

Regression Approaches to Estimation of Relative Risk: Application to Multiple Sclerosis Studies

by

Shing Fu

B.Sc., The University of Hong Kong, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Statistics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2010

© Shing Fu 2010

Abstract

Using a log link for binary response in generalized linear mixed-effects models (GLMM) allows direct estimation of the relative risk. If a random intercept is the only random effect in the conditional mean structure, the marginal mean has the same form. The fixed effects, representing the log relative risks, have the same interpretation in both the mixed-effects model and the marginal model. This leads to two approaches to estimate the relative risks, 1) maximum likelihood for the mixed-effects models and 2) the generalized estimating equations (GEE) approach for the marginal models.

In our study, we apply such log-linear models to assess the effects of neutralizing antibodies on interferon beta-1b in relapsing-remitting multiple sclerosis. The results obtained by the two approaches are compared. The relative efficiency of the GEE approach and the robustness of the GLMM approach to some forms of misspecification of the model for the random effects are studied by simulations.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Tables	vi
List of Figures	viii
Acknowledgements	ix
Dedication	x
1 Introduction	1
1.1 Multiple Sclerosis	1
1.2 The Interferon beta-1b Pivotal Trial	2
1.3 Neutralizing Antibodies in the IFNB Pivotal Trial	4
1.3.1 Overview	4
1.3.2 Review of Cross-sectional Analysis	4
1.3.3 Review of Longitudinal Analysis	5
1.4 The BEYOND Trial	8
1.5 Objectives and Outline of the Report	9
2 Methodology	11
2.1 Generalized Linear Models	11
2.2 The Quasi-likelihood Approach	16
2.3 Marginal Models for Longitudinal Data	17

Table of Contents

2.3.1	Overview	17
2.3.2	Inferential Approach: Generalized Estimating Equations	20
2.4	Generalized Linear Mixed-effect Models	26
2.4.1	Overview	26
2.4.2	Inferential Approach: Maximum Likelihood	29
2.5	Log-linear Models	32
2.5.1	Mixed-effects Models	32
2.5.2	Marginalized Models	33
3	Analysis of BEYOND Trial Data	40
3.1	Introduction	40
3.2	Basic Model Specification	46
3.3	Computational Issues	49
3.4	Time Trend	50
3.5	Neutralizing Antibody Effects	55
3.5.1	NAB Status	55
3.5.2	NAB Titer	56
3.6	Relative Efficiency of GEE to GLMM Approach on Estimating NAB Effects	64
3.7	Summary	65
4	Analysis Based on Collapsed Data	67
4.1	Introduction	67
4.2	Model Specification	69
4.3	Results	70
4.3.1	NAB Status	70
4.3.2	NAB Titer	72
4.4	Impact of Exceptionally Short Intervals	75
4.5	Summary	78

Table of Contents

5	Simulation Study	83
5.1	Introduction	83
5.2	Simulation I: Efficiency of GEE	85
5.3	Simulation II: Misspecification of Random Effects	100
5.4	Simulation III: Misspecification of the Random Effect Distribution	115
5.5	Conclusion	119
6	Conclusions and Discussion	129
	Bibliography	136

Appendices

A	Details of Model Fits for the Daily Relapse Data	139
B	Details of Model Fits for the Collapsed Data	146

List of Tables

3.1	Time of the first available NAB titer	42
3.2	Time to the first confirmed NAB+	43
3.3	NAB status dynamics	43
3.4	Number of relapses per eventually NAB+ patients	44
3.5	p-values for likelihood ratio tests comparing GLMM fits with different time trends: eventually NAB+ patients only.	54
3.6	Estimated relative risks of NAB+ to NAB− from GLMM fits with various time trends: eventually NAB+ patients only.	55
3.7	Estimated relative risks of NAB+ and sublevels to NAB−: eventually NAB+ patients only.	57
3.8	Estimated relative risk of NAB titers: eventually NAB+ pa- tients only.	62
4.1	Number of NAB titer measures	67
4.2	Number of relapse onsets per interval	68
4.3	Number of days at risk between consecutive collections of NAB serum specimens	69
4.4	Estimated relative risks of NAB+ and sublevels to NAB−: eventually NAB+ patients only, collapsed data.	73
4.5	Estimated relative risk of NAB titers: eventually NAB+ pa- tients only, collapsed data.	76
4.6	Impact of 1-day intervals on estimates of log relative rates, β	79
5.1	Results of Simulation I-1	91

List of Tables

5.2	Results of Simulation I-2: More Clusters	94
5.3	Results of Simulation I-3: Larger Cluster Size	97
5.4	Results of Simulation II-1	104
5.5	Results of Simulation II-2: More Clusters	108
5.6	Results of Simulation II-3: Larger Cluster Size	112
5.7	Results of Simulation III-1: Beta-Binomial Setup	122
5.8	Results of Simulation III-2: Triangular Distribution	126
A.1	Model fit based on eventually NAB+ patients only: linear time trend, NAB status.	140
A.2	Model fit based on eventually NAB+ patients only: linear time trend, refined NAB status.	141
A.3	Model fits based on 250 mcg arm, eventually NAB+ patients only: linear time trend, log NAB titers.	142
A.4	Model fits based on 500 mcg arm, eventually NAB+ patients only: linear time trend, log NAB titers.	144
B.1	Model fit based on collapsed data, eventually NAB+ patients only: linear time trend, NAB status.	147
B.2	Model fit based on collapsed data, eventually NAB+ patients only: linear time trend, refined NAB status.	148
B.3	Model fits based on collapsed data, 250 mcg arm, eventually NAB+ patients only: linear time trend, log NAB titers.	149
B.4	Model fits based on collapsed data, 500 mcg arm, eventually NAB+ patients only: linear time trend, log NAB titers.	151

List of Figures

1.1	Disease progression of each MS type.	2
2.1	Conditional and implied marginal mean structures of a log-linear model with random intercept and slope: random effects are evaluated at the $10^{th}, 20^{th}, \dots, 90^{th}$ percentiles.	39
3.1	Proportion of time on study as NAB− for eventually NAB+ patients.	45
3.2	Empirical annualized relapse rates for preceding 3-month periods and the corresponding loess fits for eventually NAB+ patients.	47
3.3	Annualized fitted relapse rate of GLMM for NAB− periods, natural regression splines for time trend, eventually NAB+ patients only.	52
3.4	Fitted natural cubic spline ($df = 2$) on log NAB titers of GLMM: eventually NAB+ patients only.	59
5.1	Probability density functions of random intercept b_i with mean -0.003 and variance $\sigma_b^2 = 0.316^2$ when 1) normally distributed and 2) baseline conditional probability follows beta a distribution.	121
5.2	Probability density functions of random intercept b_i with mean 0 and variance $\sigma_b^2 = 0.3^2$ when 1) normally distributed and 2) triangularly distributed.	125

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor John Petkau for his insightful guidance and constant encouragement throughout the course of my study in UBC. It is a rewarding experience for me to work under his supervision. I am also grateful to Professor Rollin Brant for being my second reader and providing valuable comments.

I would like to thank all my professors from whom I learnt a lot, in particular, Professor Harry Joe with whom I always enjoy working. Special thanks to Peggy Ng, Viena Tran, Elaine Salameh and Andrea Sollberger for their efficient assistance, and to my colleagues in the Department of Statistics for their support.

Last but not least, I am indebted to Jeanne and Ian Chen and their family for taking care of me and helping me adapt to the new environment when I first arrived in Vancouver.

To my family.

Chapter 1

Introduction

1.1 Multiple Sclerosis

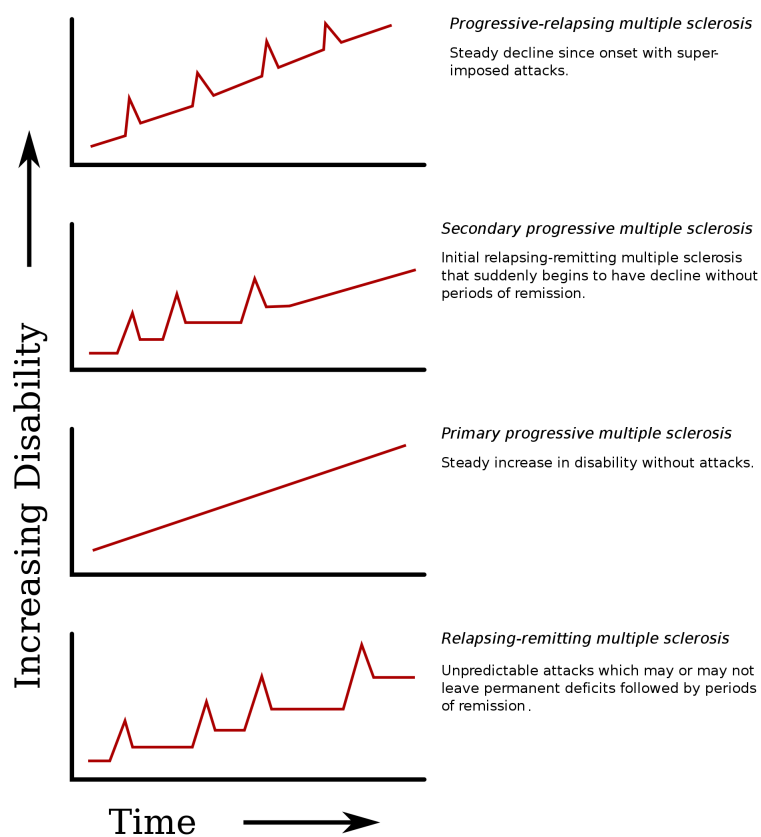
In this report, we compare two regression approaches to estimation of relative risk with application to studies of neutralizing antibodies in relapsing-remitting multiple sclerosis patients treated with interferon beta-1b. Multiple sclerosis (MS) is a chronic and often disabling disease of the central nervous system. MS is thought to be an autoimmune disease where a patient's own immune system attacks myelin in the brain and spinal cord. Myelin forms the insulating protective myelin sheaths of nerve fibres. Damage to the myelin sheaths or nerve fibres affects the transmission of nerve signals resulting in symptoms including difficulties in moving and coordination, deterioration of sensory functions, problems in bowel and bladder, among many others. Lesions in the brain and spinal cord formed at the site of inflammation where the myelin is damaged are usually monitored by magnetic resonance imaging (MRI) scans. Specific symptoms and severity of MS vary considerably across patients.

Currently there are four types of MS characterized by disease progression in term of severity of disability: 1) Relapsing-Relmitting MS (RRMS), 2) Primary Progressive MS (PPMS), 3) Secondary Progressive MS (SPMS) and 4) Progressive Relapsing MS (PRMS). Figure 1.1 shows the typical progression profiles of each type of MS.

RRMS is the most common type of MS. According to the MS Society

1.2. The Interferon beta-1b Pivotal Trial

Figure 1.1: Disease progression of each MS type.



of Canada, about 80-85% of the patients are first diagnosed with RRMS. RRMS is characterized by recurrent cycles of acute attack, also called relapse or exacerbation, followed by complete or partial recovery with absence of disease progression between relapses.

1.2 The Interferon beta-1b Pivotal Trial

There is no cure to MS to date, but disease modifying therapies exist to reduce the exacerbation rate and severity of relapse. In 1995 interferon beta-1b

1.2. *The Interferon beta-1b Pivotal Trial*

(IFNB) became the first disease modifying therapy approved for RRMS in Canada.

The efficacy of IFNB in reducing exacerbation rate in RRMS was first demonstrated by a multicenter, randomized, double blind, placebo-controlled trial conducted by the IFNB Multiple Sclerosis Study Group[1–3]. From June 1988 to May 1990, 372 MS patients with disease history longer than 1 year were enrolled in this pivotal trial: 123 patients were assigned to a placebo arm, 125 to a low dose arm of 1.6 million international units (MIU) and 124 to a high dose arm of 8 MIU of IFNB, self-administered by subcutaneous injections every other day. The primary clinical outcome was annualized exacerbation rate. An exacerbation was defined as appearance of a new symptom or worsening of an existing symptom, attributable to MS, accompanied by an appropriate new neurologic abnormality. An exacerbation should also be preceded by stability or improvement for a minimum of 30 days and last for at least 24 hours in the absence of fever. The radiological outcome was cranial MRI scans for brain lesions. All patients had a baseline MRI scan and subsequent annual MRI scans. A cohort of 52 patients at the University of British Columbia, one of the participating sites, had MRI scans every 6 weeks for 2 years.

The trial was terminated with all patients converted to high dose (8 MIU) treatment in January 1993. Total time-on-study for individual patients ranged from 3.5 to 5 years. Based on 2-year data, the annual exacerbation rate was 1.27 for the placebo group, 1.17 for the 1.6 MIU group and 0.84 for the 8 MIU group. The reduction in exacerbation rate in the 8 MIU group relative to the placebo group was 33% in the first year of study and 28% in the second year of study, both with statistical significance at the 0.05 level. For the final three years, the reduction still ranged from 24% to 30%. However, these reductions did not achieve statistical significance. A decreas-

ing trend in the annual exacerbation rate was observed over the 5-year course of study in all groups with the placebo group always having the highest rate whereas the 8 MIU group always had the lowest rate.

1.3 Neutralizing Antibodies in the IFNB Pivotal Trial

1.3.1 Overview

The presence of neutralizing antibodies (NABs) to IFNB in the serum specimens of some patients in the IFNB pivotal trial was well documented[1, 3]. Serum specimens were collected quarterly during the trial and tested for NABs by a cytopathic effect (CPE) assay. Development of NABs began mostly during the first year on study. In some patients, NABs appeared inconsistently and intermittently while in others, NABs were detected persistently subsequent to their first appearance.

The minimum detection limit of the CPE assay for NAB titers is 20 neutralizing units (NU/mL). A patient was identified as NAB+ if at least two consecutive NAB titers were at least 20 NU/mL; otherwise the patient was considered NAB-. Under this *ad hoc* rule, 2 of 123 placebo patients were eventually identified as NAB+, whereas in both the 1.6 MIU and 8 MIU arm about 35% of the patients were identified as NAB+ by the end of the trial.

1.3.2 Review of Cross-sectional Analysis

NABs were expected to diminish the efficacy of IFNB in reducing MS disease activity. To assess the impact of NABs on exacerbation rate, the annualized exacerbation rate of the NAB+ patients was compared to that of the NAB- patients on the same arm for every half-year period in a cross-

sectional analysis[4]. In the 8 MIU group, the NAB+ patients had higher estimated exacerbation rates than the NAB− patients for periods beyond 13 months on study and these exacerbation rates resembled the rates of the placebo patients in that time period. The increases were significant over the half-year periods of 19 – 24 and 25 – 30 months on study and over the longer periods of 13 – 36 and 19 – 36 months on study. In the 1.6 MIU group, there was no consistent pattern of increased exacerbation rates in the NAB+ group compared to the NAB− group. For the half-year period of 19 – 24 months on study, the exacerbation rate in the NAB+ group is even significantly lower than the NAB− group. The study claimed to have shown significant evidence of diminished treatment effect in the 8 MIU group in terms of increased risk of exacerbation associated with development of NABs.

1.3.3 Review of Longitudinal Analysis

Petkau et al.[5] pointed out that the cross-sectional analyses were inadequate as the baseline covariates, which might be predictive of exacerbation rates, were not taken into account. Also, the analyses were based on each half-year period separately and therefore were not very efficient. They suggested a longitudinal approach to analyze the effects of NABs on exacerbation rate to allow each patient to act as his own control. In their analysis, all patients were initially considered as NAB−. The subsequent NAB status for each day on study was then determined by one of several proposed *ad hoc* rules. The relapse rates during NAB+ periods of individual patients were compared to their rates during NAB− periods. This within subject comparison eliminates the main effects of baseline covariates which are potential confounders of the NAB effect. The longitudinal approach also allows straightforward incorporation of a time trend to adjust for the naturally changing exacerbation rates.

The first rule of classification of NAB status, referred to as “*once positive, always positive*”, defines the time of switching from NAB− to NAB+ by the

1.3. Neutralizing Antibodies in the IFNB Pivotal Trial

time of the first of two consecutive NAB titers of at least 20 NU/mL. Once switched to NAB+, the patient is considered to remain NAB+ until the end of the study irrespective of any of the following NAB titers. The requirement of two consecutive NAB titers higher than a cut-off value is referred to as “*confirmation required*”.

The second rule, referred to as “*all switches considered*”, allows switching back of the NAB status from NAB+ to NAB−. The definition for the time of switching from NAB− to NAB+ is the same as the first rule, which is the first time of two consecutive titers of at least 20 NU/mL. The time of switching from NAB+ back to NAB− is the time of the first subsequent NAB titer below 20 NU/mL with no confirmation required. The patient’s NAB status can switch back and forth for any number of times.

Under either rule of classification, a patient is identified as “*eventually NAB+*” if he has at least one NAB+ period; otherwise the patient is identified as “*never NAB+*”. Refining modifications applicable to both rules were also considered to take some account of the titer levels. NAB+ is divided into two sublevels, Low NAB+ and High NAB+, to examine the relative effect of different levels of NAB titers. Two modifications were considered: confirmation required or not required for switching from a lower level to a higher level. In both cases, confirmation is not required for switching from a higher level to a lower level.

The longitudinal approach facilitates direct comparison of exacerbation frequencies during NAB− periods and NAB+ periods of individuals who had at least one NAB+ period. Specifically Petkau et al.[5] considered the generalized estimating equations approach which directly estimates the effects of NABs at the population average level in terms of the relative risk of exacerbations while accounting for the correlation in the longitudinal observations.

Their analyses also adjusted for the natural trend in exacerbation rates over time by incorporating a linear effect of time.

By the time the longitudinal analysis was conducted, serum specimens were tested by a new assay, namely the MxA assay, which has the same detection limit as the CPE assay. NAB status determined by both assays was considered but in separate analyses. With the MxA assay and the “*once positive, always positive*” definition, comparing NAB+ periods to NAB– periods, the exacerbation rate increased by 28% (95% CI: –15%, 92%) in the 1.6 MIU group and by –2% (95% CI: –21%, 21%) in the 8 MIU group. With the “*all switches considered*” definition, the increase in exacerbation rate was 29% (95% CI: 0%, 67%) in the 1.6 MIU group and 18% (95% CI: 0%, 40%) in the 8 MIU group. With either definition, the increase was higher for the low dose (1.6 MIU) group than the high dose (8 MIU) group. Using the titer values obtained by the CPE assay, similar results were obtained, however, with lower statistical sensitivity. In additional analyses, no evidence indicating a carryover of NAB+ effect was demonstrated. After adjustment for the overall time trend, the relapse rates during the first NAB– period were not statistically different from the rates during subsequent NAB– periods which were preceded by at least one NAB+ period.

Petkau et al.[5] also estimated the power of the longitudinal analyses based on the estimated correlation in the repeated observations on individual patients over time. Treatment with 8 MIU IFNB reduces the exacerbation rate to about two-thirds of the placebo rate. A 50% increase in exacerbation rate associated with the switch from NAB– to NAB+ is an effect of importance as it implies that the rate during NAB+ period resembles the placebo rate. With the MxA assay, the power to detect a 50% increase in exacerbation rate when NAB status switches from negative to positive was estimated to be 96% for the “*once positive, always positive*” classification and 99.7% for

the “all switches considered” classification. These longitudinal analyses with results from the MxA assay had adequate statistical sensitivity for detecting an effect of this magnitude.

However, the NAB effects on exacerbation rates were not precisely estimated with the pivotal trial data, as demonstrated by the wide confidence intervals obtained. The estimation of NAB effects are based directly on the *eventually NAB+* patients only. The low sensitivity is due in part to the relatively small fraction of patients (about 35%) ever switching to NAB+ over the course of trial. A larger sample of *eventually NAB+* patients is required for more precise estimation of the NAB effects.

1.4 The BEYOND Trial

The Betaseron Efficacy Yielding Outcomes of a New Dose (BEYOND) trial was the largest MS phase III trial ever, involving a total of 2244 patients with RRMS[6]. Patients enrolled were followed for 2 to 3.5 years. The original objective of the BEYOND trial was to compare the efficacy of the standard dose of IFNB (250 mcg, i.e. 8 MIU) with an experimental, higher dose of IFNB (500 mcg, i.e. 16 MIU) and with 20 mg glatiramer acetate (GA), an alternative treatment to IFNB.

A total of 2244 enrolled patients were randomized to the 250 mcg IFNB group, the 500 mcg IFNB group and the GA group in the ratio 2:2:1. The two IFNB groups were given the specified dose of IFNB injected subcutaneously every other day while the GA group was given a 20 mg GA injection subcutaneously every day. Serum specimens were collected at baseline and every 6 months for patients in the two IFNB arms and serum NAB titers were measured by the MxA assay. The BEYOND trial provides, for the first time, a potentially large number of NAB+ patients for assessment of NAB

effects on RRMS disease activity so greater statistical sensitivity should be expected relative to previous studies.

1.5 Objectives and Outline of the Report

The primary objective of our study is to contrast two longitudinal approaches for comparing exacerbation rates during NAB− and NAB+ periods. The first, the generalized estimating equations (GEE) approach, has been applied in some previous longitudinal studies of NAB effects. The second is the full likelihood approach of generalized linear mixed-effects models (GLMM), which has not been utilized previously in this context due to computational difficulties. The efficiency of GEE relative to the GLMM approach is examined. An accurate model to reflect the natural diminishing relapse rates is crucial for estimating NAB effects unbiasedly. The adequacy and appropriateness of a linear time trend will be assessed. The impact of improving the fit of natural time trends by regression splines will also be examined.

When the clinical outcome is observed on a daily basis, the huge number of responses for each patient inevitably hinders the use of some standard statistical software for conducting the longitudinal analysis. The daily outcome is collapsed to a less frequently observed outcome to attempt to circumvent these computational difficulties. The loss of sensitivity due to such reductions in the resolution of the relapse outcome is also investigated. For coherence and consistency, the same set of analyses is performed separately on the 250 mcg group and the 500 mcg group.

This chapter has provided some background information on the scientific problem and described some related statistical issues. Chapter 2 provides a general review of the two longitudinal approaches to be utilized. A specific class of GLMM with log link and random intercept is described in detail

1.5. Objectives and Outline of the Report

with specific features highlighted. The regression parameters of this log-linear model can be estimated by both longitudinal approaches.

Chapter 3 presents the results of the longitudinal analyses based on daily clinical outcomes. The log-linear model described in Chapter 2 is utilized. Computational issues that arose when fitting this model to our data are discussed. The adequacy and appropriateness of a linear time trend is first assessed. When the form of the time trend is decided, the relative efficiency of the GEE approach in estimating the NAB effects is examined. The primary assessment of NAB effects is based on comparison of NAB+ periods to NAB− periods. Association between the relapse rate and the NAB titers is estimated to reveal any biological gradients of relapse rate with titer value.

Chapter 4 considers collapsing the daily relapse response to a less frequent response. The collapsed response is the number of relapses between consecutive acquisitions of serum specimens. The log-linear model discussed in Chapter 2 is applied to the collapsed count data. The loss in efficiency due to reduction in resolution of the response, as well as the efficiency of the GEE approach relative to the GLMM approach is examined.

Chapter 5 describes and presents the results of simulation studies investigating the finite sample properties of the GLMM and the GEE approaches for the type of models utilized in Chapter 3. We focus on the relative efficiency of GEE when GLMM is correctly specified and the robustness of GLMM under some form of misspecification of the random intercept. Chapter 6 summarizes the overall findings and discusses problems that remain to be investigated in the future.

