamma}_k \\ &= \prod_{k=1}^K \int \frac{\prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k) f(\boldsymbol{\gamma}_k; \boldsymbol{\xi})}{h(\boldsymbol{\gamma}_k; \boldsymbol{\xi}')} h(\boldsymbol{\gamma}_k; \boldsymbol{\xi}') d\boldsymbol{\gamma}_k \end{aligned}$$

where  $h(\boldsymbol{\gamma}_k)$  is defined as importance sampling distribution and it is known. Then

$L(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y})$  can be written as the product of expectations of  $\frac{\prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k) f(\boldsymbol{\gamma}_k; \boldsymbol{\xi})}{h(\boldsymbol{\gamma}_k; \boldsymbol{\xi}')}$

with respect to  $h(\boldsymbol{\gamma}_k; \boldsymbol{\xi}')$ . Suppose we can randomly generate  $G$  independent samples,

$\boldsymbol{\gamma}_k^1, \boldsymbol{\gamma}_k^2, \dots, \boldsymbol{\gamma}_k^G$  from  $h(\boldsymbol{\gamma}_k; \boldsymbol{\xi}')$ . Then  $L(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y})$  can be approximated by

$$\begin{aligned} L(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}) &= \prod_{k=1}^K \int \frac{\prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k) f(\boldsymbol{\gamma}_k; \boldsymbol{\xi})}{h(\boldsymbol{\gamma}_k; \boldsymbol{\xi}')} h(\boldsymbol{\gamma}_k; \boldsymbol{\xi}') d\boldsymbol{\gamma}_k \\ &\approx \prod_{k=1}^K \left[ \frac{1}{G} \sum_{g=1}^G \frac{\prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k^g) f(\boldsymbol{\gamma}_k^g; \boldsymbol{\xi})}{h(\boldsymbol{\gamma}_k^g; \boldsymbol{\xi}')} \right] \end{aligned}$$

The efficiency of the approximation depends on the choice of  $h(\boldsymbol{\gamma}_k; \boldsymbol{\xi}')$ . In practice,

$h(\boldsymbol{\gamma}_k; \boldsymbol{\xi}')$  is suggested to be chosen close to  $f(\boldsymbol{\gamma}_k; \boldsymbol{\xi})$ , and it should be feasible and efficient to draw

simulated samples. Similar with adaptive Gauss-Hermite quadrature,  $h(\boldsymbol{\gamma}_k; \boldsymbol{\xi}')$  can be chosen as the posterior distribution of random effects for  $k^{th}$  study. Let  $h(\boldsymbol{\gamma}_k; \boldsymbol{\xi}') = f(\boldsymbol{\gamma}_k; \widehat{\boldsymbol{\gamma}}_k, \widehat{\mathbf{G}}_k)$ , and  $\widehat{\boldsymbol{\gamma}}_k, \widehat{\mathbf{G}}_k$  are defined as above. Therefore,  $L(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y})$  can be approximated as

$$\begin{aligned} & L(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}) \\ &= \prod_{k=1}^K \int \frac{\prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k) f(\boldsymbol{\gamma}_k; \mathbf{0}, \mathbf{G}_k)}{f(\boldsymbol{\gamma}_k; \widehat{\boldsymbol{\gamma}}_k, \widehat{\mathbf{G}}_k)} f(\boldsymbol{\gamma}_k; \widehat{\boldsymbol{\gamma}}_k, \widehat{\mathbf{G}}_k) d\boldsymbol{\gamma}_k \\ &\approx \prod_{k=1}^K \left[ \frac{1}{G} \sum_{g=1}^G \frac{\prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k^g) f(\boldsymbol{\gamma}_k^g; \mathbf{0}, \mathbf{G}_k)}{f(\boldsymbol{\gamma}_k^g; \widehat{\boldsymbol{\gamma}}_k, \widehat{\mathbf{G}}_k)} \right] \end{aligned}$$

Again, it can be treated as the product of expectations of  $\frac{\prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k) f(\boldsymbol{\gamma}_k; \mathbf{0}, \mathbf{G}_k)}{f(\boldsymbol{\gamma}_k; \widehat{\boldsymbol{\gamma}}_k, \widehat{\mathbf{G}}_k)}$

of the normally distributed random effects  $\boldsymbol{\gamma}_k$ .

#### 4.1.4 Expectation-Maximization (EM) algorithm

In this approach, the complete data consist of the response outcome  $\mathbf{Y}_k$  and the random effects  $\boldsymbol{\gamma}_k$ ,  $k = 1, 2, \dots, K$ . If the random effects were known, we can construct the complete likelihood as

$$L_c(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}, \boldsymbol{\gamma}) = \prod_{k=1}^K \left\{ \prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k) f(\boldsymbol{\gamma}_k; \boldsymbol{\xi}) \right\}$$

However, random effects cannot be observed in practice, and instead we can estimate the expectation of  $L_c(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}, \boldsymbol{\gamma})$  based on the current estimates  $\widetilde{\boldsymbol{\beta}}_{uk}$ ,  $\widetilde{\boldsymbol{\beta}}$ , and  $\widetilde{\boldsymbol{\xi}}$ , and response outcome  $\mathbf{Y}_k$ . Therefore, the expectation of the complete log-likelihood  $l_c(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}, \boldsymbol{\gamma}, \widetilde{\boldsymbol{\beta}}_{uk}, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\xi}})$  can be estimated via

$$\begin{aligned}
& E \left( l_c(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\xi}}) \right) \\
&= E \left\{ \sum_{k=1}^K \left\{ \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \log[f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k)] + \log f(\boldsymbol{\gamma}_k; \boldsymbol{\xi}) \right\} \right\} \\
&= \sum_{k=1}^K \int \left\{ \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \log[f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k)] \right. \\
&\quad \left. + \log f(\boldsymbol{\gamma}_k; \boldsymbol{\xi}) \right\} p(\boldsymbol{\gamma}_k; \mathbf{Y}_k, \tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\xi}}) d\boldsymbol{\gamma}_k
\end{aligned} \tag{4.7}$$

where  $p(\boldsymbol{\gamma}_k; \mathbf{Y}, \tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\xi}})$  is the conditional density of the random effects given  $\tilde{\boldsymbol{\beta}}_{uk}$ ,  $\tilde{\boldsymbol{\beta}}$ , and  $\tilde{\boldsymbol{\xi}}$ , and  $\mathbf{Y}_k$ , which can be replaced by the posterior distribution of  $\boldsymbol{\gamma}_k$ , such as

$$p(\boldsymbol{\gamma}_k; \mathbf{Y}, \tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\xi}}) \propto \prod_{l=1}^{n_k} \prod_{t=1}^{p_{kl}} f(Y_{kl}(t) | \tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}_k) f(\boldsymbol{\gamma}_k; \tilde{\boldsymbol{\xi}}) \tag{4.8}$$

Once (4.7) is obtained, it is maximized as a function of  $\boldsymbol{\beta}_{uk}$ ,  $\boldsymbol{\beta}$  and  $\boldsymbol{\xi}$ , and the new estimates  $\tilde{\boldsymbol{\beta}}_{uk}$ ,  $\tilde{\boldsymbol{\beta}}$ ,  $\tilde{\boldsymbol{\xi}}$  will be updated. The algorithm is iterated between the expectation step and the maximization step until the convergence of observed likelihood function  $L(\boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\xi} | \mathbf{Y})$  or parameter estimates  $\boldsymbol{\beta}_{uk}$ ,  $\boldsymbol{\beta}$  and  $\boldsymbol{\xi}$ .

Since both parts of the integrals in (4.7) do not have closed forms, the approximation of the integral still relies on Gaussian quadrature or simulated maximum likelihood method. In the latter method, we can randomly generate  $G$  independent samples,  $\boldsymbol{\gamma}_k^1, \boldsymbol{\gamma}_k^2, \dots, \boldsymbol{\gamma}_k^G$  from (4.8). Then maximizing (4.7) is equivalent to maximizing

$$\sum_{k=1}^K \left\{ \frac{1}{G} \left[ \sum_{l=1}^{n_k} \sum_{t=1}^{p_{kl}} \log[f(Y_{kl}(t) | \boldsymbol{\beta}_{uk}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k^g)] + \log f(\boldsymbol{\gamma}_k^g; \boldsymbol{\xi}) \right] \right\}.$$

## 4.2 Inference for Fixed Effects $\beta$

In hypothesis testing, large sample tests regarding fixed effects  $\beta$  can be performed in GLMM, and related confidence intervals estimated, by using a Wald chi-square test, a likelihood ratio test or a score test. For a contrast matrix  $L_{M \times P}$  where  $M \leq P$ , we want to test the null hypothesis  $H_0: L\beta = \mathbf{0}$  versus  $H_1: L\beta \neq \mathbf{0}$ .

### 1) Wald Chi-square Test

Based on the final estimate  $\hat{\beta}$ , its estimated covariance matrix can be obtained via either the observed or expected information matrix  $I_{\hat{\beta}}$ , evaluated at  $\hat{\beta}$

$$I_{\hat{\beta}} = -\frac{\partial^2 l(\beta, \hat{\beta}_{uk}, \hat{\xi} | Y)}{\partial \beta \partial \beta^T} \Big|_{\beta = \hat{\beta}} \quad \text{or} \quad I_{\hat{\beta}} = E\left(-\frac{\partial^2 l(\beta, \hat{\beta}_{uk}, \hat{\xi} | Y)}{\partial \beta \partial \beta^T} \Big|_{\beta = \hat{\beta}}\right)$$

where  $\hat{\xi}$  is the estimate of  $\xi$ . To test the null hypothesis, the Wald statistic is

$$C^2 = (L\hat{\beta})^T \{L I_{\hat{\beta}}^{-1} L^T\}^{-1} (L\hat{\beta})$$

and compare  $C^2$  to a  $\chi_M^2$  distribution. If  $M$ , the number of rows of  $L$ , equals 1, then an approximate  $(1 - \alpha) \times 100\%$  confidence interval for  $L\beta$  is given by  $L\hat{\beta} \pm z_{1-\alpha/2} \sqrt{L I_{\hat{\beta}}^{-1} L^T}$ . Similar to linear mixed-effects model, we can convert to an approximate  $F$  distribution in the presence of a small sample size.

### 2) Likelihood Ratio Test

One alternative to the Wald test is the likelihood ratio test (LRT). Denote the maximized ML log-likelihoods under null and alternative hypotheses as  $\hat{l}_{reduced}(\tilde{\beta})$  and  $\hat{l}_{full}(\hat{\beta})$ . The LRT for two nested models can be constructed by comparing  $\hat{l}_{reduced}(\tilde{\beta})$  and  $\hat{l}_{full}(\hat{\beta})$ . The larger the

difference between  $\hat{l}_{reduced}(\tilde{\boldsymbol{\beta}})$  and  $\hat{l}_{full}(\hat{\boldsymbol{\beta}})$ , the stronger the evidence that the reduced model is inadequate. The LRT statistic is

$$G^2 = -2(\hat{l}_{reduced}(\tilde{\boldsymbol{\beta}}) - \hat{l}_{full}(\hat{\boldsymbol{\beta}}))$$

and compare  $G^2$  to a  $\chi_M^2$  distribution. If  $M$ , the number of rows of  $\mathbf{L}$ , equals 1, then an approximate  $(1 - \alpha) \times 100\%$  confidence interval for  $\mathbf{L}\boldsymbol{\beta}$  can also be obtained by inverting the LRT.

### 3) Score Test

A score test is based on the null hypothesis, and the score function is obtained at  $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$ :

$$\mathbf{u}(\tilde{\boldsymbol{\beta}}) = \frac{\partial l(\boldsymbol{\beta}, \tilde{\boldsymbol{\beta}}_{uk}, \tilde{\boldsymbol{\xi}} | \mathbf{Y})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}$$

The information matrix  $\mathbf{I}_{\boldsymbol{\beta}}$  can be either the observed or expected information matrix as defined above, but it is also evaluated at  $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$ . The score test statistic is defined as

$$s(\tilde{\boldsymbol{\beta}}) = \mathbf{u}(\tilde{\boldsymbol{\beta}})^T \mathbf{I}_{\tilde{\boldsymbol{\beta}}}^{-1} \mathbf{u}(\tilde{\boldsymbol{\beta}})$$

$s(\tilde{\boldsymbol{\beta}})$  is compared to a  $\chi_M^2$  distribution. If  $M$  equals 1, then an approximate  $(1 - \alpha) \times 100\%$  confidence interval for  $\mathbf{L}\boldsymbol{\beta}$  can also be obtained by inverting the score test.

## 4.3 Simulation Study

### 4.3.1 Simulation Design

We performed a simulation study to assess the proposed one-stage model for IPD meta-analysis with binary outcomes. The simulated data were generated in the settings of multi-center clinical trials, with a two-arm, placebo-controlled, parallel design, which aimed to detect whether the treatment could reduce a patient's total cholesterol level. The observed outcome variable is a

binary outcome, with specifying cholesterol level  $\geq 200$  mg/dL as high and  $< 200$  mg/dL as low. We designed the simulated data to have a 3-level or a 4-level structure (Figure 3-1) based on a logistic link function. For the 3-level data, we assume the clinical data were collected from multiple studies, and each study had different numbers of centers (study sites), and each center had different numbers of participants. For the 4-level data, we assumed each participant within a center had different numbers of visits, which was the visit level and leads to longitudinal data for the participant. For both 3-level and 4-level multi-center data, we simulated several participant-level variables, including treatment assignment (TRT; placebo or active treatment), baseline cholesterol level (base\_cho; mg/dL), age at enrollment, gender (male or female), race (Caucasian (Cau), African American (AA), Hispanic (His) and other), diabetes disease status (yes or no), and cardiovascular disease status (CVD; yes or no). Both base\_cho and age were centered by their population means, respectively. We generated two sets of study-level random effects, including study-level random intercept ( $r_{study}$ ) and random effect for treatment effects across studies ( $r_{study\_treatment}$ ). We imposed another two sets of random effects to the center level, including center-level random intercept ( $r_{center}$ ) and random effect for baseline cholesterol level ( $r_{center\_bc}$ ). The primary parameter of interest is the difference in treatment between treatment groups (log odds ratio) with the 3-level data. With the 4-level data structure, we generated each participant's weight (lbs) as a repeated measurement variable, and each participant might have 3 to 5 visits. Weight was centered by its population mean before it was used. We included another set of random intercepts for participant level ( $r_{participant}$ ), but removed the random effect for baseline cholesterol level ( $r_{center\_bc}$ ) and replaced the  $r_{study\_treatment}$  with the random effect for the interaction between treatment effects and visit ( $r_{study\_treatment*visit}$ ). The primary parameter of interest is the difference in slopes between treatment groups (log odds ratio) with the 4-level data. The continuous outcome variable is firstly generated based on (3.14) and (3.15) with the absolute values of coefficients divided by 10, and then it was dichotomized as a binary variable based on the cutoff level 200 mg/dL.

We developed 40 simulation scenarios by considering 3 factors, namely, the sample size of the meta-analysis data, the intra-cluster correlation coefficient (ICC), and common variables across studies. Regarding sample size, the meta-analysis data could contain 12 or 8 studies within the 3-level data structure, 10 or 6 studies within the 4-level data structure, and balance or imbalance sample size within each study. With balanced sample size, every study has a similar or a comparable number of centers and participants, while with imbalanced sample sizes, studies have larger differences in the number of centers and participants (Table 4-1).

Table 4-1: Sample size design in the simulation study with binary outcomes.

3-level data				
Balance design	balance	imbalance	balance	imbalance
No. of studies	12	8	12	8
No. of centers within a study	8-16	8-12	study 1-4: 6; study 5-8: 9; study 9-12: 12	study 1-2: 6; study 3-6: 9; study 7-8: 12
No. of participants within a center	24-70	20-30	study 1-4: 16; study 5-8: 20; study 9-12: 26	study 1-2: 16; study 3-6: 20; study 7-8: 26
4-level data				
Balance design	balance	imbalance	balance	imbalance
No. of studies	10	6	10	6
No. of centers within a study	6-10	4-8	study 1-3: 4; study 4-6: 8; study 7-10: 10	study 1-2: 3; study 3-4: 6; study 5-6: 9
No. of participants within a center	18-22	18-22	study 1-3: 16; study 4-6: 20; study 7-10: 24	study 1-2: 16; study 3-4: 20; study 5-6: 24
No. of visits of a participant			3-5	

As the value of the ICC may be associated with a participant's characteristics, we specified the ICC values for the reference group, and altered the ICC as 0.1, 0.2, 0.3, 0.4, and 0.5. Reference group is similarly defined as that in section 3.3.1. The ICC for data with binary outcomes can be calculated through a latent variable method (Raykov et al., 2015; Rodríguez et al., 2018; Hox et al.,



2017). For example, suppose with 3-level data, we have two subjects in the reference group from the same center of a study, then the ICC equals to

$$ICC = \frac{\sigma_{r\_study}^2 + \sigma_{r\_center}^2}{\sigma_{r\_study}^2 + \sigma_{r\_center}^2 + \frac{\pi^2}{3}}$$

where  $r_{study} \sim N(0, \sigma_{r\_study}^2)$ ,  $r_{center} \sim N(0, \sigma_{r\_center}^2)$ , and  $\frac{\pi^2}{3}$  represents the within-study variance when the binary data arise from a logistic distribution with a threshold. In the simulation study, we specified  $\sigma_{r\_study}^2 : \sigma_{r\_center}^2 = 1 : 2$ . As ICC ranged from 0.1, 0.2 to 0.5, we could decide the values of  $\sigma_{r\_study}^2$  and  $\sigma_{r\_center}^2$ . More details about value specification could be found in the appendix B.

