ANALYSIS OF MIXED MODELS FOR BINARY LONGITUDINAL DATA, WITH APPLICATION TO PRESCRIPTION SWITCHING PATTERNS

by

Dong Lin

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science

at

Dalhousie University
Halifax, Nova Scotia
April 2016

To Peter, Ethan and Elise

Table of Contents

List of Tables

List of Figures

Abstract

This thesis aims to study the statin use patterns of the Nova Scotia seniors population and the patients' adherence to medication by applying a generalized linear mixed effect model (abbreviated as GLMM).

Observations for a single subject will include the initial prescription and the sequence of transitions. The data can be modeled as short binary series, with transition probabilities allowed to vary by subject. In this thesis, 10 sets of parameter values were run and the results were compared using tables and box plots. Mean Squared Error (MSE) and Estimated Bias (EB) are calculated to measure how close the estimated parameters are to the true values. For each parameter set, 10 and 100 simulations were run. We can make the conclusion that the generalized linear mixed effect model works well in the application of medication use patterns and the two separate GLMM models make sense.

List of Abbreviations and Symbols Used

CNODES	Canadian Network for Observational Drug Effect Studies
DSEN	Drug Safety and Effectiveness Network
EB	Estimated Bias
GLMM	Generalized Linear Mixed Model
HDL	High Density Lipoprotein
LDL	Low Density Lipoprotein
MSE	Mean Squared Error

Acknowledgements

I would first like to thank my thesis supervisor Dr. Bruce Smith, faculty of Department of Mathematics and Statistics at Dalhousie University. He was always willing to help and steer me in the right direction when I was in need. At the same time he gave me enough independence to produce my own work. This thesis would not be possible without him.

I would also like to thank my husband for his support during the course of my Master program. I am very thankful to our two lovely kids Ethan and Elise. They keep inspiring me and motivating me along the way.

I would also like to ackowledge my reading committee for their valuable comments on this thesis.

Chapter 1

Introduction

1.1 Background on Statin Use

Cholesterol is essential chemical for animal life. It is required for the structural integrity of animal cell membranes, and it is a precursor of the production of many other molecules with important biological functions, including hormones, vitamin D, and bile acids.

About half of the body's cholesterol is synthesized from other building blocks, with about 10% of the daily production originating in the liver and another 15% in the intestines. (http://themedicalbiochemistrypage.org/cholesterol.php).

Most of the cholesterol transport in the body is via lipoproteins, which are complexes of lipids, including cholesterol, and proteins. The liver is the site of much of lipoprotein metabolism. Among other things, the liver packages cholesterol with protein forming low density lipoprotein (LDL), which is the primarly vehicle for transporting cholesterol to other areas of the body. At other locations throughout the body, excess cholesterol is packaged with proteins to form high density lipoprotein (HDL), which is transported to the liver, where the cholesterol is removed from circulation.

While cholesterol is essential for many biological functions, too much cholesterol is problematic, often leading to atheroslerosis, a buildup of cholesterol and other deposits within the artery wall, leading to reduced blood flow, and possibly serious complications such as heart disease or stroke.

LDL, as the vehicle which transports cholesterol from the liver to other areas of the body, is often referred to as "bad cholesterol", while HDL, the means of cholesterol transport away from cells, for recycling in the liver, is known as "good cholesterol".

Too high of a concentration of "bad cholesterol" is a primary risk factor for cardiovasular disease, and medication is commonly prescribed to lower the concentration of

LDL. Recent Canadian guidelines for the diagnosis and treatment of lipid imbalance as it pertains to cardiovascular disease are presented by Anderson (*et al*) [1].

Statins are a class of drugs used to lower LDL cholesterol levels by inhibiting the enzyme HMG-CoA reductase, which plays a central role in the production of cholesterol in the liver (https://en.wikipedia.org/wiki/Statin). As high LDL cholesterol levels have been associated with cardiovascular disease, statin medications are prescribed for primary prevention and secondary prevention of heart disease.

As with many pharmaceuticals, there are many possible side effects associated with statin use, some minor and some serious, including muscle pain and damage, liver damage, neurological effects, and increased risk of type 2 diabetes (http:// www.mayoclinic.org/ statin-side-effects/ art-20046013). Some of the adverse events occur with high frequency. For example, muscle adverse effects occurring in up to thirty percent of patients. In Canada, there are currently 6 statins on the market. They are atorvastatin, rosuvastatin, simvastatin, fluvastatin, lovastatin and pravastatin (cervistatin was removed from the market). The statins vary on their effect on low density lipoproteins, pharmacokinetic properties (the type of liver metabolizing enzyme, the percentage of renal excretion, and the half-life) as well as lipophilicity and adverse event profile.

The frequency with which statins are prescribed is due to both the prevalence of cardivascular disease, high degree of efficacy of statins in reducing cardiovascular risk, primarily through reduction in LDL cholesterol. Atorvistatin, the most commonly prescribed statin, is, in fact, the most widely prescribed pharmaceutial in history.

Experience is that individuals who are prescribed statins are rarely completely taken off the drugs by a physician's recommendation, and statin therapy is generally "for life" (http://www.nhs.uk/ conditions/ Cholesterol-lowering-medicines-statins/ Pages/ Introduction.aspx). However, in addition to statin use, there are other means to reduce LDL concentration, including exercise, weight loss, and dietary changes, and once LDL concentration has been reduced to low levels, then continued use at high dosage may present an unnecessary risk of adverse events. On the other hand if low dosage statin use is unsuccessful at controlling LDL level, then an increase in dosage and/or a change in statin type may be recommended. Thus statin usage patterns are of interest.

In [20], the various medication use patterns were investigated in the study, including discontinuation, restart, switch and adherence. Patients with discontinuation or nonpersistency to index medication were defined as having a gap of at least 60 days for the index medication or at least 60 days without medication before the last day of the study period for that patient. Restart was defined as the patients who refilled the index medication following discontinuation. "Switch without gap" categorized patients that overlapped or initiated a new medication within 60 days of the last index drug filled. "Switch with gap" was defined as initiating the new medication at least 60 days after the discontinuation of the index medication. Medication adherence was indicated by medication possession ratio (MPR) or proportion of days covered (PDC). The MPR was calculated as the total number of days of index medication supplied divided by the number of days in the specified time interval (360 days). The PDC was calculated as the number of days with any drug on hand divided by the number of days in the specified time interval (360 days).

Statins are classfied as high potency vs low potency. Physicians choose to prescribe a particular statin taking into account patient characteristics, disease severity and comorbidities, cost, clinical practice guidelines and other factors. Physicians may prescribe a switch to another drug of higher potency or higher dose if target LDL level is not met or switch to a lower potency or lower dose to avoid adverse outcomes (e.g. myopathy). A switch to a more hydrophilic statin (rosuvastatin, pravastatin) or of lower potency (e.g. fluvastatin) for patients with adverse effects may be useful.

1.2 Background of CNODES project

CNODES is the Canadian Network for Observational Drug Effect Studies. It is part of the Drug Safety and Effectiveness Network (DSEN), a joint initiative of Health Canada and the Canadian Institutes of Health Research. The principal aim of CNODES is to use collaborative, population-based approaches to obtain rapid answers to questions about drug safety and effectiveness.

The research of this thesis was partially funded by a grant from CNODES, with aim of studying statin use patterns of the Nova Scotia seniors population, to determine if there are opportunities to improve the benefit/risk profile of prescribing at

the population level, and the identification of factors which might improve patient adherence.

The statistical methodology used in this thesis is the generalized linear mixed effect model (abbreviated as GLMM).

Bhat *et al* [2] used logistic regression to study statin use, with a goal of indentifying the relationship between gender and statin use. Among 5,508 elderly, 47.2% of the women and 55.5% of the men reported any statin use in 2005, which indicates women were less likely than men to report any use of statins. Less than one third of the total gender difference in statin use was attributed to individual level variables such as demographics, economic status, etc.

The database for this CNODES project includes information for Nova Scotia Seniors above age 65 who are Pharmacare beneficiaries and have been prescribed statins. The study population will include patients 66 years and older who received at least one new dispensing of an eligible statin medication between April 1, 2002 and March 31, 2013. The CNODES funded analysis will be stratified by high and low dose statins, where high dose statins are defined as those which are estimated to reduce LDL cholesterol by $\geq$2 mmol/l on average ($\geq$ 40% LDL reduction). Those high dose statins are rosuvastatin $\geq 10mg$, atorvastatin $\geq 20mg$, and simvastatin $\geq 40mg$. Lower dose statins will be defined as those which are estimated to reduce LDL cholesterol by $< 2mmol/l$ on average ($<$ 40% LDL reduction).Those statins were rosuvastatin $\leq 5mg$, atorvastatin $< 20mg$, all doses of pravastatin, all doses of fluvastatin, and simvastatin $< 40mg$.

After classifying statin dosage as high or low intensity, a patient's statin usage pattern forms a binary sequence beginning with 1 where the initial prescription was high intensity, otherwise 0. In this thesis a generalized linear mixed model is proposed for the analysis of the statin use pattern, with variation among patients accomodated through the introduction of random effect terms.

Due to the confidentiality issues with the CNODES data, the analysis here is based on simulated data, with the simulation set to mimic some descriptive statistics of the CNODES data.

The layout of the remainder of this thesis is as follows. Chapter 2 provides a summary of the generalized linear model and statistical methodology, and a brief

introduction to the generalized mixed effect model. In chapter 3 the proposed mixed effect model is described, together with the associated likelihood, and a brief discussion of methods to maximize the likelihood. A small simulation study is carried out to assess the potential to precisely estimate model parameters using the approximate quantity of data available to the CNODES study. Chapter 4 provides some brief discussions about future work.

Chapter 2

Method

2.1 Linear Models

The standard linear regression models are designed to model the relationship between a continuous variable y and p explanatory variables $x_1, \dots x_p$. Denote y_i the ith observation of the response variable and $x_i = (x_{i1}, \dots, x_{ip})^T$ the vector of associated explanatory variables. The standard linear regression models can be specified as

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i \tag{2.1}$$

with $i = 1, \dots, n$. Using matrix notations, the standard linear regression can be equivalently defined as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.2}$$

where

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ x_{21} & \dots & x_{2p} \\ & \ddots & \\ x_{n1} & \dots & x_{np} \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \sim N(\mathbf{0}, \mathbf{R})$$

The coefficient vector $\boldsymbol{\beta}$ is assumed to be a fixed quantity rather than a random variable, so linear regression can be seen as a special case of fixed effect models. Another assumption of OLS is that y_i are independently sampled from a normal distribution with constant variance. This implies the variance-covariance matrix R is constant diagonal matrix.

$$\mathbf{R} = \begin{pmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \ddots & \\ & & & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{I}$$

2.2 Generalized Linear Models

The basic framework of generalized linear model was first laid out by Nelder and Wedderburn in [16]. In their original work, linear regression was extended to model data where the distribution of response variables belongs to the *exponential family*. Many distributions commonly used in practice, such as Normal, Binomial and Poisson, are special cases of the exponential family. Their work was further extended by Wedderburn in [18] to incorporate a much wider class of distributions beyond the exponential family. Their idea was to use *the quasi-likelihood* instead of the log-likelihood in model estimation. Unlike the log-likelihood function which requires a fully specified distribution, the quasi-likelihood only requires a specification of the relationship between mean and variance through a *variance function*. This allows a more flexible modeling of the data which exhibits a greater or smaller variability than expected under a known distributinon. This is called *overdispersion* and *underdispersion* respectively.