Chapter 2

Methodology

2.1 Generalized Linear Models

The class of generalized linear models (GLMs) was first developed by Nelder and Wedderburn[7] to unify and extend various regression models, such as classical linear models, logistic and probit regression models for binary observations and log-linear models for contingency tables. The following notations are established for the discussion of cross-sectional analysis in Sections 2.1 and 2.2. $\mathbf{y} = (y_1, \dots, y_n)^T$ is a $n \times 1$ vector of observations or responses which is a realization of a $n \times 1$ vector of random variables $\mathbf{Y} = (Y_1, \dots, Y_n)^T$. Each response y_i is associated with a $p \times 1$ vector of covariates \mathbf{x}_i . These vectors of covariates \mathbf{x}_i can be grouped into a $n \times p$ covariate matrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$.

The GLM model assumption has two components, a stochastic component and a systematic component. The stochastic component assumes that each component of \mathbf{Y} independently follows a distribution in the exponential family with density function $f_Y(\cdot)$ given by

$$f_{Y_i}(y_i; \theta_i, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}, \quad (2.1.1)$$

where θ_i is a location parameter and ϕ is a dispersion parameter. The dispersion parameter is known for some distributions. For example, $\phi = 1$ for the binomial and Poisson distributions, the most commonly used distributions in GLMs for binary and count responses respectively. For our scope of discussion, we assume a common constant dispersion $a(\phi) = \phi$ across observations.

2.1. Generalized Linear Models

The mean and variance of Y_i are given by

$$E(Y_i) = b'(\theta_i) = \mu_i \quad (2.1.2)$$

and

$$\text{Var}(Y_i) = \phi \cdot b''(\theta_i) = \phi \cdot v(\mu_i). \quad (2.1.3)$$

The variance is determined by the mean through the known variance function $v(\cdot)$, up to a scalar multiple represented by the dispersion parameter ϕ .

The systematic component assumes that the effects of the covariates can be represented by a linear predictor,

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}, \quad (2.1.4)$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is a $p \times 1$ vector of unknown regression coefficients. In matrix notation, $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, where $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$. A known link function $g(\cdot)$ then relates the linear predictor to the corresponding mean,

$$g(\mu_i) = \eta_i, \quad (2.1.5)$$

which completes the specification of the mean structure. The link function $g(\cdot)$ can be any monotone differentiable function but there are common choices for specific types of response variable.

The log likelihood function of a GLM has a general form,

$$l(\boldsymbol{\theta}, \phi; \mathbf{y}) = \sum_{i=1}^n l_i(\theta_i, \phi; y_i) = \sum_{i=1}^n \left\{ \frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right\}. \quad (2.1.6)$$

By recognizing from (2.1.1) and (2.1.2) that

$$\frac{\partial l_i}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{\phi} = \frac{y_i - \mu_i}{\phi},$$

and from (2.1.3) that

$$\frac{d\theta_i}{d\mu_i} = \frac{d\theta_i}{db'(\theta_i)} = \frac{1}{b''(\theta_i)} = \frac{1}{v(\mu_i)},$$

the score function, the first derivative of the log likelihood function (2.1.6), is given by

$$\begin{aligned} s(\boldsymbol{\beta}; \phi, \mathbf{y}) &= \frac{\partial l}{\partial \boldsymbol{\beta}^T} = \sum_{i=1}^n \frac{\partial l_i}{\partial \theta_i} \frac{d\theta_i}{d\mu_i} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}^T} \\ &= \sum_{i=1}^n \frac{\partial \mu_i}{\partial \boldsymbol{\beta}^T} \frac{1}{\phi \cdot v(\mu_i)} (y_i - \mu_i). \end{aligned} \quad (2.1.7)$$

From (2.1.4), we have

$$\frac{\partial \mu_i}{\partial \boldsymbol{\beta}^T} = \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}^T} = \frac{d\mu_i}{d\eta_i} \mathbf{x}_i,$$

so the score function can be re-expressed as

$$s(\boldsymbol{\beta}; \phi, \mathbf{y}) = \frac{1}{\phi} \sum_{i=1}^n \mathbf{x}_i w_i \frac{d\eta_i}{d\mu_i} (y_i - \mu_i),$$

where $w_i = \left(\frac{d\mu_i}{d\eta_i} \right)^2 \frac{1}{v(\mu_i)}$.

If the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ is an interior point in the parameter space, it is a solution to the estimating equations

$$\sum_{i=1}^n \mathbf{x}_i w_i \frac{d\eta_i}{d\mu_i} (y_i - \mu_i) = \mathbf{0},$$

which can be expressed in matrix notation as

$$\mathbf{X}^T \mathbf{W} \mathbf{\Delta} (\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0}, \quad (2.1.8)$$

with $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$, derivative matrix $\mathbf{\Delta} = \text{diag}\left(\frac{d\eta_1}{d\mu_1}, \dots, \frac{d\eta_n}{d\mu_n}\right)$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$. Under the assumption of constant dispersion, solving for the MLE of $\boldsymbol{\beta}$ does not involve the dispersion parameter ϕ .

Nelder and Wedderburn[7, 8] provided a generic model fitting algorithm, the iteratively reweighted least squares (IRWLS) method on transformed dependent variables, for solving the GLM maximum likelihood estimating equations (2.1.8). IRWLS is computationally simple and is equivalent to Fisher's scoring method first proposed by Fisher in the context of probit analysis.

Fisher's scoring method is a variant of Newton's method where the Hessian matrix is replaced by its expected value. Omitting the dispersion parameter, the expected value of the Hessian matrix is given by

$$\begin{aligned} \mathbb{E} \left[\frac{d^2 l}{d\boldsymbol{\beta} d\boldsymbol{\beta}^T} \right] &= \sum_{i=1}^n \mathbb{E} \left[(y_i - \mu_i) \frac{d}{d\boldsymbol{\beta}} \left(\mathbf{x}_i w_i \frac{d\eta_i}{d\mu_i} \right) + \mathbf{x}_i w_i \frac{d\eta_i}{d\mu_i} \frac{d}{d\boldsymbol{\beta}} (y_i - \mu_i) \right] \\ &= \sum_{i=1}^n \left[\frac{d}{d\boldsymbol{\beta}} \left(\mathbf{x}_i w_i \frac{d\eta_i}{d\mu_i} \right) \mathbb{E} (y_i - \mu_i) - \mathbf{x}_i w_i \frac{d\eta_i}{d\mu_i} \frac{d\mu_i}{d\boldsymbol{\beta}} \right] \\ &= - \sum_{i=1}^n \mathbf{x}_i w_i \frac{d\eta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{d\boldsymbol{\beta}} \\ &= - \sum_{i=1}^n \mathbf{x}_i w_i \mathbf{x}_i^T, \end{aligned}$$

which can be expressed as

$$\mathbb{E} \left[\frac{d^2 l}{d\boldsymbol{\beta} d\boldsymbol{\beta}^T} \right] = -\mathbf{X}^T \mathbf{W} \mathbf{X}.$$

2.1. Generalized Linear Models

Given the current estimates at the k^{th} iteration of Fisher's scoring method, $\hat{\boldsymbol{\beta}}^{(k)}$, $\hat{\mu}^{(k)}$, $\hat{\boldsymbol{\eta}}^{(k)}$, $\mathbf{W}^{(k)}$ and $\boldsymbol{\Delta}^{(k)}$, the updated $\hat{\boldsymbol{\beta}}^{(k+1)}$ is obtained by

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \hat{\boldsymbol{\beta}}^{(k)} + (\mathbf{X}^T \mathbf{W}^{(k)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(k)} \boldsymbol{\Delta}^{(k)} (\mathbf{y} - \hat{\boldsymbol{\mu}}^{(k)}). \quad (2.1.9)$$

Fisher's scoring method can be shown to be equivalent to IRWLS by rearranging (2.1.9) as,

$$\begin{aligned} \mathbf{X}^T \mathbf{W}^{(k)} \mathbf{X} (\hat{\boldsymbol{\beta}}^{(k+1)} - \hat{\boldsymbol{\beta}}^{(k)}) &= \mathbf{X}^T \mathbf{W}^{(k)} \boldsymbol{\Delta}^{(k)} (\mathbf{y} - \hat{\boldsymbol{\mu}}^{(k)}) \\ \Rightarrow \hat{\boldsymbol{\beta}}^{(k+1)} &= (\mathbf{X}^T \mathbf{W}^{(k)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(k)} (\mathbf{X} \hat{\boldsymbol{\beta}}^{(k)} + \boldsymbol{\Delta}^{(k)} (\mathbf{y} - \hat{\boldsymbol{\mu}}^{(k)})) \\ &= (\mathbf{X}^T \mathbf{W}^{(k)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(k)} (\hat{\boldsymbol{\eta}}^{(k)} + \boldsymbol{\Delta}^{(k)} (\mathbf{y} - \hat{\boldsymbol{\mu}}^{(k)})). \end{aligned} \quad (2.1.10)$$

This updating is thus equivalent to solving weighted least squares equations with the transformed dependent variable

$$z_i^{(k)} = \hat{\eta}_i^{(k)} + \frac{d\eta_i}{d\mu_i} (y_i - \hat{\mu}_i^{(k)})$$

and weights

$$w_i^{(k)} = \left(\frac{d\mu_i}{d\eta_i} \right)^2 \frac{1}{v(\hat{\mu}_i^{(k)})}.$$

In accordance with standard likelihood theory, $\sqrt{n} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges in distribution to normal with mean $\mathbf{0}$ and covariance matrix $\phi \cdot (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}$, as n tends to infinity.

As already noted, the MLE of $\boldsymbol{\beta}$ does not depend on the dispersion parameter ϕ . However, as $\text{Var}(\hat{\boldsymbol{\beta}})$ depends on ϕ , the dispersion parameter needs to be estimated for any statistical inference. In case of a common dispersion

2.2. The Quasi-likelihood Approach

parameter, ϕ can be consistently estimated by the moment estimator

$$\tilde{\phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{v(\hat{\mu}_i)}, \quad (2.1.11)$$

where $\hat{\mu}_i = g^{-1}(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})$ are the MLEs of the means.

2.2 The Quasi-likelihood Approach

In practice, we may be able to justify a specified mean structure and variance function for observed responses. However, there may not be any distribution in the exponential family or any actual probability distribution which has the desired mean-variance structure and range of possible response values.

Wedderburn[9] proposed the quasi-likelihood approach as an alternative to the GLM approach when inference on the regression parameters is the main interest. In the quasi-likelihood approach, the responses Y_i 's are assumed independent with mean μ_i and variance $\phi \cdot v(\mu_i)$, where $v(\cdot)$ is a known variance function and the dispersion parameter ϕ may be unknown but is assumed to be constant. The mean μ_i depends on covariates \mathbf{x}_i as in (2.1.4) and (2.1.5) for GLM. No further assumption is required regarding the distribution of Y_i .

A quasi-likelihood function, if it exists, is defined as

$$Q(\boldsymbol{\mu}, \phi; \mathbf{y}) = \sum_{i=1}^n \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi \cdot v(t)} dt,$$

which is anticipated to exhibit similar behaviour to a log-likelihood function under mild assumptions. In some cases, the quasi-likelihood function is the GLM log-likelihood function with the corresponding mean-variance

structure. The maximum quasi-likelihood estimate $\tilde{\boldsymbol{\beta}}$ of the regression parameters maximizes the quasi-likelihood function. When this estimate is an interior point solution, it solves the quasi-likelihood estimating equations which take the same form as (2.1.7) and thus can be solved by Fisher's algorithm (2.1.9) or IRWLS (2.1.10) as for GLM. The resulting quasi-likelihood estimator $\tilde{\boldsymbol{\beta}}$ is consistent and $\sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges in distribution to normal with mean $\mathbf{0}$ and covariance matrix $\phi \cdot (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}$ as n tends to infinity.

In general, the dispersion parameter can be consistently estimated by the same moment estimator in (2.1.11), given the quasi-likelihood estimates of the means, $\tilde{\boldsymbol{\mu}}_i = g^{-1}(\mathbf{x}_i^T \tilde{\boldsymbol{\beta}})$.

2.3 Marginal Models for Longitudinal Data

2.3.1 Overview

For longitudinal data, observations on a response are taken repeatedly on each subject over a period of time. The repeated measures on the same subject are likely to be statistically dependent due to unobservable or unmeasured characteristics of the subjects. An assumption of independence of the repeated measures is usually not justified. Direct application of GLM to longitudinal data is thus inappropriate. Alternative approaches which account for the dependence are needed in order to draw valid statistical inferences. Two extensions to GLM applicable to longitudinal data will be discussed and compared. First, we focus on a class of marginal models which is a straightforward extension of GLM to longitudinal data. In contrast to the class of mixed-effects models which will be discussed in the next section, this class of marginal models enable separate construction of the mean structure and the dependence structure.

2.3. Marginal Models for Longitudinal Data

Suppose m subjects are followed over time with n_i repeated measures on the i^{th} subject at times $t_{ij}, j = 1, \dots, n_i$. A total of $N = \sum_{i=1}^m n_i$ measures are recorded. Let observation y_{ij} be the realization of Y_{ij} , the response variable of the j^{th} measure on the i^{th} subject at time t_{ij} . Each response is associated with a $p \times 1$ vector of covariates \mathbf{x}_{ij} . These covariates can be invariant or varying over time. The covariate vectors for the same individual can be grouped in a $n_i \times p$ matrix, $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i})^T$. The vectors of responses $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ are independent across subjects.

A marginal model has a three-fold specification. Firstly, the marginal mean $E(Y_{ij}) = \mu_{ij}$ depends on the covariates through a known link function $g(\cdot)$ as

$$g(\mu_{ij}) = \eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta},$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is a vector of p regression coefficients. Secondly, the marginal variance $\text{Var}(Y_{ij})$ depends on the marginal mean through a known variance function $v(\cdot)$ that is scaled by a dispersion parameter,

$$\text{Var}(Y_{ij}) = \phi \cdot v(\mu_{ij}).$$

The dispersion parameter can be time dependent or modelled as a function of covariates, but for our scope of discussion, we assume a constant dispersion across subjects and times. The first two components of the model specification are the same as in GLM, whereas the third component is the extension to accommodate longitudinal data by specification of the pairwise dependence of responses of the same individual. We depict the dependence structure by pairwise correlations where the $n_i \times n_i$ correlation matrix $\mathbf{R}_i(\boldsymbol{\alpha})$ of \mathbf{Y}_i is completely specified by a $s \times 1$ vector of association parameters $\boldsymbol{\alpha}$. The variance-covariance matrix of \mathbf{Y}_i is thus given by

$$\mathbf{V}_i = \phi \cdot \mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}, \quad (2.3.1)$$

2.3. Marginal Models for Longitudinal Data

where $\mathbf{A}_i = \text{diag}(v(\mu_{i1}), \dots, v(\mu_{in_i}))$.

With only the mean structure and the variance-covariance structure specified, the joint distribution of \mathbf{Y}_i is not completely determined in general. For continuous data, the joint distribution of the longitudinal observations are often assumed to be multivariate normal which is completely specified by the first two moments. For discrete data, there is no simple analogue to the multivariate normal distribution. Higher order moment assumptions are involved in constructing a joint distribution for non-normal data which becomes impractical. For example, Bahadur[10] proposed a parameterization for multivariate binary responses $\mathbf{Y} = (y_1, \dots, y_n)^T$,

$$P(\mathbf{Y} = \mathbf{y}) = \prod_{i=1}^n \mu_i^{y_i} (1 - \mu_i)^{1-y_i} \times \left(1 + \sum_{i < j} \rho_{ij} r_i r_j + \sum_{i < j < k} \rho_{ijk} r_i r_j r_k + \dots + \rho_{1, \dots, n} r_1 \dots r_n \right),$$

where r_i is the realization of $R_i = \frac{Y_i - \mu_i}{\sqrt{\mu_i(1-\mu_i)}}$, $\rho_{ij} = \text{Corr}(Y_i, Y_j) = E(R_i R_j)$, $\rho_{ijk} = E(R_i R_j R_k), \dots$, and $\rho_{1, \dots, n} = E(R_1 \dots R_n)$. Diggle et al.[11] and Chaganty and Joe[12] pointed out that the correlations among the binary responses are constrained by the marginal means in complicated ways with the upper and lower bounds determined by the Fréchet bounds. When the marginal mean depends on some covariates, the convenient assumption that the pairwise and higher-order correlations are independent of the covariates is impossible.

When a full likelihood is not available, alternative approaches to maximum likelihood estimation must be employed for inference on regression parameters. One alternative is the generalized estimating equations (GEE) approach.

2.3.2 Inferential Approach: Generalized Estimating Equations

Liang and Zeger[13, 14] proposed the generalized estimating equations (GEE) approach as an extension of generalized linear models for longitudinal data to obtain consistent estimates of regression coefficients and their standard errors. The GEE approach does not require the assumption of an exponential family distribution for the marginal model. Therefore, it can be regarded more generally as an extension of the quasi-likelihood approach for clustered data. The GEE approach provides a framework for estimation for the class of marginal models in the previous section.

Due to the often complicated underlying stochastic process that generates the longitudinal data, the true correlation structure is usually difficult to determine. When the dependence of the mean on the covariates is of primary interest, the regression parameters β are the main target of inference and the association parameters α can be treated as nuisance parameters. In these circumstances, it would be desirable if valid inference for the regression parameters could be achieved without having to specify the correlation structure correctly.

The GEE approach gives a consistent estimate of β and the asymptotic variance-covariance matrix of this estimate regardless of the true correlation structure. The specific estimating equations of the GEE approach depend upon a so-called “working” correlation matrix, also denoted $\mathbf{R}_i(\alpha)$, but this need not be the true correlation matrix. Common choices of “working” correlation structure include,

1. **Independence** where $\mathbf{R}_i(\alpha) = I_{n_i}$, the $n_i \times n_i$ identity matrix;

2. **Exchangeable** where

$$\text{Corr}(Y_{it}, Y_{it'}) = \alpha, \quad (2.3.2)$$

for all $t \neq t'$;

3. **AR-1**, a continuous time analogue of the first-order autoregressive process, where

$$\text{Corr}(Y_{it}, Y_{it'}) = \alpha^{|t-t'|}$$

for all $t \neq t'$; and

4. **m-dependence** where

$$\text{Corr}(Y_{it}, Y_{it+l}) = \begin{cases} \alpha_l, & \text{for } l = 1, \dots, m \\ 0, & \text{for } l = m+1, \dots, n_i \end{cases}.$$

GEE can be viewed as quasi-likelihood estimating equations for longitudinal data taking the form

$$\sum_{i=1}^m \mathbf{D}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \quad (2.3.3)$$

where $\mathbf{D}_i = \frac{d\boldsymbol{\mu}_i}{d\boldsymbol{\beta}}$ is a $n_i \times p$ matrix,

$$\mathbf{D}_i = \begin{pmatrix} \partial\mu_{i1}/\partial\beta_1 & \cdots & \partial\mu_{i1}/\partial\beta_p \\ \vdots & \ddots & \vdots \\ \partial\mu_{in_i}/\partial\beta_1 & \cdots & \partial\mu_{in_i}/\partial\beta_p \end{pmatrix},$$

and $\mathbf{V}_i = \phi \cdot \mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}$ as specified in (2.3.1) but with $\mathbf{R}_i(\boldsymbol{\alpha})$ being the “working” correlation matrix. Since \mathbf{D}_i and \mathbf{V}_i do not depend on \mathbf{y}_i , the left-hand side of (2.3.3) has expectation equal to $\mathbf{0}$, so the estimating equations are unbiased irrespective of the choice of “working” correlation matrix. The

2.3. Marginal Models for Longitudinal Data

roots of the estimating equations are thus consistent if the mean structure is correctly specified so that $E(\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$.

Estimating equations (2.3.3) can be solved iteratively by a modified Fisher's scoring algorithm augmented with moment estimation of $\boldsymbol{\alpha}$ and ϕ . Regarding the left hand side of (2.3.3) as the estimating function, the expected value of its first derivative matrix is given by $-\sum_{i=1}^m \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i$. Hence, given the k^{th} step estimates $\hat{\boldsymbol{\beta}}^{(k)}$, $\hat{\boldsymbol{\alpha}}^{(k)}$ and $\hat{\phi}^{(k)}$, the updated $\hat{\boldsymbol{\beta}}^{(k+1)}$, $\hat{\boldsymbol{\alpha}}^{(k+1)}$ and $\hat{\phi}^{(k+1)}$ can be obtained by a two-stage iterative procedure:

1. Compute $\hat{\boldsymbol{\mu}}^{(k)}$, $\mathbf{D}_i^{(k)}$ and $\mathbf{V}_i^{(k)}$ based on the current estimates $\hat{\boldsymbol{\beta}}^{(k)}$, $\hat{\boldsymbol{\alpha}}^{(k)}$ and $\hat{\phi}^{(k)}$. Then, update $\hat{\boldsymbol{\beta}}^{(k+1)}$ by the modified Fisher's scoring algorithm,

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \hat{\boldsymbol{\beta}}^{(k)} + \left\{ \sum_{i=1}^m \mathbf{D}_i^{(k)T} \mathbf{V}_i^{(k)-1} \mathbf{D}_i^{(k)} \right\}^{-1} \left\{ \sum_{i=1}^m \mathbf{D}_i^{(k)T} \mathbf{V}_i^{(k)-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i^{(k)}) \right\},$$

or equivalently by

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \left\{ \sum_{i=1}^m \mathbf{D}_i^{(k)T} \mathbf{V}_i^{(k)-1} \mathbf{D}_i^{(k)} \right\}^{-1} \left\{ \sum_{i=1}^m \mathbf{D}_i^{(k)T} \mathbf{V}_i^{(k)-1} \mathbf{z}_i \right\}, \quad (2.3.4)$$

where

$$\mathbf{z}_i = \mathbf{D}_i^{(k)} \hat{\boldsymbol{\beta}}^{(k)} + (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i^{(k)}).$$

Step (2.3.4) is equivalent to weighted least squares with transformed dependent variable \mathbf{z}_i , covariate matrix \mathbf{D}_i and weight matrix \mathbf{V}_i^{-1} .

2. After obtaining the fitted mean $\hat{\boldsymbol{\mu}}^{(k+1)}$ from the updated regression coefficients $\hat{\boldsymbol{\beta}}^{(k+1)}$, the standardized residuals

$$r_{ij}^{(k+1)} = \frac{(y_{ij} - \hat{\mu}_{ij}^{(k+1)})}{\sqrt{v(\hat{\mu}_{ij}^{(k+1)})}}, \quad (2.3.5)$$

2.3. Marginal Models for Longitudinal Data

form the basis for estimation of ϕ and $\boldsymbol{\alpha}$. The association parameter, $\boldsymbol{\alpha}$, is estimated by a \sqrt{m} -consistent estimator, $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi)$, which is a function of the data, $\boldsymbol{\beta}$ and ϕ . The unknown dispersion parameter, ϕ , in $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi)$ is estimated by a \sqrt{m} -consistent estimator, $\hat{\phi}(\boldsymbol{\beta})$, which is a function of the data and $\boldsymbol{\beta}$. The unknown $\boldsymbol{\beta}$ in both $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi)$ and $\hat{\phi}(\boldsymbol{\beta})$ is replaced by $\hat{\boldsymbol{\beta}}^{(k+1)}$ evaluated in step 1.

The two steps: the updating to $\hat{\boldsymbol{\beta}}^{(k+1)}$ and the estimation of ϕ and $\boldsymbol{\alpha}$ are iterated until convergence. An initial value for $\boldsymbol{\beta}$ is usually obtained from a GLM fit pretending the observations are independent.