The covariates collected from each study could be same or different. We considered either common covariates from each study, or with distinct/unique variables. In scenarios with distinct/unique variables, we chose gender, race, CVD, and diabetes information as study-specific variables. In 3-level data, gender was only available in study no. 3 and 5, race was available in study 1, diabetes was available in study 7, and CVD was available in study 8. With 4-level data, we only considered gender and race as distinct covariates. We examined each scenario in 1000 simulation runs.

With each simulated dataset, we applied 4 test models to the data (Table 3-2). The models were a two-stage fixed-effects model, a two-stage random-effects model, a one-stage fixed-effects model, and a one-stage random-effects (proposed) model. Both two-stage methods still include center-level and participant-level random effects in the first-stage estimates, whereas the one-stage fixed-effects model ignores all random effects, and the one-stage proposed model captures all random effects. To assess the performance of the 4 meta-analytic approaches, we compare the average of mean estimates, average of standard errors from each simulation, and the coverage probability of containing the true treatment effect among 1000 simulations.

### 4.3.2 Simulation Results

The mean of the treatment effect (log odds ratios) estimates, the mean of the standard errors from each estimate, and the coverage probability from 1000 simulations for 3-level and 4-level data structure with common covariates are shown in Figures 4-1 and 4-2, with different combinations of sample size and ICC value.

For simulated data with a 3-level structure, the true value of the difference of treatment effects was specified as -0.8. As seen in Figure 4-1, the two-stage fixed-effects model, the two-stage random-effects model, and the one-stage proposed model give accurate mean estimates in most scenarios, ranging from -0.77 to -0.63. As the ICC value increases or with imbalanced design, the mean estimates become less accurate and more fluctuated. For the one-stage fixed-effects model, it yields relative larger bias regarding mean estimate in all scenarios comparing with other three models, especially when ICC increases (mean estimates range from -0.65 to -0.36). As both fixed-effects models fail to consider appropriate variability in the data, they tend to underestimate the standard errors when the ICC increases, resulting in poor coverage probabilities. The coverage probabilities of the two-stage fixed-effects model are less than 0.85. The coverage probabilities of the one-stage fixed-effects model are always less than 0.75 in all scenarios. The two-stage random-effects and one-stage proposed models behave similarly on the estimation of variance and coverage in many scenarios, but when the sample size decreases or becomes imbalanced, the two-stage random-effects model displays insufficient coverage probability (ranging from 0.823 to 0.893). The one-stage proposed model is able to appropriately estimate variation between studies and to maintain about 0.95 coverage probability in all scenarios (ranging from 0.908 to 0.967).

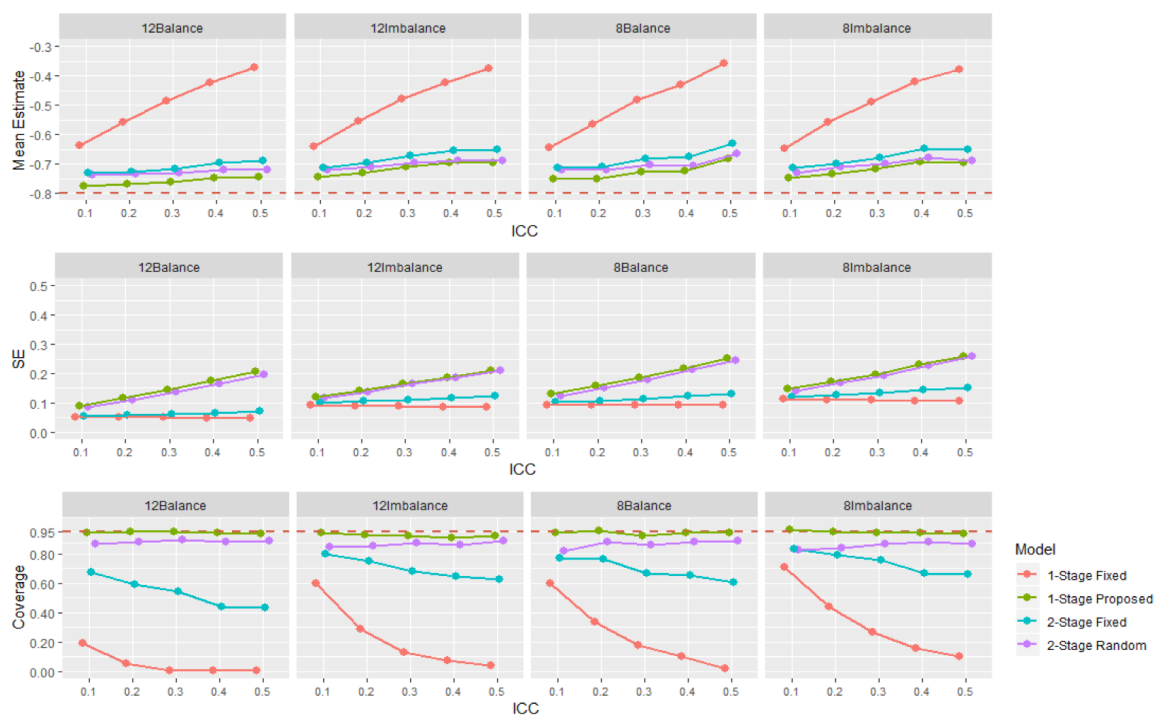


Figure 4-1: Simulation results for 3-level data structure with common covariates.

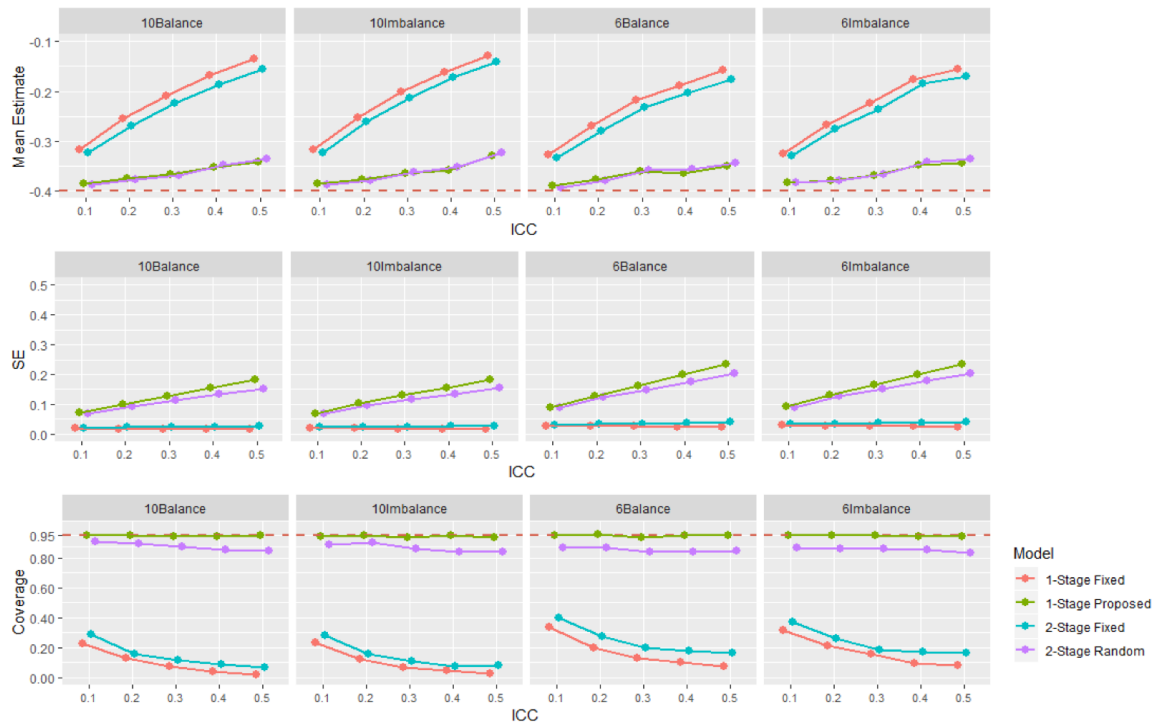


Figure 4-2: Simulation results for 4-level data structure with common covariates.

For simulations with longitudinal data (4-level structure data) and common covariates (Figure 4-2), the true value specified for the difference of slopes between groups is  $-0.4$ . Both fixed-effects models show relatively larger bias regarding mean estimate, underestimated variance estimation, and poor coverage probabilities. The coverage probabilities for both fixed-effects models are equal or lower than  $0.45$  in all scenarios. For the two-stage random-effects model and the one-stage proposed model, similar observations can be made compared with 3-level structure data. The discrepancy of coverage probabilities between the two models are more apparent comparing that with 3-level structure data, especially when the ICC value increases, sample size decreases, or with imbalanced sample size design. The coverage probabilities of the two-stage random-effects model range from  $0.837$  to  $0.912$ . The coverage probabilities of the one-stage proposed model range from  $0.937$  to  $0.955$ .

Simulation scenarios with distinct covariates display similar observations as that in the data with common covariates. Results for simulation data with distinct covariates are shown in the Figures 4-3 to 4-4.

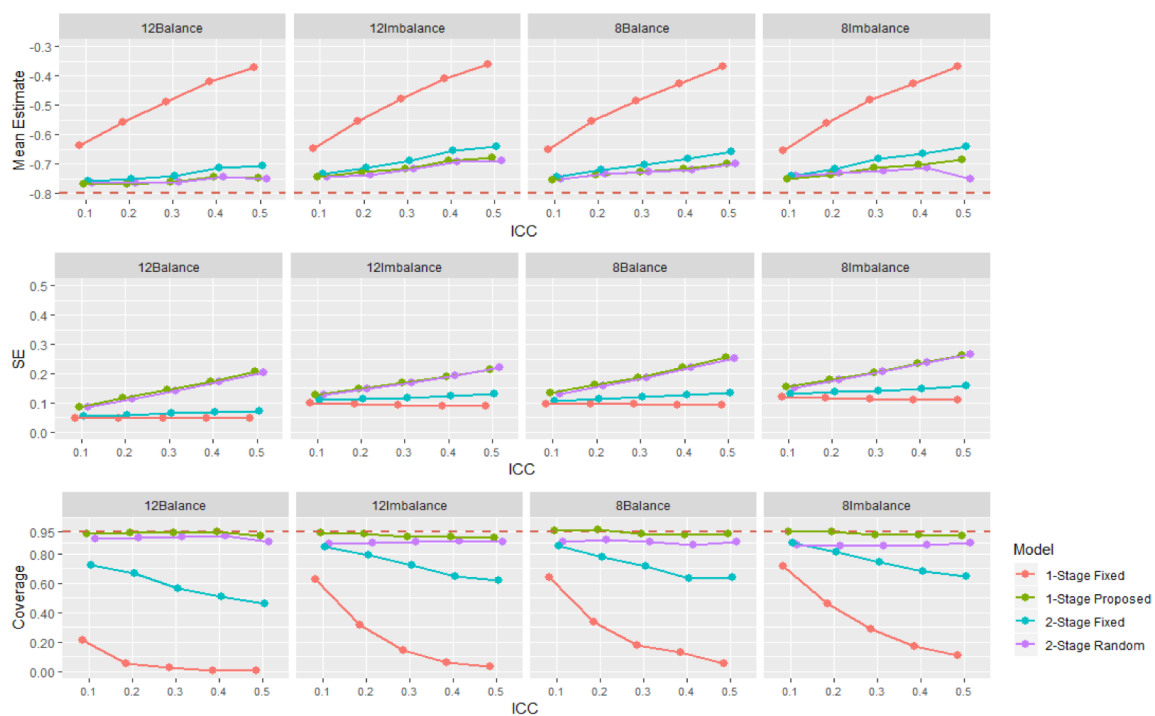


Figure 4-3: Simulation results for 3-level data structure with distinct covariates.

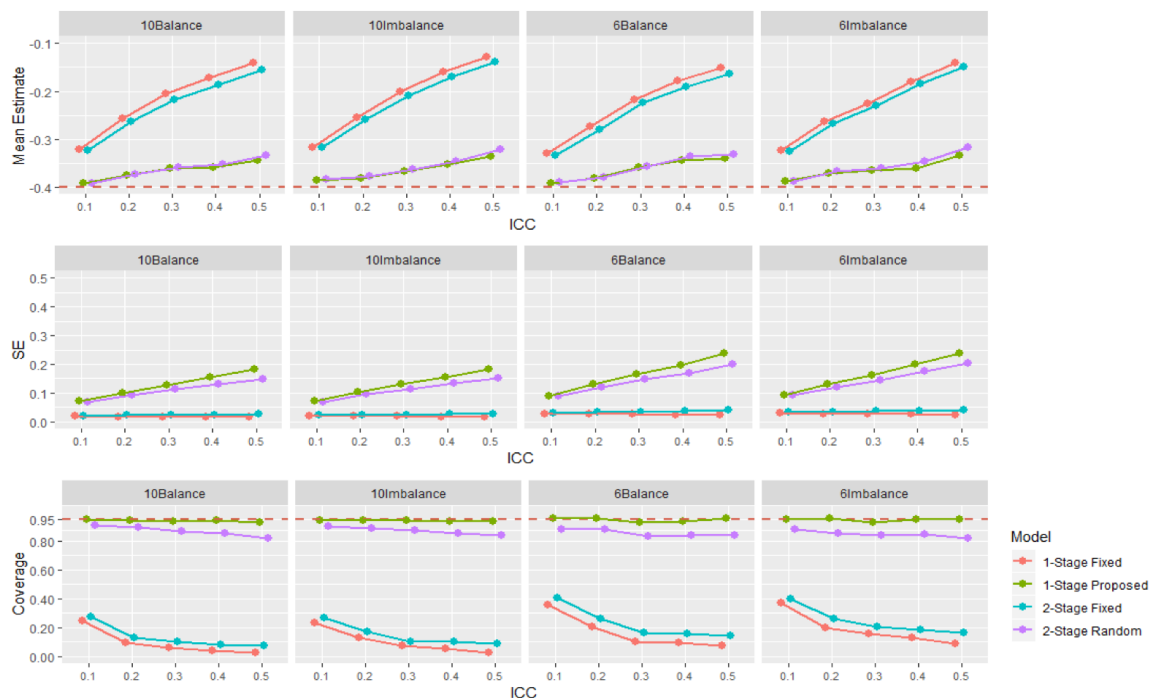


Figure 4-4: Simulation results for 4-level data structure with distinct covariates.

## 4.4 Real Data Application

### 4.4.1 Example Data - Blood Pressure Studies

We used the same 3 studies from the NHLBI to evaluate the performance of the one-stage proposed model for exponential family outcomes. The common question we tried to address was to determine the effect of reducing sodium intake on lowering blood pressure by a certain amount. The 3 studies are DASH-Sodium, PREMIER, and TOHP (phase II), and the study information and outcome variables are summarized in Table 4-2.

Table 4-2: Study information summary of the NHLBI studies.

Study	DASH-Sodium	PREMIER	TOHP (phase II)
Study period	1997-2002	1998-2004	1986-1998
Inclusion	SBP 120-159 mmHg; DBP 80-95 mmHg; free of anti-hypertensive medications	SBP 120-159 mmHg; DBP 80-95 mmHg; free of anti-hypertensive medications	SBP <140 mmHg; DBP 83-89 mmHg; free of anti-hypertensive medications
Age range	≥22	≥25	30-54
Study design	crossover RCT; DASH diet + sodium reduction	parallel RCT; DASH diet + sodium reduction	parallel RCT; weight loss + sodium reduction
Study arms used	low vs high (control) sodium level	low vs control sodium level	low vs control sodium level
Subjects used	204	541	1191
Number of centers	5	4	9
Outcome measure	at the end of 1 month	at the end of 3 months	at the end of 6 months
Binary outcome	Change of SBP ≤-10 mmHg; change of DBP ≤-5 mmHg		
Ordinal outcome	Change of SBP: ≤-10, -10 – 0, >0 mmHg; change of DBP: ≤-5, -5 – 0 >0 mmHg		

SBP: systolic blood pressure; DBP: diastolic blood pressure; RCT: randomized clinical trial.