2.2.1 Model Specification

The basic formulation of the Generalized Linear Model (GLM) consists of the following components:

- Distribution: The distribution of response variable $\mathbf{y}$ belongs to the *exponential family*.
- Link function and linear predictor: The expectation of response $\mathbf{y}$ is associated with a linear predictor $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ via

  $$\boldsymbol{\eta} = g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$$

  where $g(\cdot)$ is called the *link function* and $\boldsymbol{\mu} = \mathbb{E}(\mathbf{y})$.

Distribution

The probability density functions in the exponential family share the same format:

$$f\left(y;\boldsymbol{\theta}\right) = h(y)\exp\left(s(\boldsymbol{\theta})' \cdot T(\mathbf{y}) - A(\boldsymbol{\theta})\right) \tag{2.3}$$

where

- $\boldsymbol{\theta}$ is the parameter of interest and is sometimes referred to as the *mean parameter* (which is not necessarily the mean).
- $T(y)$ is a *sufficient statistic* of the distribution.
- $A(\boldsymbol{\theta})$ is called the *log-partition function*, which is a normalization term so that $f(y;\boldsymbol{\theta})$ is a valid density.
- $h(y)$ is called the *base measure*.

Sometimes, it is useful to reparameterize the density by letting $\boldsymbol{\theta} = s(\boldsymbol{\theta})$ and we arrive to an equivalent density

$$(y;\boldsymbol{\theta}) = h(y)\exp\left(\boldsymbol{\theta}' \cdot T(y) - B(\boldsymbol{\theta})\right) \tag{2.4}$$

The distributions specified in the format of (2.4) are said to be in *canonical form*. $\boldsymbol{\theta}$ is called the *canonical parameter* or *natural parameter*. A lot of distributions commonly used in practice belong to the exponential family, such as normal, Poisson and exponential etc.

Equations (2.3) and (2.4), provide a very general definition of the exponential family. If the parameter $\boldsymbol{\theta}$ is a s-dimensional vector, the associated distribution is said to belong to s-dimensional exponential family. In the context of GLM, we will focus on 1 and 2-dimensional exponential families. Using the same formulation as in [15], it is more intuitive and convenient to write the density as

$$(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right) \tag{2.5}$$

where a, b and c are unknown functions. If ϕ is given, (2.5) reduces to 1-dimensional exponential family with natural parameter θ. If ϕ is unknown, (2.5) is generally not an exponential family distribution. However, when $c(y,\phi) = d(y) + e(\phi)$, (2.5) defines an exponential model, which includes normal, inverse normal and gamma as special cases.

The mean and variance of the exponential family can be shown to be

$$\mathbb{E}(y) = b'(\theta) \quad \text{and} \quad \mathbb{V}(y) = b''(\theta)a(\phi) \tag{2.6}$$

where θ is the (*canonical*) *location parameter* and ϕ is the *scale* or *dispersion* parameter. In most cases, $a(\phi)$ will simply be ϕ and we have

$$\mathbb{E}(y) = b'(\theta) \quad \text{and} \quad \mathbb{V}(y) = \phi b''(\theta)$$

Link

The link function defines the relationship between $\boldsymbol{\mu}$ and the systematic components $\mathbf{X}\boldsymbol{\beta}$ as

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} \tag{2.7}$$

Note the right hand side of equation (2.7), the systematic components $\mathbf{X}\boldsymbol{\beta}$ can potentially take on any real values due to its linear form. The necessity of link function is not so obvious in OLS, in which μ ranges in $(-\infty, \infty)$. We can thus directly model μ with an identity link function, where both sides of (2.7) have the same support. However for counting data e.g a Poisson distribution, μ is always positive and one needs to find an appropriate link function $g(\cdot)$ such that $g(\mu)$ ranges in $(-\infty, \infty)$. One choice is to use $g(\mu) = \log \mu$. Another attractiveness of the log link is to enable a multiplicative effect on μ, which is often observed in counting data.

It should be noted that the link function is not unique. Different link functions will result in different results and interpretations. The choice of a particular link function often depends on the data and the scientific question to be answered. There is no prior reason to prefer a canonical link function over the others.

2.2.2 Effect Estimation

Given n observations $\mathbf{y} = (y_1, \dots y_n)'$ from an exponential family distribution, the loglikelihood can be shown to be equal to

$$l(\boldsymbol{\theta}) = \mathbf{y}'\mathbf{A}\boldsymbol{\theta} - \mathbf{1}'\mathbf{A}b(\boldsymbol{\theta}) + c(\mathbf{y}, \phi) \tag{2.8}$$

where $\mathbf{A} = diag[1/a(\phi_i)]$ and the goal is to make inference on $\boldsymbol{\beta}$. Note that the term $c(\mathbf{y}, \phi)$ in (2.8) doesn't involve $\boldsymbol{\beta}$. Although a fully specified model requires $c(\mathbf{y}, \phi)$, it is not used in the inference of $\boldsymbol{\beta}$. The specification of $\frac{y\theta - b(\theta)}{a(\phi)}$ is sufficient, which is closely related to the mean and variance (2.6). This observation motivated the theory of quasi-likelihood, in which Wedderburn showed the statistical inference of GLM for exponential family remains valid when only the first two moments are specified.

The MLE of (2.8) is commonly solved by iteratively reweighted least squares

methods, using either Newton-Raphson or Fisher scoring. The basic idea is to approximate the full likelihood using the 2nd order Taylor series expansion:

$$l\left(\boldsymbol{\theta}\right) \approx l\left(\hat{\boldsymbol{\theta}}\right) + \left.\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|'_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right) + \frac{1}{2}\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)' \left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)$$

Let $\mathbf{S} = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ be the score function and $\mathbf{H} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ be the Hessian matrix. We can re-write the above equation as

$$l\left(\boldsymbol{\theta}\right) \approx l\left(\hat{\boldsymbol{\theta}}\right) + \mathbf{S}(\hat{\boldsymbol{\theta}})' \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right) + \frac{1}{2}\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)' \mathbf{H}(\hat{\boldsymbol{\theta}}) \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)$$

Letting $\partial l\left(\boldsymbol{\theta}\right) / \partial \boldsymbol{\theta} = 0$ gives

$$\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} - \mathbf{S}(\hat{\boldsymbol{\theta}})' \mathbf{H}(\hat{\boldsymbol{\theta}})^{-1} \tag{2.9}$$

which is the basic form of Newton-Raphson method. The Fisher scoring algorithm replaces the Hessian matrix by its expectation, called information matrix

$$\mathbb{I}(\boldsymbol{\theta}) = \mathbb{E}\left[-\mathbf{H}(\boldsymbol{\theta})\right] = \mathrm{Var}[\mathbf{s}(\boldsymbol{\theta})]$$

2.3 Linear Mixed Effect Models

2.3.1 Model Specification

Linear mixed effect models (LMM) extend the standard linear regression models in two important ways. First, LMM introduce another type of effects $\mathbf{b}$ and assume $\mathbf{b}$ is sampled from a normal distribution rather than being constant. Furthermore, LMM releases the constraint of i.i.d. sampling and allows y_i to be correlated.

Linear mixed effect models can be formally defined as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon} \tag{2.10}$$

where $\mathbf{b} \sim N(\mathbf{0}, \mathbf{G})$ and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R})$. The covariance matrix $\mathbf{V}$ of $\mathbf{Y}$ is then

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}.$$

Note $\mathbf{V}$ is expressed as a linear combination of $\mathbf{G}$ and $\mathbf{R}$. It is possible to find two different pairs of $\mathbf{G}$ and $\mathbf{R}$ to derive the same $\mathbf{V}$. In other words, LMM are over-parameterized and there is potentially a model identification problem.

The unknown parameters in the linear mixed effect model (2.10) include fixed effect $\boldsymbol{\beta}$, random effect $\mathbf{b}$, variance components $\mathbf{V}$. Note the maximal possible number of parameters allowed in $\mathbf{V}$ is $\frac{q(q+1)}{2} + \frac{n(n+1)}{2}$, where q is the number of random effects. In practice, $\mathbf{V}$ often only depends on a few parameters $\boldsymbol{\theta}$.

Although the joint likelihood of $(\boldsymbol{\beta}, \mathbf{b}, \mathbf{G}, \mathbf{R})$ has a straightforward expression, the simultaneous estimation of all unknown parameters is often too difficult to be practical. Instead, parameters are split into two groups: the effects $(\boldsymbol{\beta}, \mathbf{b})$ and the covariance components $(\mathbf{G}, \mathbf{R})$. At each step, the estimation is performed separately on each group (conditional on the values from the other group). The estimated parameters are then used for the inference in the other group. The whole estimation is iteratively performed until convergence.

In the following, we will describe methods for estimating $(\boldsymbol{\beta}, \mathbf{b})$ and $(\mathbf{G}, \mathbf{R})$.

2.3.2 Effect Estimation

The standard method to estimate $(\boldsymbol{\beta}, \mathbf{b})$ is to solve a set of mixed model equations, as proposed by C.R Henderson [8, 10]. First, note the joint distribution of $\mathbf{b}$ and $\boldsymbol{\epsilon}$ is normal with mean $\mathbf{0}$ and covariance matrix

$$\mathbf{J} = \begin{pmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix}$$

The likelihood is defined as

$$L(\boldsymbol{\beta}, \mathbf{b}) = (2\pi)^{-\frac{n+q}{2}} |\mathbf{J}|^{-\frac{1}{2}} \exp -\frac{1}{2} \left([\mathbf{b}', \boldsymbol{\epsilon}'] \mathbf{J}^{-1} \begin{bmatrix} \mathbf{b} \\ \boldsymbol{\epsilon} \end{bmatrix} \right) \tag{2.11}$$

Here we assume $\mathbf{G}$ and $\mathbf{R}$ are given. Maximization of (2.11) is equivalent to minimizing its exponent

$$\begin{aligned} \mathbf{C} = [\mathbf{b}', \boldsymbol{\epsilon}'] \mathbf{J}^{-1} \begin{bmatrix} \mathbf{b} \\ \boldsymbol{\epsilon} \end{bmatrix} &= [\mathbf{b}', \boldsymbol{\epsilon}'] \begin{pmatrix} \mathbf{G}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}^{-1} \end{pmatrix} \begin{bmatrix} \mathbf{b} \\ \boldsymbol{\epsilon} \end{bmatrix} \\ &= \mathbf{b}' \mathbf{G}^{-1} \mathbf{b} + \boldsymbol{\epsilon}' \mathbf{R}^{-1} \boldsymbol{\epsilon} \\ &= \mathbf{b}' \mathbf{G}^{-1} \mathbf{b} + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) \end{aligned}$$

In order to find the minimizer of $\mathbf{C}$, we need to solve the following two equations:

$$\begin{aligned} & \frac{\partial \mathbf{C}}{\partial \boldsymbol{\beta}} = \mathbf{0} \\ \Longleftrightarrow\ & 2\mathbf{X}'\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}\right) = \mathbf{0} \\ \Longleftrightarrow\ & \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{b} = \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \end{aligned}$$

and

$$\begin{aligned} & \frac{\partial \mathbf{C}}{\partial \mathbf{b}} = \mathbf{0} \\ \Longleftrightarrow\ & 2\mathbf{G}^{-1}\mathbf{b} - 2\mathbf{Z}'\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}\right) = \mathbf{0} \\ \Longleftrightarrow\ & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta} + \left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\right)\mathbf{b} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{aligned}$$

Re-arranging the above two equations in matrix form leads to the *mixed model equations*:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \tag{2.12}$$

The solution of (2.12) is

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\ \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right) \end{bmatrix} \tag{2.13}$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{b}}$ were shown [9] to be BLUE (*Best Linear Unbiased Estimate*) and BLUP (*Best Linear Unbiased Predictor*). Note $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{b}}$ are obtained assuming $\mathbf{G}$ and $\mathbf{R}$ are known. In practise, they are often replaced with $\hat{\mathbf{G}}$ and $\hat{\mathbf{R}}$ which are the estimated covariance matrices from data. In this case, $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{b}}$ are called EBLUE (*Estimated Best Linear Unbiased Estimator*) and EBLUP (*Estimated Best Linear Unbiased Predictor*).