Under mild regularity conditions, as the number of subjects m tends to infinity, $\sqrt{m}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically p -variate normal with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_{\text{GEE}}$ given by

$$\boldsymbol{\Sigma}_{\text{GEE}} = \lim_{m \rightarrow \infty} m \cdot \left(\sum_{i=1}^m \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \right)^{-1} \left(\sum_{i=1}^m \mathbf{D}_i^T \mathbf{V}_i^{-1} \text{Cov}(\mathbf{Y}_i) \mathbf{V}_i^{-1} \mathbf{D}_i \right) \left(\sum_{i=1}^m \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \right)^{-1}. \quad (2.3.6)$$

For finite samples, the covariance matrix of $\hat{\boldsymbol{\beta}}$ can be approximated by the so-called “sandwich estimate”,

$$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}} \approx \left(\sum_{i=1}^m \hat{\mathbf{D}}_i^T \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i \right)^{-1} \left(\sum_{i=1}^m \hat{\mathbf{D}}_i^T \hat{\mathbf{V}}_i^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^T \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i \right) \left(\sum_{i=1}^m \hat{\mathbf{D}}_i^T \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i \right)^{-1}, \quad (2.3.7)$$

where the term $\text{Cov}(\mathbf{Y}_i)$ in the asymptotic covariance matrix (2.3.6) is replaced by a simple moment estimate $(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^T$, while \mathbf{D}_i and \mathbf{V}_i are also replaced by their corresponding estimates. The consistency of $\hat{\boldsymbol{\beta}}$ and the covariance matrix follows provided the mean structure is correctly specified regardless of the choice of “working” correlation structure. The validity

of these asymptotic results does not depend on the specific choice of estimators of $\boldsymbol{\alpha}$ and ϕ as long as they are \sqrt{m} -consistent. Thus, the GEE approach is robust to the “working” correlation assumption. However, higher statistical efficiency is achieved if the “working” correlation structure is closer to the true correlation structure[13].

Liang and Zeger[13, 14] also proposed the following moment estimators of $\boldsymbol{\alpha}$ and ϕ . Given the standardized residuals (2.3.5) at a certain iteration, the moment estimator of ϕ is given by

$$\hat{\phi} = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} r_{ij}^2}{N - p}.$$

The estimator of $\boldsymbol{\alpha}$ depends on the specific structure of $\mathbf{R}(\boldsymbol{\alpha})$. For example, when α represents the pairwise correlation in an exchangeable correlation structure as defined in (2.3.2), for given ϕ , it can be estimated by

$$\hat{\alpha} = \frac{\phi \cdot \sum_{i=1}^m \sum_{j>k} r_{ij} r_{ik}}{\sum_{i=1}^m \frac{1}{2} n_i (n_i - 1) - p}.$$

The sandwich estimator of variance may exhibit bias when the number of clusters is small, e.g. $m < 30$. Paik[15] proposed the jackknife variance estimator as an improvement to the sandwich estimator. This jackknife estimator is defined as

$$\tilde{\Sigma}_{\hat{\beta}} = \frac{m-p}{m} \sum_{i=1}^m \left(\hat{\beta}_{-i} - \hat{\beta} \right) \left(\hat{\beta}_{-i} - \hat{\beta} \right)^T, \quad (2.3.8)$$

where $\hat{\beta}_{-i}$ is the estimate obtained for β when the i^{th} subject is removed from the dataset. In a simulation study, the jackknife variance estimator was shown to have superior properties over the sandwich variance estimator. This jackknife estimator is sometimes referred to as the fully iterated jackknife estimator. In each iteration, the model is fitted by fully iterated

Newton method to the dataset with one cluster removed. Since the model needs to be fitted as many times as the number of clusters, this jackknife estimator is computationally intensive when the number of clusters is moderate or large. Halekoh et al.[16] implemented the sandwich estimator, the fully iterated jackknife estimator as well as an approximate jackknife estimator and an one-step jackknife estimator in the R package **geepack**. For the one-step jackknife estimator, each $\hat{\beta}_{-i}$ is obtained by one-step Newton iteration. Their simulations and other previous simulations indicate that the latter two computationally less demanding versions of jackknife estimates are in many cases accurate approximates to the fully iterated jackknife variance estimate.

This class of marginal models is a suitable modelling approach in many clinical trials and observational studies when the average effects of treatments or exposures in the population of interest are the primary target of inference. The regression parameters are directly interpretable as the population average effects. This interpretation is invariant to the specification of the correlation structure, or more generally the association structure.

GEE permits consistent point estimation and interval estimation under weak assumptions. Hypothesis tests for contrasts of β are based on the resulting Wald-type statistics. Suppose we are interested in testing $H_0 : \mathbf{C}\beta = \mathbf{0}$ against $H_0 : \mathbf{C}\beta \neq \mathbf{0}$, with \mathbf{C} being a contrast matrix of rank q . The Wald-type statistic

$$W^2 = \hat{\beta}^T \mathbf{C}^T \left(\mathbf{C} \hat{\Sigma}_{\hat{\beta}} \mathbf{C}^T \right)^{-1} \mathbf{C} \hat{\beta} \quad (2.3.9)$$

asymptotically follows a χ^2 -distribution with q degrees of freedom. The Wald-type test can be used to compare nested models if the reduced model can be obtained by setting some contrasts in the full model to zero. However, likelihood-based tools such as likelihood ratio tests are not available. There

are also no widely accepted criteria, such as the Akaike Information Criterion (AIC), for comparing non-nested models. Thus, model comparisons are sometimes difficult, especially for non-nested models.

2.4 Generalized Linear Mixed-effect Models

2.4.1 Overview

Generalized linear mixed-effect models (GLMM) is another extension of generalized linear models for repeated measures and longitudinal data. In GLMM, some of the regression coefficients are assumed to vary across subjects to reflect natural between-subject heterogeneity due to unmeasured and unobservable factors.

Given a $q \times 1$ vector of random effects \mathbf{b}_i , Y_{ij} is assumed to have a conditional distribution in the exponential family with density function, $f_Y(\cdot)$, as (2.1.1). Conditional on the random effects, the Y_{ij} 's are usually assumed to be independent of one another. This is sometimes referred to as the conditional independence assumption. Though the conditional independence assumption is not necessarily required, it is invariably adopted for discrete responses since direct modelling of correlation structure is difficult.

The conditional mean of Y_{ij} is modelled as

$$g(E[Y_{ij}|\mathbf{b}_i]) = \eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{u}_{ij}^T \mathbf{b}_i,$$

where $g(\cdot)$ is a known link function and the $q \times 1$ vector \mathbf{u}_{ij} is a subset of the $p \times 1$ vector of covariates \mathbf{x}_{ij} . The random effects \mathbf{b}_i are assumed to be independently and identically distributed with density $f_{\mathbf{b}}(\cdot)$, having mean $E(\mathbf{b}_i) = \mathbf{0}$ and depending on a set of parameters $\mathbf{V}_{\mathbf{b}}$. The random effects distribution is often assumed to be multivariate normal for mathematical

convenience. In this case, \mathbf{V}_b represents the covariance matrix of the random effects.

When the conditional independence assumption is made, association among the repeated measures is not modelled directly. Instead, within subject correlation is entirely induced by the random effects. In the case of linear mixed-effects models where the continuous responses are assumed to be conditionally independent and normal with common variance σ^2 and an identity link $g(t) = t$ is used, the induced marginal covariance structure is given by

$$\text{Cov}(\mathbf{Y}_i) = \mathbf{U}_i \mathbf{V}_b \mathbf{U}_i^T + \sigma^2 \mathbf{I}_i,$$

where $\mathbf{U}_i = (\mathbf{u}_{i1} \cdots \mathbf{u}_{in_i})^T$ is the $n_i \times q$ matrix of covariates corresponding to the random effects and \mathbf{I}_i is the $n_i \times n_i$ identity matrix. For discrete responses, a nonlinear link would usually be used and the induced covariance matrix is often mathematically intractable.

The regression coefficients β are usually referred to as the fixed effects. In the general case, they can be interpreted as the effects of the covariates on an individual whose random effects \mathbf{b}_i are equal to $\mathbf{0}$. In some cases the random effects are assumed only for the time independent covariates, representing heterogeneity at baseline levels. In that case, the fixed effects for the time dependent covariates are interpreted as the effects of the covariates at the subject level. These common effects across individuals do not necessarily represent the effects at population level. Therefore, mixed-effects models are sometimes called subject-specific models, in contrast the marginal models in Section 2.3 which are referred to as population-averaged models.

When an identity link is used, the conditional mean structure of GLMM,

$$\text{E}[Y_{ij} | \mathbf{b}_i] = \mathbf{x}_{ij}^T \beta + \mathbf{u}_{ij}^T \mathbf{b}_i,$$

and its implied marginal mean structure,

$$E[Y_{ij}] = \mathbf{x}_{ij}^T \boldsymbol{\beta},$$

happen to have the same functional form for $\boldsymbol{\beta}$. In general when nonlinear links are adopted, the implied marginal mean structure,

$$E[Y_{ij}] = E_b[g^{-1}(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{u}_{ij}^T \mathbf{b}_i)],$$

and the conditional mean structure have different functional forms. The functional form of the implied marginal mean structure depends on the distribution of the random effects and the link function. In many cases, such as when a logit link and normal random effects are assumed, an explicit analytic form may not exist. As a consequence, the regression coefficients $\boldsymbol{\beta}$ may not have a straightforward interpretation in terms of the marginal mean structure.

This issue can be illustrated by an example. Suppose given a random intercept b_{i0} and a random slope b_{i1} , the responses Y_{ij} are independent with conditional means $\mu_{ij}^{(C)}$. which depend on the conditional linear predictor $\eta_{ij}^{(C)}$ through a log link,

$$\log \left(\mu_{ij}^{(C)} \right) = \eta^{(C)} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) x_{ij}. \quad (2.4.1)$$

The marginal mean $\mu_{ij}^{(M)}$ can be obtained by taking expectation over the distribution of the random effects,

$$\begin{aligned} \mu_{ij}^{(M)} &= E \left[\mu_{ij}^{(C)} \right] = E \left[\exp \{ (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) x_{ij} \} \right] \\ &= \exp \{ \beta_0 + \beta_1 x_{ij} \} \cdot E \left[\exp \{ b_{i0} + b_{i1} x_{ij} \} \right] \\ &= \exp \{ \beta_0 + \beta_1 x_{ij} \} \cdot M_{\mathbf{b}_0, \mathbf{b}_1} (1, x_{ij}), \end{aligned}$$

where $M_{b_0, b_1}(\cdot, \cdot)$ is the moment-generating function (MGF) of the joint distribution of \mathbf{b}_0 and \mathbf{b}_1 . If \mathbf{b}_0 and \mathbf{b}_1 are statistically independent, the joint MGF is the product of the marginal MGFs $M_{\mathbf{b}_0}(\cdot)$ and $M_{\mathbf{b}_1}(\cdot)$. If the random effects are assumed to follow independent normal distributions with mean 0 and variance σ_0^2 and σ_1^2 respectively, the log transformed marginal mean happens to have a simple analytic form

$$\log(\mu_{ij}^{(M)}) = \left(\beta_0 + \frac{1}{2}\sigma_0^2\right) + \beta_1 x_{ij} + \frac{1}{2}\sigma_1^2 x_{ij}^2, \quad (2.4.2)$$

which is no longer a linear but a quadratic function in x_{ij} . The impact of x_{ij} on the marginal mean $\mu_{ij}^{(M)}$ depends not only on β_1 but also on the variability of the random slope σ_1^2 . The regression coefficient β_1 does not have a simple interpretation in terms of the population averaged effect of the predictor x . Figure 2.1 shows the conditional means as well as the implied marginal means for parameter values $\beta_0 = -5, \sigma_0 = 2, \beta_1 = 1.5, \sigma_1 = 1$. For the conditional means, the random intercept and random slope are evaluated at the 10th, 20th, ..., 80th and 90th percentiles.

2.4.2 Inferential Approach: Maximum Likelihood

Maximum likelihood estimation is the standard inferential approach for GLMM. The marginal likelihood function for the parameters $\boldsymbol{\beta}$, ϕ and $\mathbf{V}_{\mathbf{b}}$ is

$$L(\boldsymbol{\beta}, \mathbf{V}_{\mathbf{b}}, \phi; \mathbf{y}_1, \dots, \mathbf{y}_m) = \prod_{i=1}^m \int \left\{ \prod_{j=1}^{n_i} f_Y(y_{ij} | \mathbf{b}_i; \mathbf{x}_i, \boldsymbol{\beta}, \phi) \right\} f_{\mathbf{b}}(\mathbf{b}_i; \mathbf{V}_{\mathbf{b}}) d\mathbf{b}_i.$$

The likelihood function does not have a closed form in general. Numerical methods are required to evaluate the integral that yields the marginal likelihood of \mathbf{y}_i

$$L_i(\boldsymbol{\beta}, \mathbf{V}_{\mathbf{b}}, \phi; \mathbf{y}_i) = \int \left\{ \prod_{j=1}^{n_i} f_Y(y_{ij} | \mathbf{b}_i; \mathbf{x}_i, \boldsymbol{\beta}, \phi) \right\} f_{\mathbf{b}}(\mathbf{b}_i; \mathbf{V}_{\mathbf{b}}) d\mathbf{b}_i, \quad (2.4.3)$$

which can also be expressed as an expectation of the conditional likelihood over the distribution of the random effects, $E \left[\prod_{j=1}^{n_i} f_Y(y_{ij} | \mathbf{b}_i; \mathbf{x}_i, \boldsymbol{\beta}, \phi) \right]$. If the random effects are assumed to be normally distributed and the number of random effects q is small, for example fewer than 5, the marginal likelihood (2.4.3) is usually approximated by Gauss-Hermite quadrature. Using a fixed set of K quadrature points and weights $\{(\mathbf{z}_1, w_1), \dots, (\mathbf{z}_K, w_K)\}$, the numerical approximation is simply a weighted sum of the conditional likelihood evaluated at scaled quadrature points, $\mathbf{b}_i = \mathbf{V}_b^{1/2} \mathbf{z}_k$,

$$L_i(\boldsymbol{\beta}, \mathbf{V}_b, \phi; \mathbf{y}_i) \approx \sum_{k=1}^K w_k \cdot \left\{ \prod_{j=1}^{n_i} f_Y(y_{ij} | \mathbf{b}_i = \mathbf{V}_b^{1/2} \mathbf{z}_k; \boldsymbol{\beta}) \right\}.$$

The accuracy of the approximation is higher with more quadrature points. Adaptive Gaussian quadrature is an alternative numerical approximation which shifts the centers of the Gaussian approximation from $\mathbf{0}$ to the posterior mode of the random effects. Higher accuracy is achieved with the same number of quadrature points than with Gauss-Hermite quadrature. Numerical optimization, such as the quasi-Newton method with numerical derivatives, is then applied to locate the maximum of the marginal likelihood function.

Given the MLEs of $\boldsymbol{\beta}$, \mathbf{V}_b and ϕ , inference on the fixed effects $\boldsymbol{\beta}$ follows standard likelihood theory. Hypothesis tests for nested models can be carried out by the likelihood ratio (LR) test, in addition to the Wald-type test (2.3.9) which is the sole option for the GEE approach. The LR test statistic,

$$G^2 = 2(l_{\text{full}} - l_{\text{reduced}}), \quad (2.4.4)$$

follows asymptotically a χ^2 -distribution with s degrees of freedom, where l_{full} and l_{reduced} are the maximum log likelihood under the full model and the reduced model respectively and s is the number of restrictions on the

parameters of the full model for obtaining the reduced model. The LR test is sometimes preferred to the Wald-type test for higher statistical sensitivity and its property of invariance to parameterization.

Testing of variance components by the LR test are non-standard problems. Under the null hypothesis, some variance components are putatively zero, falling on the boundary of the parameter space. In general, the asymptotic distributions of the test statistics (2.4.4) are mixtures of χ^2 -distributions[17]. The form of mixture depends on the number of parameters of interest on the boundary and in the interior space under null hypothesis, and the number of nuisance parameters on the boundary and in the interior space. For the simplest case when a single variance component is the only parameter of interest and all other nuisance parameters lie in the interior space, the asymptotic distribution of the LR test statistic is a 50:50 mixture of χ_0^2 and χ_1^2 distributions.

Random effects \mathbf{b}_i of individual subjects can be predicted by the empirical Bayes estimates,

$$\hat{\mathbf{b}}_i = E \left[\mathbf{b}_i | \mathbf{Y}_i = \mathbf{y}_i; \boldsymbol{\beta} = \hat{\boldsymbol{\beta}}, \mathbf{V}_{\mathbf{b}} = \hat{\mathbf{V}}_{\mathbf{b}}, \phi = \hat{\phi} \right],$$

with the posterior mean of the random effects evaluated as if all the parameters conditioned on are known with values equal to the corresponding plug-in estimates $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{V}}_{\mathbf{b}}$ and $\hat{\phi}$. The integration yielding the conditional expectation typically needs to be evaluated by numerical methods.

The maximum likelihood estimates of the variance components of regression models are well-known to be underestimates in small samples. For example, in simple linear models, the maximum likelihood estimate of the variance is obtained by dividing the residual sum of squares by n , the number of observations. An unbiased estimate of this variance is obtained by replacing n in the denominator of the maximum likelihood estimator by the

residual degrees of freedom. For general linear models or linear mixed-effects models, the bias adjustment is more than a simple adjustment to the denominator. This motivated a variant of maximum likelihood estimation, namely restricted maximum likelihood (REML) estimation, first introduced by Patterson and Thompson[18]. One explanation of the underestimation is that maximum likelihood does not take into account the fact that β is also estimated, while the estimator of the variance components is not orthogonal to the estimator of β . The main idea of REML is therefore to estimate the variance components by maximum likelihood upon a partition of the data whose distribution does not depend on β but only on the variance components. In the case of general linear model, one way to partition the data is to use the ordinary least squares residuals for estimation of the variance components. After obtaining the REML estimates of the variance components, β is estimated by the usual general least squares approach.

The notion of REML generalizes to GLMM, but the estimation algorithm is less straightforward than for linear models. For example, McCulloch considered a class of probit normal models for binary data and described a version of REML implemented with the EM algorithm[19]. As the bias of maximum likelihood for variance components diminishes when the sample size is substantially larger than the dimension of β , the difference between maximum likelihood and REML becomes less important[20].

2.5 Log-linear Models

2.5.1 Mixed-effects Models

In this section, we describe a class of log-linear models with random intercept which will be utilized for data analysis in the following chapters. Given a univariate random effect b_i , the responses Y_{ij} are assumed to be conditionally independent, following some distribution in the exponential family (2.1.1)

2.5. Log-linear Models

with conditional mean $E(Y_{ij}|b_i) = \mu_{ij}^{(C)}$. The conditional mean depends on the random intercept and a $p \times 1$ vector of covariates \mathbf{x}_{ij} through a log link,

$$\log(\mu_{ij}^{(C)}) = b_i + \eta_{ij} = b_i + \beta_0 + \mathbf{x}_{ij}^T \boldsymbol{\beta},$$

where β_0 is the overall intercept and $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects. The random intercept is assumed to follow some distribution with density function $f_b(\cdot)$ having mean $E(b_i) = 0$ and variance $\text{Var}(b_i) = \sigma_b^2$.

Conditional on the random effect, the regression coefficient corresponding to a continuous covariate can be interpreted as the log relative risk for a binary response or the log relative rate for a count response for every unit increment of the covariate, adjusted for all other covariates. When the covariate is binary, representing two exposures or treatments, the regression coefficient can be interpreted as the adjusted log relative risk or relative rate comparing the two exposures or treatments. The likelihood-based approach for GLMM described in Section 2.4 can be applied to analyze data using this log-linear random intercept model.

2.5.2 Marginalized Models

The marginal mean structure for the log-linear mixed-effects model can be obtained by taking the expectation of the conditional mean over the distribution of the random intercept. Denoting the marginal mean as

$$\mu_{ij}^{(M)} = E(Y_{ij}) = E[E(Y_{ij}|b_i)] = E(\mu_{ij}^{(C)}),$$

2.5. Log-linear Models

the marginal mean structure is given by

$$\begin{aligned}
 \log \left(\mu_{ij}^{(M)} \right) &= \eta_{ij}^{(M)} = \log \left(\mathbb{E} \left[\exp(b_i + \beta_0 + \mathbf{x}_{ij}^T \boldsymbol{\beta}) \right] \right) \\
 &= \log \left(\mathbb{E} \left[e^{b_i} \right] \right) + \beta_0 + \mathbf{x}_{ij}^T \boldsymbol{\beta} \\
 &= \log \left(M_b(1) \right) + \beta_0 + \mathbf{x}_{ij}^T \boldsymbol{\beta} \\
 &= \log \left(M_b(1) \right) + \eta_{ij}
 \end{aligned} \tag{2.5.1}$$

where $M_b(\cdot)$ is the MGF of the random intercept b_i . The constant term $\log(M_b(1))$ can be absorbed into the intercept, yielding

$$\log \left(\mu_{ij}^{(M)} \right) = \beta_0^* + \mathbf{x}_{ij}^T \boldsymbol{\beta}. \tag{2.5.2}$$

In this model, the marginal mean structure has the same functional form as the conditional mean structure. It follows that the regression parameters $\boldsymbol{\beta}$ have the same interpretation at the individual level as well as at the population averaged level regardless of the distribution of the random intercept. The only exception is the intercept term which depends on the distribution of the random intercept as the MGF of the random effect enters the intercept term in the implied marginal mean structure. In the special case of a normal random effect distribution with variance σ_b^2 ,

$$\beta_0^* = \beta_0 + \frac{1}{2} \sigma_b^2. \tag{2.5.3}$$

Binary response

When the response Y_{ij} is binary, the conditional and marginal distribution are both Bernoulli with mean $\mu_{ij}^{(C)}$ and $\mu_{ij}^{(M)}$ respectively. The marginal variance is simply

$$\text{Var}(Y_{ij}) = \mu_{ij}^{(M)} (1 - \mu_{ij}^{(M)}).$$

2.5. Log-linear Models

The marginal covariance of two distinct responses on the same individual, Y_{ij} and Y_{ik} where $j \neq k$, is given by

$$\begin{aligned}
 \text{Cov}(Y_{ij}, Y_{ik}) &= \text{Cov}(E[Y_{ij}|b_i], E[Y_{ik}|b_i]) + E(\text{Cov}[Y_{ij}, Y_{ik}|b_i]) \\
 &= \text{Cov}(\mu_{ij}^{(C)}, \mu_{ik}^{(C)}) \\
 &= E(\mu_{ij}^{(C)} \mu_{ik}^{(C)}) - \mu_{ij}^{(M)} \mu_{ik}^{(M)} \\
 &= E(\exp\{2b_i + \eta_{ij} + \eta_{ik}\}) - \mu_{ij}^{(M)} \mu_{ik}^{(M)} \\
 &= \frac{M_b(2)}{M_b^2(1)} \mu_{ij}^{(M)} \mu_{ik}^{(M)} - \mu_{ij}^{(M)} \mu_{ik}^{(M)} \\
 &= \left(\frac{M_b(2)}{M_b^2(1)} - 1 \right) \mu_{ij}^{(M)} \mu_{ik}^{(M)}. \tag{2.5.4}
 \end{aligned}$$

The marginal correlation is thus

$$\begin{aligned}
 \text{Corr}(Y_{ij}, Y_{ik}) &= \frac{\text{Cov}(Y_{ij}, Y_{ik})}{\sqrt{\text{Var}(Y_{ij})\text{Var}(Y_{ik})}} \\
 &= \frac{\left(\frac{M_b(2)}{M_b^2(1)} - 1 \right) \mu_{ij}^{(M)} \mu_{ik}^{(M)}}{\sqrt{\mu_{ij}^{(M)} (1 - \mu_{ij}^{(M)}) \mu_{ik}^{(M)} (1 - \mu_{ik}^{(M)})}},
 \end{aligned}$$

which depends upon the random effect distribution and the marginal means in an obscure way. In general, the pairwise correlation increases with the marginal mean. When the $\mu_{ij}^{(M)}$'s are small, the marginal correlation is approximately

$$\text{Corr}(Y_{ij}, Y_{ik}) \approx \left(\frac{M_b(2)}{M_b^2(1)} - 1 \right) \sqrt{\mu_{ij}^{(M)} \mu_{ik}^{(M)}}.$$

For a fixed individual i , if the marginal mean $\mu_{it}^{(M)}$ does not vary much over time so that $\sqrt{\mu_{ij}^{(M)} \mu_{ik}^{(M)}} \approx \mu_i^{(M)}$, an exchangeable correlation structure is a close approximation to the above implied marginal correlation. If no baseline covariates are included in the model such that $\mu_i^{(M)}$ does not vary across individuals, a common exchangeable correlation structure is a reasonable

choice of “working” correlation structure. In the special case when there is no covariate effect, the correlation matrix has exactly an exchangeable structure.