For the purpose of this real data application, we focus on the comparison of the sodium intake reduction group versus control, advice only, or usual care. DASH-Sodium had a crossover randomization design, and participants from DASH-Sodium were used as a self-comparison. The measurement times of the 3 studies were not the same across studies, and we use the closest measurement times between studies, which were the end of 1 month, 3 months, and 6 months for DASH-Sodium, PREMIER, and TOHP, respectively. We focused on the change of SBP and the change of DBP from the baseline to their corresponding selected end time point of each study as the final outcome variables. For both outcome variables, we considered binary and ordinal outcome variables. For binary outcomes, we used -10 mmHg and -5 mmHg as the cutoff point for the change of SBP and the change of DBP, respectively. For the ordinal outcomes, we categorized the change of SBP as ≤-10 mmHg, -10 to 0 mm, and >0 mmHg; we categorized the change of DBP as ≤-5 mmHg, -5 to 0 mmHg, and >0 mmHg.

We considered participant-level variables, study-level variables, and random effects considered, which we included in the one-stage proposed model (Table 4-3). For comparison, we calculated the individual treatment effects estimated from each study. We also fitted the two-stage fixed-effects model, the two-stage random-effects model, and the one-stage fixed-effects model to compare with the one-stage proposed model. We conducted all analyses in SAS 9.4.

Table 4-3: Model fitting with NHLBI studies.

Covariates	DASH-sodium	PREMIER	TOHP (phase II)
Participant-level			
Sodium level (low or control)	X	X	X
Age	X	X	X
Female	X	X	X
Race	X	X	X
Baseline SBP/DBP	X	X	X
Baseline weight	X	X	X
Study level			
SS_mean	X	X	X
SS_std	X	X	X
Random effect			
Study_treatment	X	X	X
Center	X	X	X

SBP: systolic blood pressure; DBP: diastolic blood pressure.

#### 4.4.2 Results

For the binary outcomes of SBP and DBP, the individual treatment effect estimates from each study along with the pooled estimates from the proposed model and other meta-analysis models are presented in Figures 4-5 to 4-8.



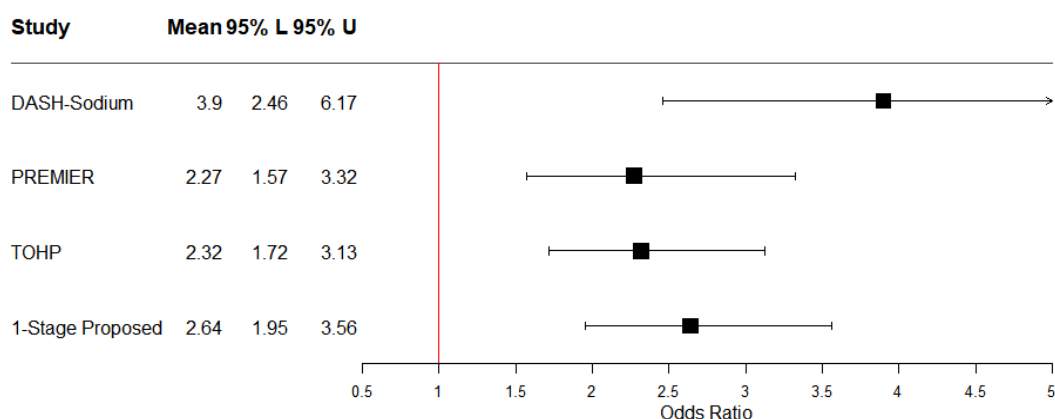


Figure 4-5: Individual estimates and pooled estimates with the binary SBP outcome.

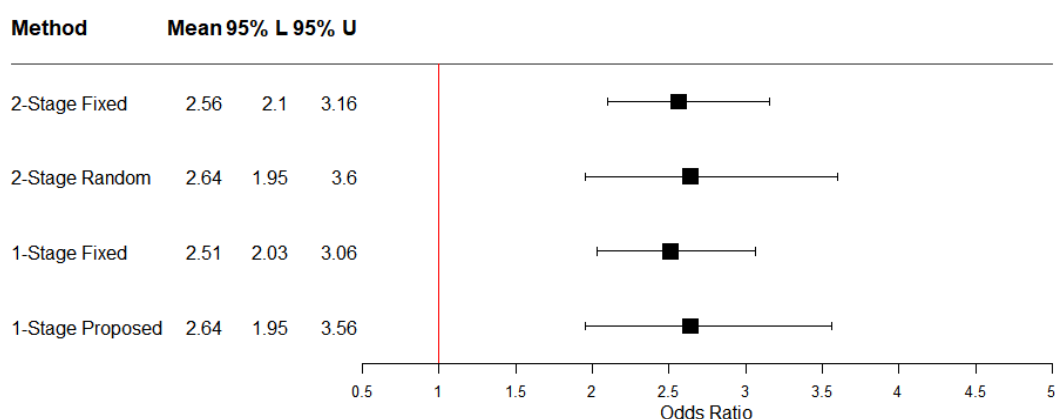


Figure 4-6: Meta-analysis methods comparison with the binary SBP outcome.

For the binary outcome of SBP, all estimates yielded statistically significant results. Sodium intake reduction had a significant impact on the change of SBP  $\leq$ -10 mmHg. As seen from the individual estimates from each study, DASH-Sodium (odds ratio (OR)=3.90, 95% CI (2.46, 6.17)) shows the largest OR, followed by TOHP (OR=2.32, 95% CI (1.72, 3.13)) and PREMIER (OR=2.27, 95% CI (1.57, 3.32)). The overall treatment effect estimate obtained from the one-stage proposed model was (OR=2.64, 95% CI (1.95, 3.56)). Heterogeneity of the treatment effect existed between studies, and the one-stage proposed model was able to integrate the information from the three studies, and gave an appropriate overall estimate. Comparing the estimate from the one-stage proposed model with other tested models, both two-stage and one-stage fixed-effects models

yielded more similar ORs to the TOHP's individual estimate, and with narrower 95% confidence intervals (two-stage fixed-effects: OR=2.56, 95%CI (2.10, 3.16); one-stage fixed-effects: OR=2.51, 95% CI (2.03, 3.06)). Since the TOHP study has the largest sample size among all the three studies, both fixed-effects models tended to be affected by the TOHP study, and failed to appropriately accommodate the variation between studies. Though the two-stage random-effects model showed very similar results as the one-stage proposed model, with OR=2.64 and 95% CI as (1.95, 3.60), as seen in the simulation study, the two-stage random effects model might lead to insufficient coverage probability as the sample size becomes smaller or the sample sizes becomes inequivalent between studies.

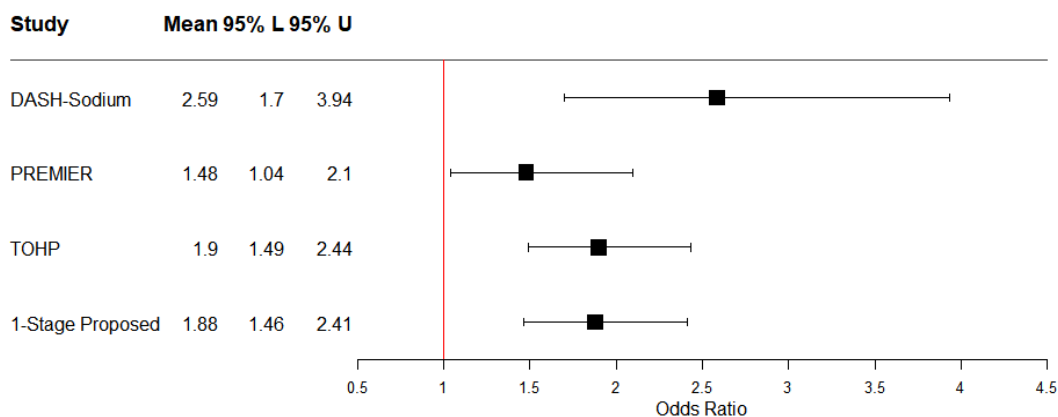


Figure 4-7: Individual estimates and pooled estimates with the binary DBP outcome.

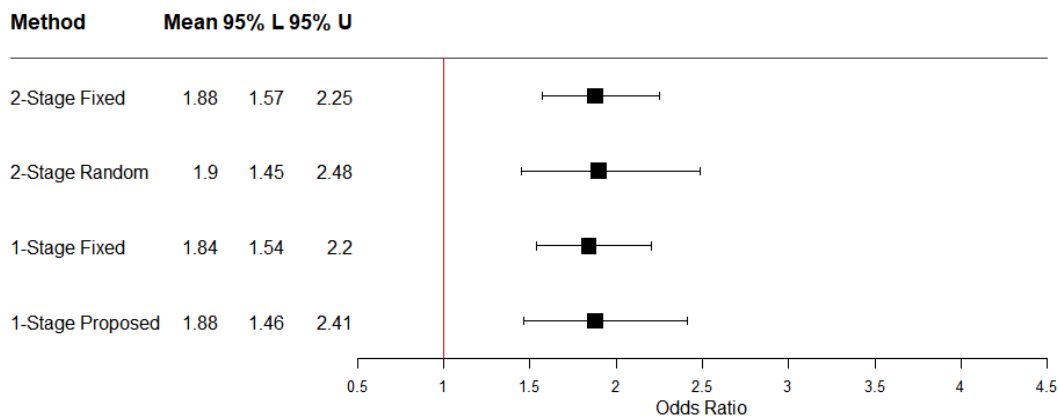


Figure 4-8: Meta-analysis methods comparison with the binary DBP outcome.

With the binary outcome of DBP, the impact of sodium intake reduction on the change of DBP  $\leq$ -5 mmHg was smaller compared with that on the change of SBP  $\leq$ -10 mmHg. The largest individual OR estimate was from DASH-Sodium (OR=2.59, 95% CI (1.70, 3.94)), followed by the TOHP (OR=1.90, 95% CI (1.49, 2.44)) and PREMIER (OR=1.48, 95% CI (1.04, 2.10)). The overall OR estimate the from the one-stage proposed model was 1.88 with 95% CI as (1.46, 2.41). Since the TOHP study had the largest sample size and an intermediate individual estimate among the three studies, the OR estimates from the four test models were similar regarding the change of DBP, but both fixed-effects models still showed tighter 95% CIs compared with the two random-effects models (two-stage fixed-effects: OR=1.88, 95% CI (1.57, 2.25); two-stage random-effects: OR=1.90, 95% CI (1.45, 2.48); one-stage fixed-effects: OR=1.84, 95% CI (1.54, 2.20)).

Results for ordinal SBP and DBP outcomes are shown in Figures 4-9 to 4-12. For the ordinal outcome of the change of SBP, a larger discrepancy exists between individual estimates, and the TOHP showed a more extreme estimates, then both fixed-effects models were influenced by this study. In this situation, both random-effects models showed more appropriate pooling estimates regarding mean and variation. For the ordinal outcome of the change of DBP, it showed the similar pattern as discussed above.

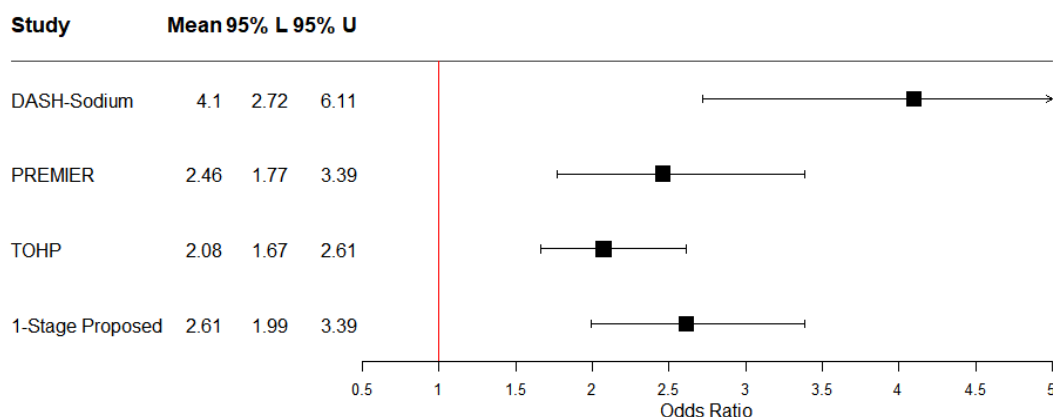


Figure 4-9: Individual estimates and pooled estimates with the ordinal SBP outcome.

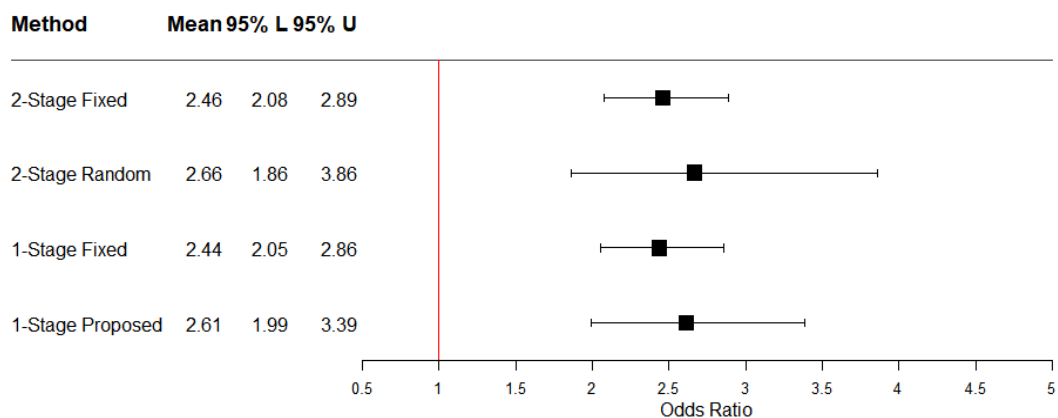


Figure 4-10: Meta-analysis methods comparison with the ordinal SBP outcome.

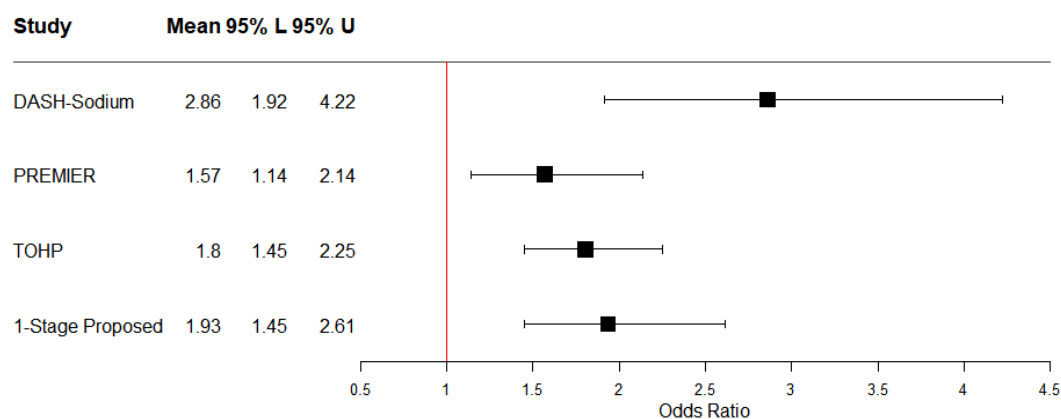


Figure 4-11: Individual estimates and pooled estimates with the ordinal DBP outcome.