2.3.3 Covariance Components Estimation

Different methods have been proposed to estimate covariance components $\mathbf{V}(\boldsymbol{\theta})$, including ML (Maximum Likelihood), REML(Residual/restricted Maximum Likelihood), MIVQUE(minimum variance quadratic unbiased estimation) etc. Among them, the REML approach is arguably the best. In this section, we will review methods based on ML and REML.

The likelihood can be obtained directly from (2.10) using the fact that $\mathbf{y}$ follows a normal distribution with covariance $\mathbf{V}(\boldsymbol{\theta})$. Here we write $\boldsymbol{\theta}$ in parentheses following $\mathbf{V}$ to emphasize the dependence of $\mathbf{V}$ on $\boldsymbol{\theta}$.

Instead of maximizing the likelihood directly, in practice the -2 log likelihood

$$-2l\left(\boldsymbol{\theta}\right)=\log\left|\mathbf{V}(\boldsymbol{\theta})\right|+\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)^{'}\mathbf{V}(\boldsymbol{\theta})^{-1}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right) \tag{2.14}$$

is minimized where $\hat{\boldsymbol{\beta}}$ is estimated from (2.13).

Maximum likelihood estimator has several nice limiting statistical properties, such as consistency and asymptotic normality as sample size increases to infinity. However, it is well known that MLE can yield biased estimate in certain cases. As a simple example, suppose we have n observations $y_1,\ldots,y_n$ sampled from a normal distribution $N\left(\mu,\sigma^2\right)$ with both μ and σ^2 unknown.

It can be shown that the maximum likelihood estimators are

$$\hat{\mu}=\frac{\sum_{i=1}^{n}y_n}{n}$$
$$\hat{\sigma}^2=\frac{\sum_{i=1}^{n}\left(y_i-\hat{\mu}\right)^2}{n}$$

with

$$\mathbb{E}(\hat{\mu})=\mu$$
$$\mathbb{E}(\hat{\sigma}^2)=\frac{n-1}{n}\sigma^2$$

In this case, $\hat{\sigma}^2$ underestimates σ^2 on average and the biasedness vanishes as sample size increases.

The reason of biasedness is transparent by noticing $\hat{\mu}$ is used in replacement of true mean μ in $\hat{\sigma}^2$. If μ is known, $\hat{\sigma}^2$ becomes $\sum_{i=1}^{n}\left(y_i-\mu\right)^2/n$ with $\mathbb{E}(\hat{\sigma}^2|\mu)=\sigma^2$. In other words, the biasness comes from using $\hat{\mu}^2$ as the true mean and treating it as a constant without accounting for randomness in $\hat{\sigma}^2$. Since $\hat{\mu}$ is calculated from the data, it is a better representation of the sample than μ in such a way that the total deviation of sampled observations from the mean is minimal. This is why $\hat{\sigma}^2$ in general underestimates σ^2.

The biasedness of MLE is ignorable with a relatively large sample size but becomes troublesome in small samples. REML (Restricted/Residual Maximum Likelihood) was designed to remedy this challenge. Instead of working on the log likelihood of $\mathbf{y}\sim N(\mathbf{X}\boldsymbol{\beta},\mathbf{V}(\boldsymbol{\theta}))$ directly, REML finds a transformation matrix $\mathbf{K}$ such

that $\mathbb{E}(\mathbf{K}'\mathbf{y}) = 0$ and $\mathbf{K}'\mathbf{y} \sim N(\mathbf{0}, \mathbf{K}'\mathbf{V}(\boldsymbol{\theta})\mathbf{K})$. Let $\mathbf{V}_R = \mathbf{K}'\mathbf{V}(\boldsymbol{\theta})\mathbf{K}$ and the -2 log likelihood of $\mathbf{K}'\mathbf{y}$ is

$$-2l_R(\boldsymbol{\theta}) = \log|\mathbf{V}_R(\boldsymbol{\theta})| + \mathbf{y}'\mathbf{K}\mathbf{V}_R(\boldsymbol{\theta})^{-1}\mathbf{K}'\mathbf{y} \tag{2.15}$$

which removes the fixed effects $\boldsymbol{\beta}$ from its expression and the MLE of (2.15) becomes unbiased. Harville [7] suggested using the $n - p$ linearly independently rows of $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{1}\mathbf{X}'$ as $\mathbf{K}'$, where p is the rank of $\mathbf{X}$. The -2 log likelihood can be simplied as

$$-2l_R(\boldsymbol{\theta}) = \log|\mathbf{V}(\boldsymbol{\theta})| + \log\left|\mathbf{X}'\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{X}\right| + \mathbf{r}'\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{r} \tag{2.16}$$

where $\mathbf{r} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ is the residual.

In practice, the inference of linear mixed effects models is performed iteratively by minimizing REML (2.16) using the Newton-Raphson algorithm and solving for the fixed effects from (2.13). Lindstrom and Bates [13] provided good explanations for preferring Newton-Raphson to EM based approaches [6, 12].

2.4 Generalized Linear Mixed Effect Models

2.4.1 Model Specification

Generalized Linear Mixed Effect Models (GLMM) can be considered as a hybrid of GLM and LMM and their definition is a straightforward extension of the GLM.

$$g(\mathbb{E}[\mathbf{y}|\mathbf{b}]) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} \tag{2.17}$$

with

$$\mathbf{b} \sim N(\mathbf{0}, \mathbf{G}).$$

The conditional distribution of $\mathbf{y}|\mathbf{b}$ belongs to the exponential family and g is the link function. Conditional on the random effects, GLMM is essentially the same as GLM, but the GLMM is typically much more difficult to deal with computationally due to the random effects.

The GLMM used to model the statin switching data is described in detail in the following chapter.

2.4.2 Parameter Estimation

The likelihood of a GLMM is

$$L(\boldsymbol{\theta}) = \int_{\mathbf{b}} f(\mathbf{y}|\mathbf{b}) f(\mathbf{b}) d\mathbf{b} \tag{2.18}$$

When both random effects and the conditional distribution of $\mathbf{y}|\mathbf{b}$ are normal, the marginal distribution is normal and (2.17) reduces to a linear mixed model. When the conditional distribution of $\mathbf{y}|\mathbf{b}$ is not normal, the marginal distribution generally doesn't have a closed form and statistical inference can be challenging. The difficulty is that the integral in (2.18) must be approximated.

The commonly used inference methods for GLMM fall into two categories. One group of methods are based on *linearization* of the mean function, which approximate $g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b})$ as a linear function of $\boldsymbol{\beta}$ and $\mathbf{b}$ using a Taylor series expansion. The standard estimation method for linear mixed models can then be applied to solve for $\boldsymbol{\beta}$. The most well known linearization methods include penalized quasi-likelihood (PQL) [4] and pseudo-likelihood (PL) or restricted pseudo-likelihood [19]. However, it has been shown that these two methods can produce significant bias for binomial GLMM estimates when the number of observations within each cluster is small [5].

The second type of methods directly optimize the likelihood (2.18) using numerical integration techniques such as adaptive Gauss quandrature [17], Monte Carlo integration or the EM algorithm [6, 3]. These methods are typically computationally expensive.

2.5 Software Implementation

In SAS, the **glimmix** procedure can be used to fit GLMMs, and several estimation methods are implemented including an adaptive Gauss-Hermite quadrature, Laplace approximation, residual likelihood or a maximum likelihood based pseudo-likelihood techniques.

Several R functions can be used to fit GLMMs, including **nlme()** in the *nlme* package and **glmer()** in the *lme4* package. Both functions are mainly based on Laplace approximation for parameter estimation, with options to use penalized iteratively reweighted least squares or adaptive Gauss-Hermite quadrature in certain simple cases.

In this thesis we have used the **glmer()** function in the R *lme4* package, with default settings, in which the Laplace approximation to the integral is maximized.

Kim *et al* [11] provides a review of methods for approximating the marginal likelihood of a GLMM, including a simulation study to assess the performance of different methods, with a focus on the logistic GLMM. Their conclusion is that SAS GLIMMIX Laplace and SuperMix Gaussian quadrature perform the best among selected packages, including SAS NLMIXED, R lme4, Stata xtmelogit, etc.

Chapter 3

Simulation results

In this chapter a simulation study is carried out to assess the potential to estimate the parameters of a switching model for statin usage.

Due to the confidentiality of the CNODES data, simulated data are used, with the hope that the simulated data will mimic some aspects of the CNODES data. In these data, switches from low to high or high to low intensity statins are rare events. For the simulation, we have assumed here that probabilities of the four transitions (low to low, low to high, high to low and high to high) are, respectively, .763, .014, .01, and .213, which are representative probabilities. The vast majority of transitions are to the same state (high to high or low to low), with only 2.4% being state switches. We have assumed that the number of switches for a subject follows a discretised version of a gamma distribution with mean 50 and standard deviation 9. On average 2.4% of transitions will be from low to high or high to low, so on average, a subject will experience just over 1 switch from low to high or high to low. The 99'th percentile of the gamma distribution is slightly larger than 70, so about 1/100 subjects for whom data are simulated will have 70 or more transitions, with on average, just under 1.7 state changes.

3.1 A generalized linear mixed effect model for statin switching

Observations for a single subject will include the initial prescription, and the sequence of transitions. The state is a binary variable, with 0 indicative of low potency and 1 indicative of high potency, so the data can be modeled as short binary time series. To reflect differences among patients, the transition probabilities are allowed to vary by subject.

Let Z_i denote the prescriber covariates of the physician who prescribed the index statin for subject i, X_i denote covariates associated with the switching probabilities for the i'th subject, and $\tau_i = (\tau_{i0}, \tau_{i1})$ are a pair of random effects for subject i.