Count response

When the response Y_{ij} is a count, the conditional distribution of Y_{ij} given b_i is usually assumed to be Poisson, with mean $\mu_{ij}^{(C)}$ and variance equal to the mean. By observing in (2.5.1) that

$$\exp(\eta_{ij}) = \frac{\mu_{ij}^{(M)}}{M_b(1)},$$

the variance of the marginal distribution is

$$\begin{aligned} \text{Var}(Y_{ij}) &= \text{Var}(\mathbb{E}[Y_{ij}|b_i]) + \mathbb{E}(\text{Var}[Y_{ij}|b_i]) \\ &= \text{Var}(\mu_{ij}^{(C)}) + \mathbb{E}(\mu_{ij}^{(C)}) \\ &= \text{Var}(\exp\{b_i + \eta_{ij}\}) + \mu_{ij}^{(M)} \\ &= \exp\{2\eta_{ij}\} \cdot \text{Var}(\exp\{b_i\}) + \mu_{ij}^{(M)} \\ &= (\mathbb{E}(\exp\{2b_i\}) - \mathbb{E}(\exp\{b_i\})^2) \left[\frac{\mu_{ij}^{(M)}}{M_b(1)} \right]^2 + \mu_{ij}^{(M)} \\ &= \left(\frac{M_b(2)}{M_b^2(1)} - 1 \right) \mu_{ij}^{(M)^2} + \mu_{ij}^{(M)}. \end{aligned}$$

When the random intercept follows normal distribution with variance σ_b^2 , the marginal variance,

$$\text{Var}(Y_{ij}) = \left(e^{\sigma_b^2} - 1 \right) \mu_{ij}^{(M)^2} + \mu_{ij}^{(M)},$$

depends on the marginal mean $\mu_{ij}^{(M)}$ and the variance component σ_b^2 . For $\sigma_b^2 > 0$, the marginal variance is greater than the marginal mean. The overdispersion of Y_{ij} with respect to Poisson is due to the extra variation in-

2.5. Log-linear Models

roduced by the random intercept representing baseline heterogeneity which persists throughout the course of the longitudinal study. Since the constant dispersion parameter ϕ can be eliminated in the estimating equations (2.3.3) and the sandwich estimator of variance (2.3.8), when an independence or a fixed “working” correlation matrix is used, the GEE estimator of β and the sandwich estimator of variance are robust to overdispersion. However, when a more complicated “working” correlation structure is used, some association parameters α need to be estimated. The consistent estimator of α may involve ϕ . In this case, we have to allow for and estimate the overdispersion ϕ when GEE with Poisson variance function is used for this marginal model.

The marginal covariance in (2.5.4) does not depend the distributional form of the response but only the marginal means of the responses and the distributional form of the random effect. The marginal correlation of the counts Y_{ij} and Y_{ik} is thus

$$\begin{aligned} \text{Corr}(Y_{ij}, Y_{ik}) &= \frac{\text{Cov}(Y_{ij}, Y_{ik})}{\sqrt{\text{Var}(Y_{ij})\text{Var}(Y_{ik})}} \\ &= \frac{\left(\frac{M_b(2)}{M_b^2(1)} - 1\right) \mu_{ij}^{(M)} \mu_{ik}^{(M)}}{\sqrt{\left(\left[\frac{M_b(2)}{M_b^2(1)} - 1\right] \mu_{ij}^{(M)^2} + \mu_{ij}^{(M)}\right) \left(\left[\frac{M_b(2)}{M_b^2(1)} - 1\right] \mu_{ik}^{(M)^2} + \mu_{ik}^{(M)}\right)}} \\ &= \frac{1}{\sqrt{\left(1 + 1/\left[\frac{M_b(2)}{M_b^2(1)} - 1\right] \mu_{ij}^{(M)}\right) \left(1 + 1/\left[\frac{M_b(2)}{M_b^2(1)} - 1\right] \mu_{ik}^{(M)}\right)}}, \end{aligned}$$

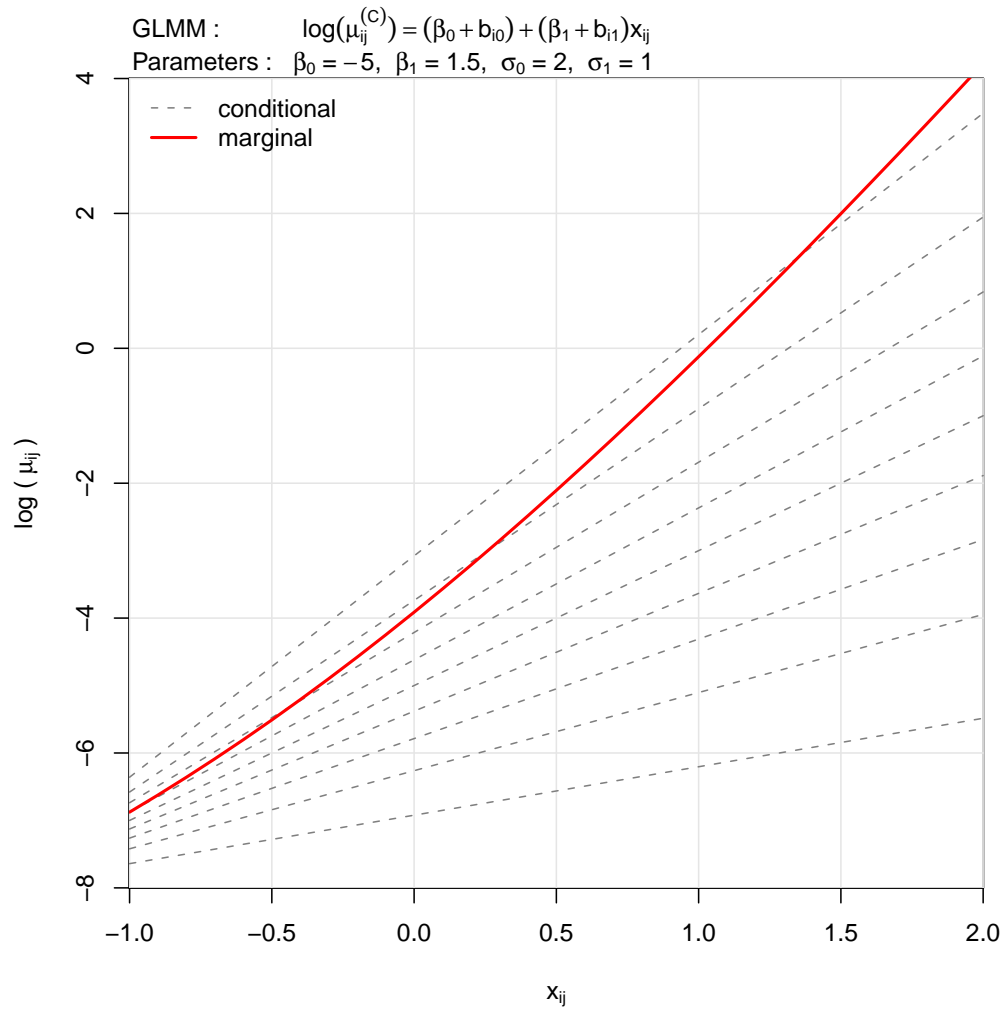
which depends on the random effect distribution and the marginal means in a complicated way. When the within subject covariate effects are small such that $\mu_{ij}^{(M)} \approx \mu_{ik}^{(M)}$ for $j \neq k$, an exchangeable structure may provide a good approximation to the true correlation structure. When the between subject covariate effects are also small such that the $\mu_i^{(M)}$'s are similar across subjects, a common exchangeable structure with a common pairwise correlation of ob-

servations across subjects and time may be a reasonable approximation. In the extreme case when there are no covariate effects at all, the marginal correlation matrix has exactly an exchangeable structure. In general, direct utilization of the exact marginal correlation structure involves some difficulties. The estimation of the component $\xi = \frac{M_b(2)}{M_b^2(1)}$ is non-standard and most generic GEE model-fitting algorithms do not support the use of such a customized “working” correlation. When a simple “working” correlation structure better than the independence structure is sought, the exchangeable structure may be applied as an effort to partially account for the within-subject correlation.

The GLMM log-linear random intercept model and its derived marginal model are applied to analyze the BEYOND trial data. The GLMMs are fitted by MLE, whereas the marginal models are fitted by GEE with an exchangeable correlation structure. Chapter 3 presents the results when the response is a binary indicator of relapse onset on the corresponding day on study. Chapter 4 considers an alternative form of response obtained by collapsing the daily binary responses to counts over time intervals.

2.5. Log-linear Models

Figure 2.1: Conditional and implied marginal mean structures of a log-linear model with random intercept and slope: random effects are evaluated at the 10th, 20th, ..., 90th percentiles.



Chapter 3

Analysis of BEYOND Trial Data

3.1 Introduction

In the BEYOND trial, 897 patients were randomized to the 250 mcg IFNB group, 899 patients were randomized to the 500 mcg IFNB group and 448 patients randomized to the 20 mg GA group. Serum samples were collected on most of the patients at baseline and semiannually in the two IFNB groups. Serum samples were tested for NABs to IFNB. NAB titers were measured by the MxA assay. Among the 1773 patients who had at least one NAB titer measure, 320 out of 888 in the 250 mcg IFNB arm and 340 out of 885 in the 500 mcg IFNB arm had at least one NAB titer of 20 NU/mL or higher.

Variables of Interest

Onset of relapse. The clinical outcomes of interest are the dates of onset of new relapses. Whether a new relapse started or not was recorded for every day on study for each individual patient. By definition, a new relapse cannot start before the patient recovers from an ongoing relapse. Therefore, during the periods of ongoing relapses patients were not subjected to the risk of new relapses. We call the time between the second day and the last day of a relapse inclusively *days-not-at-risk* and other time during the trial *days-at-risk*. Inference on the probability of a new relapse starting on a particular day should be based exclusively on observations on *days-at-risk*. Unfortunately,

the ending dates of relapses were not recorded in the BEYOND trial. Hence, we adopt an *ad hoc* rule which assumes that all relapses last for exactly 30 days. The 29 days following the start of each relapse are excluded from all our reported analyses.

NAB titers. The NAB titers of serum samples collected at baseline and every 6 months afterward were measured by the MxA assay. The MxA assay has a detection limit of 20 NU/mL. NABs to IFNB are regarded as absent if the NAB titer is less than the detection limit. In the BEYOND dataset, some NAB measures are missing at baseline. In all our analyses in this and the next chapter, we assume the titer values corresponding to any missing NAB measures at baseline are less than 20 NU/mL.

NAB status. In our analyses, the NAB status for each day is determined by the “*all switches considered*” rule with “*confirmation required*” for both switching from NAB− to NAB+ as well as from NAB+ back to NAB−. The requirement of confirmation is due a belief that the reliability of the NAB titers is low. Detailed description of this *ad hoc* classification rule is given in Section 1.3.3. The cutoff for switching between NAB− and NAB+ is chosen to be 20 NU/mL, the detection limit of the MxA assay.

Refined NAB status. A refinement of the NAB status classification is also considered. The NAB+ periods are further divided into periods of Low NAB+, Medium NAB+ and High NAB+ to explore the usefulness of levels of NAB titer in predicting strength of NAB effects. Confirmation is not required for switchings between sublevels of NAB+. The NAB+ periods are classified as *Low NAB+* for NAB titers in the interval $[20\text{NU/mL}, 100\text{NU/mL})$, *Medium NAB+* for NAB titers in the interval $[100\text{NU/mL}, 400\text{NU/mL})$ and *High NAB+* for NAB titers equal to or higher than 400 NU/mL.

3.1. Introduction

Table 3.1 summarizes the scheduled clinical visits where the first NAB serum titer became available for individual patients. Some patients did not have their first NAB titer available until some time after baseline. Since NABs are not expected to develop in the very early part of treatment, or to be present at baseline before the treatment is administered, we assume that the NAB titers for these patients are 0 NU/mL for all visits before NAB titers are available.

Table 3.1: Time of the first available NAB titer

Groups	Baseline/Day 1	Week 26	Week 52	Week 78	End of study
250 mcg	863	24	1	0	0
500 mcg	863	19	1	1	1

Using the above NAB classification rule, 248 of the 897 (28%) patients in the 250 mcg group have at least one NAB+ period while the remaining 639 patients never have any NAB+ periods. In the 500 mcg group, 262 of the 899 (29%) patients have at least one NAB+ period while the remaining 623 patients have no NAB+ periods.

Table 3.2 shows the distribution of the time to the first confirmed NAB+ among *eventually NAB+* patients. In both treatment arms, the majority of the *eventually NAB+* became NAB+ in the first one and a half years on study. More than half of the *eventually NAB+* patients were already NAB+ by around 6 months on study.

Table 3.3 shows the distribution of the number of NAB+ periods for each patient. All *eventually NAB+* patients had only one NAB+ period. About 85% of the *eventually NAB+* patients in both groups, had an initial NAB− period follow by a NAB+ period which extended to the end of their follow-

3.1. Introduction

Table 3.2: Time to the first confirmed NAB+

Groups	Time [†] (Year)					
	0.0	0.5	1.0	1.5	2.0	2.5
250 mcg	2	144	72	18	7	0
500 mcg	1	173	56	27	4	1

[†]Rounded to the nearest 0.5 year.

up. The remaining 15% first had a NAB− period, followed by a NAB+, then by a conversion back to NAB−. Two exceptional cases in the 250 mcg group and one in the 500 mcg group had NAB+ periods beginning from baseline.

Table 3.3: NAB status dynamics

All Switches considered; confirmed NAB+ and confirmed NAB−

NAB Dynamics		250 mcg	500 mcg
Eventually NAB+	− +	215	229
	− + −	31	32
	+	1	1
	+ −	1	0
	subtotal	248	262
Never NAB+	−	639	623
Total		887	885

The total number of days at risk for the *eventually NAB+* patients in the 250 mcg group ranges from 254 to 1267 days. The *never NAB+* patients in the 250 mcg group have days at risk ranging from 15 to 1280 days. For

3.1. Introduction

the 500 mcg group, the range for *eventually NAB+* patients is 378 to 1242 days, whereas the range for *never NAB+* patients is 7 to 1197 days. In both IFNB treated groups, the average proportion of time on study as NAB– for individual *eventually NAB+* patients is about 36%. Figures 3.1a and 3.1b show the distribution of the proportion of NAB– periods

The total number of relapses during the course of trial for individuals in the 250 mcg group ranges from 0 to 6 with an average of 0.86 for *eventually NAB+* patients and ranges from 0 to 7 with an average of 0.78 for *never NAB+* patients . The annualized crude relapse rate is 0.362 for the *eventually NAB+* group and 0.356 for the *never NAB+* group. In the 500 mcg group, the total number of relapses ranges from 0 to 5 with an average of 0.77 for *eventually NAB+* patients and from 0 to 8 with an average of 0.72 for *never NAB+* patients. The annualized crude relapse rate is 0.330 for the *eventually NAB+* group and 0.334 for the *never NAB+* group.

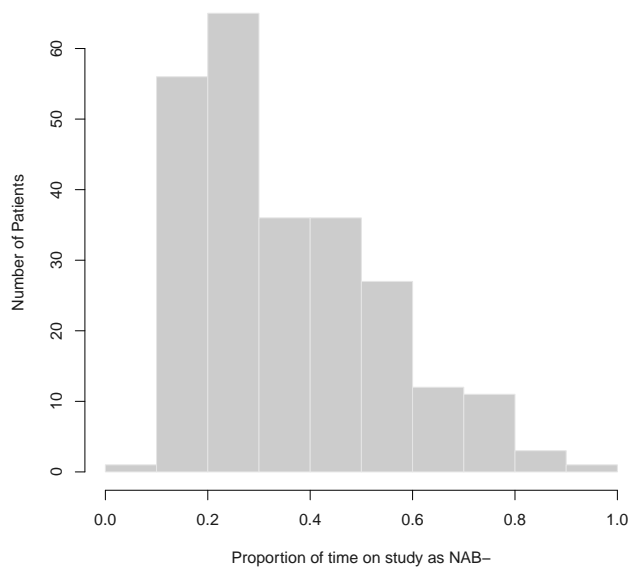
Table 3.4 is a frequency table of the total number of relapses per patient throughout the course of study for the *eventually NAB+* patients. A similar distribution for the number of relapses is observed for both IFNB treated groups. For the 250 mcg and the 500 mcg groups respectively, 51% and 56% of the patients did not experience any relapses during the study; 26% and 23% had one relapse; and 23% and 21% had two or more relapses.

Table 3.4: Number of relapses per eventually NAB+ patients

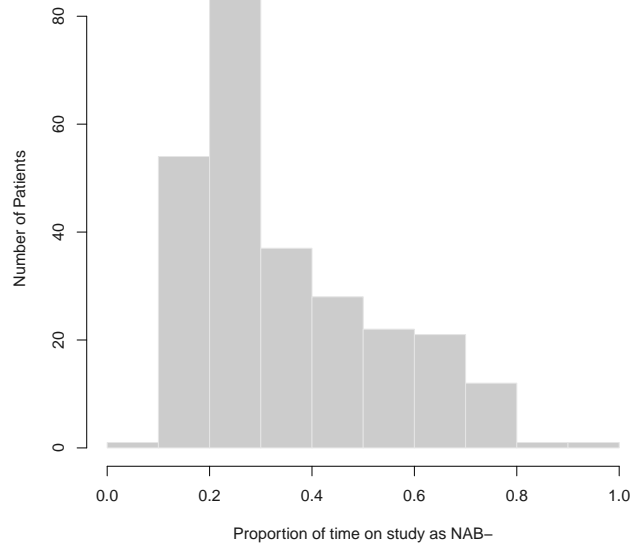
Groups	Number of relapses						
	0	1	2	3	4	5	6
250 mcg	126	65	36	11	7	2	1
500 mcg	147	60	32	15	7	1	0

Figure 3.1: Proportion of time on study as NAB− for eventually NAB+ patients.

(a) The 250 mcg group



(b) The 500 mcg group



The NAB effect in our report refers to the relative risk of relapse onset if a patient is NAB+ instead of NAB− on that day. The NAB effect is assessed by comparing the relapse risk during the NAB+ periods to the risk during the NAB− periods within individuals. Therefore only *eventually NAB+* patients provide the relevant information for estimating NAB effects. All our analyses are based only on the *eventually NAB+* groups.

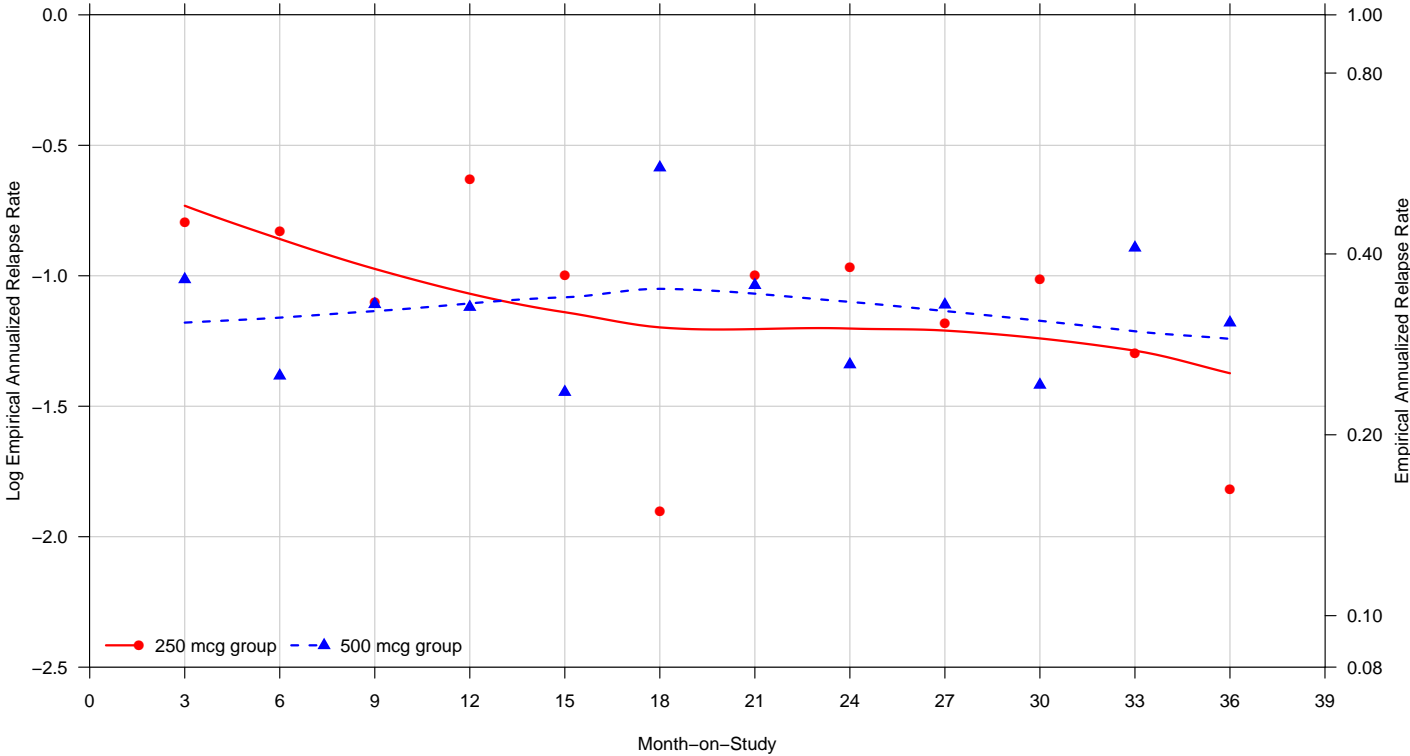
3.2 Basic Model Specification

We construct a mixed effect log-binomial regression model for the daily probability of a relapse onset. The response Y_{it} is a binary indicator of whether a relapse started at time t for the i^{th} patient. Conditional on the random effects, Y_{it} is assumed to follow a Bernoulli distribution with probability $\mu_{it}^{(C)}$. The conditional independence assumption is adopted.

The primary target of inference is the NAB effect quantified as the percentage increase in relapse risk on any given day on study during NAB+ periods relative to NAB− periods. This succinct summary of relative risk can be directly estimated by relating the daily relapse probability $\mu_{it}^{(C)}$ to a linear predictor through a log link. In general, we consider linear predictors for the daily relapse probability which includes additive components of 1) a NAB effect, 2) the time trend of the natural progression of relapse frequency over the course of trial, and 3) a random intercept representing inherent between subject variation in MS disease activity.