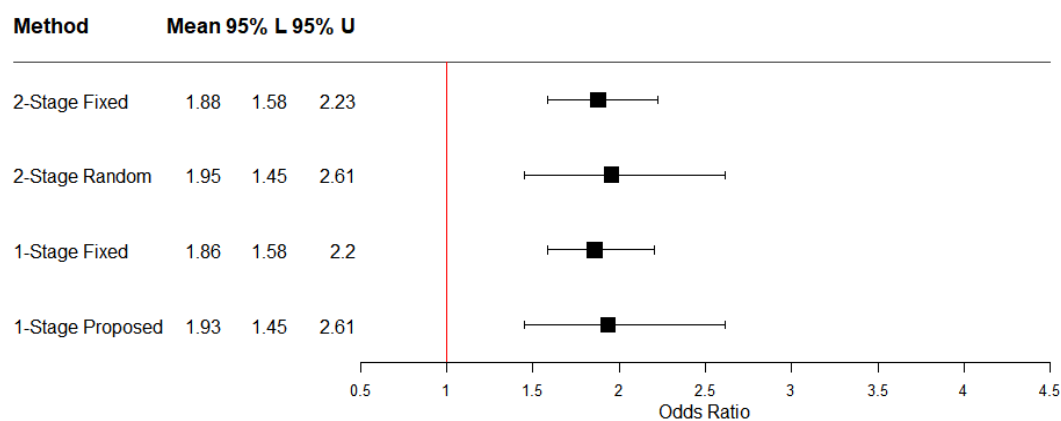


Figure 4-12: Meta-analysis methods comparison with the ordinal DBP outcome.

## Chapter 5

### Summary and Future Work

We propose a one-stage multi-level mixed-effect model for IPD meta-analysis which logically incorporates the hierarchy structure of meta-analysis data, maximally utilizes the information from each study, and provides more flexibility to address different research questions. The proposed model is firstly developed for data with a continuous outcome in chapter 3, and then extended to outcomes from an exponential family in chapter 4. Besides continuous outcomes, binary, categorical or count outcomes also are very common in practical research, and the extension of application from continuous outcomes to exponential family outcomes greatly improve the application range.

One-stage IPD meta-analysis methods are based on exact modeling of the likelihood function, and thus avoid the biases that can arise in two-stage methods. Allowing both fixed effects and random effects at each level helps in identifying different sources of variation. The proposed model is applicable to either observational studies or randomized studies with or without repeated measurements. In both simulation studies, we investigated 40 simulation scenarios, respectively, including three different factors, different sample sizes, different ICC values, and whether studies have common or distinct covariates. The four tested models presented small biases regarding the mean estimate in most scenarios when we specified the treatment effects of each study came from a same normal distribution. In practice, however, it is likely that treatment effects from each study do not share the same distribution, which would be more complicated to test in a simulation study. Nonetheless, we still clearly could notice the different abilities of appropriately estimating the variation between methods. Both fixed-effects models fail to correctly capture the variability in the data, and result in poorer coverage probabilities, especially when the between-study variance

increases. The one-stage fixed-effects model performs worst as it totally ignores any random effects in the model. The two-stage random-effects model though displays similar mean and variance estimates compared with the one-stage proposed model in most cases, the one-stage proposed model is always able to provide about 0.95 coverage probability in most cases, especially when sample size decreases, sample size become imbalanced between studies, or the ICC value increases. In the real data applications, large or smaller amounts of heterogeneity of the effect size existed across studies, and the one-stage proposed model reasonably accounts for the heterogeneity compared with both of the fixed-effects models. Though little difference was observed between the two-stage random-effects model and the one-stage proposed model, as shown through the simulation study, more apparent differences on coverage probability between the two models was observed, especially when the ICC increased. Through both simulation studies and the real data examples, fixed-effects models showed certain limitations and problems. They tended to underestimate the variation of the meta-analytic data, and their results are usually dominated by the studies with larger scale.

Though the one-stage proposed model has complex specifications, it still has the form of a LMM or GLMM. Thus, the implementation of the model is straightforward and is usually available in primary programming languages. Since the solution is based on maximizing the likelihood function by iteratively updating the parameters until convergence, it usually is more computationally demanding than two-stage methods, and it may encounter a convergence problem on occasion. In the main simulation studies conducted, we did not encounter convergence issues with the 3-level structure data, and only experienced about 5% of the simulated data sets that failed to converge with the 4-level structure data. In chapter 4, the calculation involves the approximation to the integration of random effects, and we used PQL to approximate the integration of random effects, since quasi-likelihood methods usually have a higher chance and a faster speed to achieve the approximation. Also, other researchers have shown that, for binary data with relatively small

number of repeated measurements, PQL may result in substantial bias as the accuracy depends on the distribution assumption of the pseudo-data. Other techniques mentioned previously are also recommended if the quasi-likelihood method fails. The inclusion of random effects needs to be done cautiously. Inappropriately incorporating redundant random effects may lead to misleading results and convergence failure as the number of parameters increases or the variance components are non-identifiable. For example, if there are only limited eligible studies, the inclusion of study-level random effects should be restricted. Sufficient explorations, such as graphical techniques, should be conducted before adding random effects to the model. Including more random effects may also lead to an identifiability problem when estimating variance-covariance parameters. The identifiability problem indicates that the sources of variation cannot be correctly differentiated, and results in incorrect variance-covariance estimates.

Another consideration is about the selection of covariates. The selection of covariates from each study should also follow major variable selection principles. Covariates included should be associated with the outcome variable no matter statistically or scientifically. The form of the variable included in the model should be evaluated. The model can be applied to longitudinal data, but in practice, the measurement time points may or may not be the same across studies. If the lengths of follow-up time and time points are very different from each other, it may not be appropriate to simply pool studies together, as the slopes of each study are actually different between studies, and it is not proper to pool them together. Selecting time points within a similar follow-up time period for each study is recommended.

Bayesian methods have been investigated to develop the method for meta-analysis. When applying Bayesian methods, the prior distributions for effect size and between-study variance need to be decided. A previous study already shows the pooled effect size is sensitive the choice to or prior distribution for the between-study variance, even with different noninformative or vague priors (Lambert et al., 2005). Other researchers have tried to utilize existing meta-analysis results

to generate the informative empirical priors for unknown parameters for both continuous and binary outcomes (Riley et al., 2008; Rhodes et al., 2014; Turner et al., 2012). Bayesian meta-analysis methods have been extended to bivariate or multivariate meta-analysis. It has been suggested that vague priors should be avoided through simulation studies, and sensible or empirical priors are recommended (Burke et al., 2016). Bayesian methods have also been developed for meta-analysis with mixed types of outcomes, such as continuous and binary, which are usually based on a two-stage method framework (Bujkiewicz et al., 2013; Wei et al., 2013, Riley et al., 2015).

In summary, the proposed model can be applied to IPD meta-analysis data with cross-sectional data or longitudinal data, with continuous or exponential family outcomes, which attempts to maximally utilize the information and hierarchy feature of the data. The model could be evaluated in more aspects, such as more extreme sample size settings, other type of outcomes from an exponential family, the impact of different integration approximation methods. Besides continuous outcomes and outcomes from an exponential family, time-to-event outcomes are also very common in biomedical research, so the extension of the proposed model to survival outcome is necessary.



## Appendix A

### Chapter 3

#### A.1 Simulation Study

```

/** 3-level Data Simulation ***/
/** Generate data ***/
proc format;
value treat 0='Placebo' 1='Active';
value ny     0='No' 1='Yes';
value race  1='Caucasian' 2='African American' 3='Hispanic' 4='Other';
run;

/*ICC=0.1   vs=6.25      vc=18.75      v=225 vs_trt=vs/2 vc_bc=vc/2*/
/*ICC=0.2   vs=14.0625  vc=42.1875  v=225 vs_trt=vs/2 vc_bc=vc/2*/
/*ICC=0.3   vs=32.14286 vc=64.28571  v=225 vs_trt=vs/2 vc_bc=vc/2*/
/*ICC=0.4   vs=50       vc=100      v=225 vs_trt=vs/2 vc_bc=vc/2*/
/*ICC=0.5   vs=75       vc=150      v=225 vs_trt=vs/2 vc_bc=vc/2*/

%macro simulation(n);

%do sample=1 %to 4;
%do icc=1 %to 5;
%do common=1 %to 1;
%do i=1 %to &n.;

/*Participant-level covariates + study-level var*/
data simulated_data;

sample=&sample.;
icc=&icc.;
common=&common.;

if sample in (1,3) then num_studies=12;
else num_studies=8;

intercept=220;
beta_treatment=-8;
beta_basecholesterol=0.2;
beta_age=0.2;
beta_female=-4;
beta_female_3=-3;
beta_female_5=-4;
beta_AfrAm=4;
beta_Hisp=6;
beta_Other=-0.8;
beta_diabetes=6;
beta_CVD=6;
beta_ss_mean=-0.3;
beta_ss_std=1.5;

```

```

beta_order=-0.5;

do study=1 to num_studies by 1;
  if sample=1 then num_centers=round(8+(8*rand('uniform')));
  *range: 8-16;
  else if sample=2 then num_centers=round(8+(4*rand('uniform')));
  *range: 8-12;
  else if sample=3 then
num_centers=(study<5)*6+(study>=5)*(study<9)*9+(study>=9)*12;
  else
num_centers=(study<3)*6+(study>=3)*(study<7)*9+(study>=7)*12;

  order=study; *sequential order of the study = study #;

  if icc=1 then
    do;
      r1=rand('normal',0,sqrt(6.25)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(6.25/2)); *study-level random
trt effect;
    end;
  else if icc=2 then
    do;
      r1=rand('normal',0,sqrt(14.0625)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(14.0625/2)); *study-level
random trt effect;
    end;
  else if icc=3 then
    do;
      r1=rand('normal',0,sqrt(32.14286)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(32.14286/2)); *study-level
random trt effect;
    end;
  else if icc=4 then
    do;
      r1=rand('normal',0,sqrt(50)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(50/2)); *study-level random trt
effect;
    end;
  else
    do;
      r1=rand('normal',0,sqrt(75)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(75/2)); *study-level random trt
effect;
    end;

  do center=1 to num_centers by 1;
    if sample=1 then
sample_size=round(24+(46*rand('uniform'))); *range: 24-70;
    else if sample=2 then
sample_size=round(20+(10*rand('uniform'))); *range: 20-30;

```

```

        else if sample=3 then
sample_size=(study<5)*16+(study>=5)*(study<9)*20+(study>=9)*26;
        else
sample_size=(study<3)*16+(study>=3)*(study<7)*20+(study>=7)*26;

        if mod(sample_size,2)=1 then sample_size=sample_size+1;
*guarantee equal sizes for two arms;

        if icc=1 then
do;
r2=rand('normal',0,sqrt(18.75));
r2_bc=rand('normal',0,sqrt(18.75/2));
end;
        else if icc=2 then
do;
r2=rand('normal',0,sqrt(42.1875));
r2_bc=rand('normal',0,sqrt(42.1875/2));
end;
        else if icc=3 then
do;
r2=rand('normal',0,sqrt(64.28571));
r2_bc=rand('normal',0,sqrt(64.28571/2));
end;
        else if icc=4 then
do;
r2=rand('normal',0,sqrt(100));
r2_bc=rand('normal',0,sqrt(100/2));
end;
        else
do;
r2=rand('normal',0,sqrt(150));
r2_bc=rand('normal',0,sqrt(150/2));
end;

do participant=1 to sample_size by 1;

if (participant<=sample_size/2) then treatment=0;
else treatment=1;

age=round(rand('normal',50,8));

basecholesterol=round(rand('normal',240,3));

if common=1 then do;
    rand_female=rand('uniform');
if (0.00<=rand_female<0.60) then female=0;
    if (0.60<=rand_female) then female=1;

    rand_race=rand('uniform');
    if (0.00<=rand_race<0.50) then race=1;
    if (0.50<=rand_race<0.70) then race=2;
    if (0.70<=rand_race<0.90) then race=3;
    if (0.90<=rand_race) then race=4;

    rand_diabetes=rand('uniform');

```

```

        if (0.00<=rand_diabetes<0.60) then diabetes=0;
        if (0.60<=rand_diabetes)          then diabetes=1;

        rand_CVD=rand('uniform');
        if (0.00<=rand_CVD<0.70) then CVD=0;
        if (0.70<=rand_CVD)          then CVD=1;
    end;
    else do;
        female=0;
        rand_female=rand('uniform');
        if (0.00<=rand_female<0.60 and study in (3,5)) then
female=0;
            if (0.60<=rand_female and study in (3,5))      then
female=1;

                race=0;
                rand_race=rand('uniform');
                if (0.00<=rand_race<0.50 and study=1) then race=1;
                if (0.50<=rand_race<0.70 and study=1) then race=2;
                if (0.70<=rand_race<0.90 and study=1) then race=3;
                if (0.90<=rand_race and study=1)      then race=4;

                diabetes=0;
                rand_diabetes=rand('uniform');
                if (0.00<=rand_diabetes<0.60 and study=7) then
diabetes=0;
                    if (0.60<=rand_diabetes and study=7)      then
diabetes=1;

                        CVD=0;
                        rand_CVD=rand('uniform');
                        if (0.00<=rand_CVD<0.70 and study=8) then CVD=0;
                        if (0.70<=rand_CVD and study=8)      then CVD=1;
                    end;

                    format treatment treat. female ny. race race. diabetes CVD
ny.;
                output;
            end;

        end;

    end;

end;

run;

/*Study-level variables: ss_mean ss_std*/
proc sort data=simulated_data out=sort (keep=study num_centers center
sample_size) nodupkey;
by study center;
run;

proc means data=sort mean var noprint;
var sample_size;
class study;

```

```

output out=mean_var mean=ss_mean std=ss_std;
run;

data simulated_data2;
merge simulated_data mean_var;
by study;
if study=. then delete;
drop _type_ _freq_;
run;

/*Generate outcome variable: serum_cholesterol*/
proc sql noprint;
select mean(basecholesterol),mean(age) into :mean_bc, :mean_age from
simulated_data2;
quit;

data simulated_data2;
set simulated_data2;

cntr_bc=basecholesterol-&mean_bc.;
cntr_age=age-&mean_age.;

if common=1 then do;
mean=intercept+(cntr_bc*beta_basecholesterol)+(treatment*beta_treatment
)+(cntr_age*beta_age)+(female*beta_female)
+((race=2)*beta_AfrAm)+((race=3)*beta_Hisp)+((race=4)*beta_Other)
+(diabetes*beta_diabetes)+(CVD*beta_CVD)
+(ss_mean*beta_ss_mean)+(ss_std*beta_ss_std)+(order*beta_order)
+r1+(treatment*r1_trt)+r2+(cntr_bc*r2_bc);
end;
else do;
mean=intercept+(cntr_bc*beta_basecholesterol)+(treatment*beta_treatment
)+(cntr_age*beta_age)
+((study=3)*female*beta_female_3)+((study=5)*female*beta_female_5
)
+((study=1)*(race=2)*beta_AfrAm)+((study=1)*(race=3)*beta_Hisp)+((
(study=1)*(race=4)*beta_Other)
+((study=7)*diabetes*beta_diabetes)+((study=8)*CVD*beta_CVD)
+(ss_mean*beta_ss_mean)+(ss_std*beta_ss_std)+(order*beta_order)
+r1+(treatment*r1_trt)+r2+(cntr_bc*r2_bc);
end;

serum_cholesterol=round(rand('normal',mean,15));

run;

/*two-stage methods*/
%if &sample.=1 or &sample.=3 %then %do;
    %let nums=12;
    %end;
%else %do;
    %let nums=8;
    %end;

```

```

%do s=1 %to &nums.;

data dat&s.;
set simulated_data2;
if study=&s.;
run;

proc mixed data=dat&s. covtest;
class treatment center;
model serum_cholesterol=treatment cntr_bc cntr_age female race diabetes
CVD /solution;
random intercept cntr_bc/ type=vc subject=center;
ods exclude all;
ods output SolutionF=S&s.;
run;

data NS&s.;
studyno=&s.;
set S&s.;
if treatment=1;
est=StdErr**2;
keep studyno Estimate EST;
run;

%if &s.=1 %then %do;
data trt;
set NS&s.;
run;
%end;
%else %do;
proc append base=trt data=NS&s.;
run;
%end;

%end;

/*Fixed-effects model*/
data trt;
set trt;
w=1/est;
wy=w*Estimate;
run;

data out_temp;
stop;
run;

proc sql;
create table out_temp as
select sum(wy)/sum(w) as treatment_overall, sqrt(1/sum(w)) as se
from trt;
quit;

data out_temp;
no=&i.;

```

```

set out_temp;
lower=treatment_overall-1.96*se;
upper=treatment_overall+1.96*se;
run;

%if &i.=1 %then %do;
data d.fixed_s&sample.i&icc.c&common.;
set out_temp;
run;
%end;
%else %do;
proc append base=d.fixed_s&sample.i&icc.c&common. data=out_temp;
run;
%end;

/*Random-effects model*/
data trt;
set trt;
wy2=w*estimate**2;
w2=w**2;
run;

proc sql;
select ((sum(wy2)-sum(wy)**2/sum(w))-(&nums.-1))/(sum(w)-
sum(w**2)/sum(w)) into: b_var from trt;
quit;

data trt;
set trt;
w_r=1/(est+&b_var.);
wy_2=w_r*Estimate;
run;

data out_temp;
stop;
run;

proc sql;
create table out_temp as
select sum(wy_2)/sum(w_r) as treatment_overall, sqrt(1/sum(w_r)) as se
from trt;
quit;

data out_temp;
no=&i.;
set out_temp;
lower=treatment_overall-1.96*se;
upper=treatment_overall+1.96*se;
between_trt=&b_var.;
run;

%if &i.=1 %then %do;
data d.random_s&sample.i&icc.c&common.;
set out_temp;