Where 1 indicates a high potency statin and 0 indicates a low potency statin, it is assumed that $Y_{i,1}$ the index prescription for subject i depends only on the prescriber characteristics, but not the patient characteristics, and has a Bernoulli distribution with

$$P(Y_{i,1} = 1|Z_i, X_i, \tau_i) = \frac{e^{\alpha+Z_i'\gamma}}{1 + e^{\alpha+Z_i'\gamma}} \tag{3.1}$$

or equivalently

$$logit(P(Y_{i,1} = 1|Z_i, X_i, \tau_i)) = \alpha + Z_i'\gamma \tag{3.2}$$

The four transition probabilities are denoted as P_{00} (from low to low), P_{01} (from low to high), P_{11} (from high to high) and P_{10} (from high to low). We assume two random effects, with the τ_1 a random effect associated with transitions from high intensity to low intensity, and τ_0 a random effect associated with transistions from low to high intensity. Where X_i is the fixed effect covariate vector for subject i and τ_{i1} and τ_{i0} are the random effects for subject i, we assume that

$$p_{i00} = \frac{\exp(\mu_0 + X_i\beta_0 + \tau_{i0})}{1 + \exp(\mu_0 + X_i\beta_0 + \tau_{i0})} \tag{3.3}$$

$$p_{i01} = \frac{1}{1 + \exp(\mu_0 + X_i\beta_0 + \tau_{i0})} \tag{3.4}$$

$$p_{i11} = \frac{\exp(\mu_1 + X_i\beta_1 + \tau_{i1})}{1 + \exp(\mu_1 + X_i\beta_1 + \tau_{i1})} \tag{3.5}$$

$$p_{i10} = \frac{1}{1 + \exp(\mu_1 + X_i\beta_1 + \tau_{i1})} \tag{3.6}$$

Let Y_{ij} be the j'th observation for subject i, which is either 0 (low potency), or 1 (high potency). If subject i has n_i measurements including the index prescription, then under an assumption of conditional independence, and assuming that the fixed effect covariates Z_i are associated only with the physician making the index prescription, and that the fixed effects X_i and that the random effects $\tau_i = (\tau_{i0}, \tau_{i1})$ are associated only with the transitions, the joint probability of $Y_{i1}, Y_{i2}, \ldots, Y_{i,n_i}$ is given by

$$\begin{aligned} P(Y_{i1}, Y_{i2}, \ldots, Y_{i,n_i}|Z_i, X_i, \tau_i) = {} & P(Y_{i,n_i}|Y_{i,n_{i-1}}, X_i, \tau_i)P(Y_{i,n_{i-1}}|Y_{i,n_{i-2}}, X_i, \tau_i) \\ & \ldots P(Y_{i,3}|Y_{i,2}, X_i, \tau_i)P(Y_{i,2}|Y_{i,1}, X_i, \tau_i)P(Y_{i,1}|Z_i) \end{aligned} \tag{3.7}$$

Here the Markov property has been assumed, whereby future states depend only on the current state and not on the sequence of events that precede it. Where n_{ijk} denotes the number of transitions from state j to state k for individual i, and p_{ijk} be the probability of transitions from state j to state k for the individual i, as given in (3.3 - 3.6), the conditional likelihood for subject i is

$$P(Y_{i1}, Y_{i2}, \ldots, Y_{i,n_i}|Z_i, X_i, \tau_i) = P(Y_{i,1}|Z_i) \prod_{j=0}^{1} \prod_{k=0}^{1} p_{ijk}^{n_{ijk}} \tag{3.8}$$

The conditional likelihood over all the I individuals factors into two pieces (the initial prescription and the transitions) as follows (3.9). It is clear that the n_{ijk} are sufficient statistics for the transition model parameters. So the data storage required for the simulation study might be substantially reduced by retaining only the 4 values for each subject.

$$P(Y_1, Y_2, \ldots, Y_n|Z, X, \tau) = \prod_{i=1}^{I} P(Y_{i,1}|Z_i) \prod_{i=1}^{I} \prod_{j=0}^{1} \prod_{k=0}^{1} p_{ijk}^{n_{ijk}} \tag{3.9}$$

The unconditional likelihood is the integral of the conditional likelihood above (3.9) with respect to the random subject effects, so it includes a $2I$ dimensional integral. Because the initial prescriber characteristics do not depend on the subject effects, the piece $\prod_{i=1}^{I} P(Y_{i,1}|Z_i)$ factors out of the integral, and the unconditional likelihood is as follows:

$$\begin{aligned}
P(Y|X, Z) &= \prod_{i=1}^{I} p(Y_{i,1}|Z_i) \int_{i=\tau_1,\ldots,\tau_I} \cdots \int \prod_{i=1}^{I} \prod_{j=0}^{1} \prod_{k=0}^{1} p_{ijk}^{n_{ijk}} \cdot f(\tau) d\tau \\
&= \prod_{i=1}^{I} p(Y_{i,1}|Z_i) \int_{\tau_{10}} p_{100}^{n_{100}} \cdot p_{101}^{n_{101}} \cdot f(\tau_{10}) d\tau_{10} \int_{\tau_{11}} p_{110}^{n_{110}} \cdot p_{111}^{n_{111}} \cdot f(\tau_{11}) d\tau_{11} \\
&\cdots \int_{\tau_{I0}} p_{I00}^{n_{I00}} \cdot p_{I01}^{n_{I01}} \cdot f(\tau_{I0}) d\tau_{I0} \int_{\tau_{I1}} p_{I10}^{n_{I10}} \cdot p_{I11}^{n_{I11}} \cdot f(\tau_{I1}) d\tau_{I1}
\end{aligned}$$

Substituting $P_{00}, P_{01}, P_{10}, P_{11}$ above with the equations of the transition probabilities (3.3–3.6), we get:

$$
\begin{aligned}
P(Y|X,Z) = &\prod_{i=1}^{I} P(Y_{i,1}|Z_i)\cdot \\
&\int_{\tau_{10}} \left[\frac{\exp(\mu_0 + X_1\beta_0 + \tau_{10})}{1+\exp(\mu_0 + X_1\beta_0 + \tau_{10})}\right]^{n_{100}} \cdot \left[\frac{1}{1+\exp(\mu_0 + X_1\beta_0 + \tau_{10})}\right]^{n_{101}} \cdot f(\tau_{10})d\tau_{10} \\
&\cdot \int_{\tau_{11}} \left[\frac{\exp(\mu_1 + X_1\beta_1 + \tau_{11})}{1+\exp(\mu_1 + X_1\beta_1 + \tau_{11})}\right]^{n_{110}} \cdot \left[\frac{1}{1+\exp(\mu_1 + X_1\beta_1 + \tau_{11})}\right]^{n_{111}} \cdot f(\tau_{11})d\tau_{11} \\
&\qquad\qquad \cdots \\
&\int_{\tau_{I0}} \left[\frac{\exp(\mu_0 + X_I\beta_0 + \tau_{I0})}{1+\exp(\mu_0 + X_I\beta_0 + \tau_{I0})}\right]^{n_{I00}} \cdot \left[\frac{1}{1+\exp(\mu_0 + X_I\beta_0 + \tau_{I0})}\right]^{n_{I01}} \cdot f(\tau_{I0})d\tau_{I0} \\
&\cdot \int_{\tau_{I1}} \left[\frac{\exp(\mu_1 + X_I\beta_1 + \tau_{I1})}{1+\exp(\mu_1 + X_I\beta_1 + \tau_{I1})}\right]^{n_{I10}} \cdot \left[\frac{1}{1+\exp(\mu_1 + X_I\beta_1 + \tau_{I1})}\right]^{n_{I11}} \cdot f(\tau_{I1})d\tau_{I1}
\end{aligned}
$$

There are multiple procedures available for fitting GLMM in R. One popular package is lme4, which implements the Gauss-Hermite quadrature to approximate the log-likelihood using numerical integration. The default setting is Laplace approximation when only one quadrature point is used. Another available package is ZELIG, and the third option is the glmmML package. This thesis used the lme4 pacakge to fit the GLMM model.

3.2 A simulation study

The *R* statistical package is used for the simulation, and the code is attached in the appendix A. In the simulation, the initial prescribing state (whether a patient starts with low or high potency) is assumed to follow a Bernoulli distribution.

In terms of the associated random subject effect, the same subject could have different subject effects for switches from high potency drugs and switches from low potency drugs. Our notation is that τ_{i0} indicates the random effect for subject i when switching from low potency drugs, and τ_{i1} is the random effect for subject i when switching from high potency drugs. The random effects are assumed to be i.i.d. according to normal distributions with means 0 and standard deviations σ_{τ_0} and σ_{τ_1}. As above, the fixed effect regression coefficients associated with switches

are different between the groups of starting from state 0 and state 1. They are β_0 and β_1 respectively.

The literature [2] suggests that gender is an important fixed effect for the switching distribution. In the simulation we have used gender as the only fixed effect covariate X, and we have assumed P(male)=.52, P(female)=.48, based on the CNODES data.

The simulation code is included in the Appendix. The "Functions" module generates the output as potency of statins, gender, transition probablity from 0, transition probability from 1 and subject effect. The output produced by the "Functions" module is used as input to the "GLMM-batch" module. The code is written in the way that parameter values can be read from an Excel spreadsheet, so a user can have the great flexibility to control the number and values of parameter sets.

In this thesis, 10 different sets of parameters were run. The simulation is computer intensive, and because of this, data were generated for only 100 subjects. Two sets of results are presented, one with simulation batch size 100, and a second with simulation batch size of 10, with the intent to show the convergence of estimates with increasing batch size. As mentioned above, on average 2.4% of transitions will be from low to high or high to low, and on average a subject will experience just over 1 switch from low to high or high to low.

The simulation study used 10 sets of parameter values as listed in Table 3.1. For each simulation batch and each parameter configuration, data for 100 subjects was simulated, and parameters were estimated using the glmer function in the R library lme4, using default settings for glmer. For a given parameter configuration, one simulated data set consists of 100 subjects, with on average, 50 observations per subject, or 5000 generated data points. The resulting computational requirements are fairly substantial, both in terms of data storage and cpu usage, and for this reason, the number of simulation batches was kept moderately small, at 100.

The simulation parameters are $P(male) = P(X_i = 1)$, μ_0 and β_0 - the fixed effect parameters associated with transitions from state 0, $\sigma^2_{\tau_0}$ - the variance of the random effect associated with transitions from 0; μ_1 and β_1 - the fixed effect parameters associated with transitions from state 1, and $\sigma^2_{\tau_1}$ - the variance of the random effect associated with transitions from 1. We have not used any physician prescribing

covariates for the index prescription, but rather have considered two values for the probability that the index prescription is high intensity, $P(Y_1 = 1) = .223$ and $P(Y_1 = 1) = .530$.

For each of the 10 parameter configurations, the results are divided into 2 groups. Within each group, all of the parameters are fixed, except for the standard deviation of the random subject effect, and in this way, the impact of the subject effects can be seen. In order to reduce the number of displays, the standard deviations of the two random effects, σ_{τ_0} and σ_{τ_1} have been fixed at the same value for a particular parameter configuration.

Table 3.1: Parameters

index	$P(male)$	$P(Y_1 = 1)$	μ_0	β_0	σ_{τ_0}	μ_1	β_1	σ_{τ_1}
1	0.52	0.223	-4	5	0.1	-3	0.02	0.1
2	0.52	0.223	-4	5	1	-3	0.02	1
3	0.52	0.223	-4	5	2	-3	0.02	2
4	0.52	0.223	-4	5	3	-3	0.02	3
5	0.52	0.223	-4	5	4	-3	0.02	4
6	0.52	0.53	-4	0.08	1	-1.5	0.2	1
7	0.52	0.53	-4	0.08	2	-1.5	0.2	2
8	0.52	0.53	-4	0.08	3	-1.5	0.2	3
9	0.52	0.53	-4	0.08	4	-1.5	0.2	4
10	0.52	0.53	-4	0.08	5	-1.5	0.2	5

To assess the accuracy of the parameter estimates, the mean squared errors (MSE) have been calculated for each of the parameter estimates: μ_0 and β_0, the intercept and slope for transitions from the low intensity state, and μ_1 and β_1, the intercept and slope for transitions from the high intensity state. The estimated bias (EB) is also calculated for each parameter, by averaging the differences between the estimate and the true parameter value.

The results are shown in Tables 3.2, 3.3, 3.4 and 3.5, respectively, for transitions

from 0 and the transitions from 1, for 10 vs 100 simulation batches respectively.

We expect that as the number of simulation batches increase by a factor of c then the MSE will decrease by a factor of $1/c$, and that as the standard deviation of the random subject effects increases, the MSE will increase. There is no such expectation regarding the estimated bias.