Figure 3.2 shows the log empirical annualized relapse rates for preceding 3-month periods separately for the two IFNB treated groups. For month 36 to 39, the empirical relapse rate is zero in both groups; whereas for month 39 to 42, the empirical relapse rate is zero for the 500 mcg group. Since relapse rates are presented on log scale, these two periods are not included in Fig-

Figure 3.2: Empirical annualized relapse rates for preceding 3-month periods and the corresponding loess fits for eventually NAB+ patients.



3.2. Basic Model Specification

ure 3.2. The loess fit to the log empirical relapse rate illustrates a decreasing time trend for the 250 mcg group which is reasonably linear. On the other hand, no distinguishable time trend is suggested in the 500 mcg group.

We consider an initial model with a simple mean structure,

$$\log \left(\mu_{it}^{(C)} \right) = b_i + \beta_0 + \beta_1 t + \beta_2 \text{NAB}_{it}, \quad (3.2.1)$$

where b_i is a random intercept from $\text{normal}(0, \sigma_b^2)$, β_0 is the intercept; β_1 is the slope of the linear time trend; and β_2 represents the NAB effect with NAB_{it} being the binary covariate of NAB status corresponding to Y_{it} coded as 0 for NAB– and 1 for NAB+. The model assumes that the log daily relapse risk is changed by a constant level during the NAB+ period. When the NAB status switches from NAB+ back to NAB–, the log daily relapse risk returns to the initial NAB– level adjusted for the time trend without any NAB effects carried over. As described in Section 2.5.2, the marginalized model derived from the initial mixed-effects model 3.2.1 has mean structure,

$$\log \left(\mu_{it}^{(M)} \right) = \beta_0^* + \beta_1 t + \beta_2 \text{NAB}_{it}, \quad (3.2.2)$$

where the intercept $\beta_0^* = \beta_0 + \frac{1}{2}\sigma_b^2$ as mentioned in (2.5.3). The mixed-effects model and the marginalized model share the same regression coefficients β_1 and β_2 .

For analysis of the daily relapse data presented in this chapter, the log-binomial random intercept models are fitted by maximum likelihood, whereas the marginalized models are fitted by GEE with an exchangeable “working” correlation structure (GEE-EXCH) with variance $\text{Var}(Y_{ij}) = \phi \mu_{ij}^{(M)} (1 - \mu_{ij}^{(M)})$ for binary responses. Given the estimated regression coefficients from either approach, the relative risk of NAB+ periods versus NAB– periods can be estimated by $\widehat{\text{RR}} = \exp \left(\hat{\beta}_2 \right)$. The standard error of the estimated relative

risk can be approximated by the delta method, $SE(\widehat{RR}) \approx \widehat{RR} \cdot SE(\hat{\beta}_2)$. An approximate $(1 - \alpha) \times 100\%$ confidence interval for the relative risk can be obtained by transforming a symmetric confidence interval for β_2 , $\widehat{RR} \cdot \exp(\pm z_{1-\alpha/2} SE(\hat{\beta}_2))$, where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)^{th}$ upper quantile of a standard normal distribution.

3.3 Computational Issues

Various R packages and routines are available for fitting GLMMs. Packages we are aware of up to the time this report was written include

- 1) function `glmmPQL()` in package `MASS`,
- 2) package `glmmADMB`,
- 3) package `MCMCglmm`,
- 4) package `glmmML`,
- 5) package `glmmAK`,
- 6) function `glmer()` in package `lme4`, and
- 7) function `glmm()` in package `repeated`.

Not all of these packages suit our purposes: 1 – 3) do not perform maximum likelihood estimation and 4 – 5) do not have the log link available for the binomial family. Only `glmer(lme4)` and `glmm(repeated)` perform log-binomial regression with adaptive Gauss quadrature implemented. Among the two model fitting routines, `glmer(lme4)` can handle most specifications of random effects, while `glmm(repeated)` can only accommodate a random intercept. Neither of these routines provides an option of REML as an alternative to maximum likelihood for estimation of the variance components

except for the case of linear mixed-effects models.

Although patients in the BEYOND trial were only followed for a moderate length of 2 to 3.5 years, since relapse onset was monitored on a daily basis, each patient has a long vector of binary responses. The average number of days at risk for *eventually NAB+* patients is 871 in the 250 mcg group and 854 in the 500 mcg group. The size of our dataset also poses a challenge for the memory management of the model fitting routines and resources of the platform **R** runs on. The presumably more powerful `glmer(lme4)` cannot handle our large dataset, while `glmm(repeated)` still manages to fit the models, despite taking a substantial amount of time. The MLE of GLMMs in this chapter and Chapter 4 are obtained by `glmm(repeated)` using 15 quadrature points with default convergence criteria and control parameters.

The same problem is encountered for the GEE approach. The two standard of excellence **R** packages for GEE, `gee` and `geepack`, were unable to handle our data. With the permission of Mr. Rick White, we used his GEE model fitting **C** routine, which can handle our large dataset, for all our reported GEE model fitting in this chapter, as well as in Chapter 4.

3.4 Time Trend

The proposed longitudinal analysis of NAB effects automatically takes into account the confounding effects of baseline covariates by making the comparison of NAB− time periods to NAB+ time periods within individual patients. The natural trend of relapse risk over time becomes the most important factor remain to be adjusted for. Accuracy in modelling this time trend is essential to the soundness of the estimation of the NAB effect.

Based on the patterns in Figure 3.2, we propose an initial linear time

3.4. Time Trend

trend in the log transformed daily relapse risk. The linear trend on the log scale implies that the relapse risk changes exponentially over time. Adequacy of the linearity assumption is assessed by comparing GLMM fits with alternative time trends modelled by natural cubic splines with degrees of freedom equals to 2, 3 or 4 to the linear time trend. The linear time trend can be regarded as a natural cubic spline with 1 degree of freedom. NAB status is included as a covariate in all the models so that the impact of including a more complicated time trend on the estimation of NAB effects can be assessed.

Fitted time trends are visualized in Figures 3.3a and 3.3b for the 250 mcg group and 500 mcg group respectively. The “annualized” fitted relapse rates are obtained by multiplying the fitted daily relapse probabilities by a factor of 365.25. For the 250 mcg group, the 2 and 3 degrees of freedom splines seem to suggest more curvature for the diminishing trend in relapse risk than the exponential decay implied by linear time trend in log relapse risk. The wigglyness of the fitted curve of the 4 degrees of freedom spline suggests overfitting. In contrast, the linearity assumption in the 500 mcg group appears to be representing the trend of relapse risk over time well as the higher degree of freedom splines almost coincide with the linear trend. Signs of overfitting are observed for the 3 and 4 degrees of freedom splines. All the fitted time trends for the 500 mcg group suggest that the change over the 3.5 year period is minimal. Comparing the two IFNB treated groups, the “annualized” fitted relapse rates in the 250 mcg group are higher than the rates in the 500 mcg group during the 3.5 years on study. Analyses are expected to have lower sensitivity for the 500 mcg group than the 250 mcg group due to the lower relapse risk.

Results of likelihood ratio tests comparing the fitted splines and linear time trends are presented in Table 3.5. Among the natural cubic splines we

Figure 3.3: Annualized fitted relapse rate of GLMM for NAB– periods, natural regression splines for time trend, eventually NAB+ patients only.

(a) The 250 mcg group

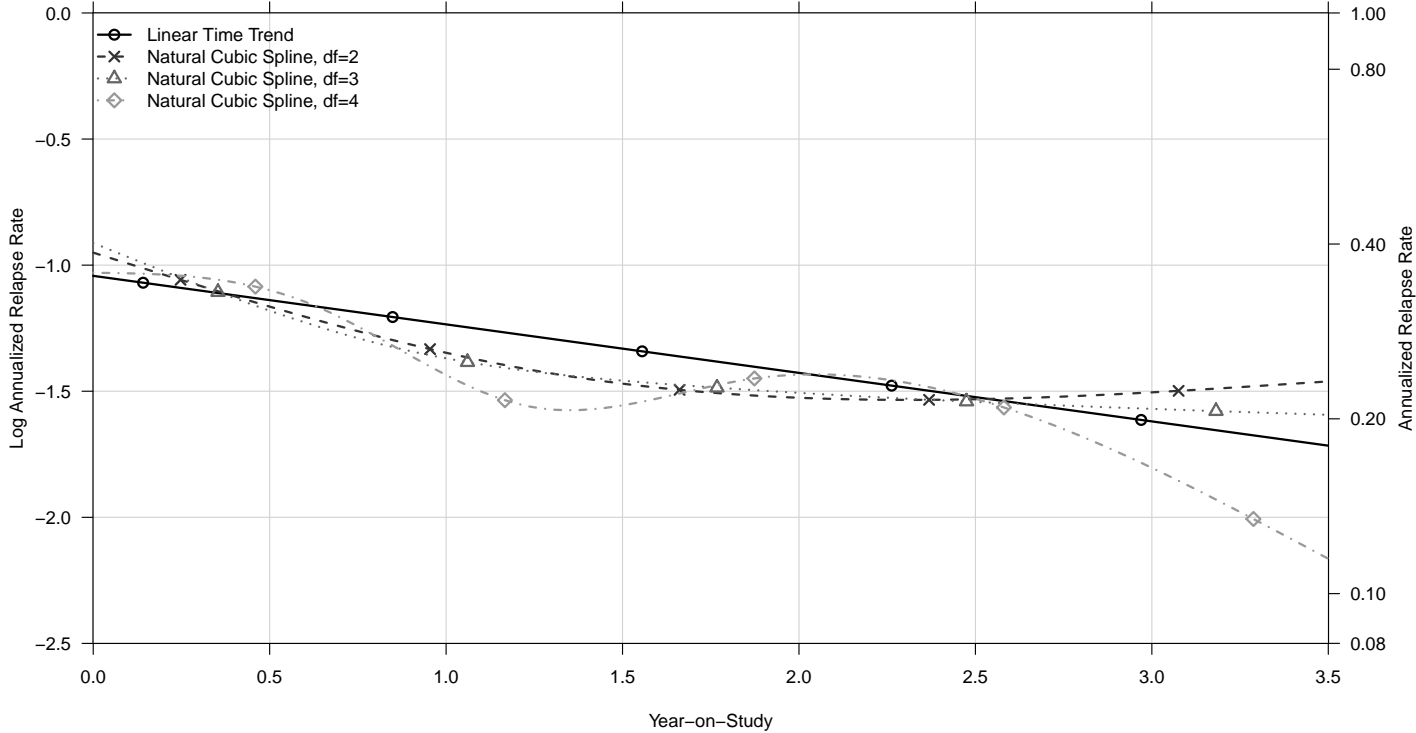
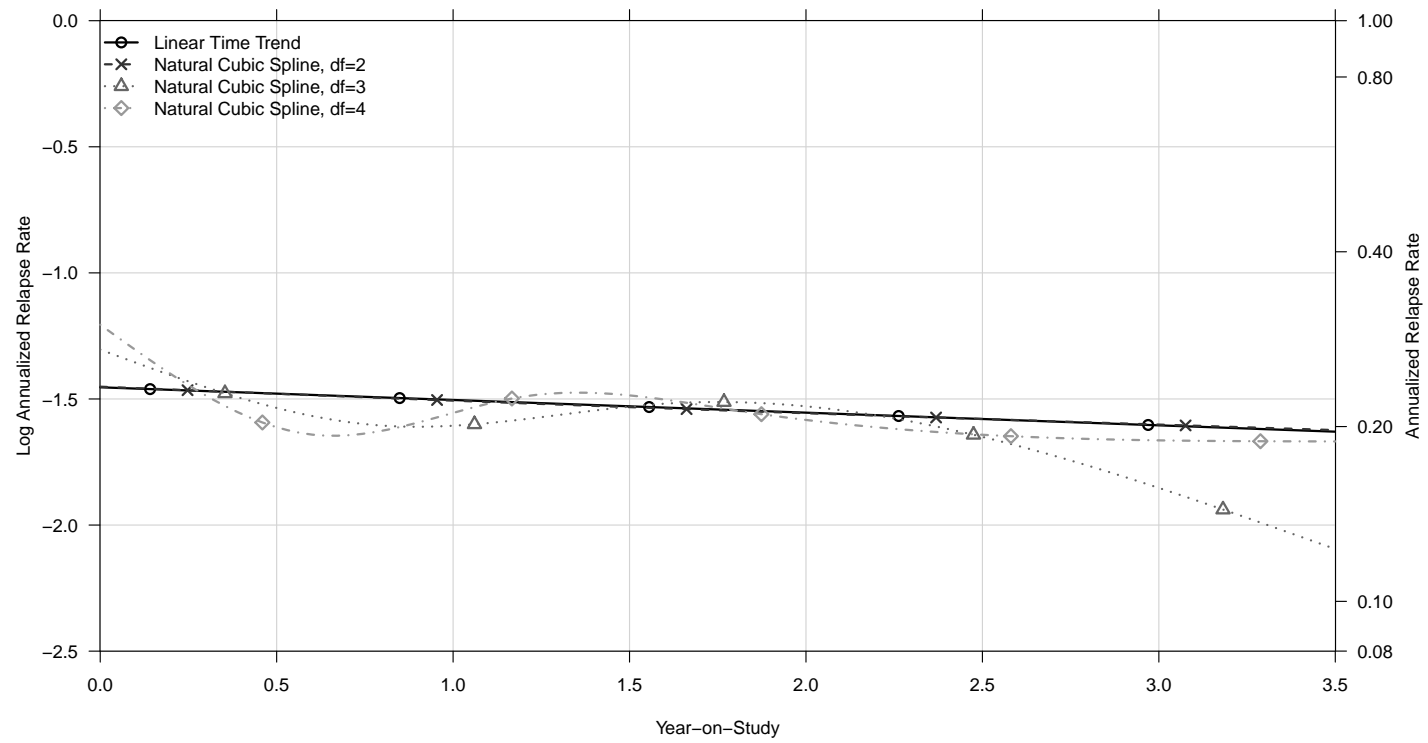


Figure 3.3: – Continued

(b) The 500 mcg group



3.4. Time Trend

3.4. Time Trend

considered, none fits the data significantly better than the linear trend in either IFNB treated group.

Table 3.5: p-values for likelihood ratio tests comparing GLMM fits with different time trends: eventually NAB+ patients only.

All Switches considered; confirmed NAB+, confirmed NAB–

Null hypothesis, H_0		Alternative hypothesis, H_1					
		250 mcg group			500 mcg group		
		Natural Splines			Natural Splines		
		df = 2	df = 3	df = 4	df = 2	df = 3	df = 4
Natural Splines	df = 1 [†]	0.40	0.69	0.55	0.97	0.65	0.72
	df = 2	–	0.81	0.49	–	0.36	0.52
	df = 3	–	–	0.24	–	–	0.49

[†] natural spline with 1 degree of freedom is equivalent to linear.

The estimated relative risks of NAB+ to NAB– time periods under each of the time trend models considered are presented in Table 3.6. With the linear time trend, the estimated increase in relapse risk of NAB+ periods is 10% in the 250 mcg group, but this doubles to about 20% when the time trends are fitted with splines. The increase in relapse risk is about 30% in the 500 mcg group under all the different models considered for the time trend. The estimated NAB effects are always stronger in the 500 mcg group, though the effects are statistically insignificant in both groups under all the time trends considered. The standard errors of the estimated relative risks in the 500 mcg group are larger than those in the 250 mcg group, as expected given the lower relapse risks in the 500 mcg group revealed in Figures 3.3a and 3.3b.

For the 500 mcg group, as revealed by the fitted splines and the likelihood

3.5. Neutralizing Antibody Effects

Table 3.6: Estimated relative risks of NAB+ to NAB− from GLMM fits with various time trends: eventually NAB+ patients only.

All Switches considered; confirmed NAB+, confirmed NAB−

	250 mcg group				500 mcg group			
	$\widehat{\text{RR}}$	SE	95% CI		$\widehat{\text{RR}}$	SE	95% CI	
Linear	1.10	0.19	(0.79,	1.54)	1.29	0.24	(0.90,	1.84)
Natural Spline, df=2	1.21	0.24	(0.81,	1.79)	1.29	0.27	(0.86,	1.94)
Natural Spline, df=3	1.21	0.25	(0.81,	1.81)	1.34	0.29	(0.87,	2.05)
Natural Spline, df=4	1.24	0.25	(0.83,	1.84)	1.33	0.29	(0.87,	2.03)

$\widehat{\text{RR}}$: estimated relative risk.

ratio tests, the assumption of a linear time trend is apparently reasonable and adequately represents the natural change of relapse risks over time. For the 250 mcg group, the fitted splines suggest that a quadratic time trend might be more appropriate. However, no substantial evidence of lack of fit is indicated by the likelihood ratio tests for either group. In both groups, the standard error of the estimated relative risk increases with the degrees of freedom of the regression splines, illustrating a tradeoff between bias and variance. Based on the results of the likelihood ratio tests, we decide to adopt the parsimonious model of a linear trend in log relapse risk over time for the subsequent analyses.

3.5 Neutralizing Antibody Effects

3.5.1 NAB Status

Tables 3.7a and 3.7b present the estimates and standard errors of the relative risks obtained by GEE-EXCH and GLMM based on the initial specification of mean structure (3.2.1). Results based on the refined NAB status of Low/Medium/High NAB+ are also included.

3.5. Neutralizing Antibody Effects

Similar patterns and levels of estimated NAB effects are given by GEE-EXCH and GLMM for both IFNB treated groups. For the 250 mcg group, no significant NAB effect is detected. The estimated increase in relapse risk is about 10% comparing NAB+ to NAB-. The NAB effects do not differ substantially across the three sublevels of NAB+. Comparing to NAB-, the estimated increases in relapse risk are less than 10% for Low NAB+, about 15% for Medium NAB+ and less than 5% for High NAB+.

A different pattern is observed for the 500 mcg group. The time adjusted increase in relapse risk of NAB+ period relative to NAB- period is estimated to be about 30%. The increases in relapse risk are about 10%, 55% and 60% for Low, Medium and High NAB+ periods respectively. The NAB effects are statistically significant at the 0.05 level for Medium and High NAB+ based on the GEE-EXCH model fit. Although the results are not quite statistically significant (Low: $p = 0.65$, Medium: $p = 0.05$, High: $p = 0.05$) using the GLMM approach, they still provide evidence suggesting presence of NAB effects in the 500 mcg group. The suggested NAB effects also appear to increase when NAB titer raises from low to medium level, but the relative risks are similar for medium and high levels.

3.5.2 NAB Titer

The direct use of NAB titer in estimating NAB effect on relapse risk is also explored by replacing the NAB status predictor NAB_{it} in model (3.2.1) by a log titer term $\log\left(\frac{Titer_{it}}{20}\right) \cdot I\{Titer_{it} \geq 20\}$, where $Titer_{it}$ is the latest NAB titer available as of time t for the i^{th} patient. Natural regression splines are fitted on the log titers, serving to discover the form of association between relapse risk and NAB titers. Figures 3.4a and 3.4b show the fitted natural cubic splines with 2 degrees of freedom and histograms of the titer values. As illustrated by the fitted splines, there is little suggestion of an NAB effect in the 250 mcg group, whereas in the 500 mcg group, the relapse risks appear

Table 3.7: Estimated relative risks of NAB+ and sublevels to NAB−: eventually NAB+ patients only.

(a) The 250 mcg group

All Switches considered; confirmed NAB+, unconfirmed L/M/H and confirmed NAB−

	GEE-EXCH				GLMM			
	$\widehat{\text{RR}}$	SE	95% CI		$\widehat{\text{RR}}$	SE	95% CI	
NAB+ vs. NAB−	1.092	0.172	(0.802,	1.487)	1.102	0.188	(0.788,	1.541)
Low NAB+ vs. NAB−	1.075	0.181	(0.774,	1.495)	1.094	0.199	(0.766,	1.564)
Med NAB+ vs. NAB−	1.158	0.256	(0.751,	1.786)	1.150	0.254	(0.747,	1.772)
High NAB+ vs. NAB−	1.035	0.270	(0.621,	1.725)	1.023	0.277	(0.602,	1.738)

(b) The 500 mcg group

All Switches considered; confirmed NAB+, unconfirmed L/M/H and confirmed NAB−

	GEE-EXCH				GLMM			
	$\widehat{\text{RR}}$	SE	95% CI		$\widehat{\text{RR}}$	SE	95% CI	
NAB+ vs. NAB−	1.293	0.222	(0.923,	1.811)	1.287	0.236	(0.898,	1.843)
Low NAB+ vs. NAB−	1.111	0.223	(0.749,	1.647)	1.099	0.228	(0.732,	1.650)
Med NAB+ vs. NAB−	1.541	0.339	(1.001,	2.370)	1.547	0.345	(1.000,	2.394)
High NAB+ vs. NAB−	1.602	0.379	(1.007,	2.548)	1.593	0.384	(0.993,	2.557)

$\widehat{\text{RR}}$: estimated relative risk.
Titters for NAB status — Low: [20, 100), Med: [100, 400), High: [400, ∞).

to increase with NAB titers until reaching a certain maximum limit.

As revealed by the plots of the fitted splines in Figures 3.4a and 3.4b, a quadratic trend in log titer value might represents the form association between relapse risk and NAB titer more accurately than a linear trend. However, a quadratic fit would complicate the interpretation. A model which captures the basic shape of this exposure-response relationship but retains simple interpretation of NAB effects is desired. Therefore, a piecewise linear model on log titers which implies a constant relative risk for doubled titers over a range of titers from 20 NU/mL to a cutoff k NU/mL is proposed for more formal analysis. The conditional mean structure of the piecewise linear model

$$\log(\mu_{it}^{(C)}) = b_i + \beta_0 + \beta_1 t + \beta_2 \log\left(\frac{\min\{\text{Titer}_{it}, k\}}{20}\right) \cdot I\{\text{Titer}_{it} \geq 20\}, \quad (3.5.1)$$

marginalizes to

$$\log(\mu_{it}^{(M)}) = \beta_0^* + \beta_1 t + \beta_2 \log\left(\frac{\min\{\text{Titer}_{it}, k\}}{20}\right) \cdot I\{\text{Titer}_{it} \geq 20\}, \quad (3.5.2)$$

where k is a prespecified titer cutoff level. Presuming that β_2 is positive, the time adjusted relapse risk reaches its maximum at the cutoff level. The relative risk for an arbitrary titer value compared to any titer ≤ 20 NU/mL is $\left(\frac{\min\{\text{Titer}_{it}, k\}}{20}\right)^{\beta_2}$. When the titer values are between 20 and k NU/mL, the relative risk corresponding to doubling the NAB titer is 2^{β_2} . Given the estimated regression coefficient $\hat{\beta}_2$, for titers less than cutoff, the standard error of $\widehat{\text{RR}} = \left(\frac{\text{Titer}_{it}}{20}\right)^{\hat{\beta}_2}$ can be approximated by $\log\left(\frac{\text{Titer}_{it}}{20}\right) \widehat{\text{RR}} \cdot SE(\hat{\beta}_2)$.

Four values for the cutoff k are considered: 100 NU/mL, 400 NU/mL, 1000 NU/mL and ∞ NU/mL, where the last case reduces to a linear trend in log titers. Estimated relative risks based on log titers are presented in

Figure 3.4: Fitted natural cubic spline (df = 2) on log NAB titers of GLMM: eventually NAB+ patients only.