```

```

run;
%end;
%else %do;
proc append base=d.random_s&sample.i&icc.c&common. data=out_temp;
run;
%end;

/**** LMM ****/
/**** 3-level models ****/

%if &common.=1 %then %do;

/*Proposed model*/
title 'Full model';
proc mixed data=simulated_data2 covtest;
class female(ref='No') race(ref='Caucasian') diabetes(ref='No')
CVD(ref='No') center study;
model serum_cholesterol=treatment cntr_bc cntr_age female race diabetes
CVD ss_mean ss_std order/solution cl;
random intercept cntr_bc/ type=vc subject=center(study);
random intercept treatment/ type=vc subject=study;
ods exclude all;
ods output SolutionF=SolutionF1;
ods output CovParms=CovParms1;
run;

%out(SolutionF=SolutionF1,CovParms=CovParms1,num=1,sample=&sample.,icc=
&icc.,common=&common.);

/*Model 1*/
title 'Model 1';
proc mixed data=simulated_data2 covtest;
class female(ref='No') race(ref='Caucasian') diabetes(ref='No')
CVD(ref='No') center study;
model serum_cholesterol=treatment cntr_bc cntr_age female race diabetes
CVD /solution cl;
ods exclude all;
ods output SolutionF=SolutionF2;
ods output CovParms=CovParms2;
run;

%out(SolutionF=SolutionF2,CovParms=CovParms2,num=2,sample=&sample.,icc=
&icc.,common=&common.);
%end;

%else %do;
/**** LMM ****/
/**** unique 3-level models ****/

data simulated_data2;
set simulated_data2;
AfrAm=0;Hispanic=0;Other=0;
if race=2 then AfrAm=1;
if race=3 then Hispanic=1;
if race=4 then Other=1;

```



```

run;

/*Proposed model*/
title 'Full model';
proc mixed data=simulated_data2 covtest;
class center study;
model serum_cholesterol=treatment cnr_bc cnr_age female(study)
AfrAm(study) Hisp(study) Other(study)
                                diabetes(study) CVD(study) ss_mean
ss_std order/solution cl;
random intercept cnr_bc/ type=vc subject=center(study);
random intercept treatment/ type=vc subject=study ;
ods exclude all;
ods output SolutionF=SolutionF1;
ods output CovParms=CovParms1;
run;

%out (SolutionF=SolutionF1,CovParms=CovParms1,num=1,sample=&sample.,icc=
&icc.,common=&common.);

/*Model 1*/
title 'Model 1';
proc mixed data=simulated_data2 covtest;
class center study;
model serum_cholesterol=treatment cnr_bc cnr_age female(study)
AfrAm(study) Hisp(study) Other(study)
                                diabetes(study) CVD(study)

/solution cl;
ods exclude all;
ods output SolutionF=SolutionF2;
ods output CovParms=CovParms2;
run;

%out (SolutionF=SolutionF2,CovParms=CovParms2,num=2,sample=&sample.,icc=
&icc.,common=&common.);

%end;

%end;
%end;
%end;
%end;

%mend simulation;

%simulation(n=1000);

/** 4-level Data Simulation ***/
/** Generate data ***/
proc format;
value treat 0='Placebo' 1='Active';
value ny    0='No' 1='Yes';
value race  1='Caucasian' 2='African American' 3='Hispanic' 4='Other';

```

```

run;

/*ICC=0.1   vs=4.166667   vc=8.333333   vp=12.5     v=225 vs_trt=vs/2 */
/*ICC=0.2   vs=9.375     vc=18.75    vp=28.125  v=225 vs_trt=vs/2 */
/*ICC=0.3   vs=16.071429 vc=32.142857 vp=48.21429 v=225 vs_trt=vs/2 */
/*ICC=0.4   vs=25        vc=50        vp=75      v=225 vs_trt=vs/2 */
/*ICC=0.5   vs=37.5      vc=75        vp=112.5   v=225 vs_trt=vs/2 */

%macro simulation(n);

%do sample=1 %to 4;
%do icc=1 %to 5;
%do common=1 %to 2;
%do i=1 %to &n.;

/*Participant-level covariates + study-level var*/
data simulated_data;

sample=&sample.;
icc=&icc.;
common=&common.;

if sample in (1,3) then num_studies=10;
else num_studies=6;

intercept=220;
beta_basecholesterol=0.2;
beta_age=0.2;
beta_female=-4;
beta_female_3=-3;
beta_female_5=-4;
beta_AfrAm=4;
beta_Hisp=6;
beta_Other=-0.8;
beta_diabetes=6;
beta_CVD=6;
beta_ss_mean=-0.3;
beta_ss_std=1.5;
beta_order=-0.5;
beta_weight=0.4;
beta_visit=-1;
beta_visit_trt=-4;

do study=1 to num_studies by 1;
  if sample=1 then num_centers=round(6+(4*rand('uniform')));
  *range: 6-10;
  else if sample=2 then num_centers=round(4+(4*rand('uniform')));
  *range: 4-8;

  else if sample=3 then
num_centers=(study<4)*4+(study>=4)*(study<7)*8+(study>=7)*12;
  else num_centers=(study<3)*3+(study>=3)*(study<5)*6+(study>=5)*9;

  order=study; *sequential order of the study = study #;

```

```

        if icc=1 then
            do;
                r1=rand('normal',0,sqrt(4.166667)); *study-level random
intercept;
                r1_trt=rand('normal',0,sqrt(4.166667/2)); *study-level
random trt effect;
            end;
        else if icc=2 then
            do;
                r1=rand('normal',0,sqrt(9.375)); *study-level random
intercept;
                r1_trt=rand('normal',0,sqrt(9.375/2)); *study-level random
trt effect;
            end;
        else if icc=3 then
            do;
                r1=rand('normal',0,sqrt(16.071429)); *study-level random
intercept;
                r1_trt=rand('normal',0,sqrt(16.071429/2)); *study-level
random trt effect;
            end;
        else if icc=4 then
            do;
                r1=rand('normal',0,sqrt(25)); *study-level random
intercept;
                r1_trt=rand('normal',0,sqrt(25/2)); *study-level random trt
effect;
            end;
        else
            do;
                r1=rand('normal',0,sqrt(37.5)); *study-level random
intercept;
                r1_trt=rand('normal',0,sqrt(37.5/2)); *study-level random
trt effect;
            end;

        do center=1 to num_centers by 1;
            if sample=1 then
                sample_size=round(24+(46*rand('uniform'))); *range: 24-70;
            else if sample=2 then
                sample_size=round(20+(10*rand('uniform'))); *range: 20-30;
            else if sample=3 then
                sample_size=(study<5)*16+(study>=5)*(study<9)*20+(study>=9)*26;
            else
                sample_size=(study<3)*16+(study>=3)*(study<7)*20+(study>=7)*26;

            if mod(sample_size,2)=1 then sample_size=sample_size+1;
*guarantee equal sizes for two arms;

            if icc=1 then
                do;
                    r2=rand('normal',0,sqrt(8.333333));
                end;
            else if icc=2 then

```

```

do;
r2=rand('normal',0,sqrt(18.75));
end;
else if icc=3 then
do;
r2=rand('normal',0,sqrt(32.142857));
end;
else if icc=4 then
do;
r2=rand('normal',0,sqrt(50));
end;
else
do;
r2=rand('normal',0,sqrt(75));
end;

do participant=1 to sample_size by 1;

    rand_visit=rand('uniform');
    if (0.00<=rand_visit<0.80) then visit_num=5;
    if (0.80<=rand_visit<0.95) then visit_num=4;
    if (0.95<=rand_visit) then visit_num=3; *range: 3-
5;

if (participant<=sample_size/2) then treatment=0;
else treatment=1;

age=round(rand('normal',50,8));

basecholesterol=round(rand('normal',240,3));

if common=1 then do;
    rand_female=rand('uniform');
if (0.00<=rand_female<0.60) then female=0;
    if (0.60<=rand_female) then female=1;

    rand_race=rand('uniform');
    if (0.00<=rand_race<0.50) then race=1;
    if (0.50<=rand_race<0.70) then race=2;
    if (0.70<=rand_race<0.90) then race=3;
    if (0.90<=rand_race) then race=4;

    rand_diabetes=rand('uniform');
    if (0.00<=rand_diabetes<0.60) then diabetes=0;
    if (0.60<=rand_diabetes) then diabetes=1;

    rand_CVD=rand('uniform');
    if (0.00<=rand_CVD<0.70) then CVD=0;
    if (0.70<=rand_CVD) then CVD=1;
end;
else do;
    female=0;
    rand_female=rand('uniform');
if (0.00<=rand_female<0.60 and study in (3,5)) then
female=0;

```

```

                                if (0.60<=rand_female and study in (3,5))      then
female=1;

                                race=0;
                                rand_race=rand('uniform');
                                if (0.00<=rand_race<0.50 and study=1) then race=1;
                                if (0.50<=rand_race<0.70 and study=1) then race=2;
                                if (0.70<=rand_race<0.90 and study=1) then race=3;
                                if (0.90<=rand_race and study=1)      then race=4;

                                diabetes=0;
                                rand_diabetes=rand('uniform');
                                if (0.00<=rand_diabetes<0.60 and study=7) then
diabetes=0;
                                if (0.60<=rand_diabetes and study=7)      then
diabetes=1;

                                CVD=0;
                                rand_CVD=rand('uniform');
                                if (0.00<=rand_CVD<0.70 and study=8) then CVD=0;
                                if (0.70<=rand_CVD and study=8)      then CVD=1;
                                end;

                                if icc=1 then
                                do;
                                r3=rand('normal',0,sqrt(12.5));
                                end;
                                else if icc=2 then
                                do;
                                r3=rand('normal',0,sqrt(28.125));
                                end;
                                else if icc=3 then
                                do;
                                r3=rand('normal',0,sqrt(48.21429));
                                end;
                                else if icc=4 then
                                do;
                                r3=rand('normal',0,sqrt(75));
                                end;
                                else
                                do;
                                r3=rand('normal',0,sqrt(112.5));
                                end;

                                do visit=1 to visit_num by 1;
                                if treatment=0 then
weight=round(rand('normal',150,3));
                                else weight=round(rand('normal',150-visit*1.5,3));
                                output;
                                end;

                                format treatment treat. female ny. race race. diabetes CVD
ny.;
                                output;
                                end;

```

```

    end;
end;

run;

/*Study-level variables: ss_mean ss_std*/
proc sort data=simulated_data out=sort (keep=study num_centers center
sample_size) nodupkey;
by study center;
run;

proc means data=sort mean var noprint;
var sample_size;
class study;
output out=mean_var mean=ss_mean std=ss_std;
run;

data simulated_data2;
merge simulated_data mean_var;
by study;
if study=. then delete;
drop _type_ _freq_;
run;

/*Generate outcome variable: serum_cholesterol*/
proc sql noprint;
select mean(basecholesterol),mean(age),mean(weight)
into :mean_bc, :mean_age, :mean_weight from simulated_data2;
quit;

data simulated_data2;
set simulated_data2;

cntr_bc=basecholesterol-&mean_bc.;
cntr_age=age-&mean_age.;
cntr_weight=weight-&mean_weight.;

if common=1 then do;
mean=intercept+(cntr_bc*beta_basecholesterol)+(cntr_age*beta_age)+(fema
le*beta_female)
+((race=2)*beta_AfrAm)+((race=3)*beta_Hisp)+((race=4)*beta_Other)
+(diabetes*beta_diabetes)+(CVD*beta_CVD)
+(ss_mean*beta_ss_mean)+(ss_std*beta_ss_std)+(order*beta_order)
+(cntr_weight*beta_weight)+(visit*beta_visit)+(visit*treatment*be
ta_visit_trt)
+r1+(visit*treatment*r1_trt)+r2+r3;
end;
else do;
mean=intercept+(cntr_bc*beta_basecholesterol)+(cntr_age*beta_age)
+((study=3)*female*beta_female_3)+((study=5)*female*beta_female_5
)
+((study=1)*(race=2)*beta_AfrAm)+((study=1)*(race=3)*beta_Hisp)+((
study=1)*(race=4)*beta_Other)
+((study=7)*diabetes*beta_diabetes)+((study=8)*CVD*beta_CVD)
+(ss_mean*beta_ss_mean)+(ss_std*beta_ss_std)+(order*beta_order)

```

```

        +(cntr_weight*beta_weight)+(visit*beta_visit)+(visit*treatment*be
ta_visit_trt)
        +r1+(visit*treatment*r1_trt)+r2+r3;
end;

serum_cholesterol=round(rand('normal',mean,15));

run;

/*two-stage methods*/
%if &sample.=1 or &sample.=3 %then %do;
    %let nums=10;
    %end;
%else %do;
    %let nums=6;
    %end;

%do s=1 %to &nums.;

data dat&s.;
set simulated_data2;
if study=&s.;
run;

proc mixed data=dat&s. covtest;
class treatment center participant;
model serum_cholesterol=cntr_bc cntr_age female race diabetes CVD
cntr_weight visit visit*treatment/solution;
random intercept/ type=vc subject=participant(center);
random intercept/ type=vc subject=center;
ods exclude all;
ods output SolutionF=S&s.;
run;

data NS&s.;
studyno=&s.;
set S&s.;
if treatment=1;
est=StdErr**2;
keep studyno Estimate EST;
run;

%if &s.=1 %then %do;
data trt;
set NS&s.;
run;
%end;
%else %do;
proc append base=trt data=NS&s.;
run;
%end;

%end;

```

```

/*Fixed-effects model*/
data trt;
set trt;
w=1/est;
wy=w*Estimate;
run;

data out_temp;
stop;
run;

proc sql;
create table out_temp as
select sum(wy)/sum(w) as treatment_overall, sqrt(1/sum(w)) as se
from trt;
quit;

data out_temp;
no=&i.;
set out_temp;
lower=treatment_overall-1.96*se;
upper=treatment_overall+1.96*se;
run;

%if &i.=1 %then %do;
data d.fixed_s&sample.i&icc.c&common.;
set out_temp;
run;
%end;
%else %do;
proc append base=d.fixed_s&sample.i&icc.c&common. data=out_temp;
run;
%end;

/*Random-effects model*/
data trt;
set trt;
wy2=w*estimate**2;
w2=w**2;
run;

proc sql;
select ((sum(wy2)-sum(wy)**2/sum(w))-(&nums.-1))/(sum(w)-
sum(w**2)/sum(w)) into: b_var from trt;
quit;

data trt;
set trt;
w_r=1/(est+&b_var.);
wy_2=w_r*Estimate;
run;

data out_temp;
stop;
run;

```



```

proc sql;
create table out_temp as
select sum(wy_2)/sum(w_r) as treatment_overall, sqrt(1/sum(w_r)) as se
from trt;
quit;

data out_temp;
no=&i.;
set out_temp;
lower=treatment_overall-1.96*se;
upper=treatment_overall+1.96*se;
between_trt=&b_var.;
run;

%if &i.=1 %then %do;
data d.random_s&sample.i&icc.c&common.;
set out_temp;
run;
%end;
%else %do;
proc append base=d.random_s&sample.i&icc.c&common. data=out_temp;
run;
%end;

/**** LMM ****/
/**** 4-level models ****/

%if &common.=1 %then %do;

/*Proposed model*/
title 'Full model';
proc mixed data=simulated_data2 covtest;
class female(ref='No') race(ref='Caucasian') diabetes(ref='No')
CVD(ref='No') treatment participant center study;
model serum_cholesterol=cntr_bc cntr_age female race diabetes CVD
ss_mean ss_std order
                                cntr_weight visit
visit*treatment/solution cl;
random intercept / type=vc subject=participant(center study);
random intercept / type=vc subject=center(study);
random intercept visit*treatment/ type=vc subject=study;
ods exclude all;
ods output SolutionF=SolutionF1;
ods output CovParms=CovParms1;
run;

%out(SolutionF=SolutionF1,CovParms=CovParms1,num=1,sample=&sample.,icc=
&icc.,common=&common.);

/*Model 1*/
title 'Model 1';
proc mixed data=simulated_data2 covtest;