Table 3.2: MSE and EB for μ_0 and β_0 (10 simulations)

index	μ_0	β_0	σ_{τ_0}	$\hat{MSE} - \mu_0$	$\hat{EB} - \mu_0$	$\hat{MSE} - \beta_0$	$\hat{EB} - \beta_0$
1	-4	5	0.1	0.0302	0.0207	0.0433	-0.0649
2	-4	5	1	0.0698	-0.0960	0.1325	0.1227
3	-4	5	2	0.1805	-0.0470	0.5345	0.2806
4	-4	5	3	0.7628	-0.6097	1.4409	0.8890
5	-4	5	4	1.4475	-0.5399	3.0655	0.8684
6	-4	0.08	1	0.0343	-0.0430	0.0945	0.0560
7	-4	0.08	2	0.0699	-0.0264	0.1161	-0.1815
8	-4	0.08	3	0.2982	-0.0908	0.5004	0.1377
9	-4	0.08	4	1.5096	-0.9020	1.7761	0.2976
10	-4	0.08	5	7.6672	-1.8809	2.1266	0.7930

Table 3.3: MSE and estimated bias for μ_0 and β_0 (100 simulations)

index	μ_0	β_0	σ_{τ_0}	$\hat{MSE} - \mu_0$	$\hat{EB} - \mu_0$	$\hat{MSE} - \beta_0$	$\hat{EB} - \beta_0$
1	-4	5	0.1	0.0474	-0.0239	0.0756	0.0423
2	-4	5	1	0.0576	0.0041	0.1251	0.0351
3	-4	5	2	0.1777	0.0166	0.4043	-0.0015
4	-4	5	3	0.4685	-0.1768	1.1015	0.4110
5	-4	5	4	1.3151	-0.4754	4.1907	0.9364
6	-4	0.08	1	0.0646	-0.0449	0.1027	0.0168
7	-4	0.08	2	0.1612	-0.0704	0.2259	-0.0156
8	-4	0.08	3	0.6278	-0.148	0.6179	0-0.0389
9	-4	0.08	4	1.1300	-0.4343	1.7530	0.2273
10	-4	0.08	5	5.8974	-1.0983	2.6484	-0.0422

Table 3.4: MSE and EB for μ_1 and β_1 (10 simulation batches)

index	μ_1	β_1	σ_{τ_1}	$\hat{MSE} - \mu_1$	$\hat{EB} - \mu_1$	$\hat{MSE} - \beta_1$	$\hat{EB} - \beta_1$
1	-4	5	0.1	0.0563	-0.0066	0.0693	-0.0022
2	-4	5	1	0.0768	-0.0825	0.0960	0.0087
3	-4	5	2	0.3390	-0.2024	0.6036	0.3543
4	-4	5	3	0.7691	-0.5914	1.3066	0.8465
5	-4	5	4	0.3445	-0.2443	0.3127	0.1528
6	-4	0.08	1	0.0429	-0.0284	0.0618	0.0211
7	-4	0.08	2	0.1697	0.0097	0.2327	-0.0527
8	-4	0.08	3	0.3841	0.3032	0.4284	-0.2726
9	-4	0.08	4	0.7260	0.1966	1.9964	-0.4994
10	-4	0.08	5	0.8699	-0.3401	1.0291	0.3868

Table 3.5: MSE and EB for μ_1 and β_1 (100 simulation batches)

index	μ_1	β_1	σ_{τ_1}	$\hat{MSE} - \mu_1$	$\hat{EB} - \mu_1$	$\hat{MSE} - \beta_1$	$\hat{EB} - \beta_1$
1	-4	5	0.1	0.0431	-0.0160	0.0560	-0.0029
2	-4	5	1	0.0750	-0.0146	0.1166	0.0080
3	-4	5	2	0.2755	-0.0005	0.4259	-0.0038
4	-4	5	3	0.4901	-0.0065	0.7449	-0.0556
5	-4	5	4	1.2410	-0.0438	1.2031	0.0063
6	-4	0.08	1	0.0540	-0.0086	0.1121	-0.0151
7	-4	0.08	2	0.1337	-0.0446	0.2731	0.0475
8	-4	0.08	3	0.2380	0.0318	0.4400	0.0123
9	-4	0.08	4	0.8209	0.0568	1.5245	-0.1096
10	-4	0.08	5	1.2324	0.1632	2.1418	-0.1810

The following 16 figures show boxplots of the estimated parameters. For each plot, all parameters are held constant, except the standard deviation of the subject effect.

For example, in Figure 3.1 the values of the first four parameters from Table 3.1 are fixed at $P(male) = .53$, $P(Y_1 = 1) = .223$, $\mu_0 = -4$, $\beta_0 = 5$, $\mu_1 = -3$, $\beta_1 = .02$, and as index changes from 1 through 5, the common random effect standard deviation $\sigma_{\tau_0} = \sigma_{\tau_1}$ is set to .1, 1, 2, 3, and 4 respectively. Only 10 simulation batches were used in generating this plot. Figure 3.2 plots the same quantities, but based on 100 simulation batches.

The associated pair of plots, 3.3 and 3.4, based on 10 and 100 simulation batches respectively, fix the parameter values at $P(male) = .53$, $P(Y_1 = 1) = .530$, $\mu_0 = -4$, $\beta_0 = 0.08$, $\mu_1 = -1.5$, $\beta_1 = .2$, and set the common random effect standard deviation $\sigma_{\tau_0} = \sigma_{\tau_1}$ to 1, 2, 3, 4, and 5, as specified in the last five rows of Table 3.1.

In each plot, the true value of the parameter of focus (μ_0, β_0, μ_1 or β_1) is indicated by the horizontal red line.

As can be seen in the plots, when all other parameters are held constant, the estimated variation in the parameter estimates increases as the standard deviation of the subject effects $\sigma_{\tau_0} = \sigma_{\tau_1}$ increases.