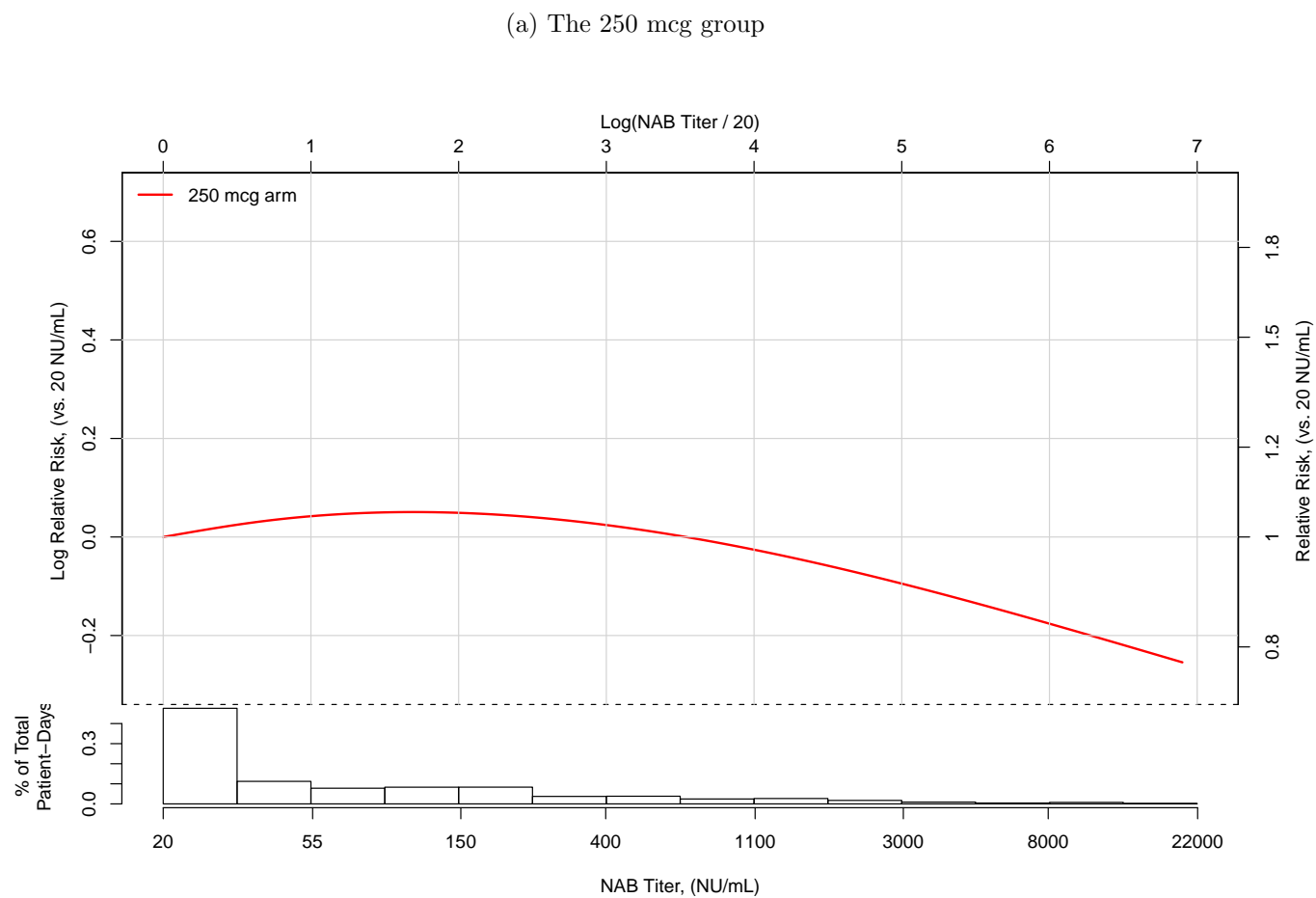
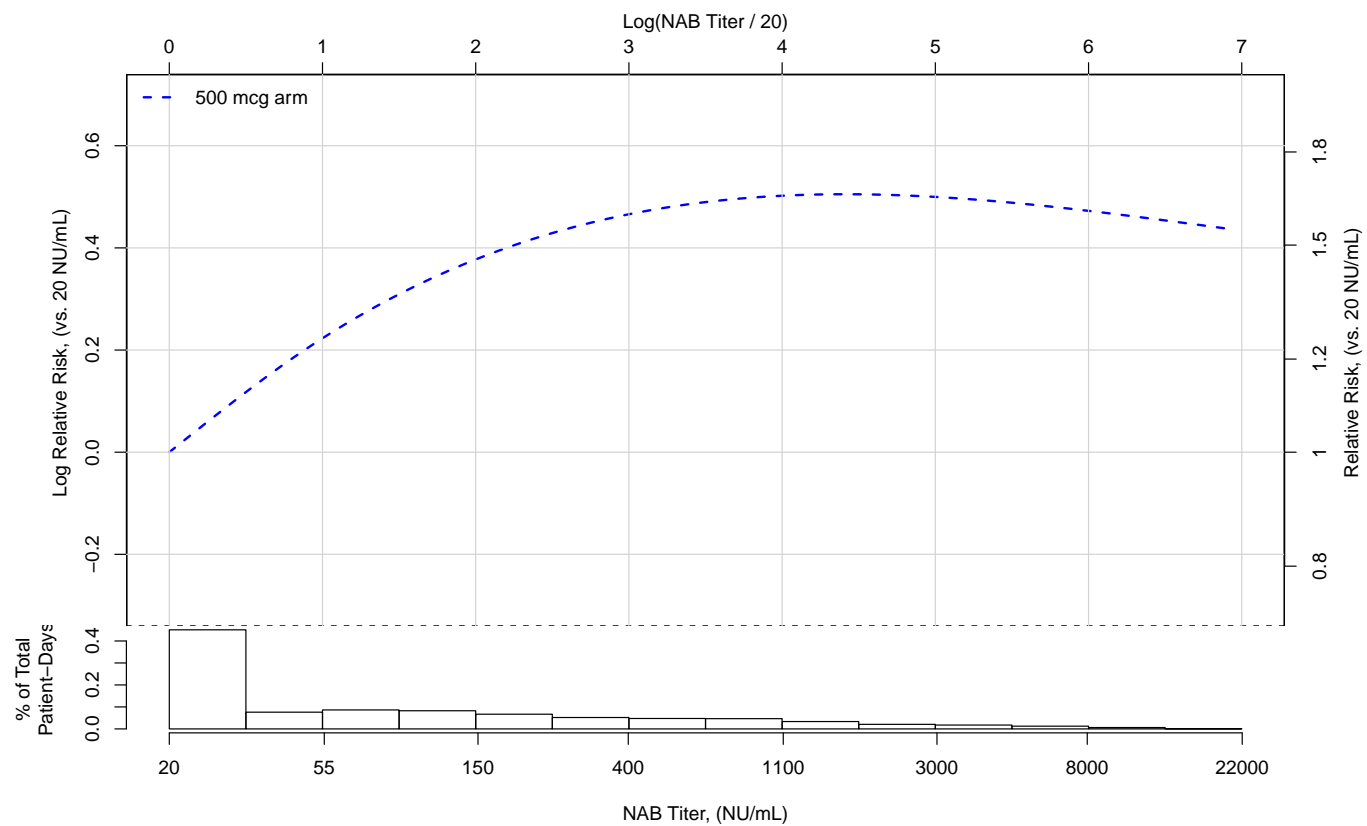


Figure 3.4: – Continued

(b) The 500 mcg group



3.5. Neutralizing Antibody Effects

Tables 3.8a and 3.8b. The pattern of the results given by GEE-EXCH and GLMM resemble each other closely in both IFNB treated groups. Estimated relative risks from the two approaches are effectively the same.

For the 250 mcg group, no significant associations between NAB titers and relapse risks are found regardless of the titer cutoff. For the model with no cutoff ($k = \infty$), $\hat{\beta}_2$ is -0.002 . For models with finite cutoff, $\hat{\beta}_2$'s are positive ranging from 0.002 to 0.028 . The estimated increase in relapse risk for doubled titers are less than 2% in all cases, and the maximum increases in relapse risk achievable at the cutoff are less than 5%.

For the 500 mcg group, for all four models, $\hat{\beta}_2$ is positive and the association between NAB titers and relapse risk is statistically significant. For cutoffs at 100, 400 and 1000 NU/mL, the estimated increases in relapse risks for doubled titers are about 18%, 12% and 10% respectively, whereas the estimated increases in relapse risk at cutoff are 48%, 63% and 71% respectively. With no cutoff, the estimated relative risks for doubled titers is about 8%.

Tables A.1a to A.4 in Appendix A show complete results of the model fits for all the models in this section. For GEE-EXCH, the estimated correlation α is 0.001 in every model for both the 250 mcg and 500 mcg groups. With essentially zero estimated correlation, the same model fits would be obtained with an independent “working” correlation structure.

The estimated marginal intercept $\hat{\beta}_0^*$ obtained from the GEE-EXCH fit is compared to the estimated conditional intercept $\hat{\beta}_0$ and variance components $\hat{\sigma}_b^2$ obtained from the GLMM fit with reference to their theoretical relationship (2.5.3). The derived marginal intercepts $\hat{\beta}_0 + \frac{1}{2}\hat{\sigma}_b^2$ are consistently larger than the estimated marginal intercepts $\hat{\beta}_0^*$ in all the fitted models. The difference ranges from 0.042 to 0.055.

Table 3.8: Estimated relative risk of NAB titers: eventually NAB+ patients only.

	(a) The 250 mcg group					
	GEE-EXCH			GLMM		
	\widehat{RR}	SE	95% CI	\widehat{RR}	SE	95% CI
cutoff= ∞						
Doubled titer before cutoff	0.999	0.038	(0.928, 1.075)	0.994	0.038	(0.922, 1.072)
cutoff=1000						
Doubled titer before cutoff	1.004	0.043	(0.923, 1.092)	0.999	0.043	(0.918, 1.088)
At cutoff 1000 NU/mL vs. 20 NU/mL	1.022	0.248	(0.635, 1.644)	0.996	0.243	(0.617, 1.608)
cutoff=400						
Doubled titer before cutoff	1.002	0.049	(0.910, 1.103)	0.996	0.050	(0.904, 1.098)
At cutoff 400 NU/mL vs. 20 NU/mL	1.007	0.214	(0.664, 1.528)	0.984	0.212	(0.646, 1.500)
cutoff=100						
Doubled titer before cutoff	1.019	0.077	(0.879, 1.183)	1.014	0.079	(0.870, 1.181)
At cutoff 100 NU/mL vs. 20 NU/mL	1.045	0.184	(0.740, 1.476)	1.032	0.187	(0.724, 1.472)

Table 3.8: – Continued

	(b) The 500 mcg group							
	GEE-EXCH				GLMM			
	\widehat{RR}	SE	95% CI		\widehat{RR}	SE	95% CI	
cutoff= ∞								
Doubled titer before cutoff	1.078	0.038	(1.005,	1.156)	1.079	0.037	(1.009,	1.153)
cutoff=1000								
Doubled titer before cutoff	1.100	0.047	(1.012,	1.196)	1.100	0.043	(1.018,	1.189)
At cutoff 1000 NU/mL vs. 20 NU/mL	1.712	0.411	(1.069,	2.741)	1.714	0.382	(1.108,	2.652)
cutoff=400								
Doubled titer before cutoff	1.119	0.057	(1.014,	1.235)	1.120	0.053	(1.020,	1.229)
At cutoff 400 NU/mL vs. 20 NU/mL	1.626	0.355	(1.060,	2.493)	1.630	0.334	(1.091,	2.435)
cutoff=100								
Doubled titer before cutoff	1.181	0.096	(1.006,	1.385)	1.182	0.093	(1.013,	1.380)
At cutoff 100 NU/mL vs. 20 NU/mL	1.470	0.278	(1.015,	2.131)	1.475	0.270	(1.029,	2.112)

3.6 Relative Efficiency of GEE to GLMM Approach on Estimating NAB Effects

Comparing the two longitudinal approaches applied, when the conditional models are the correct models, the GLMM approach has the highest asymptotic efficiency. However, the GEE approach, as a method of moments approach, is more robust to model misspecification as it only requires correct specification of the mean structure and independence between subjects. The efficiency of the GEE approach relative to the GLMM approach in our finite sample case is investigated. Specifically the efficiency in estimating NAB effects is our primary interest. We define relative efficiency of GEE to GLMM as the ratio of the variance of GLMM to the variance of GEE estimated relative risks.

The pattern observed based on the estimates in Tables 3.7a, 3.7b, 3.8a and 3.8b is not consistent. Based on NAB status, GEE-EXCH demonstrated a higher efficiency; the relative efficiencies are 119.5% and 113.0% for the 250 mcg and the 500 mcg groups respectively. For the 250 mcg group, the relative efficiency of GEE-EXCH is 120.9%, 98.4% and 105.3% for estimating the effects of Low NAB+, Medium NAB+ and High NAB+ versus NAB— respectively. For the 500 mcg group, the relative efficiency of GEE-EXCH is 104.5%, 103.6% and 102.7% for estimating the effects of Low NAB+, Medium NAB+ and High NAB+ versus NAB— respectively. Based on NAB titers, comparisons are made on all the estimated relative risks presented in Tables 3.8a and 3.8b. In the 250 mcg group, the relative efficiencies are close to 100%, ranging from 96.0% to 105.3%. However, in the 500 mcg group, the estimates given by GEE-EXCH are less precise with relative efficiencies ranging from 83.7% to 94.8%

Comparison based on the variance of the estimated regression coefficients

3.7. Summary

representing the NAB effects results in similar conclusions. The relative efficiencies are 113.0% and 113.7% for the 250 mcg and the 500 mcg groups respectively for estimating an overall effects associated with the binary NAB status. For the 250 mcg group, the relative efficiencies are 104.5% for Low, 103.6% for Medium and 102.7% for High NAB+ versus NAB-. For the 500 mcg group, the relative efficiencies are 106.1% for Low, 102.8% for Medium and 103.7% for High NAB+ versus NAB-. GEE-EXCH returned more precise estimates of NAB effects when they are based on NAB status. While based on NAB titers, the relative efficiency of GEE-EXCH is close to 100%, ranging from 101.4% to 105.8% for the 250 mcg group. Considerable loss of efficiency is noticed, however, for the 500 mcg group with relative efficiency of GEE-EXCH ranging from 86.0% to 91.3%. The average loss of efficiency in this group with the direct use of NAB titer is about 10%.

3.7 Summary

The analysis of the BEYOND trial suggests some effects of NABs to IFNB on RRMS patients' relapse risks. The additional analysis based on NAB titers sheds some light on how the level of NABs is associated with an increase in the risk of relapse. Some positive associations between NAB titers and frequency of relapse are suggested. Regarding the efficiency of the GEE and GLMM approaches, no consistent evidence of one being superior to the other is found.

Recall that the model for this daily relapse data cannot be fitted with most generic GEE or GLMM routines in **R**. An alternative to avoid this computational difficulty is to use another configuration of the relapse data. Since NAB titers were measured every 6 months, we may consider collapsing the daily relapse data and performing our analysis based on the number of relapses between consecutive NAB titer measures. This reduction in the

3.7. *Summary*

resolution of the relapse data and the potentially less precise estimation of the time trend may affect the statistical power in detecting NAB effects. The next chapter presents analyses based on collapsed data and attempts to answer the questions concerning impacts on the sensitivity for detecting NAB effects.

Chapter 4

Analysis Based on Collapsed Data

4.1 Introduction

In the BEYOND trial data, each patient has about 800 longitudinal binary indicators of relapse onset. The dimension of this response data causes some computational difficulties in fitting GLMMs and marginal models. For most model fitting procedures in R to be applicable, the response vectors need to be reduced to a more manageable size. Since NAB serum specimens were measured normally every half-year, much less frequently than the daily record of relapse onset, we consider collapsing the daily relapse data to the number of relapse onsets between consecutive NAB measures. This collapsing reduces the average number of longitudinal observations to about 6. Table 4.1 shows the distribution of the number of NAB serum specimens collected, which is equivalent to the number of collapsed intervals. In both the 250 mcg and 500 mcg groups, most patients had 5 to 7 collapsed intervals.

Table 4.1: Number of NAB titer measures

	2	3	4	5	6	7	8
250 mcg group	1	1	15	75	84	58	14
500 mcg group	0	5	24	80	89	57	7

4.1. Introduction

Table 4.2 shows the number of relapse onsets per interval. For the 250 mcg group, about 86% of the intervals had no relapse onsets, about 13% had one relapse onset and the remaining 1% had two relapse onsets. For the 500 mcg group, about 88% of the intervals had no relapse onsets, about 11% had one relapse onset and the remaining 1% had two relapse onsets. Only about 1% of the intervals in each group has two relapse onsets, which was the largest number of relapse onsets observed in a single interval, the aggregation of the events from the daily data is minimal. Some loss of precision in estimation due to collapsing is anticipated but not to a large extent.

Table 4.2: Number of relapse onsets per interval

	0	1	2	Total
250 mcg group	1261	188	13	1462
500 mcg group	1315	168	17	1500

Table 4.3 shows the distribution of the lengths of the intervals at risk corresponding to the collapsed observations. For the 250 mcg group, the lengths of the intervals at risk range from 1 to 544 days, whereas for the 500 mcg group, the lengths range from 1 to 675 days. In both groups, the number of days at risk clustered at around 1 day and 180 days. In the 250 mcg group, 200 out of 1462 are 1-day intervals, whereas in the 500 mcg group, 205 out of 1500 are 1-day intervals. Only one of these 1-day intervals had a relapse onset in each group. These exceptionally short intervals correspond to the end of study visit for some of the patients when the protocol specified that a final NAB specimen be taken.

4.2. Model Specification

Table 4.3: Number of days at risk between consecutive collections of NAB serum specimens

	Days at risk				
	0 – 49	50 – 99	100 – 149	150 – 199	200+
250 mcg group	215	82	64	1063	38
500 mcg group	225	72	60	1100	43

4.2 Model Specification

For the collapsed data, the response Y_{ij} of the i^{th} patient is the number of relapse onsets during the period starting from t_j , the day the j^{th} NAB titer was measured, to $t_{j+1} - 1$, the day before the next NAB titer measurement inclusively. Given a random effect b_i , the counts Y_{ij} are assumed to be conditionally independent following a Poisson distribution with mean $\mu_{ij}^{(C)} = \lambda_{ij}s_{ij}$, where s_{ij} is the length of the period at risk expressed in years. The conditional mean depends on the covariates through a log link,

$$\log(\mu_{ij}^{(C)}) = \beta_0 + b_i + \beta_1 t_j + \beta_2 \text{NAB}_{ij} + \log(s_{ij}).$$

The relative relapse rate for NAB+ time periods versus NAB– time periods is represented by $\exp(\beta_2)$. Since the NAB status remains unchanged between consecutive NAB titer measures, the NAB status NAB_{ij} corresponding to Y_{ij} is readily obtainable from the daily data in Chapter 3. The random intercept b_i is assumed to be a random sample from a normal distribution with mean 0 and variance σ_b^2 .

The marginal mean structure of this GLMM is only a shift of the intercept from the conditional mean structure,

$$\log(\mu_{ij}^{(M)}) = \beta_0^* + b_i + \beta_1 t_j + \beta_2 \text{NAB}_{ij} + \log(s_{ij}),$$

as shown in (2.5.2). The Poisson regression model with random intercept for count data is fitted by maximum likelihood, whereas the marginalized model is fitted by the GEE approach with an exchangeable “working” correlation structure (GEE-EXCH). Although the marginal distribution of Y_{ij} is not Poisson, we assume the marginal variance is proportional to the mean, $\text{Var}(Y_{ij}) = \phi \mu_{ij}^{(M)}$, in our GEE analyses. The dispersion parameter ϕ allows for anticipated overdispersion with respect to Poisson. With the dimension of the data reduced as indicated, both `glmer()` of the `lme4` package and `glmm()` of the `repeated` package can fit the GLMM model. For the GEE approach, both `gee()` of the `gee` package and `geeglm()` of the `geepack` package can handle the collapsed data. For coherence and ease of comparison of results based on the daily relapse data and the collapsed data, the same model fitting routines in R as in Chapter 3 are implemented for the collapsed data.

4.3 Results

4.3.1 NAB Status

Tables 4.4a and 4.4b show the estimated relative rates obtained by the GLMM and the GEE approach. Complete results of these model fits are given in Tables B.1a to B.2b, in Appendix B. In the 250 mcg group, estimated NAB effects given by the two approaches based on the collapsed relapse data are very similar. The time adjusted increase in annualized relapse rate associated with switching from NAB− to NAB+ is about 15%. For NAB+ sublevels, the increases with respect to NAB− are about 13% for Low NAB+, 22% for Medium NAB+ and 10% for High NAB+. We define the relative efficiency of GEE-EXCH by the ratio of the variance of the GLMM to that of the GEE estimate. Based on the estimates of the log relative rates, $\hat{\beta}_2$, the relative efficiency of GEE-EXCH ranges from 93.2% to 109.6% in the 250 mcg group and from 81.3% to 94.8% in the 500 mcg group. Based on estimates of relative rates in Table 4.4a, the relative efficiency of

4.3. Results

GEE-EXCH ranges from 91.7% to 107.4% for the 250 mcg group and from 79.4% to 90.4% for the 500 mcg group. The estimated relative rates obtained by either approach have similar precision for the 250 mcg group but some noticeable loss of efficiencies for GEE-EXCH is observed for the 500 group.

In the 250 mcg group, no convincing NAB effects is demonstrated by either approach, just as in the analyses based on the daily data. The estimated relative rates based on the collapsed data are higher than the corresponding estimates based on the daily data. The pattern of the relative rates with respect to low, medium and high levels of NABs is similar. In the 500 mcg group, estimated increases in relapse rate given by the two approaches are similar, with those given by GLMM always slightly smaller. The estimated increases in relapse rate associated with switching from NAB– to NAB+ are about 29% and 29% from GEE-EXCH and GLMM, respectively. For the NAB sublevels, the estimated increases in relapse rates relative to NAB– are 11% and 10% for Low NAB+, 54% and 55% for Medium NAB+ and 60% and 59% for High NAB+. GLMM has a higher efficiency in this case with the relative efficiency of GEE-EXCH ranging from 79.4% to 90.4%. None of the estimated NAB effects are statistically significant at the 0.05 level. The estimated relative rates given by both approaches closely resemble the corresponding estimated relative risks based on the daily relapse data.

From the model fits in Tables B.1a to B.2b, the derived marginal intercepts $\hat{\beta}_0 + \frac{1}{2}\hat{\sigma}_b^2$ estimated from the GLMM fits are larger than the estimated marginal intercepts $\hat{\beta}_0^*$ by about 0.06 for the 250 mcg group and about 0.11 for the 500 mcg group. For the GEE approach, the estimated exchangeable correlation $\hat{\alpha}$ ranges from 0.050 to 0.054 for the 250 mcg group and from 0.105 to 0.107 for the 500 mcg group.

The precision in estimating the log relative rate β_2 is similar for the col-

lapsed and the daily data for GLMM. In the 250 mcg group, the variances of the estimated log relative rates ($\hat{\beta}_2$) based on the collapsed data are about 3% to 4% larger than the variances based on the daily relapse data. In the 500 mcg group, the variances of $\hat{\beta}_2$ based on the collapsed data are about -1% to 2% larger than those based on the daily relapse data. However, a more severe loss of efficiency due to collapsing is observed for GEE-EXCH. In the 250 mcg group, the variances based on the collapsed data are about 10% to 16% larger than the variances based on the daily relapse data. In the 500 mcg group, the variances are inflated by about 17% to 20%.