```



```

                                cntr_weight visit
visit*treatment/solution cl;
repeated / type=vc subject=participant(center study);
ods exclude all;
ods output SolutionF=SolutionF2;
ods output CovParms=CovParms2;
run;

%out(SolutionF=SolutionF2,CovParms=CovParms2,num=2,sample=&sample.,icc=
&icc.,common=&common.);
%end;

%end;
%end;
%end;
%end;

%mend simulation;

%simulation(n=1000);

```

## A.2 Real Data Application

```

title '1-stage proposed';
proc glimmix data=BP_change2 method=RSPL;
class cohort study;
model SBP_binary(event='1')=treatment age sex race weight_base SBP_base
ss_mean ss_std
                                /dist=binary

link=logit solution cl;
random intercept / type=vc subject=cohort(study);
random treatment / type=vc subject=study;
ods exclude none;
covtest / wald;
run;

title '1-stage fixed';
proc glimmix data=BP_change2 method=RSPL;
class cohort study;
model SBP_binary(event='1')=treatment age sex race weight_base SBP_base
                                /dist=binary

link=logit solution cl;
ods output ParameterEstimates=SolutionF2;
ods output CovParms=CovParms2;
run;

%macro twostage;
%do s=1 %to 3;

```

```

data dat&s.;
set BP_change2;
if study=&s.;
run;

proc glimmix data=dat&s. method=RSPL;
class cohort;
model SBP_binary(event='1')=treatment age sex race weight_base SBP_base
      /dist=binary;

link=logit solution cl;
random intercept/ type=vc subject=cohort;
ods exclude all;
ods output ParameterEstimates=S&s.;
run;

data NS&s.;
studyno=&s.;
set S&s.;
if effect="treatment";
est=StdErr**2;
keep studyno Estimate EST;
run;

%if &s.=1 %then %do;
data trt;
set NS&s.;
run;
%end;
%else %do;
proc append base=trt data=NS&s.;
run;
%end;

%end;

/*Fixed-effects model*/
data trt;
set trt;
w=1/est;
wy=w*Estimate;
run;

data out_temp;
stop;
run;

proc sql;
create table out_temp as
select sum(wy)/sum(w) as treatment_overall, sqrt(1/sum(w)) as se
from trt;
quit;

data twofixed;
set out_temp;
lower=treatment_overall-1.96*se;

```

```

upper=treatment_overall+1.96*se;
run;

/*Random-effects model*/
data trt;
set trt;
wy2=w*estimate**2;
w2=w**2;
run;

proc sql;
select ((sum(wy2)-sum(wy)**2/sum(w))-(3-1))/(sum(w)-sum(w**2)/sum(w))
into: b_var from trt;
quit;

data trt;
set trt;
w_r=1/(est+&b_var.);
wy_2=w_r*Estimate;
run;

data out_temp;
stop;
run;

proc sql;
create table out_temp as
select sum(wy_2)/sum(w_r) as treatment_overall, sqrt(1/sum(w_r)) as se
from trt;
quit;

data tworandom;
set out_temp;
lower=treatment_overall-1.96*se;
upper=treatment_overall+1.96*se;
between_trt=&b_var.;
run;

%mend twostage;

%twostage;

```

## Appendix B

### Chapter 4

#### B.1 Simulation Study

```

/** 3-level Data Simulation  */
/** Generate data  */
proc format;
value treat 0='Placebo' 1='Active';
value ny     0='No' 1='Yes';
value race  1='Caucasian' 2='African American' 3='Hispanic' 4='Other';
run;

/*ICC=0.1  vs=0.1218470  vc=0.2436939  v=pi^2/3 vs_trt=vs/2
vc_bc=vc/2*/
/*ICC=0.2  vs=0.2741557  vc=0.5483114  v=pi^2/3 vs_trt=vs/2
vc_bc=vc/2*/
/*ICC=0.3  vs=0.4699812  vc=0.9399623  v=pi^2/3 vs_trt=vs/2
vc_bc=vc/2*/
/*ICC=0.4  vs=0.7310818  vc=1.4621636  v=pi^2/3 vs_trt=vs/2
vc_bc=vc/2*/
/*ICC=0.5  vs=1.0966227  vc=2.1932454  v=pi^2/3 vs_trt=vs/2
vc_bc=vc/2*/
/*vs:vc=1:2 vs+vc=(pi^2/3*icc)/(1-icc)*/

%macro simulation(n);

%do sample=1 %to 4;
%do icc=1 %to 5;
%do common=1 %to 2;
%do i=1 %to &n.;

/*Participant-level covariates + study-level var*/
data simulated_data;

sample=&sample.;
icc=&icc.;
common=&common.;

if sample in (1,3) then num_studies=12;
else num_studies=8;

beta_treatment=-0.8;
beta_basecholesterol=0.02;
beta_age=0.02;
beta_female=-0.4;
beta_female_3=-0.3;
beta_female_5=-0.4;
beta_AfrAm=0.4;

```

```

beta_Hisp=0.6;
beta_Other=-0.08;
beta_diabetes=0.6;
beta_CVD=0.6;
beta_ss_mean=-0.03;
beta_ss_std=0.15;
beta_order=-0.05;

do study=1 to num_studies by 1;
  if sample=1 then num_centers=round(8+(8*rand('uniform')));
  *range: 8-16;
  else if sample=2 then num_centers=round(8+(4*rand('uniform')));
  *range: 8-12;
  else if sample=3 then
num_centers=(study<5)*6+(study>=5)*(study<9)*9+(study>=9)*12;
  else
num_centers=(study<3)*6+(study>=3)*(study<7)*9+(study>=7)*12;

  order=study; *sequential order of the study = study #;

  if icc=1 then
    do;
      r1=rand('normal',0,sqrt(0.1218470)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(0.1218470/2)); *study-level
random trt effect;
    end;
  else if icc=2 then
    do;
      r1=rand('normal',0,sqrt(0.2741557)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(0.2741557/2)); *study-level
random trt effect;
    end;
  else if icc=3 then
    do;
      r1=rand('normal',0,sqrt(0.4699812)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(0.4699812/2)); *study-level
random trt effect;
    end;
  else if icc=4 then
    do;
      r1=rand('normal',0,sqrt(0.7310818)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(0.7310818/2)); *study-level
random trt effect;
    end;
  else
    do;
      r1=rand('normal',0,sqrt(1.0966227)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(1.0966227/2)); *study-level
random trt effect;
    end;
end;

```

```

do center=1 to num_centers by 1;
  if sample=1 then
sample_size=round(24+(46*rand('uniform'))); *range: 24-70;
  else if sample=2 then
sample_size=round(20+(10*rand('uniform'))); *range: 20-30;
  else if sample=3 then
sample_size=(study<5)*16+(study>=5)*(study<9)*20+(study>=9)*26;
  else
sample_size=(study<3)*16+(study>=3)*(study<7)*20+(study>=7)*26;

  if mod(sample_size,2)=1 then sample_size=sample_size+1;
*guarantee equal sizes for two arms;

  if icc=1 then
do;
r2=rand('normal',0,sqrt(0.2436939));
r2_bc=rand('normal',0,sqrt(0.2436939/2));
end;
  else if icc=2 then
do;
r2=rand('normal',0,sqrt(0.5483114));
r2_bc=rand('normal',0,sqrt(0.5483114/2));
end;
  else if icc=3 then
do;
r2=rand('normal',0,sqrt(0.9399623));
r2_bc=rand('normal',0,sqrt(0.9399623/2));
end;
  else if icc=4 then
do;
r2=rand('normal',0,sqrt(1.4621636));
r2_bc=rand('normal',0,sqrt(1.4621636/2));
end;
  else
do;
r2=rand('normal',0,sqrt(2.1932454));
r2_bc=rand('normal',0,sqrt(2.1932454/2));
end;

do participant=1 to sample_size by 1;

if (participant<=sample_size/2) then treatment=0;
  else treatment=1;

age=round(rand('normal',50,8));

basecholesterol=round(rand('normal',240,3));

if common=1 then do;
  rand_female=rand('uniform');
if (0.00<=rand_female<0.60) then female=0;
  if (0.60<=rand_female) then female=1;

  rand_race=rand('uniform');

```



```

        if (0.00<=rand_race<0.50) then race=1;
        if (0.50<=rand_race<0.70) then race=2;
        if (0.70<=rand_race<0.90) then race=3;
        if (0.90<=rand_race)      then race=4;

        rand_diabetes=rand('uniform');
        if (0.00<=rand_diabetes<0.60) then diabetes=0;
        if (0.60<=rand_diabetes)      then diabetes=1;

        rand_CVD=rand('uniform');
        if (0.00<=rand_CVD<0.70) then CVD=0;
        if (0.70<=rand_CVD)      then CVD=1;
    end;
else do;
    female=0;
    rand_female=rand('uniform');
    if (0.00<=rand_female<0.60 and study in (3,5)) then
female=0;
        if (0.60<=rand_female and study in (3,5))      then
female=1;

        race=0;
        rand_race=rand('uniform');
        if (0.00<=rand_race<0.50 and study=1) then race=1;
        if (0.50<=rand_race<0.70 and study=1) then race=2;
        if (0.70<=rand_race<0.90 and study=1) then race=3;
        if (0.90<=rand_race and study=1)      then race=4;

        diabetes=0;
        rand_diabetes=rand('uniform');
        if (0.00<=rand_diabetes<0.60 and study=7) then
diabetes=0;
        if (0.60<=rand_diabetes and study=7)      then
diabetes=1;

        CVD=0;
        rand_CVD=rand('uniform');
        if (0.00<=rand_CVD<0.70 and study=8) then CVD=0;
        if (0.70<=rand_CVD and study=8)      then CVD=1;
    end;

    format treatment treat. female ny. race race. diabetes CVD
ny.;
    output;

    end;

    end;

end;

run;

/*Study-level variables: ss_mean ss_std*/

```

```

proc sort data=simulated_data out=sort (keep=study num_centers center
sample_size) nodupkey;
by study center;
run;

proc means data=sort mean var noprint;
var sample_size;
class study;
output out=mean_var mean=ss_mean std=ss_std;
run;

data simulated_data2;
merge simulated_data mean_var;
by study;
if study=. then delete;
drop _type_ _freq_;
run;

/*Generate outcome variable: serum_cholesterol_cat*/
proc sql noprint;
select mean(basecholesterol),mean(age) into :mean_bc, :mean_age from
simulated_data2;
quit;

data simulated_data2;
set simulated_data2;

cntr_bc=basecholesterol-&mean_bc.;
cntr_age=age-&mean_age.;

if common=1 then do;
e=exp((cntr_bc*beta_basecholesterol)+(treatment*beta_treatment)+(cntr_a
ge*beta_age)+(female*beta_female)
+((race=2)*beta_AfrAm)+((race=3)*beta_Hisp)+((race=4)*beta_Other)
+(diabetes*beta_diabetes)+(CVD*beta_CVD)
+(ss_mean*beta_ss_mean)+(ss_std*beta_ss_std)+(order*beta_order)
+r1+(treatment*r1_trt)+r2+(cntr_bc*r2_bc));
end;
else do;
e=exp((cntr_bc*beta_basecholesterol)+(treatment*beta_treatment)+(cntr_a
ge*beta_age)
+((study=3)*female*beta_female_3)+((study=5)*female*beta_female_5
)
+((study=1)*(race=2)*beta_AfrAm)+((study=1)*(race=3)*beta_Hisp)+
(study=1)*(race=4)*beta_Other)
+((study=7)*diabetes*beta_diabetes)+((study=8)*CVD*beta_CVD)
+(ss_mean*beta_ss_mean)+(ss_std*beta_ss_std)+(order*beta_order)
+r1+(treatment*r1_trt)+r2+(cntr_bc*r2_bc));
end;

P=e/(1+e);
serum_cholesterol_cat=rand('BERN',P);

run;

```

```

/*two-stage methods*/
%if &sample.=1 or &sample.=3 %then %do;
    %let nums=12;
    %end;
%else %do;
    %let nums=8;
    %end;

%do s=1 %to &nums.;

data dat&s.;
set simulated_data2;
if study=&s.;
run;

proc glimmix data=dat&s. method=RSPL;
class treatment(ref='Placebo') female race diabetes CVD center;
model serum_cholesterol_cat(event='1')=treatment cntr_bc cntr_age
female race diabetes CVD /dist=binary link=logit solution;
random intercept cntr_bc/ type=vc subject=center;
ods exclude all;
ods output ParameterEstimates=S&s.;
run;

data NS&s.;
studyno=&s.;
set S&s.;
if treatment=1;
est=StdErr**2;
keep studyno Estimate EST;
run;

%if &s.=1 %then %do;
data trt;
set NS&s.;
run;
%end;
%else %do;
proc append base=trt data=NS&s.;
run;
%end;

%end;

/*Fixed-effects model*/
data trt;
set trt;
w=1/est;
wy=w*Estimate;
run;

data out_temp;
stop;
run;

```

```

proc sql;
create table out_temp as
select sum(wy)/sum(w) as treatment_overall, sqrt(1/sum(w)) as se
from trt;
quit;

data out_temp;
no=&i.;
set out_temp;
lower=treatment_overall-1.96*se;
upper=treatment_overall+1.96*se;
run;

%if &i.=1 %then %do;
data d.fixed_s&sample.i&icc.c&common.;
set out_temp;
run;
%end;
%else %do;
proc append base=d.fixed_s&sample.i&icc.c&common. data=out_temp;
run;
%end;

/*Random-effects model*/
data trt;
set trt;
wy2=w*estimate**2;
w2=w**2;
run;

proc sql;
select ((sum(wy2)-sum(wy)**2/sum(w))-(&nums.-1))/(sum(w)-
sum(w**2)/sum(w)) into: b_var from trt;
quit;

data trt;
set trt;
w_r=1/(est+&b_var.);
wy_2=w_r*Estimate;
run;

data out_temp;
stop;
run;

proc sql;
create table out_temp as
select sum(wy_2)/sum(w_r) as treatment_overall, sqrt(1/sum(w_r)) as se
from trt;
quit;

data out_temp;
no=&i.;

```

```

set out_temp;
lower=treatment_overall-1.96*se;
upper=treatment_overall+1.96*se;
between_trt=&b_var.;
run;

%if &i.=1 %then %do;
data d.random_s&sample.i&icc.c&common.;
set out_temp;
run;
%end;
%else %do;
proc append base=d.random_s&sample.i&icc.c&common. data=out_temp;
run;
%end;

/**** GLMM ****/
/**** 3-level models ****/

%if &common.=1 %then %do;

/*Proposed model*/
title 'Full model';
proc glimmix data=simulated_data2 method=RSPL;
class treatment(ref='Placebo') female(ref='No') race(ref='Caucasian')
diabetes(ref='No') CVD(ref='No') center study;
model serum_cholesterol_cat(event='1')=treatment cntr_bc cntr_age
female race diabetes CVD ss_mean ss_std order
                                /dist=binary
link=logit solution cl;
random intercept cntr_bc/ type=vc subject=center(study);
random intercept treatment/ type=vc subject=study;
covtest / wald;
ods exclude all;
ods output ParameterEstimates=SolutionF1;
ods output CovParms=CovParms1;
run;

%out(SolutionF=SolutionF1,CovParms=CovParms1,num=1,sample=&sample.,icc=
&icc.,common=&common.);

/*Model 1*/
title 'Model 1';
proc glimmix data=simulated_data2 method=RSPL;
class treatment(ref='Placebo') female(ref='No') race(ref='Caucasian')
diabetes(ref='No') CVD(ref='No') center study;
model serum_cholesterol_cat(event='1')=treatment cntr_bc cntr_age
female race diabetes CVD /dist=binary link=logit solution cl;
covtest / wald;
ods exclude all;
ods output ParameterEstimates=SolutionF2;
ods output CovParms=CovParms2;
run;