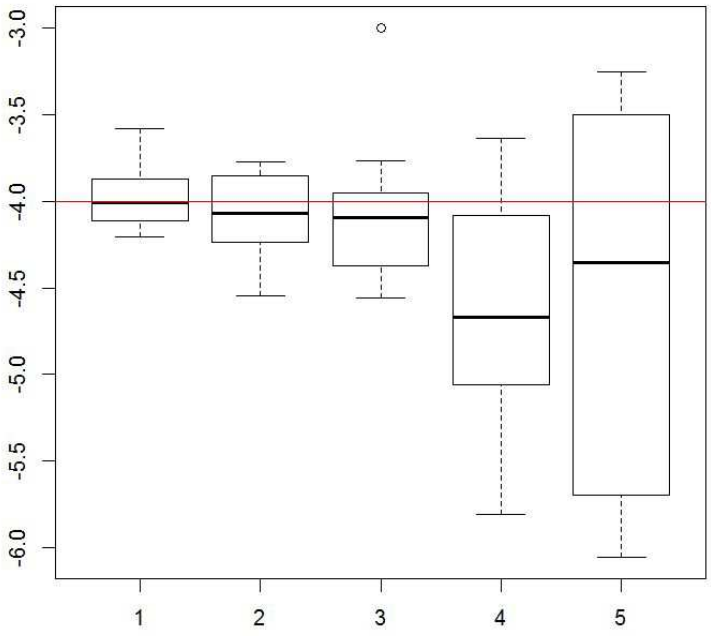

Figure 3.1: $\hat{\mu_0}$ - index 1 to 5 for 10 simulation runs

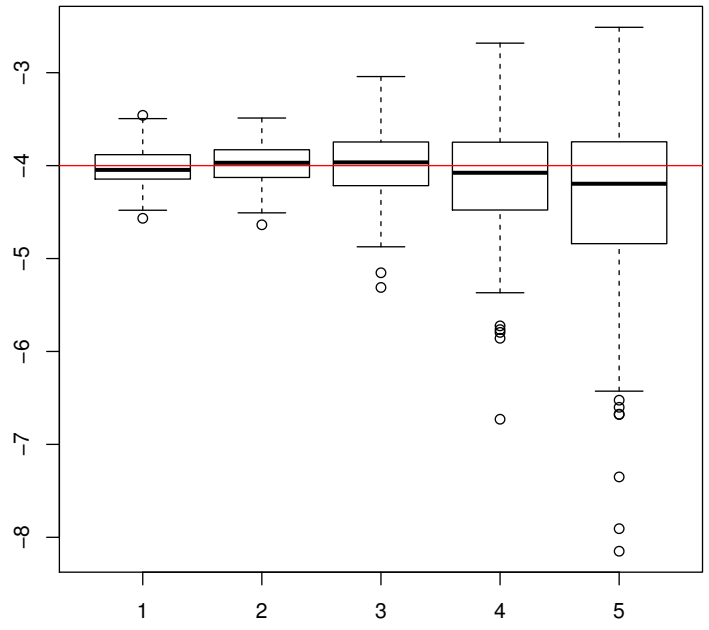

Figure 3.2: $\hat{\mu_0}$ - index 1 to 5 for 100 simulation runs

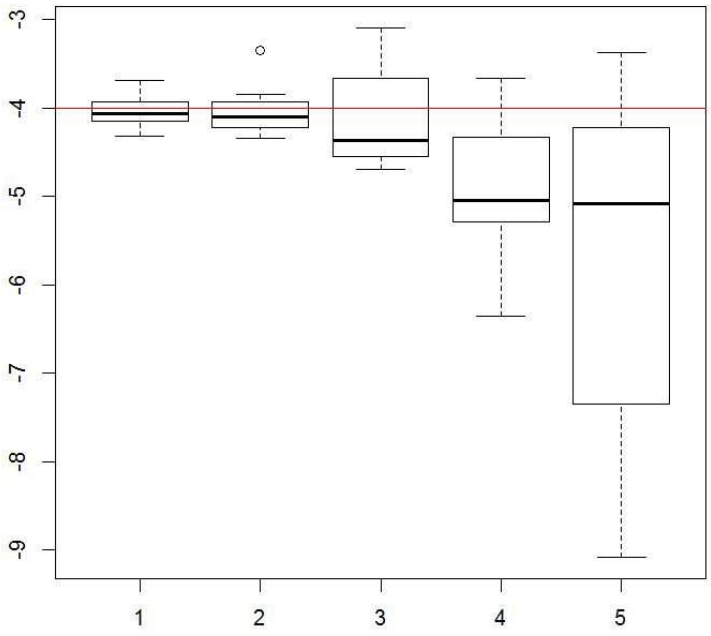

Figure 3.3: $\hat{\mu_0}$ - index 6 to 10 for 10 simulation runs

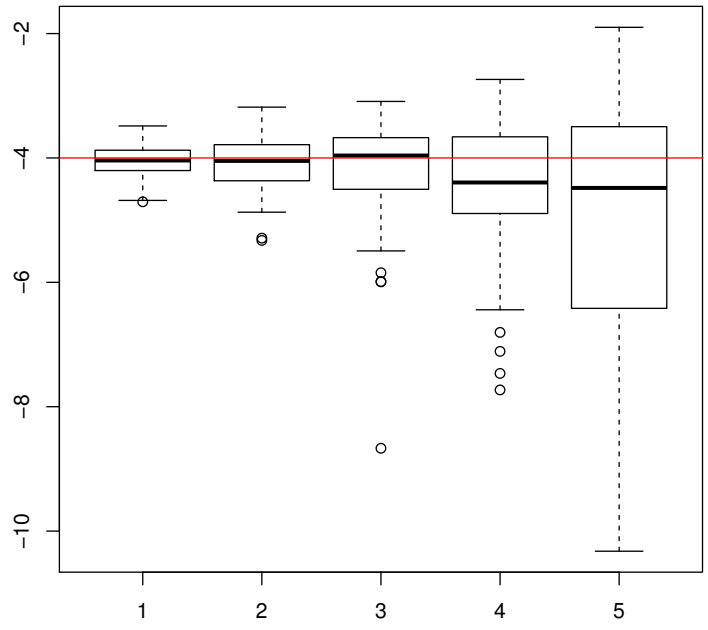

Figure 3.4: $\hat{\mu_0}$ - index 6 to 10 for 100 simulation runs

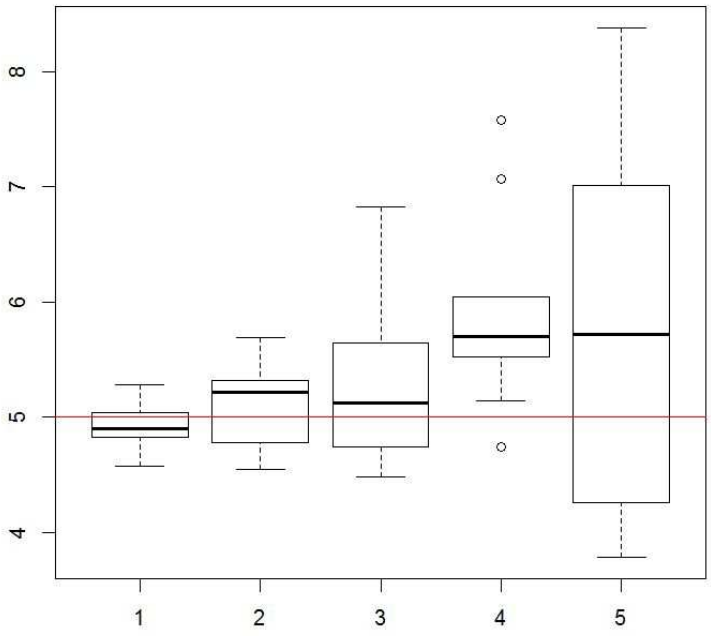

Figure 3.5: $\hat{\beta}_0$ - index 1 to 5 for 10 simulation runs

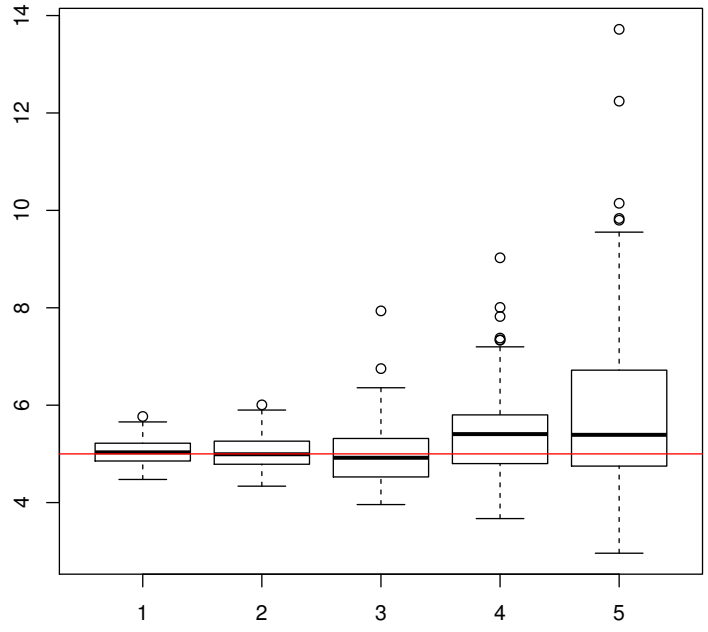

Figure 3.6: $\hat{\beta}_0$ - index 1 to 5 for 100 simulation runs

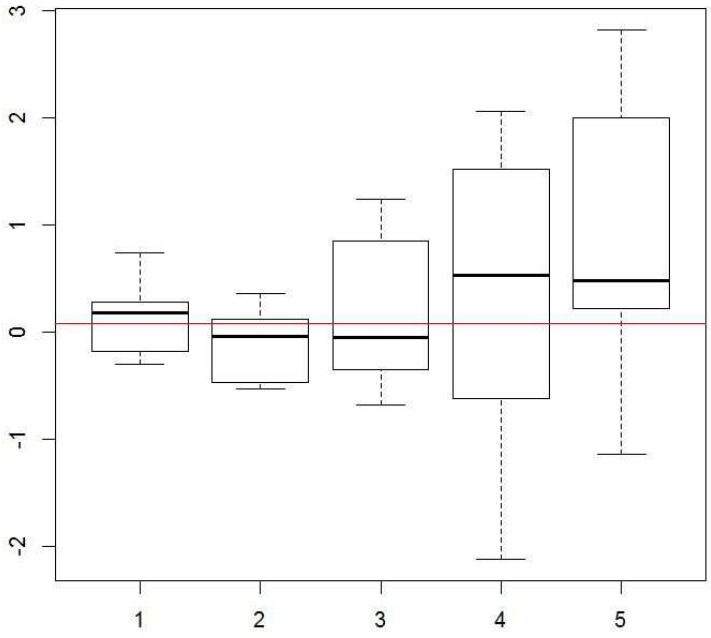

Figure 3.7: $\hat{\beta}_0$ - index 6 to 10 for 10 simulation runs

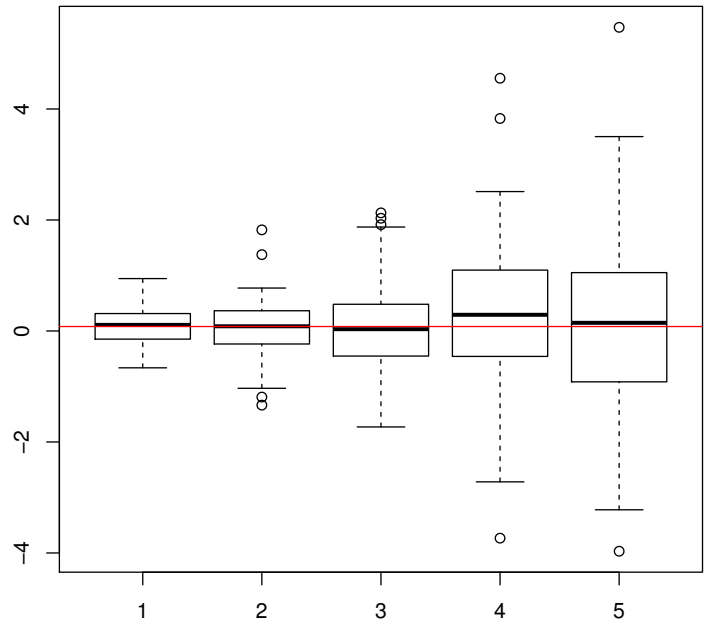

Figure 3.8: $\hat{\beta}_0$ - index 6 to 10 for 100 simulation runs

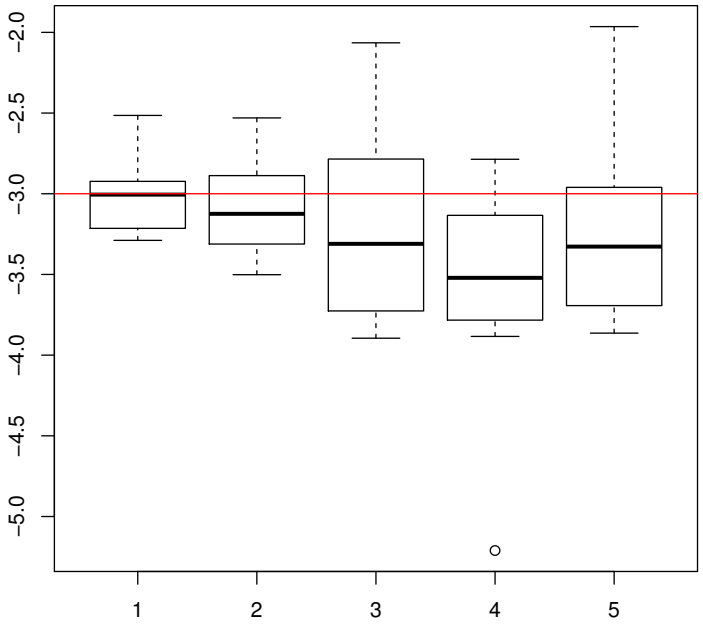

Figure 3.9: $\hat{\mu_1}$ - index 1 to 5 for 10 simulation runs

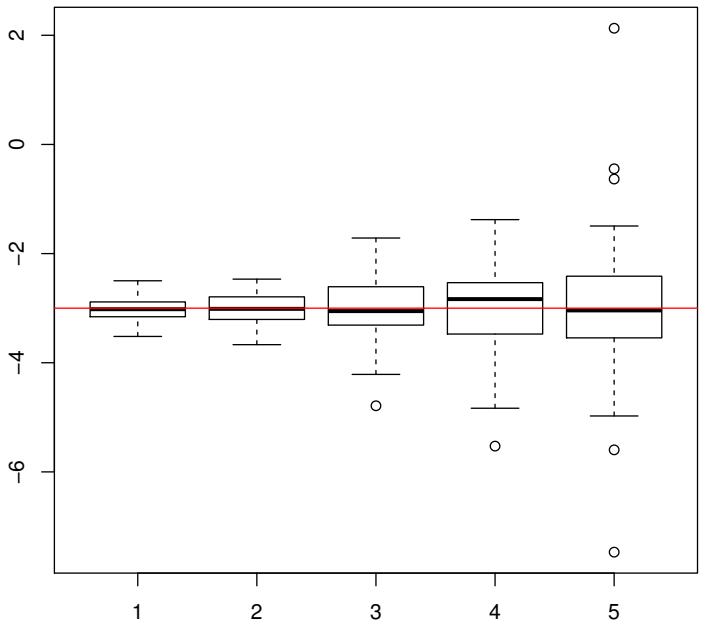

Figure 3.10: $\hat{\mu_1}$ - index 1 to 5 for 100 simulation runs

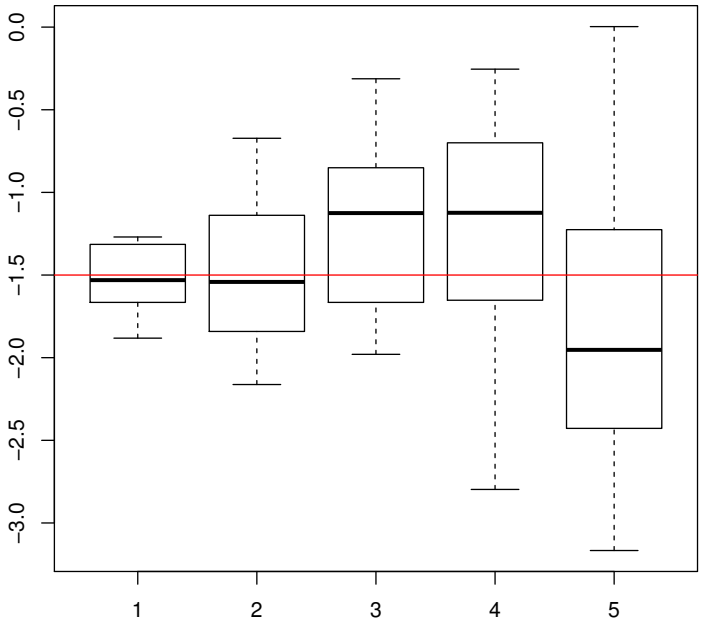

Figure 3.11: $\hat{\mu_1}$ - index 6 to 10 for 10 simulation runs

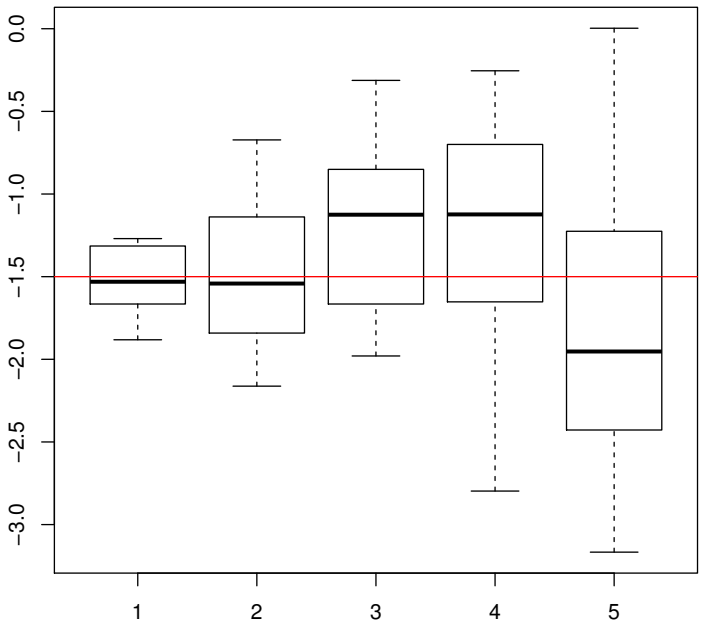

Figure 3.12: $\hat{\mu_1}$ - index 6 to 10 for 100 simulation runs

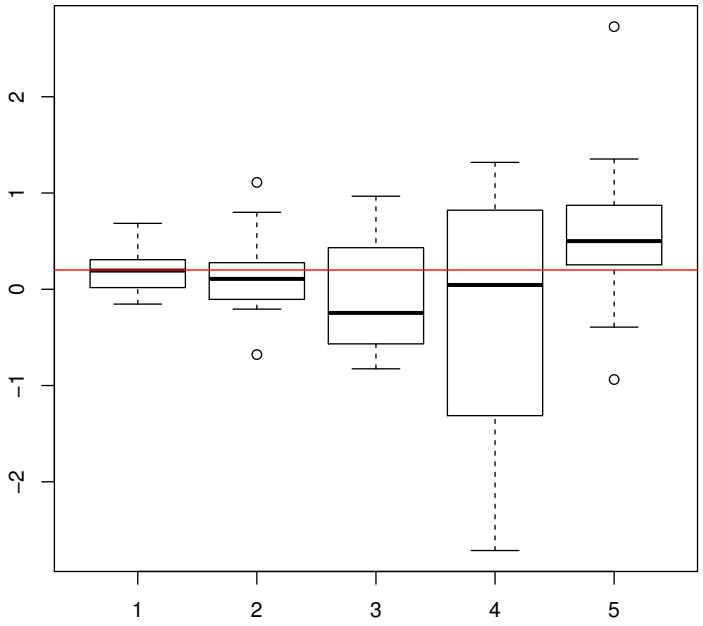

Figure 3.13: $\hat{\beta}_1$ - index 1 to 5 for 10 simulation runs

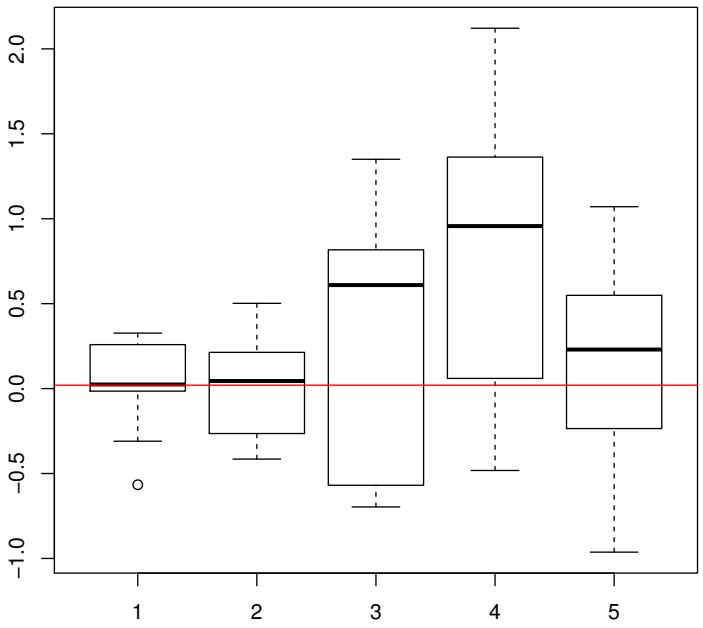

Figure 3.14: $\hat{\beta}_1$ - index 1 to 5 for 10 simulation runs

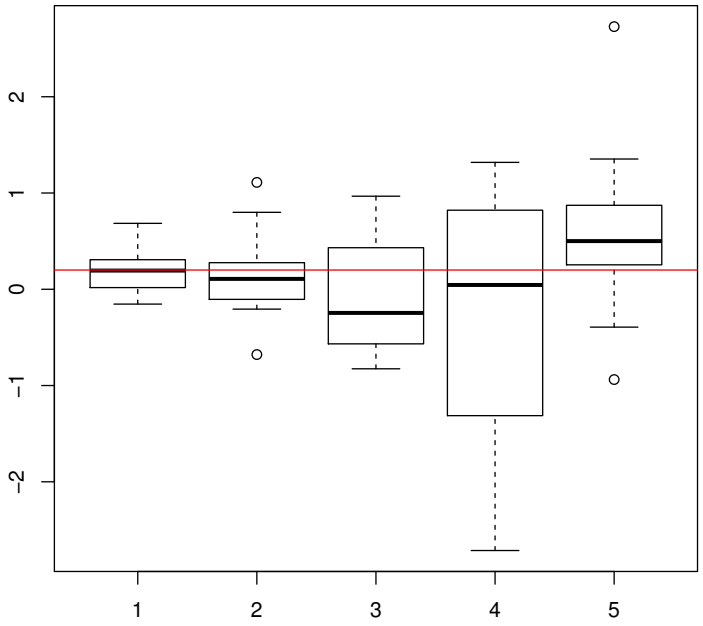

Figure 3.15: $\hat{\beta}_1$ - index 6 to 10 for 10 simulation runs

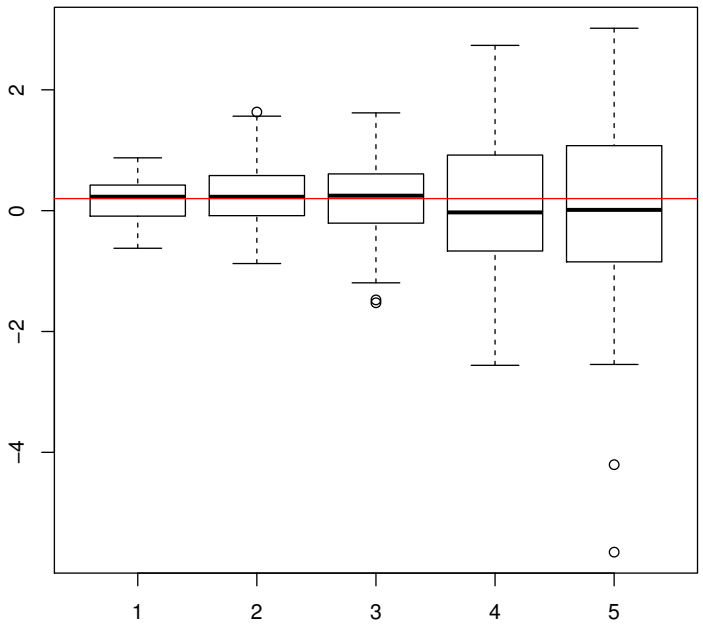

Figure 3.16: $\hat{\beta}_1$ - index 6 to 10 for 100 simulation runs

From this chapter we can see, the average mean squared errors are fairly small when comparing the true parameter values with the estimated. Also, the discrepancy between the true and estimated parameter values gets bigger as the variations of subject effect increase when holding all the other parameter values constant. The expecation is that the MSE should decrease in proportion to the inverse of the number of simulation batches. This is not evident from either the tables nor the figures above, and this is likely due to the high degree of the Monte Carlo variability when using only 10 simulation batches. In general, we can make the conclusion that the Generalized Linear Mixed Effect model works well in the application of medication use patterns and the two separate GLMM models make sense in this case.

Chapter 4

Discussion