The comparison of precision based on estimated relative rates $\hat{RR} = \exp(\hat{\beta}_2)$ shows similar but amplified patterns. For GLMM, the inflation of variance due to collapsing is about 10% to 20% in the 250 mcg group and about -3% to 3% in the 500 mcg group. For GEE-EXCH, the variances are inflated by about 23% to 28% in the 250 mcg group and about 17% to 29% in the 500 mcg group.

4.3.2 NAB Titer

The NAB effect based on NAB titers is also examined by replacing the NAB status covariate, NAB_{ij} , by log NAB titers with a cutoff at k NU/mL, $\log\left(\frac{\min\{\text{Titer}_{ij}, k\}}{20}\right)$. Results based on NAB titers are presented in Tables 4.5a and 4.5b. Similar patterns as with the daily relapse data are observed with the collapsed data.

For the 250 mcg group, similar results are given by GLMM and GEE-EXCH with the estimates obtained by GLMM always slightly larger. The estimated increases in relapse rate for doubled titers are about 3% for GEE-EXCH and 4% for GLMM when a cutoff at 100 NU/mL is used. For higher cutoffs, the estimates are less than 1.5%. The estimated increases in relapse rate at the cutoffs relative to titer value ≤ 20 NU/mL are less than or equal

Table 4.4: Estimated relative risks of NAB+ and sublevels to NAB−: eventually NAB+ patients only, collapsed data.

(a) The 250 mcg group							
<i>All Switches considered; confirmed NAB+, unconfirmed L/M/H and confirmed NAB−</i>							
	GEE-EXCH				GLMM		
	\widehat{RR}	SE	95% CI		\widehat{RR}	SE	95% CI
NAB+ vs. NAB−	1.092	0.172	(0.802,	1.487)	1.102	0.188	(0.788, 1.541)
Low NAB+ vs. NAB−	1.075	0.181	(0.774,	1.495)	1.094	0.199	(0.766, 1.564)
Med NAB+ vs. NAB−	1.158	0.256	(0.751,	1.786)	1.150	0.254	(0.747, 1.772)
High NAB+ vs. NAB−	1.035	0.270	(0.621,	1.725)	1.023	0.277	(0.602, 1.738)

(b) The 500 mcg group							
<i>All Switches considered; confirmed NAB+, unconfirmed L/M/H and confirmed NAB−</i>							
	GEE-EXCH				GLMM		
	\widehat{RR}	SE	95% CI		\widehat{RR}	SE	95% CI
NAB+ vs. NAB−	1.293	0.222	(0.923,	1.811)	1.287	0.236	(0.898, 1.843)
Low NAB+ vs. NAB−	1.111	0.223	(0.749,	1.647)	1.099	0.228	(0.732, 1.650)
Med NAB+ vs. NAB−	1.541	0.339	(1.001,	2.370)	1.547	0.345	(1.000, 2.394)
High NAB+ vs. NAB−	1.602	0.379	(1.007,	2.548)	1.593	0.384	(0.993, 2.557)

\widehat{RR} : estimated relative risk.

Titers for NAB status — Low: [20, 100), Med: [100, 400), High: [400, ∞).

4.3. Results

8.5%. No convincing association between NAB titers and relapse rates is demonstrated; none of the increases are statistically significant at the 0.05 level. Compared to results based on the daily relapse data, estimated relative rates based on the collapsed data are always slightly higher.

For the 500 mcg group, results based on the two approaches also resemble each other but with the estimates based on the GEE approach always slightly larger. For doubled titers, the estimated increases in relapse rate are about 20% for GEE-EXCH and 18% for GLMM for a cutoff at 100 NU/mL, about 13% and 12% for a cutoff at 400 NU/mL, about 11% and 10% for a cutoff at 1000 NU/mL and about 9% and 8% for no cutoff. The maximum increases achievable at cutoff with respect to titer ≤ 20 NU/mL are about 54% for GEE-EXCH and 47% for GLMM for cutoff at 100 NU/mL, about 70% and 64% for cutoff at 400 NU/mL and about 81% and 74% for cutoff at 1000 NU/mL. Regardless of the cutoff level chosen, the NAB effects are statistically significant at the 0.05 level, indicating some evidence of a positive association between NAB titers and relapse rate. The efficiency in estimating the relative rates are similar for the two approaches in the 250 mcg group, whereas in the 500 mcg group, the GLMM approach gives somewhat more precise estimates in terms of the relative rates.

Complete results of the model fits are listed in Tables B.3 and B.4 in Appendix B. Comparing the estimated marginal intercept $\hat{\beta}_0^*$ to the derived marginal intercept $\hat{\beta}_0 + \frac{1}{2}\hat{\sigma}_b^2$, the latter is larger by about 0.06 for the 250 mcg group and about 0.10 to 0.11 for the 500 mcg group. For the GEE approach, the estimated exchangeable correlation $\hat{\alpha}$ ranges from 0.051 to 0.052 for the 250 mcg group and from 0.097 to 0.098 for the 500 mcg group.

The loss of efficiency due to collapsing is assessed by comparing the variances of the estimated log relative risks ($\hat{\beta}_2$) based on the collapsed data to

the variances of the estimates based on the daily relapse data. For GLMM, the variances of the estimates based on the collapsed data are about 0% to 3% larger than the variances based on the daily relapse data. For GEE-EXCH, the variances of the estimates based on the collapsed data are about 5% to 10% larger than the variances of the estimates based on the daily relapse data. A consistently larger loss of efficiency due to collapsing is observed for GEE-EXCH compared to GLMM.

4.4 Impact of Exceptionally Short Intervals

The analyses of the collapsed data presented in the previous section are based on all intervals. In the 250 mcg group, 1262 of the 1462 intervals have length longer than 1 day. Among the 200 1-day intervals, 87% are NAB+ and 1 of the NAB- 1-day intervals had a relapse onset. In the 500 mcg group, 1295 of the 1500 intervals have length longer than 1 day. Among the 205 1-day intervals, 86% are NAB+ and 1 of the NAB+ 1-day intervals had a relapse onset. Since the empirical relapse rate in these 2 exceptionally short intervals with relapse onset is 1 per day, we are concerned that they may have excessive impact on the estimated rates. On the other hand, negligible effect on the standard error is expected if these 1-day intervals are excluded from the dataset as they contribute little to the total exposure time in the complete dataset. The impacts of these 1-day intervals are investigated in this section.

Tables 4.6a and 4.6b show the fitted log relative rates with and without the 1-day intervals for both IFNB treatment groups. For the 250 mcg group, the estimates based on all intervals are consistently smaller than the estimates based only on intervals longer than one day for both GEE-EXCH and GLMM. The increase in $\hat{\beta}$ due to exclusion of the 1-day intervals ranges from 17% to 65%. The standard errors of the estimates are almost the same under

Table 4.5: Estimated relative risk of NAB titers: eventually NAB+ patients only, collapsed data.

	(a) The 250 mcg group							
	GEE-EXCH				GLMM			
	\widehat{RR}	SE	95% CI		\widehat{RR}	SE	95% CI	
cutoff= ∞								
Doubled titer before cutoff	1.004	0.039	(0.929,	1.084)	1.006	0.040	(0.931,	1.086)
cutoff=1000								
Doubled titer before cutoff	1.010	0.046	(0.925,	1.104)	1.012	0.045	(0.929,	1.103)
At cutoff 1000 NU/mL vs. 20 NU/mL	1.058	0.270	(0.642,	1.743)	1.071	0.266	(0.659,	1.742)
cutoff=400								
Doubled titer before cutoff	1.008	0.052	(0.911,	1.116)	1.011	0.051	(0.916,	1.116)
At cutoff 400 NU/mL vs. 20 NU/mL	1.037	0.233	(0.668,	1.610)	1.048	0.229	(0.683,	1.608)
cutoff=100								
Doubled titer before cutoff	1.031	0.082	(0.881,	1.206)	1.036	0.082	(0.887,	1.210)
At cutoff 100 NU/mL vs. 20 NU/mL	1.073	0.199	(0.746,	1.544)	1.085	0.199	(0.757,	1.556)

Table 4.5: – Continued

	(b) The 500 mcg group							
	GEE-EXCH				GLMM			
	\widehat{RR}	SE	95% CI		\widehat{RR}	SE	95% CI	
cutoff= ∞								
Doubled titer before cutoff	1.085	0.041	(1.008,	1.168)	1.082	0.037	(1.012,	1.158)
cutoff=1000								
Doubled titer before cutoff	1.110	0.051	(1.015,	1.214)	1.103	0.044	(1.021,	1.192)
At cutoff 1000 NU/mL vs. 20 NU/mL	1.805	0.465	(1.089,	2.991)	1.740	0.390	(1.121,	2.699)
cutoff=400								
Doubled titer before cutoff	1.133	0.062	(1.017,	1.261)	1.122	0.053	(1.022,	1.231)
At cutoff 400 NU/mL vs. 20 NU/mL	1.714	0.407	(1.076,	2.728)	1.644	0.338	(1.099,	2.459)
cutoff=100								
Doubled titer before cutoff	1.204	0.109	(1.009,	1.438)	1.181	0.093	(1.012,	1.377)
At cutoff 100 NU/mL vs. 20 NU/mL	1.539	0.323	(1.020,	2.323)	1.470	0.269	(1.028,	2.103)

both the complete and the reduced data set, with those based on the latter being very slightly larger in general. The conclusion on the NAB effects is not affected by exclusion of 1-day intervals as the p-values are still quite high, larger than about 0.3.

For the 500 mcg group, the estimates and their standard errors are very similar for model fits based on all intervals and those based only on intervals longer than one day for both GEE-EXCH and GLMM. For GEE-EXCH, the standard errors are always slightly smaller based on intervals longer than one day, whereas for GLMM, the standard errors are almost identical.

It is worth noting that, when the 1-day intervals are excluded, the correlation estimated by GEE-EXCH increases from 0.05 to 0.15 for the 250 mcg group and from 0.11 to about 0.12 to 0.13 for the 500 mcg group. The higher correlation may partly compensate for the increase in variance due to reduced total follow-up time. On the other hand, for GLMM, the estimated variance components are similar for both versions of the data.

Although exclusion of the 1-day intervals from the model fits results in increases in estimated log relative rates in the 250 mcg group, the general conclusions about the NAB effects are not altered. On the contrary, the model fits in the 500 mcg group are not noticeably affected.

4.5 Summary

Collapsing the daily relapse data to less frequent responses greatly reduces the computational burden of model estimation at a potential cost of reduced statistical efficiency. For our particular cases, the loss of efficiency for estimating β_2 is small for GLMM; the variances of $\hat{\beta}_2$ increased by less than 5%. It appears that collapsing the sparse daily response vector, consisting mostly

Table 4.6: Impact of 1-day intervals on estimates of log relative rates, β .

(a) The 250 mcg group

All Switches considered; confirmed NAB+, unconfirmed L/M/H and confirmed NAB−

	All Intervals				Intervals > 1 day			
	$\hat{\beta}$	SE	z-score	P value	$\hat{\beta}$	SE	z-score	P value
GEE-EXCH								
NAB+ vs. NAB−	0.142	0.167	0.85	0.40	0.171	0.167	1.02	0.31
Low NAB+ vs. NAB−	0.125	0.181	0.70	0.49	0.146	0.179	0.81	0.42
Med NAB+ vs. NAB−	0.201	0.232	0.87	0.39	0.249	0.233	1.07	0.29
High NAB+ vs. NAB−	0.094	0.274	0.34	0.73	0.155	0.275	0.56	0.57
GLMM								
NAB+ vs. NAB−	0.138	0.174	0.79	0.43	0.182	0.177	1.03	0.30
Low NAB+ vs. NAB−	0.122	0.185	0.66	0.51	0.164	0.188	0.87	0.38
Med NAB+ vs. NAB−	0.195	0.224	0.87	0.38	0.241	0.226	1.07	0.29
High NAB+ vs. NAB−	0.097	0.275	0.35	0.72	0.156	0.279	0.56	0.58

Table 4.6: – Continued

(b) The 500 mcg group

All Switches considered; confirmed NAB+, unconfirmed L/M/H and confirmed NAB–

	All Intervals				Intervals > 1 day			
	$\hat{\beta}$	SE	z-score	P value	$\hat{\beta}$	SE	z-score	P value
GEE-EXCH								
NAB+ vs. NAB–	0.269	0.188	1.43	0.15	0.263	0.183	1.44	0.15
Low NAB+ vs. NAB–	0.103	0.218	0.48	0.63	0.105	0.212	0.49	0.62
Med NAB+ vs. NAB–	0.476	0.239	1.99	0.05	0.461	0.236	1.95	0.05
High NAB+ vs. NAB–	0.502	0.256	1.96	0.05	0.496	0.252	1.97	0.05
GLMM								
NAB+ vs. NAB–	0.240	0.183	1.31	0.19	0.239	0.185	1.30	0.20
Low NAB+ vs. NAB–	0.084	0.207	0.41	0.68	0.074	0.209	0.36	0.72
Med NAB+ vs. NAB–	0.433	0.222	1.95	0.05	0.438	0.223	1.96	0.05
High NAB+ vs. NAB–	0.469	0.243	1.93	0.05	0.480	0.245	1.96	0.05

Titers for NAB status — Low: [20, 100), Med: [100, 400), High: [400, ∞).

4.5. Summary

of long strings of 0's, does not result in much loss of resolution. Thus the efficiencies for estimating the NAB effect are similar in both configurations of data.

For the GEE approach, greater loss of efficiency due to collapsing is reported; the variances of $\hat{\beta}_2$ increased by 10% to 20%. The variable length of intervals adds further variation in the mean of the responses which determine the strength of the pairwise correlations of the responses. An exchangeable “working” correlation structure approximates the true correlation structure less accurately in the case of the collapsed data than in the case of the daily relapse data. This is likely the cause of the higher loss of efficiency for GEE-EXCH than for GLMM.

Overall, the pattern of estimated relative rates based on the collapsed data resembles that based on the daily relapse data. Analyses based on the two configurations of data arrived at the same overall conclusions for the NAB effects.

For the daily relapse data in Chapter 3, the relative efficiency of GEE-EXCH for estimating the log relative risks β_2 in terms of NAB status is 100% to 108% for the 250 mcg group and 101% to 106% for the 500 mcg group. The relative efficiency demonstrated is slightly lower when the relapse rate is modelled as a function of NAB titers: 100% to 104% for the 250 mcg group and 93% to 96% for the 500 mcg group. For the collapsed data considered in this chapter, the relative efficiency of GEE-EXCH for estimating the log relative rates β_2 is 97% to 104% for the 250 mcg group and 93% to 98% for the 500 mcg group, when NAB status is utilized. With NAB titers, the relative efficiency for estimating β_2 is 97% to 100% for the 250 mcg group and 86% to 93% for the 500 mcg group.

4.5. Summary

The estimated relative efficiency is always higher for the 250 mcg group compared to the 500 mcg group, for daily relapse data compared to collapsed data and for using NAB status compared to using NAB titers. However, the reported figures do not provide a clear picture of whether one of the two estimation approaches is generally more efficient than the other.

As an attempt to clarify this issue, in Chapter 5, simulation studies contrasting the GEE approach to the GLMM approach are reported. In particular, scenarios with binary and sparse responses are considered. The efficiency of GEE approaches with independent (GEE-IND) and exchangeable (GEE-EXCH) correlation structure are compared to GLMM when both GEE and GLMM are correctly specified. Situations when GEE is correctly specified but GLMM is misspecified are also considered.

Chapter 5

Simulation Study

5.1 Introduction

In the BEYOND data, *eventually NAB+* patients had an average of 0.7 to 0.8 relapse onsets during the course of the study. The empirical annualized relapse risk was about 0.33 to 0.36. The average proportion of time on study as NAB– was about 36%. Assuming a random intercept model, the estimated variance component was about 0.8². In this chapter, we consider simulation studies comparing the small sample properties of the GEE and GLMM approaches used in Chapter 3 to analyze the BEYOND data. The simulations are designed such that the generated data capture some salient features of the daily relapse data. Thus the results of the simulation are more relevant to our data analyses in Chapter 3. The number of daily relapse observations for each patient in the BEYOND trial is about 800. Conducting simulations with this cluster size is not feasible with the computing resources currently available to us. Thus, much smaller cluster sizes are considered in our simulations. A typical simulated cluster will have roughly one event, as for the BEYOND dataset. The simulations do not fully represent the scenario of the BEYOND data, nevertheless, they shed some light on the performance of the GLMM and GEE approaches for clustered binary data when the event is rare.

Specifically, the efficiency of GEE-EXCH and GEE-IND are compared to GLMM when (1) the observations follow the distribution modelled by GLMM and (2) when the marginal and conditional mean structures are the same but some aspect of the distribution of the random effect is misspecified

in GLMM. Three simulation studies are conducted. The first study examines primarily the finite sample efficiency of GEE when the GLMM is the true model. The relative efficiency is defined as the reciprocal of the ratio of empirical variances. The finite sample biases are also examined. The second and third studies examine the impact of two forms of model misspecification on GEE and GLMM.

In all the simulations, clustered binary responses are generated from the log-linear models for binary response described in Section 2.5¹. The Bernoulli probability is associated with a binary covariate while no time trend is considered. As shown previously in Section 2.5, the only difference between the marginal and conditional mean structure is the intercept term. The estimated regression coefficients for the binary covariate obtained by GLMM and GEE are thus directly comparable in all three simulation studies.

The model fitting routine for GLMM implemented for the simulations is `glmer()` of the R package `lme4`, which is computationally faster than `glmm()` in the R package `repeated`. In `glmer()`, the random effect is assumed to be normal, the variance component is estimated by maximum likelihood and fifteen quadrature points are used for the adaptive Gaussian quadrature. For the GEE approach, the function `gee()` in the R package `gee` is utilized. The standard errors are estimated by the robust sandwich estimator of variance. The moment estimates for the correlation (when an exchangeable “working” correlation is used) and the sandwich estimator of variance proposed by Liang and Zeger[13] are implemented in `gee()`.

¹The mixed-effects models for the second simulation study in Section 5.3 are exceptions with some changes to the random intercept which does not alter the marginal mean structure.

5.2 Simulation I: Efficiency of GEE

Assume there are M independent subjects, each with n_i repeated observations. Given a random intercept b_i , the binary responses Y_{ij} are independent Bernoullis with conditional mean $\pi_{ij}^{(C)}$ modelled as,

$$\log\left(\pi_{ij}^{(C)}\right) = \beta_0 + b_i + \beta_1 x_{ij},$$

where x_{ij} is a binary covariate. The random intercept b_i follows a normal distribution with mean 0 and standard deviation $\sigma_b = 0.3$. Since a log link and a normal random effect are used, the simulated Bernoulli probabilities are not bounded above by 1. This smaller variance component than that estimated for the BEYOND data (≈ 0.8) is therefore chosen, so that the simulated probabilities would very seldom exceed one. In fact, for all three simulation studies, all our simulated probabilities are within 0 and 1.

Combinations of the following covariate and parameter settings are considered in all simulations unless otherwise specified:

A Number of observations per subject:

1. $n_i = 15$, for all i .
2. n_i uniformly distributed over the interval $[14, 15]$.
3. n_i uniformly distributed over the interval $[12, 15]$.

B Baseline probability: $\beta_0 = \log(1/15)$, so that a typical patient (with $b_i \approx 0$) has roughly one event on average among the repeated observations in the absence of the covariate effect.

C Effect of the binary covariate x :

1. $\beta_1 = \log(1.0) = 0$.
2. $\beta_1 = \log(1.3) \approx 0.26$.

3. $\beta_1 = \log(1.5) \approx 0.41$.

D Distribution of covariate x for individual subject: For each subject, the covariate x_{ij} starts with a sequence of 0's followed by a sequence of 1's:

1. The first 30% of x_{ij} values are 0 and the subsequent 70% are 1.
2. The proportion of x_{ij} values that equal 0 is distributed uniformly from 20% to 50%.

When $\beta_1 = 0$ and $\sigma_b = 0.3$,

$$P\left(0.04 < \pi_{ij}^{(C)} < 0.18\right) \approx 0.95$$

and

$$P\left(0.03 < \pi_{ij}^{(C)} < 0.22\right) \approx 0.99.$$

For $\beta_1 = 0$, the within-subject correlation is about 0.007. For $\beta_1 = \log(1.5)$, the correlation ranges from 0.007 when both $x = 0$ to 0.011 when both $x = 1$.

The number of simulated replicates is 1000 for each setting. The model fitting algorithms used are `glmer(lme4)` for GLMM and `gee(gee)` for GEE-exchangeable (GEE-EXCH) and GEE-independent (GEE-IND). Sometimes the GLMM model fitting algorithm, `glmer()`, complains about boundary problems. The actual cause of the error is still unknown. The algorithm basically complains that some fitted probabilities are numerically zero. Occasionally, this error can be solved by reducing the number of quadrature points to as low as 3. For the same data sets, this problem disappears when the log link is replaced by the logit link. Simulated replicates with this type of error are replaced until no errors are reported. The number of replicates replaced is also recorded. No errors were encountered with `gee()` in any of our reported simulation studies even for the replicates causing a problem for `glmer()`.

Results

Tables 5.1a to 5.1c show the results of this simulation. Table 5.1a shows the estimated biases, the empirical standard errors and the average standard errors: the average robust standard errors for GEE estimates and the average model-based standard errors for GLMM. Estimates obtained by GEE with exchangeable and independent “working” correlation structures are almost identical. Similar patterns of estimated bias in β_1 are demonstrated by GLMM and GEE. The estimated bias is consistently small, usually within ± 0.015 . Exceptions include the three cases where the proportion of $x_{ij} = 0$ within cluster is 30% and $\beta_1 = 0$ (RR = 1.0), where the estimated biases range from 0.020 to 0.026 and the last of the cases where the proportion of $x_{ij} = 0$ within cluster ranges from 20% to 50% and $\beta_1 = \log(1.5)$, where the estimated biases are about -0.044 . For GEE, no specific patterns of discrepancy between the average robust standard errors and the empirical standard errors are observed; differences are within about -4% to 6% . For GLMM, the average model-based standard errors are consistently larger than the empirical standard errors by up to about 8% except in one case where the average model-based standard error is smaller by about 2% .

As shown in Table 5.1b, the efficiencies of GEE-EXCH and GEE-IND relative to GLMM are consistently larger than 1 by a slight margin across all 18 cases considered. The GEE approaches are as efficient as GLMM in this simulation.

Table 5.1c reports the average estimated variance component and the numbers of simulated replicates replaced due to errors reported by `glmer`. The variance component is underestimated by about 10% to 30% as about 14% to 23% of the estimated variance components are numerically zero. The proportion of numerically zero estimated variance components increases with the variability of cluster size in general. Moderate to large number of sim-

ulated replicates are replaced, with the number increasing with the relative risk and the variability in cluster size. For the cases with relative risk 1.5 and cluster size ranging from 12 to 15, more than 1000 replicates were replaced, inducing substantial bias in the types of samples included in the simulation.

More Clusters

Since the responses and covariates are binary, the pattern of observations on each cluster can be represented by a 2×2 table. In our simulation, the Bernoulli probabilities are very low and the vector of covariate values are similar across clusters, or identical in some cases. Hence, there may be very little variation in terms of the pattern of observations across clusters. The predicted random effects for clusters sharing the same pattern should intuitively be identical, so the estimation of the variance component is essentially based on a few groups of predicted random effects taking the same values. This partly explains the reason for poor estimation of the variance component.