```

```

%out (SolutionF=SolutionF2,CovParms=CovParms2,num=2,sample=&sample.,icc=
&icc.,common=&common.);
%end;

%else %do;
/** GLMM **/
/** unique 3-level models **/

data simulated_data2;
set simulated_data2;
AfrAm=0;Hispanic=0;Other=0;
if race=2 then AfrAm=1;
if race=3 then Hisp=1;
if race=4 then Other=1;
run;

/*Proposed model*/
title 'Full model';
proc glimmix data=simulated_data2 method=RSPL;
class center study;
model serum_cholesterol_cat(event='1')=treatment cnr_bc cnr_age
female(study) AfrAm(study) Hisp(study) Other(study)
diabetes(study) CVD(study) ss_mean
ss_std order/dist=binary link=logit solution cl;
random intercept cnr_bc/ type=vc subject=center(study);
random intercept treatment/ type=vc subject=study;
covtest / wald;
ods exclude all;
ods output ParameterEstimates=SolutionF1;
ods output CovParms=CovParms1;
run;

%out (SolutionF=SolutionF1,CovParms=CovParms1,num=1,sample=&sample.,icc=
&icc.,common=&common.);

/*Model 1*/
title 'Model 1';
proc glimmix data=simulated_data2 method=RSPL;
class center study;
model serum_cholesterol_cat(event='1')=treatment cnr_bc cnr_age
female(study) AfrAm(study) Hisp(study) Other(study)
diabetes(study) CVD(study)

/dist=binary link=logit solution cl;
covtest / wald;
ods exclude all;
ods output ParameterEstimates=SolutionF2;
ods output CovParms=CovParms2;
run;

%out (SolutionF=SolutionF2,CovParms=CovParms2,num=2,sample=&sample.,icc=
&icc.,common=&common.);
%end;

%end;
%end;

```

```

%end;
%end;

%mend simulation;

%simulation(n=1000)

/**** 4-level Data Simulation ****/
/**** Generate data ****/
proc format;
value treat 0='Placebo' 1='Active';
value ny     0='No' 1='Yes';
value race  1='Caucasian' 2='African American' 3='Hispanic' 4='Other';
run;

/*ICC=0.1  vs=0.1015391  vc=0.1218470  vp=0.1421548  v=pi^2/3
vs_trt=vs/2 */
/*ICC=0.2  vs=0.2284631  vc=0.2741557  vp=0.3198483  v=pi^2/3
vs_trt=vs/2 */
/*ICC=0.3  vs=0.3916510  vc=0.4699812  vp=0.5483114  v=pi^2/3
vs_trt=vs/2 */
/*ICC=0.4  vs=0.6092348  vc=0.7310818  vp=0.8529288  v=pi^2/3
vs_trt=vs/2 */
/*ICC=0.5  vs=0.9138523  vc=1.0966227  vp=1.2793932  v=pi^2/3
vs_trt=vs/2 */

%macro simulation(n);

%do sample=1 %to 4;
%do icc=1 %to 5;
%do common=1 %to 2;
%do i=1 %to &n.;

/*Participant-level covariates + study-level var*/
data simulated_data;

sample=&sample.;
icc=&icc.;
common=&common.;

if sample in (1,3) then num_studies=10;
else num_studies=6;

beta_basecholesterol=0.02;
beta_age=0.02;
beta_female=-0.4;
beta_female_3=-0.3;
beta_female_5=-0.4;
beta_AfrAm=0.4;
beta_Hisp=0.6;
beta_Other=-0.08;
beta_diabetes=0.6;

```

```

beta_CVD=0.6;
beta_ss_mean=-0.03;
beta_ss_std=0.15;
beta_order=-0.05;
beta_weight=0.04;
beta_visit=-0.1;
beta_visit_trt=-0.4;

do study=1 to num_studies by 1;
  if sample=1 then num_centers=round(6+(4*rand('uniform')));
  *range: 6-10;
  else if sample=2 then num_centers=round(4+(4*rand('uniform')));
  *range: 4-8;

  else if sample=3 then
num_centers=(study<4)*4+(study>=4)*(study<8)*8+(study>=8)*10;
  else num_centers=(study<3)*3+(study>=3)*(study<5)*6+(study>=5)*9;

  order=study; *sequential order of the study = study #;

  if icc=1 then
    do;
      r1=rand('normal',0,sqrt(0.1015391)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(0.1015391/2)); *study-level
random trt effect;
    end;
  else if icc=2 then
    do;
      r1=rand('normal',0,sqrt(0.2284631)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(0.2284631/2)); *study-level
random trt effect;
    end;
  else if icc=3 then
    do;
      r1=rand('normal',0,sqrt(0.3916510)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(0.3916510/2)); *study-level
random trt effect;
    end;
  else if icc=4 then
    do;
      r1=rand('normal',0,sqrt(0.6092348)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(0.6092348/2)); *study-level
random trt effect;
    end;
  else
    do;
      r1=rand('normal',0,sqrt(0.9138523)); *study-level random
intercept;
      r1_trt=rand('normal',0,sqrt(0.9138523/2)); *study-level
random trt effect;
    end;
  end;
end;

```



```

end;

do center=1 to num_centers by 1;
  if sample=1 then sample_size=round(18+(4*rand('uniform')));
  *range: 18-22;
  else if sample=2 then
  sample_size=round(18+(4*rand('uniform'))); *range: 18-22;
  else if sample=3 then
  sample_size=(study<4)*16+(study>=4)*(study<8)*20+(study>=8)*24;
  else
  sample_size=(study<3)*16+(study>=3)*(study<5)*20+(study>=5)*24;

  if mod(sample_size,2)=1 then sample_size=sample_size+1;
  *guarantee equal sizes for two arms;

  if icc=1 then
  do;
  r2=rand('normal',0,sqrt(0.1218470));
  end;
  else if icc=2 then
  do;
  r2=rand('normal',0,sqrt(0.2741557));
  end;
  else if icc=3 then
  do;
  r2=rand('normal',0,sqrt(0.4699812));
  end;
  else if icc=4 then
  do;
  r2=rand('normal',0,sqrt(0.7310818));
  end;
  else
  do;
  r2=rand('normal',0,sqrt(1.0966227));
  end;

do participant=1 to sample_size by 1;

  rand_visit=rand('uniform');
  if (0.00<=rand_visit<0.80) then visit_num=5;
  if (0.80<=rand_visit<0.95) then visit_num=4;
  if (0.95<=rand_visit) then visit_num=3; *range: 3-
5;

if (participant<=sample_size/2) then treatment=0;
else treatment=1;

age=round(rand('normal',50,8));

basecholesterol=round(rand('normal',240,3));

if common=1 then do;
  rand_female=rand('uniform');
if (0.00<=rand_female<0.60) then female=0;
  if (0.60<=rand_female) then female=1;

```

```

    rand_race=rand('uniform');
    if (0.00<=rand_race<0.50) then race=1;
    if (0.50<=rand_race<0.70) then race=2;
    if (0.70<=rand_race<0.90) then race=3;
    if (0.90<=rand_race)      then race=4;

    rand_diabetes=rand('uniform');
    if (0.00<=rand_diabetes<0.60) then diabetes=0;
    if (0.60<=rand_diabetes)      then diabetes=1;

    rand_CVD=rand('uniform');
    if (0.00<=rand_CVD<0.70) then CVD=0;
    if (0.70<=rand_CVD)      then CVD=1;
end;
else do;
    female=0;
    rand_female=rand('uniform');
    if (0.00<=rand_female<0.60 and study in (3,5)) then
        if (0.60<=rand_female and study in (3,5))      then
            female=1;

            race=0;
            rand_race=rand('uniform');
            if (0.00<=rand_race<0.50 and study=1) then race=1;
            if (0.50<=rand_race<0.70 and study=1) then race=2;
            if (0.70<=rand_race<0.90 and study=1) then race=3;
            if (0.90<=rand_race and study=1)      then race=4;

            diabetes=0;
            rand_diabetes=rand('uniform');
            if (0.00<=rand_diabetes<0.60) then diabetes=0;
            if (0.60<=rand_diabetes)      then diabetes=1;

            CVD=0;
            rand_CVD=rand('uniform');
            if (0.00<=rand_CVD<0.70) then CVD=0;
            if (0.70<=rand_CVD)      then CVD=1;
        end;
    end;

    if icc=1 then
    do;
    r3=rand('normal',0,sqrt(0.1421548));
    end;
    else if icc=2 then
    do;
    r3=rand('normal',0,sqrt(0.3198483));
    end;
    else if icc=3 then
    do;
    r3=rand('normal',0,sqrt(0.5483114));
    end;
    else if icc=4 then

```

```

do;
  r3=rand('normal',0,sqrt(0.8529288));
end;
else
do;
  r3=rand('normal',0,sqrt(1.2793932));
end;

do visit=1 to visit_num by 1;
  if treatment=0 then
weight=round(rand('normal',150,3));
  else weight=round(rand('normal',150-visit*1.5,3));
  output;
end;

format treatment treat. female ny. race race. diabetes CVD
ny.;
output;
end;
end;
end;

run;

/*Study-level variables: ss_mean ss_std*/
proc sort data=simulated_data out=sort (keep=study num_centers center
sample_size) nodupkey;
by study center;
run;

proc means data=sort mean std noprint;
var sample_size;
class study;
output out=mean_var mean=ss_mean std=ss_std;
run;

data simulated_data2;
merge simulated_data mean_var;
by study;
if study=. then delete;
drop _type_ _freq_;
run;

/*Generate outcome variable: serum_cholesterol*/
proc sql noprint;
select mean(basecholesterol),mean(age),mean(weight)
into :mean_bc, :mean_age, :mean_weight from simulated_data2;
quit;

data simulated_data2;
set simulated_data2;

cntr_bc=basecholesterol-&mean_bc.;
cntr_age=age-&mean_age.;
cntr_weight=weight-&mean_weight.;

```

```

if common=1 then do;
e=exp((cntr_bc*beta_basecholesterol)+(cntr_age*beta_age)+(female*beta_f
emale)
      +((race=2)*beta_AfrAm)+((race=3)*beta_Hisp)+((race=4)*beta_Other)
+ (diabetes*beta_diabetes)+(CVD*beta_CVD)
      +(ss_mean*beta_ss_mean)+(ss_std*beta_ss_std)+(order*beta_order)
      +(cntr_weight*beta_weight)+(visit*beta_visit)+(visit*treatment*be
ta_visit_trt)
      +r1+(visit*treatment*r1_trt)+r2+r3);
end;
else do;
e=exp((cntr_bc*beta_basecholesterol)+(cntr_age*beta_age)
      +((study=3)*female*beta_female_3)+((study=5)*female*beta_female_5
)
      +((study=1)*(race=2)*beta_AfrAm)+((study=1)*(race=3)*beta_Hisp)+
(study=1)*(race=4)*beta_Other)
      +(diabetes*beta_diabetes)+(CVD*beta_CVD)
      +(ss_mean*beta_ss_mean)+(ss_std*beta_ss_std)+(order*beta_order)
      +(cntr_weight*beta_weight)+(visit*beta_visit)+(visit*treatment*be
ta_visit_trt)
      +r1+(visit*treatment*r1_trt)+r2+r3);
end;

P=e/(1+e);
serum_cholesterol_cat=rand('BERN',P);

run;

/*two-stage methods*/
%if &sample.=1 or &sample.=3 %then %do;
    %let nums=10;
    %end;
%else %do;
    %let nums=6;
    %end;

%do s=1 %to &nums.;

data dat&s.;
set simulated_data2;
if study=&s.;
run;

proc glimmix data=dat&s. method=RSPL;
class treatment(ref='Placebo') female race diabetes CVD center;
model serum_cholesterol_cat(event='1')= cntr_bc cntr_age female race
diabetes CVD cntr_weight visit visit*treatment/dist=binary link=logit
solution;
random intercept/ type=vc subject=participant(center);
random intercept/ type=vc subject=center;
ods exclude all;
ods output ParameterEstimates=S&s.;
run;

```

```

data NS&s.;
studyno=&s.;
set S&s.;
if treatment=1;
est=StdErr**2;
keep studyno Estimate EST;
run;

%if &s.=1 %then %do;
data trt;
set NS&s.;
run;
%end;
%else %do;
proc append base=trt data=NS&s.;
run;
%end;

%end;

/*Fixed-effects model*/
data trt;
set trt;
w=1/est;
wy=w*Estimate;
run;

data out_temp;
stop;
run;

proc sql;
create table out_temp as
select sum(wy)/sum(w) as treatment_overall, sqrt(1/sum(w)) as se
from trt;
quit;

data out_temp;
no=&i.;
set out_temp;
lower=treatment_overall-1.96*se;
upper=treatment_overall+1.96*se;
run;

%if &i.=1 %then %do;
data d.fixed_s&sample.i&icc.c&common.;
set out_temp;
run;
%end;
%else %do;
proc append base=d.fixed_s&sample.i&icc.c&common. data=out_temp;
run;
%end;

```

```

/*Random-effects model*/
data trt;
set trt;
wy2=w*estimate**2;
w2=w**2;
run;

proc sql;
select ((sum(wy2)-sum(wy)**2/sum(w))-(&nums.-1))/(sum(w)-
sum(w**2)/sum(w)) into: b_var from trt;
quit;

data trt;
set trt;
w_r=1/(est+&b_var.);
wy_2=w_r*Estimate;
run;

data out_temp;
stop;
run;

proc sql;
create table out_temp as
select sum(wy_2)/sum(w_r) as treatment_overall, sqrt(1/sum(w_r)) as se
from trt;
quit;

data out_temp;
no=&i.;
set out_temp;
lower=treatment_overall-1.96*se;
upper=treatment_overall+1.96*se;
between_trt=&b_var.;
run;

%if &i.=1 %then %do;
data d.random_s&sample.i&icc.c&common.;
set out_temp;
run;
%end;
%else %do;
proc append base=d.random_s&sample.i&icc.c&common. data=out_temp;
run;
%end;

/**** LMM ****/
/**** 4-level models ****/

%if &common.=1 %then %do;

/*Proposed model*/
title 'Full model';
proc glimmix data=simulated_data2 method=RSPL;

```

```

class female(ref='No') race(ref='Caucasian') diabetes(ref='No')
CVD(ref='No') participant center study;
model serum_cholesterol_cat(event='1')=cntr_bc cntr_age female race
diabetes CVD ss_mean ss_std order
                                cntr_weight visit
visit*treatment/dist=binary link=logit solution cl;
random intercept / type=vc subject=participant(center study);
random intercept / type=vc subject=center(study);
random intercept visit*treatment/ type=vc subject=study;
covtest / wald;
ods exclude all;
ods output ParameterEstimates=SolutionF1;
ods output CovParms=CovParms1;
run;

%out(SolutionF=SolutionF1,CovParms=CovParms1,num=1,sample=&sample.,icc=
&icc.,common=&common.);

/*Model 1*/
title 'Model 1';
proc glimmix data=simulated_data2 method=RSPL;
class female(ref='No') race(ref='Caucasian') diabetes(ref='No')
CVD(ref='No') participant center study;
model serum_cholesterol_cat(event='1')=cntr_bc cntr_age female race
diabetes CVD
                                cntr_weight visit
visit*treatment/dist=binary link=logit solution cl;
random _residual_ / type=vc subject=participant(center study);
covtest / wald;
ods exclude all;
ods output ParameterEstimates=SolutionF2;
ods output CovParms=CovParms2;
run;

%out(SolutionF=SolutionF2,CovParms=CovParms2,num=2,sample=&sample.,icc=
&icc.,common=&common.);

%end;

%else %do;
/** LMM ***/
/** unique 3-level models ***/

data simulated_data2;
set simulated_data2;
AfrAm=0;Hispanic=0;Other=0;
if race=2 then AfrAm=1;
if race=3 then Hispanic=1;
if race=4 then Other=1;
run;

/*Proposed model*/
title 'Full model';
proc glimmix data=simulated_data2 method=RSPL;

```

```

class diabetes(ref='No') CVD(ref='No') participant center study;
model serum_cholesterol_cat(event='1')=cntr_bc cntr_age female(study)
AfrAm(study) Hisp(study) Other(study)
                                diabetes CVD ss_mean ss_std order
                                cntr_weight visit
visit*treatment/dist=binary link=logit solution cl;
random intercept / type=vc subject=participant(center study);
random intercept / type=vc subject=center(study);
random intercept visit*treatment/ type=vc subject=study;
covtest / wald;
ods exclude all;
ods output ParameterEstimates=SolutionF1;
ods output CovParms=CovParms1;
run;

%out(SolutionF=SolutionF1,CovParms=CovParms1,num=1,sample=&sample.,icc=
&icc.,common=&common.);

/*Model 1*/
title 'Model 1';
proc glimmix data=simulated_data2 method=RSPL;
class diabetes(ref='No') CVD(ref='No') participant center study;
model serum_cholesterol_cat(event='1')=cntr_bc cntr_age female(study)
AfrAm(study) Hisp(study) Other(study)
                                diabetes CVD
                                cntr_weight visit
visit*treatment/dist=binary link=logit solution cl;
random _residual_ / type=vc subject=participant(center study);
covtest / wald;
ods exclude all;
ods output ParameterEstimates=SolutionF2;
ods output CovParms=CovParms2;
run;

%out(SolutionF=SolutionF2,CovParms=CovParms2,num=2,sample=&sample.,icc=
&icc.,common=&common.);

%end;

%end;

%end;

%end;

%end;

%mend simulation;

%simulation(n=1000);