In this thesis, we proposed a simple method to study the statin use patterns of Nova Scotia seniors population and the patients' adherence to medication. The switching patterns of statin usage were modeled using a Markov model, which was reduced to two independent generalized linear models. Subsequently, a simulation study was conducted to show that this approach was able to produce consistent estimates of transition probablities and effects of confouding factors. On the other hand, the variance and covariance structure of random effects were much harder to estimate. The variance parameter estimate was shown to converge very slowly to true values.

In our model, the switching of statin types (low dose to high dose and vice versa) was assumed to be independent. This is a somewhat over-simplified assumption. In reality, the adherence to a certain statin type is expected to be correlated for an individual patient. To introduce correlation, a common parameter might be introduced in equation (3.4) and (3.6). For example, one can enforce patient effects in both processes to only differ by signs i.e.

$$P_{01} = \frac{1}{1 + \exp(x\beta_0 + \tau_0)} \tag{4.1}$$

$$P_{10} = \frac{1}{1 + \exp(x\beta_1 - \tau_0)} \tag{4.2}$$

In this case, the transitions to and from a certain statin level are negatively correlated. In this thesis, I am able to use the *lme4* package due to the independence assumption on the two random effects. However, even with a small change like above, a custom software has to be written to implement the generalized linear mixed model.

The model in this thesis assumes separate "physician" covariates for the index prescription, and subject level covariates for the switching part of the model. In reality, however, it is likely that the index prescription will be related to both physician and patient characteristics. For example, if the severity of disease is high, then the index prescription is almost invariably going to be high intensity.

Given our limitation on time and resources, this thesis was really "proof of concept" to try to develop a model which might be appropriate for the CNODES data. I was trying to assess the bias and variability in the parameter estimates. Based on the simulations, the bias in the fixed effect parameters is small, the MSE increases with the variability of the subject effects and is expected to decrease with larger sample size. Given the large sample size (close to 20,000 patients) of the CNODES data, it is expected much less variability in the estimates based on the real data. In the future, this model will be tested on the real data, to further explore the convergence properties of the model, especially for the variance covariance structure of the random effects. The choice of estimation methods will also greatly impact the precision of estimates. It is thus interesting to also evaluate performance of various estimation methods with various parameter settings and covariance structures. The calculation of sample size and confidence interval are also desirable for practical applications.

Appendices

Appendix A

Simulation code in R

```
# true model is
# logit(p) = int + beta*gender + tau
#  with tau ~ N(0,tau.std);
# equivalently p = exp(x)/(1+exp(x)) where x = int + beta*gender + tau

#n00 is the number of transitions from 0 to 0
#n01 is the number of transitions from 0 to 1
#n10 is the number of transitions from 1 to 0
#n11 is the number of transitions from 1 to 1

##Module1: Function

inv.logit = function(x){return(exp(x)/(1+exp(x)))};
logit = function(x){return(log(x/(1-x)))};

sim.model=function(par){
  # generate gender from a binomial distribution
  male = rbinom(1,1,par$male.p);

  # generate 2 random effects for p01 and p10
  tau0.i = rnorm(1,0,par$tau0.std);
  tau1.i = rnorm(1,0,par$tau1.std);

  x0 = par$int0 + par$beta0*male + tau0.i;
  p01 = exp(x0)/(1+exp(x0))
  p00 = 1 - p01;
```

```
  x1 = par$int1 + par$beta1*male + tau1.i;
  p10 = exp(x1)/(1+exp(x1));
  p11 = 1 - p10;

  return(c(p00,p01,p10,p11,tau0.i,tau1.i,male))
}

######### function to fit glmm #########
glmm.fit = function(fname){
   dat=read.table(fname,header=TRUE);
   dat$id = factor(dat$id)
   dat$statin = factor(dat$statin)

   ret = glmer(switch~factor(male)+(1|id),family=binomial,data=dat);
   return(c(fixef(ret), attr(VarCorr(ret)$id,"stddev")));
}

######### simulate data #########
simulate=function(nsubj=3,par){

# generate number of observations and keep nobs > 1
nobs= trunc(rgamma(nsubj,shape=alpha,scale=beta)+.5) + 1
nobs = subset(nobs,nobs>1);

mnobs = max(nobs);
nsubj = length(nobs);

data=matrix(NA,nrow=nsubj,ncol=mnobs);
```

```
# define output list
ntrans = matrix(0,nrow=nsubj,ncol=4);
colnames(ntrans) = c("n00","n01","n10","n11");

# true parameters
par.true = matrix(0,nrow=nsubj,ncol=7);
colnames(par.true) = c("p00","p01","p10","p11","tau0","tau1","male");

for (i in 1:nsubj){

    # simulate transition probability for subject i
    par.true[i,] = sim.model(par);

    p0 = c(par.true[i,"p00"],par.true[i,"p01"]);
    p1 = c(par.true[i,"p10"],par.true[i,"p11"]);

    # simulate initial state x0
    old=rbinom(1,1,par$x0.p)
    data[i,1] = old;

    for (j in 2:nobs[i])
    {
      if(old==0)
      {
        new=sample(c(0,1),1,prob=p0)
       if (new==0){
           ntrans[i,"n00"] = ntrans[i,"n00"] + 1}
        if (new==1){
           ntrans[i,"n01"] = ntrans[i,"n01"] + 1}
      }
      else{
```

```
        new=sample(c(0,1),1,prob=p1)
        if (new==0){
            ntrans[i,"n10"] = ntrans[i,"n10"] + 1}
        if (new==1){
            ntrans[i,"n11"] = ntrans[i,"n11"] + 1}
      }
      old = new;
      data[i,j] = new;
    } # end j in 2:nobs[i] loop


} # for each subject i

# long format data for p0 and p1
data0 = NULL;
data1 = NULL;
for (i in 1:nsubj){
   n00 = ntrans[i,'n00'];
   n01 = ntrans[i,'n01'];
   n0=n00+n01;
   d0 = data.frame(statin=c(rep(0,n00), rep(1,n01)),
       male=rep(par.true[i,"male"],n0),
       p00=rep(par.true[i,"p00"],n0),
       p01=rep(par.true[i,"p01"],n0),
       tau0=rep(par.true[i,"tau0"],n0), id=rep(i,n0));

   n10 = ntrans[i,'n10'];
   n11 = ntrans[i,'n11'];
   n1=n10+n11;
   d1 = data.frame(statin=c(rep(0,n10), rep(1,n11)),
   male=rep(par.true[i,"male"],n1),
   p10=rep(par.true[i,"p10"],n1),
```

```
   p11=rep(par.true[i,"p11"],n1),
   tau1=rep(par.true[i,"tau1"],n1), id=rep(i,n1));

   d0$switch=d0$statin;
   data0 = rbind(data0,d0);

   d1$switch=ifelse(d1$statin==1,0,1);
   data1 = rbind(data1,d1);
}
 # calculate raw transition probability
 ptrans=sweep(ntrans,1,nobs-1,'/');
 colnames(ptrans) = c("p00","p01","p10","p11");

 return(list(data=data,nobs=nobs,ntrans=ntrans,
  ptrans=ptrans,par.true=par.true,data0=data0,data1=data1))
} # function end;


##Module 2: Simulation;
parameter=read.table("parameter.csv",header=TRUE,sep=',');

# switches has mean 50, sd 9
beta=9^2/50;
alpha=50/beta;

source('functions.R');


for(i in 1:nrow(parameter)){
  for(j in 1:parameter$nsample[i]){
     # ret = simulate(nsubj=5,parameter[i,]);
     dat = simulate(nsubj=100,parameter[i,]);
```

```
    #To concatenate text, you use the paste() function;
    filename = paste("data\\d0_p",parameter$index[i],"_",j,sep="");
    write.table(dat$data0,file=filename,row.names = FALSE);

    filename = paste("data\\d1_p",parameter$index[i],"_",j,sep="");
    write.table(dat$data1,file=filename,row.names = FALSE);
  }
}


##Module 3: GLMM model
rm(list=ls())
library(lme4);

source('functions.R');

parameter=read.table("parameter.csv",header=TRUE,sep=',');

# define output matrix
out0 = matrix(NA,nrow=sum(parameter$nsample),ncol=4);
colnames(out0)= c("e.int0","e.beta0","e.tau.std0","sample");

out1 = out0;
colnames(out1)= c("e.int1","e.beta1","e.tau.std1","sample");

idx = 1;

# fit glmm on all data
for(i in 1:nrow(parameter)){
  for(j in 1:parameter$nsample[i]){
   # d0
   filename = paste("data\\d0_p",parameter$index[i],"_",j,sep="");
```

```
   out0[idx,] =  c(glmm.fit(filename),j);

   # d1
   filename = paste("data\\d1_p",parameter$index[i],"_",j,sep="");
   out1[idx,] =  c(glmm.fit(filename),j);

   idx = idx + 1;
  }
}

result0 = data.frame(cbind(apply(parameter,2,rep,parameter$nsample),out0));
result0 = subset(result0,select=-c(int1,beta1,tau1.std))
result1 = data.frame(cbind(apply(parameter,2,rep,parameter$nsample),out1));
result1 = subset(result1,select=-c(int0,beta0,tau0.std))

write.table(result0,file='result\\result0.txt',row.names=FALSE);
write.table(result1,file='result\\result1.txt',row.names=FALSE);

#mean squared error, and estimated bias(SE, sum of error);
#for index 1 (the first row of parameters), start state is 0;

for (j in 1:10){
SSE0_j=0;
SE0_j=0;
for (i in (1+10*(j-1)):(10+10*(j-1))){
SSE0_j=SSE0_j+(result0[i,8]-result0[i,5])^2;
      SE0_j=SE0_j+(result0[i,8]-result0[i,5])
      };
MSE_int0_j=SSE0_j/result0[1,4];
ME_int0_j=SE0_j/result0[1,4];
print(MSE_int0_j);
print(ME_int0_j);
```

```
};


for (j in 1:10){
SSE0_j=0;
SE0_j=0;
for (i in (1+10*(j-1)):(10+10*(j-1))){
SSE0_j=SSE0_j+(result0[i,9]-result0[i,6])^2;
      SE0_j=SE0_j+(result0[i,9]-result0[i,6])
};
MSE_beta0_j=SSE0_j/result0[1,4];
ME_beta0_j=SE0_j/result0[1,4];
print(MSE_beta0_j);
print(ME_beta0_j);
};

#for index 1, start state 1;
for (j in 1:10){
SSE1_j=0;
SE1_j=0;
for (i in (1+10*(j-1)):(10+10*(j-1))){
SSE1_j=SSE1_j+(result1[i,8]-result1[i,5])^2;
      SE1_j=SE1_j+(result1[i,8]-result1[i,5])
      };
MSE_int1_j=SSE1_j/result1[1,4];
ME_int1_j=SE1_j/result1[1,4];
print(MSE_int1_j);
print(ME_int1_j);
};


for (j in 1:10){
```

```
SSE1_j=0;
SE1_j=0;
for (i in (1+10*(j-1)):(10+10*(j-1))){
SSE1_j=SSE1_j+(result1[i,9]-result1[i,6])^2;
      SE1_j=SE1_j+(result1[i,9]-result1[i,6])
};
MSE_beta1_j=SSE1_j/result1[1,4];
ME_beta1_j=SE1_j/result1[1,4];
#print(MSE_beta1_j);
print(ME_beta1_j);
};
```

Bibliography

[1] Todd J Anderson, Jean Grégoire, Robert A Hegele, Patrick Couture, GB John Mancini, Ruth McPherson, Gordon A Francis, Paul Poirier, David C Lau, Steven Grover, et al. 2012 update of the canadian cardiovascular society guidelines for the diagnosis and treatment of dyslipidemia for the prevention of cardiovascular disease in the adult. *Canadian Journal of Cardiology*, 29(2):151–167, 2013.

[2] S. Bhattacharjee, P. A. Findley, and U. Sambamoorthi. Understanding gender differences in statin use among elderly medicare beneficiaries. *Drugs Aging*, 29:971–980, 2012.

[3] J.G. Booth and J.P. Hobert. Maximizing generalized linear mixed model likelihoods with an automated monte carlo em algorithm. *Journal of the Royal Statistical Society, Series B*, 61:265285, 1999.

[4] N.E. Breslow and D.G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, (88):925, 1993.

[5] N.E. Breslow and X. Lin. Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika*, 82(1):8191, 1995.

[6] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.

[7] D.A. Harville. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72:320–338, 1977.

[8] C.R. Henderson. Estimation of genetic parameters. *The Annals of Mathematical Statistics*, 21:309–310, 1950.

[9] C.R. Henderson. Selection index and expected genetic advance. *Statistical Genetics and Plant Breeding*, Nat. Acad. Sci., Nat. Res. Council, Publication 982, Washington, D. C.:141–163, 1963.

[10] C.R. Henderson. *Applications of Linear Models in Animal Breeding*. University of Guelph, 1984.

[11] Yoonsang Kim, Young-Ku Choi, and Sherry Emery. Logistic regression with multiple random effects: A simulation study of estimation methods and statistical packages. *Am Stat*, 67(3), 2013.

[12] N.M. Laird, N. Lange, and D. Stram. Maximum likelihood computations with repeated measures: Application of the em algorithm. *Journal of the American Statistical Association*, 82(397):97–105, 1987.

[13] M.J. Lindstrom and D.M. Bates. Newton-raphson and em algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83:1014–1022, 1988.

[14] Micha Mandel and Rebecca A. Betensky. Estimating time-to-event from longitudinal ordinal data using random-effects markov models: application to multiple sclerosis progression. *Biostatistics*, 9(4):750764, 2008.

[15] P. McCullagh and J.A. Nelder. *Generalized Linear Models*. New York: Chapman & Hall, 2nd edition, 1989.

[16] J.A. Nelder and R.W.M. Wedderburn. Maximum likelihood computations with repeated measures: Application of the em algorithm. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384, 1972.

[17] J.C. Pinheiro and D.M. Bates. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4:1235, 1995.

[18] R.W.M. Wedderburn. Quasi-likelihood functions, generalized linear models, and the gaussnewton method. *Biometrika*, 61(3):439447, 1974.

[19] Russ Wolfingera and Michael O'connellb. Generalized linear mixed models a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, 48:233–243, 1993.

[20] L. Yen, J. Wu, and P. Hodgkins. Medication use patterns and predictors of nonpersistence and nonadherence with oral 5-aminosalicylic acid therapy in patients with ulcerative colitis. *Journal of Managed Care Pharmacy*, 18:701–712, 2012.