```



## B.2 Real Data Application

```

/***** SBP *****/
/*binary*/
title '1-stage proposed';
proc glimmix data=BP_change2 method=RSPL;
class cohort study;
model SBP_binary(event='1')=treatment age sex race weight_base SBP_base
ss_mean ss_std
                                                    /dist=binary

link=logit solution cl;
random intercept / type=vc subject=cohort(study);
random treatment / type=vc subject=study;
ods exclude none;
covtest / wald;
run;

title '1-stage fixed';
proc glimmix data=BP_change2 method=RSPL;
class cohort study subject;
model SBP_binary(event='1')=treatment age sex race weight_base SBP_base
                                                    /dist=binary

link=logit solution cl;
ods exclude none;
ods output ParameterEstimates=SolutionF2;
ods output CovParms=CovParms2;
run;

%macro twostage_binary(BP);
%do s=1 %to 3;

data dat&s.;
set BP_change2;
if study=&s.;
run;

proc glimmix data=dat&s. method=RSPL;
class cohort;
model &BP._binary(event='1')=treatment age sex race weight_base
&BP._base
                                                    /dist=binary

link=logit solution cl;
random intercept/ type=vc subject=cohort;
ods exclude all;
ods output ParameterEstimates=S&s.;
run;

data NS&s.;
studyno=&s.;
set S&s.;
if effect="treatment";
est=StdErr**2;
keep studyno Estimate EST;
run;

```

```

%if &s.=1 %then %do;
data trt;
set NS&s.;
run;
%end;
%else %do;
proc append base=trt data=NS&s.;
run;
%end;

%end;

/*Fixed-effects model*/
data trt;
set trt;
w=1/est;
wy=w*Estimate;
run;

data out_temp;
stop;
run;

proc sql;
create table out_temp as
select sum(wy)/sum(w) as treatment_overall, sqrt(1/sum(w)) as se
from trt;
quit;

data twofixed;
set out_temp;
lower=treatment_overall-1.96*se;
upper=treatment_overall+1.96*se;
run;

/*Random-effects model*/
data trt;
set trt;
wy2=w*estimate**2;
w2=w**2;
run;

proc sql;
select ((sum(wy2)-sum(wy)**2/sum(w))-(3-1))/(sum(w)-sum(w**2)/sum(w))
into: b_var from trt;
quit;

data trt;
set trt;
w_r=1/(est+&b_var.);
wy_2=w_r*Estimate;
run;

data out_temp;

```

```

stop;
run;

proc sql;
create table out_temp as
select sum(wy_2)/sum(w_r) as treatment_overall, sqrt(1/sum(w_r)) as se
from trt;
quit;

data tworandom;
set out_temp;
lower=treatment_overall-1.96*se;
upper=treatment_overall+1.96*se;
between_trt=&b_var.;
run;

%mend twostage_binary;

%twostage_binary(SBP);

/*ordinal*/
proc glimmix data=BP_change2 method=laplace;
class cohort study;
model SBP_ordinal(descending)=treatment age sex race weight_base
SBP_base ss_mean ss_std
/dist=multinomial
link=cumlogit solution cl;
random intercept / type=vc subject=cohort(study) ;
random treatment / type=vc subject=study ;
ods exclude none;
covtest / wald;
run;

title '1-stage fixed';
proc glimmix data=BP_change2 method=RSPL;
class cohort study ;
model SBP_ordinal(descending)=treatment age sex race weight_base
SBP_base
/dist=multinomial
link=cumlogit solution cl;
ods exclude none;
ods output ParameterEstimates=SolutionF2;
ods output CovParms=CovParms2;
run;

%macro twostage_ordinal(BP);
%do s=1 %to 3;

data dat&s.;
set BP_change2;
if study=&s.;
run;

proc glimmix data=dat&s. method=RSPL;

```

```

class cohort;
model &BP._ordinal(descending)=treatment age sex race weight_base
&BP._base
                                                    /dist=multinomial

link=cumlogit solution cl;
random intercept/ type=vc subject=cohort;
ods exclude all;
ods output ParameterEstimates=S&s.;
run;

data NS&s.;
studyno=&s.;
set S&s.;
if effect="treatment";
est=StdErr**2;
keep studyno Estimate EST;
run;

%if &s.=1 %then %do;
data trt;
set NS&s.;
run;
%end;
%else %do;
proc append base=trt data=NS&s.;
run;
%end;

%end;

/*Fixed-effects model*/
data trt;
set trt;
w=1/est;
wy=w*Estimate;
run;

data out_temp;
stop;
run;

proc sql;
create table out_temp as
select sum(wy)/sum(w) as treatment_overall, sqrt(1/sum(w)) as se
from trt;
quit;

data twofixed;
set out_temp;
lower=treatment_overall-1.96*se;
upper=treatment_overall+1.96*se;
run;

/*Random-effects model*/
data trt;

```

```

set trt;
wy2=w*estimate**2;
w2=w**2;
run;

proc sql;
select ((sum(wy2)-sum(wy)**2/sum(w))-(3-1))/(sum(w)-sum(w**2)/sum(w))
into: b_var from trt;
quit;

data trt;
set trt;
w_r=1/(est+&b_var.);
wy_2=w_r*Estimate;
run;

data out_temp;
stop;
run;

proc sql;
create table out_temp as
select sum(wy_2)/sum(w_r) as treatment_overall, sqrt(1/sum(w_r)) as se
from trt;
quit;

data tworandom;
set out_temp;
lower=treatment_overall-1.96*se;
upper=treatment_overall+1.96*se;
between_trt=&b_var.;
run;

%mend twostage_ordinal;

%twostage_ordinal(SBP);

/***** DBP *****/
/*binary*/
title '1-stage proposed';
proc glimmix data=BP_change2 method=RSPL;
class cohort study;
model DBP_binary(event='1')=treatment age sex race weight_base DBP_base
ss_mean ss_std
/dist=binary
link=logit solution cl;
random intercept / type=vc subject=cohort(study);
random treatment / type=vc subject=study;
ods exclude none;
covtest / wald;
run;

title '1-stage fixed';

```

```

proc glimmix data=BP_change2 method=RSPL;
class cohort study subject;
model DBP_binary(event='1')=treatment age sex race weight_base DBP_base
                                         /dist=binary

link=logit solution cl;
ods exclude none;
ods output ParameterEstimates=SolutionF2;
ods output CovParms=CovParms2;
run;

%twostage_binary(DBP);

/*ordinal*/
proc glimmix data=BP_change2 method=RSPL;
class cohort study;
model DBP_ordinal(descending)=treatment age sex race weight_base
DBP_base ss_mean ss_std
                                         /dist=multinomial

link=cumlogit solution cl;
random intercept / type=vc subject=cohort(study) ;
random treatment / type=vc subject=study ;
ods exclude none;
covtest / wald;
run;

title '1-stage fixed';
proc glimmix data=BP_change2 method=RSPL;
class cohort study ;
model DBP_ordinal(descending)=treatment age sex race weight_base
DBP_base
                                         /dist=multinomial

link=cumlogit solution cl;
*random _residual_ / type=vc subject=cohort(study);
ods exclude none;
ods output ParameterEstimates=SolutionF2;
ods output CovParms=CovParms2;
run;

%twostage_ordinal(DBP);

```

## Bibliography

- Appel, L. J., Champagne, C. M., Harsha, D. W., Cooper, L. S., Obarzanek, E., Elmer, P. J., ... & Young, D. R. (2003). Effects of comprehensive lifestyle modification on blood pressure control: main results of the PREMIER clinical trial. *JAMA: Journal of the American Medical Association*.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to Meta-Analysis*: John Wiley & Sons.
- Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421), 9. doi: 10.2307/2290687
- Burke, D. L., Bujkiewicz, S., & Riley, R. D. (2018). Bayesian bivariate meta-analysis of correlated effects: Impact of the prior distributions on the between-study correlation, borrowing of strength, and joint inferences. *Statistical methods in medical research*, 27(2), 428-450.
- Bujkiewicz, S., Thompson, J. R., Sutton, A. J., Cooper, N. J., Harrison, M. J., Symmons, D. P., & Abrams, K. R. (2013). Multivariate meta-analysis of mixed outcomes: a Bayesian approach. *Statistics in Medicine*, 32(22), 3926-3943.
- Diggle, P. (2002). *Analysis of Longitudinal Data*: Oxford University Press.
- Higgins, J. P., Whitehead, A., Turner, R. M., Omar, R. Z., & Thompson, S. G. (2001). Meta-analysis of continuous outcome data from individual patients. *Stat Med*, 20(15), 2219-2241. doi: 10.1002/sim.918
- Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.

- Jennrich, R. I., & Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, 42(4), 805-820. doi: Doi 10.2307/2530695
- Jones, A. P., Riley, R. D., Williamson, P. R., & Whitehead, A. (2009). Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. *Clin Trials*, 6(1), 16-27. doi: 10.1177/1740774508100984
- Karim, M. R., & Zeger, S. L. (1992). Generalized linear models with random effects; Salamander mating revisited. *Biometrics*, 48(2), 631. doi: 10.2307/2532317
- Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4), 963-974. doi: Doi 10.2307/2529876
- Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., & Jones, D. R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in medicine*, 24(15), 2401-2428.
- Lindstrom, M. J., & Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83(404), 1014-1022 %@ 0162-1459.
- Mathew, T., & Nordstrom, K. (2010). Comparison of one-step and two-step meta-analysis models using individual patient data. *Biom J*, 52(2), 271-287. doi: 10.1002/bimj.200900143
- McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92(437), 162. doi: 10.2307/2291460
- McCulloch, C. E., & Neuhaus, J. M. (2001). *Generalized Linear Mixed Models*: Wiley Online Library.
- Moore, T. J., Svetkey, L. P., Karanja, N., & Jenkins, M. (2011). *The DASH diet for hypertension: lower your blood pressure in 14 days--without drugs*. Pocket Books.



- Raykov, T., & Marcoulides, G. A. (2015). Intraclass correlation coefficients in hierarchical design studies with discrete response variables: A note on a direct interval estimation procedure. *Educational and psychological measurement*, 75(6), 1063-1070.
- Rhodes, K. M., Turner, R. M., & Higgins, J. P. (2015). Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *Journal of clinical epidemiology*, 68(1), 52-60.
- Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *Bmj*, 340, c221. doi: 10.1136/bmj.c221
- Riley, R. D., Lambert, P. C., Staessen, J. A., Wang, J., Gueyffier, F., Thijs, L., & Bouillon-Buonafina, F. (2008). Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med*, 27(11), 1870-1893. doi: 10.1002/sim.3165
- Riley, R. D., Price, M. J., Jackson, D., Wardle, M., Gueyffier, F., Wang, J., White, I. R. (2015). Multivariate meta-analysis using individual participant data. *Res Synth Methods*, 6(2), 157-174. doi: 10.1002/jrsm.1129
- Rodriguez, G., & Elo, I. (2003). Intra-class correlation in random-effects models for binary data. *The Stata Journal*, 3(1), 32-46.
- Rondeau, V., Michiels, S., Liqueur, B., & Pignon, J. P. (2008). Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Stat Med*, 27(11), 1894-1910. doi: 10.1002/sim.3161
- Sacks, F. M., Svetkey, L. P., Vollmer, W. M., Appel, L. J., Bray, G. A., Harsha, D., ... & Karanja, N. (2001). Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. *New England journal of medicine*, 344(1), 3-10.
- Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika*, 6(5), 309-316.

- Simmonds, M. C., Higgins, J. P., & Stewart, L. A. (2013). Random-effects meta-analysis of time-to-event data using the expectation-maximisation algorithm and shrinkage estimators. *Res Synth Methods*, 4(2), 144-155. doi: 10.1002/jrsm.1067
- Simmonds, M. C., Tierney, J., Bowden, J., & Higgins, J. P. (2011). Meta-analysis of time-to-event data: a comparison of two-stage methods. *Res Synth Methods*, 2(3), 139-149. doi: 10.1002/jrsm.44
- Smith, C. T., Williamson, P. R., & Marson, A. G. (2005). Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Stat Med*, 24(9), 1307-1319. doi: 10.1002/sim.2050
- Trials of Hypertension Prevention Collaborative Research Group. (1997). Effects of weight loss and sodium reduction intervention on blood pressure and hypertension incidence in overweight people with high normal blood pressure: the Trials of Hypertension Prevention, Phase II. *Arch. Intern. Med.*, 157, 657-667.
- Trikalinos, T. A., & Olkin, I. (2012). Meta-analysis of effect sizes reported at multiple time points: a multivariate approach. *Clin Trials*, 9(5), 610-620. doi: 10.1177/1740774512453218
- Tuerlinckx, F., Rijmen, F., Verbeke, G., & De Boeck, P. (2006). Statistical inference in generalized linear mixed models: a review. *Br J Math Stat Psychol*, 59(Pt 2), 225-255. doi: 10.1348/000711005X79857
- Turner, R. M., Omar, R. Z., Yang, M., Goldstein, H., & Thompson, S. G. (2000). A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat Med*, 19(24), 3417-3432.
- Turner, R. M., Davey, J., Clarke, M. J., Thompson, S. G., & Higgins, J. P. (2012). Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International journal of epidemiology*, 41(3), 818-827.

- Vonesh, E., & Chinchilli, V. M. (1996). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*: CRC Press.
- Wei, Y., & Higgins, J. P. (2013). Bayesian multivariate meta-analysis with multiple outcomes. *Statistics in Medicine*, 32(17), 2911-2934.
- Whitehead, A., Omar, R. Z., Higgins, J. P., Savaluny, E., Turner, R. M., & Thompson, S. G. (2001). Meta-analysis of ordinal outcomes using individual patient data. *Stat Med*, 20(15), 2243-2260. doi: 10.1002/sim.919
- Wolfinger, R., & O'Connell, M. (1993). Generalized linear mixed models a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, 48(3-4), 233-243. doi: 10.1080/00949659308811554
- World Health Organization. (2015). *Q&A on Hypertension*. Retrived from <https://www.who.int/features/qa/82/en/>

## VITA

### Ying Zhang

#### EDUCATION

- Ph.D. in Biostatistics August 2019  
PENNSYLVANIA STATE UNIVERSITY, Hershey, PA
- M.S. in Biostatistics May 2015  
DUKE UNIVERSITY, Durham, NC
- B.S. in Biostatistics June 2013  
SOUTHERN MEDICAL UNIVERSITY, Guangzhou, GD, P.R. China

#### PUBLICATIONS

- Zaorsky, N. G., **Zhang, Y.**, Tuanquin L., Bluethmann, S. M., Park, H. S., & Chinchilli, V. M. (2018). Suicide Among Cancer Patients. *Nature Communications*. (Accepted)
- Kokolus, K. M., **Zhang, Y.**, Sivik, J. M., Schmeck, C., Zhu, J., Repasky, E. A., ... & Schell, T. D. (2018). Beta blocker use correlates with better overall survival in metastatic melanoma patients and improves the efficacy of immunotherapies in mice. *OncImmunity*, 7(3), e1405205.
- Stinchcombe, T. E., **Zhang, Y. (co-first author)**, Vokes, E. E., Schiller, J. H., Bradley, J. D., Kelly, K., ... & Govindan, R. (2017). Pooled Analysis of Individual Patient Data on Concurrent Chemoradiotherapy for Stage III Non–Small-Cell Lung Cancer in Elderly Patients Compared with Younger Patients Who Participated in US National Cancer Institute Cooperative Group Studies. *Journal of Clinical Oncology*, 2017 May 11: JCO-2016.
- Pang, H. H., Wang, X., Stinchcombe, T. E., Wong, M. L., Cheng, P., Ganti, A. K., Sargent, D.J., **Zhang, Y.**, ... & Redman, M. W. (2016). Enrollment Trends and Disparity Among Patients with Lung Cancer in National Clinical Trials, 1990 to 2012. *Journal of Clinical Oncology*, 34(33), 3992-3999.
- Kuder, C. H., Weivoda, M. M., **Zhang, Y.**, Zhu, J., ... & Hohl, R. J. (2015). 3-Deoxyschweinfurthin B Lowers Cholesterol Levels by Decreasing Synthesis and Increasing Export in Cultured Cancer Cell Lines. *Lipids*, 50(12), 1195-1207.
- Wang, X., Gu, L., **Zhang, Y.**, Sargent, D. J., Richards, W., Ganti, A. K., ... & Pang, H. (2015). Validation of Survival Prognostic Models for Non-small-cell Lung Cancer in Stage-and Age-Specific Groups. *Lung Cancer*, 90(2), 281-287.