The same simulations were repeated with the number of subjects M increased from 250 to 1000 to see if this leads to better estimation of the variance components. Results are reported in Tables 5.2a to 5.2c. GEE-EXCH and GEE-IND again give essentially identical results. GLMM and GEE show similar patterns of estimated biases, which are small, less than 0.012. The only exceptions are the two cases where $\beta_1 = \log(1.5)$ and n_i is uniformly distributed from 12 to 15, where the estimated bias is about -0.031 for the case where the proportion of $x_{ij} = 0$ is 30% and about -0.036 for the case where the proportion of $x_{ij} = 0$ ranges from 20% to 50%. For GLMM, the differences of the model-based standard errors from the empirical standard errors range from about -2% to 7% , whereas those for GEE range from about -4% to 5% . The estimated relative efficiencies of GEE-EXCH are consistently greater than 1 but by less than 1%. Similar relative

efficiencies are observed for GEE-IND.

For the variance component, less than 4% of the estimates are numerically 0. A decreasing trend of the average estimated variance component with respect to increasing relative risk is observed, which was also observed in the previous simulation with $M = 250$. The numbers of replicates replaced are small for relative risk 1.0. However, the numbers of replacements are excessive when the relative risk and the variation in cluster size increases. The number of replicates replaced is larger than 10000 for the case with $RR = 1.5$ and cluster size varying from 12 to 15. The simulation results may not be reliable for these cases.

Larger Cluster Size

A third simulation under the scenario of this section was conducted, where the efficiency of GEE is compared to GLMM hopefully in the absence of possible effects from poor estimation of variance components. The number of clusters, M , is 250. The cluster sizes are distributed uniformly from 80 to 100, and $\beta_0 = \log(0.05)$ so that about 5 events are expected in a typical cluster with size 100 in the absence of covariate effect ($\beta_1 = 0$). The larger β_0 leads to higher within cluster correlation. In the absence of the covariate effect, the within cluster correlation is about 0.005. Results are presented in Tables 5.3a to 5.3c.

In all cases for both GLMM and GEE, the estimated biases are negligible, less than 0.01, and the average standard errors and the empirical standard errors are in general agreement. The relative efficiency of GEE-EXCH is consistently slightly larger than 1, but by less than 1%. The relative efficiency of GEE-IND ranges from about 98% to 100%. Table 5.3c shows that the variance component is accurately estimated and no estimates are numerically 0. However, a modest number of simulated replicates are still replaced due

to the boundary problem in `glmer()`.

Summary

In this section, we considered a scenario of clustered data with sparse binary events and a binary covariate which varies within cluster. Clustered data is a more general type of data which encompasses longitudinal data, repeated measurement and familial data. In contrast to longitudinal data, a natural chronological ordering of the observations may not necessarily be relevant to clustered data. In this setting, GEE-EXCH demonstrates the same efficiency as GLMM in estimating the log relative risk β_1 . The GEE-IND approach also demonstrates high efficiency. The GEE-IND and the GEE-EXCH approaches are almost equivalent in the first two simulation settings. In the third setting where a larger cluster size is considered, the relative efficiency of GEE-IND is still close to 100%, although clearly less than that of GEE-EXCH.

Table 5.1: Results of Simulation I–1

(a) Comparison of estimated log relative risk, β_1

Number of subjects, $M = 250$

$\beta_0 = \log(1/15)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GLMM-MLE			GEE-EXCH			GEE-IND		
			Bias	SE $^*_{(M)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$
30%	1.0	15	0.022	0.165	0.159	0.022	0.161	0.159	0.022	0.161	0.159
		14–15	0.026	0.166	0.158	0.026	0.162	0.158	0.026	0.162	0.158
		12–15	0.020	0.167	0.157	0.020	0.163	0.157	0.020	0.163	0.157
	1.3	15	0.009	0.161	0.165	0.009	0.157	0.164	0.009	0.157	0.165
		14–15	0.006	0.161	0.159	0.005	0.158	0.158	0.005	0.158	0.158
		12–15	0.007	0.162	0.160	0.007	0.159	0.160	0.007	0.159	0.160
	1.5	15	0.005	0.159	0.156	0.004	0.156	0.155	0.004	0.156	0.155
		14–15	-0.001	0.159	0.156	-0.002	0.156	0.156	-0.002	0.156	0.156
		12–15	-0.012	0.160	0.157	-0.013	0.158	0.156	-0.013	0.158	0.156
20%–50%	1.0	15	0.003	0.127	0.127	0.004	0.124	0.127	0.004	0.124	0.127
		14–15	0.006	0.128	0.119	0.006	0.125	0.119	0.006	0.125	0.119
		12–15	-0.002	0.131	0.123	-0.002	0.128	0.122	-0.002	0.128	0.122
	1.3	15	0.003	0.121	0.114	0.003	0.118	0.114	0.003	0.118	0.114
		14–15	-0.004	0.122	0.119	-0.005	0.119	0.119	-0.005	0.119	0.119
		12–15	-0.015	0.124	0.119	-0.016	0.121	0.119	-0.016	0.121	0.119
	1.5	15	-0.001	0.118	0.114	-0.002	0.116	0.114	-0.002	0.116	0.114
		14–15	-0.013	0.119	0.118	-0.014	0.116	0.118	-0.014	0.116	0.118
		12–15	-0.044	0.121	0.113	-0.045	0.119	0.112	-0.045	0.119	0.112

* Averaged model-based standard error.

† Averaged robust standard error.

‡ Empirical standard error, i.e. sample standard deviation of simulation replicates.

5.2. Simulation I: Efficiency of GEE

Table 5.1: – Continued

(b) Relative efficiency[†] to maximum likelihood for estimating β_1

Number of subjects, $M = 250$

$\beta_0 = \log(1/15)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GEE-EXCH	GEE-IND
30%	1.0	15	1.005	1.004
		14–15	1.005	1.004
		12–15	1.002	1.004
	1.3	15	1.004	1.001
		14–15	1.004	1.004
		12–15	1.004	1.005
	1.5	15	1.005	1.004
		14–15	1.003	1.004
		12–15	1.004	1.004
20%–50%	1.0	15	1.005	1.005
		14–15	1.006	1.006
		12–15	1.006	1.007
	1.3	15	1.005	1.006
		14–15	1.004	1.001
		12–15	1.005	1.001
	1.5	15	1.004	1.005
		14–15	1.004	1.004
		12–15	1.003	1.003

[†]Relative efficiency defined as reciprocal of the ratio of empirical variances.

5.2. Simulation I: Efficiency of GEE

Table 5.1: – Continued

(c) Estimation of variance component in GLMM and computation issues of `glmer()`

Number of subjects, $M = 250$
 $\beta_0 = \log(1/15)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	Average $\hat{\sigma}_b$	Prop. $\hat{\sigma}_b < 0.01$	Replicates replaced
30%	1.0	15	0.271	18.7%	112
		14–15	0.268	21.0%	108
		12–15	0.255	22.8%	160
	1.3	15	0.261	14.0%	158
		14–15	0.260	13.7%	206
		12–15	0.248	17.7%	320
	1.5	15	0.251	11.2%	300
		14–15	0.257	11.7%	433
		12–15	0.234	16.1%	1169
20%–50%	1.0	15	0.270	19.6%	78
		14–15	0.272	18.3%	102
		12–15	0.261	22.0%	150
	1.3	15	0.269	14.9%	158
		14–15	0.265	13.9%	208
		12–15	0.250	19.5%	358
	1.5	15	0.254	13.0%	365
		14–15	0.243	15.6%	565
		12–15	0.216	23.2%	1744

Table 5.2: Results of Simulation I-2: More Clusters

(a) Comparison of estimated log relative risk, β_1

Number of subjects, $M = 1000$

$\beta_0 = \log(1/15)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GLMM-MLE			GEE-EXCH			GEE-IND		
			Bias	SE $^*_{(M)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$
30%	1.0	15	0.003	0.082	0.079	0.003	0.080	0.078	0.003	0.080	0.078
		14-15	0.004	0.082	0.080	0.004	0.080	0.080	0.004	0.080	0.080
		12-15	0.006	0.083	0.078	0.006	0.081	0.077	0.006	0.081	0.077
	1.3	15	0.003	0.080	0.075	0.002	0.078	0.075	0.002	0.078	0.075
		14-15	0.005	0.080	0.082	0.004	0.079	0.082	0.004	0.079	0.082
		12-15	-0.001	0.081	0.081	-0.001	0.079	0.080	-0.001	0.079	0.080
	1.5	15	-0.003	0.079	0.077	-0.004	0.078	0.077	-0.004	0.078	0.077
		14-15	-0.007	0.079	0.078	-0.008	0.078	0.078	-0.008	0.078	0.078
		12-15	-0.031	0.080	0.075	-0.032	0.078	0.075	-0.031	0.078	0.075
20%-50%	1.0	15	0.002	0.063	0.061	0.002	0.062	0.061	0.002	0.062	0.061
		14-15	0.003	0.064	0.062	0.003	0.062	0.062	0.003	0.063	0.062
		12-15	0.000	0.066	0.062	0.000	0.064	0.062	0.000	0.064	0.062
	1.3	15	-0.002	0.060	0.059	-0.003	0.059	0.059	-0.003	0.059	0.059
		14-15	-0.002	0.061	0.061	-0.003	0.060	0.061	-0.003	0.060	0.061
		12-15	-0.010	0.062	0.058	-0.010	0.061	0.058	-0.010	0.061	0.058
	1.5	15	-0.003	0.059	0.055	-0.004	0.058	0.055	-0.004	0.058	0.055
		14-15	-0.011	0.059	0.057	-0.012	0.058	0.057	-0.012	0.058	0.057
		12-15	-0.036	0.061	0.057	-0.037	0.059	0.057	-0.037	0.059	0.057

* Averaged model-based standard error.

† Averaged robust standard error.

‡ Empirical standard error, i.e. sample standard deviation of simulation replicates.

5.2. Simulation I: Efficiency of GEE

Table 5.2: – Continued

(b) Relative efficiency[†] to maximum likelihood for estimating β_1

Number of subjects, $M = 1000$

$\beta_0 = \log(1/15)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GEE-EXCH	GEE-IND
30%	1.0	15	1.007	1.009
		14–15	1.006	1.005
		12–15	1.005	1.007
	1.3	15	1.004	1.005
		14–15	1.006	1.007
		12–15	1.005	1.005
	1.5	15	1.006	1.006
		14–15	1.005	1.006
		12–15	1.004	1.002
20%–50%	1.0	15	1.004	1.002
		14–15	1.006	1.003
		12–15	1.006	1.007
	1.3	15	1.006	1.007
		14–15	1.006	1.004
		12–15	1.004	1.001
	1.5	15	1.005	1.003
		14–15	1.004	1.000
		12–15	1.002	0.999

[†]Relative efficiency defined as reciprocal of the ratio of empirical variances.

5.2. Simulation I: Efficiency of GEE

Table 5.2: – Continued

(c) Estimation of variance component in GLMM and computation issues of `glmer()`

Number of subjects, $M = 1000$
 $\beta_0 = \log(1/15)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	Average $\hat{\sigma}_b$	Prop. $\hat{\sigma}_b < 0.01$	Replicates replaced
30%	1.0	15	0.324	1.7%	36
		14–15	0.323	2.8%	36
		12–15	0.323	3.6%	67
	1.3	15	0.316	0.6%	102
		14–15	0.313	0.9%	162
		12–15	0.309	0.8%	640
	1.5	15	0.302	0.2%	612
		14–15	0.300	0.4%	1258
		12–15	0.280	1.6%	11624
20%–50%	1.0	15	0.328	2.3%	16
		14–15	0.330	2.2%	36
		12–15	0.325	3.6%	63
	1.3	15	0.318	0.7%	141
		14–15	0.313	1.2%	234
		12–15	0.305	1.9%	908
	1.5	15	0.302	0.5%	1008
		14–15	0.296	0.9%	2234
		12–15	0.267	3.7%	24633

Table 5.3: Results of Simulation I–3: Larger Cluster Size

(a) Comparison of estimated log relative risk, β_1

Number of subjects, $M = 250$
 $\beta_0 = \log(5/100)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GLMM-MLE			GEE-EXCH			GEE-IND		
			Bias	SE $^*_{(M)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$
30%	1.0	100	0.007	0.076	0.073	0.007	0.075	0.073	0.007	0.075	0.073
		90–100	0.009	0.076	0.077	0.009	0.075	0.077	0.009	0.075	0.077
		80–100	0.003	0.077	0.074	0.003	0.076	0.074	0.003	0.076	0.074
	1.3	100	0.001	0.074	0.072	0.001	0.073	0.072	0.001	0.074	0.073
		90–100	0.003	0.075	0.072	0.003	0.074	0.072	0.004	0.074	0.072
		80–100	0.003	0.075	0.073	0.002	0.074	0.073	0.002	0.074	0.073
20%–50%	1.0	100	0.004	0.057	0.057	0.004	0.056	0.057	0.004	0.057	0.057
		90–100	-0.002	0.058	0.057	-0.001	0.057	0.057	-0.002	0.058	0.057
		80–100	0.002	0.059	0.058	0.002	0.058	0.058	0.002	0.059	0.058
	1.3	100	-0.001	0.055	0.053	-0.001	0.054	0.053	-0.001	0.054	0.053
		90–100	0.000	0.055	0.054	0.000	0.055	0.054	0.000	0.055	0.054
		80–100	-0.001	0.056	0.056	-0.001	0.055	0.056	-0.002	0.056	0.056

* Averaged model-based standard error.

† Averaged robust standard error.

‡ Empirical standard error, i.e. sample standard deviation of simulation replicates.

5.2. Simulation I: Efficiency of GEE

Table 5.3: – Continued

(b) Relative efficiency[†] to maximum likelihood for estimating β_1

Number of subjects, $M = 250$

$\beta_0 = \log(5/100)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GEE-EXCH	GEE-IND
30%	1.0	100	1.002	0.997
		90–100	1.003	0.995
		80–100	1.002	0.988
	1.3	100	1.002	0.984
		90–100	1.001	0.994
		80–100	1.001	0.988
20%–50%	1.0	100	1.002	0.999
		90–100	1.003	0.994
		80–100	1.001	0.996
	1.3	100	1.001	1.000
		90–100	1.004	1.003
		80–100	1.002	0.995

[†]Relative efficiency defined as reciprocal of the ratio of empirical variances.

5.2. Simulation I: Efficiency of GEE

Table 5.3: – Continued

(c) Estimation of variance component in GLMM and computation issues of `glmer()`

Number of subjects, $M = 250$
 $\beta_0 = \log(5/100)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	Average $\hat{\sigma}_b$	Prop. $\hat{\sigma}_b < 0.01$	Replicates replaced
30%	1.0	100	0.300	0%	272
		90–100	0.299	0%	362
		80–100	0.297	0%	440
	1.3	100	0.296	0%	159
		90–100	0.297	0%	221
		80–100	0.299	0%	334
20%–50%	1.0	100	0.296	0%	127
		90–100	0.297	0%	155
		80–100	0.296	0%	164
	1.3	100	0.297	0%	103
		90–100	0.295	0%	104
		80–100	0.298	0%	161

5.3 Simulation II: Misspecification of Random Effects

In our second scenario, we examine the impact of a form of misspecification of the random effects. All the features of this simulation study are the same as in Section 5.2 except the assumptions concerning the random effects. The conditional mean Y_{ij} is now modelled as

$$\log \left(\pi_{ij}^{(C)} \right) = \beta_0 + b_{ij} + \beta_1 x_{ij}.$$

The random effects b_{ij} are independent across subjects. Within a subject, b_{ij} follows an AR-1 process,

$$b_{ij} = \rho b_{ij-1} + \sqrt{1 - \rho^2} \epsilon_{ij},$$

for $j = 2, \dots, n_i$, where b_{i1} and ϵ_{ij} are *i.i.d.* $N(0, \sigma_b^2)$. The former conditional independence assumption is violated in the sense that within a cluster, the events tend to be clustered together, although the effect may not be apparent for binary data even with a substantially long sequence of observations.

The marginal distribution of any individual Y_{ij} does not depend on the latent stochastic process of the random effects. Since b_{ij} follows a normal distribution with zero mean and variance σ_b^2 , the marginal distribution of Y_{ij} is the same as in Section 5.2. The marginal correlation of Y_{ij} and Y_{ik} is

$$\text{Corr}(Y_{ij}, Y_{ik}) = \frac{[\exp(\rho^{|j-k|} \sigma_b^2) - 1] \mu_{ij}^{(M)} \mu_{ik}^{(M)}}{\sqrt{\mu_{ij}^{(M)} (1 - \mu_{ij}^{(M)}) \mu_{ik}^{(M)} (1 - \mu_{ik}^{(M)})}},$$

so the within-subject observations have a banded correlation structure for $\beta_1 = 0$.

We consider two choices of ρ , 0.1 and 0.5. For $\rho = 0.1$, and $\beta_1 = \log(1.5)$, the lag-1 correlation of Y_{ij} and Y_{ij+1} ranges from 0.0007 when both x 's are 0 to 0.0011 when both x 's are 1. The lag-2 and lag-3 correlation are about 0.0001 and less than 0.0001 respectively. For $\rho = 0.5$, and $\beta_1 = \log(1.5)$, the lag-1 correlation ranges from 0.0035 when both x 's are 0 to 0.0054 when both x 's are 1. The lag-2 correlation ranges from 0.0017 to 0.0027 and the lag-3 correlation ranges from 0.0008 to 0.0013.

Results

Tables 5.4a to 5.4d present the results for the sets of simulations with 250 subjects in each of the 1000 simulated replicates. The results given by GEE-EXCH and GEE-IND are almost identical which is in coherence with the negligible true marginal correlations. In fact, the estimated biases, the average standard errors and the empirical standard errors from all three approaches are very close to each other. The estimated biases for estimation of β_1 lie in the range of -0.031 to 0.022 , which converts to biases of less than 3% in estimation of the relative risk, $\exp(\beta_1)$. The average standard errors are in general agreement with the empirical standard errors with no specific pattern of discrepancy. The largest disagreements are observed in the case where $\beta_1 = \log(1.5)$, the proportion of the binary covariate x being zero ranges from 20% to 50% within cluster and the cluster sizes n_i range from 12 to 15; for both $\rho = 0.1$ and 0.5, for all three approaches, the average standard error is larger than the empirical standard error by about 7%.

The simulation results show that the GEE approaches are as efficient as the GLMM approach. For both GEE-EXCH and GEE-IND, the estimated relative efficiencies are consistently larger than 1 by a slight margin of less than 0.005 in all settings considered, except one instance for GEE-IND in which the estimated relative efficiency is 0.999.

For GLMM, about half of the estimated variance components are numerically zero, though the interpretation of this estimated variance component in the context of the true model is unclear. The number of simulated replicates replaced ranged from 4 to 763 times across the 24 cases considered. The frequency increases with β_1 as well as with variation in cluster size and the variation in the within cluster proportion of x taking the value zero.

More Clusters

Tables 5.5a to 5.5d show the results of the same set of simulations with the number of subjects increased from 250 to 1000. The bias for estimation of β_1 is smaller than 0.01 in general except in four cases where $RR = 1.5$ and cluster size ranges from 12; to 15. In these cases, the biases range from -0.010 to -0.030 , however, the number of replicates replaced ranges from about 2400 to 5469. For the standard errors, the percentage differences between the average standard error and the empirical standard error are within 6.5%.

The estimated relative efficiency of GEE-EXCH and GEE-IND is consistently slightly larger than 100% but by no more than 0.3%. For the settings with relative risk equal to 1.0, no more than 5 replicates are replaced in each case. For the settings with relative risk equal to 1.5, the number of replicates replaced increases with the variation in cluster size, ranging from about 100 to about 5000. Regarding the estimated variance component, 37.5% to 56.1% are numerically zero for low autocorrelation $\rho = 0.1$ and 48.1% to 69.4% are numerically zero for high autocorrelation $\rho = 0.5$.

Larger Cluster Size

Tables 5.6a to 5.6c present results for selected cases with 250 subjects and larger and constant cluster size of $n_i = 100$ in each simulated replicate. We considered combinations of 2 levels of log relative risk $\beta_1 = 0$ and $\log(1.3)$ and

2 levels of autocorrelation $\rho = 0.1$ and 0.5 . The same conclusions concerning the pattern of bias, the validity of the robust standard error for the GEE approaches and the model-based standard error for GLMM, and the relative efficiency of GEE are apparent. For this setting with cluster size $n_i = 100$, the numbers of replicates replaced are negligible. The observed proportions of estimated variance components numerically zero are still around 50%. The average estimated variance component ranges from 0.044 to 0.065. However, the estimation of the fixed effect is not affected by the estimated variance components being severely underestimated.

Summary

The GEE and GLMM approaches are in general robust to this investigated form of misspecification for the latent stochastic process of the random effect. The percentage of estimated variance component numerically equal to zero is higher, about 50%, compared to about 0% to 23% in Section 5.2 when the model is correctly specified. The number of replicates replaced is also much smaller in this section. Compared to Section 5.2, more or less the same patterns are revealed for bias and the accuracy of the standard error. The efficiency of GEE-EXCH and GEE-IND are almost the same as GLMM probably because the correlation within cluster is almost zero.

Table 5.4: Results of Simulation II-1

(a) Comparison of estimated log relative risk, β_1 , when the time-dependent random effect follows an AR-1 process, $\rho = 0.1$.

Autocorrelation, $\rho = 0.1$

Number of subjects, $M = 250$
 $\beta_0 = \log(1/15)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GLMM			GEE-EXCH			GEE-IND		
			Bias	SE $^*_{(M)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$
30%	1.0	15	0.013	0.163	0.161	0.013	0.161	0.161	0.013	0.161	0.161
		14–15	0.019	0.164	0.161	0.019	0.162	0.161	0.019	0.162	0.161
		12–15	0.022	0.165	0.163	0.022	0.163	0.163	0.022	0.163	0.163
	1.5	15	0.015	0.158	0.164	0.014	0.157	0.164	0.014	0.157	0.164
		14–15	0.009	0.158	0.158	0.009	0.157	0.158	0.009	0.157	0.157
		12–15	-0.010	0.158	0.157	-0.010	0.157	0.156	-0.010	0.157	0.156
20%–50%	1.0	15	-0.003	0.125	0.124	-0.003	0.123	0.124	-0.003	0.123	0.124
		14–15	0.003	0.127	0.130	0.003	0.126	0.129	0.003	0.126	0.129
		12–15	0.002	0.129	0.126	0.002	0.128	0.126	0.002	0.128	0.126
	1.5	15	0.003	0.116	0.113	0.003	0.116	0.113	0.003	0.116	0.113
		14–15	-0.007	0.117	0.115	-0.007	0.116	0.115	-0.007	0.116	0.115
		12–15	-0.025	0.120	0.113	-0.025	0.119	0.113	-0.025	0.119	0.113
Continued on Next Page...											

Table 5.4: – Continued

(b) Comparison of estimated log relative risk, β_1 , when the time-dependent random effect follows an AR-1 process, $\rho = 0.5$.

Autocorrelation, $\rho = 0.5$

Prop. $x_{ij} = 0$	RR	n_i	GLMM			GEE-EXCH			GEE-IND		
			Bias	SE $^*_{(M)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$
30%	1.0	15	0.010	0.163	0.156	0.010	0.161	0.156	0.010	0.161	0.156
		14–15	0.011	0.164	0.157	0.011	0.163	0.156	0.011	0.163	0.156
		12–15	0.005	0.164	0.153	0.005	0.162	0.153	0.005	0.163	0.153
	1.5	15	0.005	0.157	0.157	0.005	0.156	0.157	0.005	0.156	0.157
		14–15	0.009	0.158	0.156	0.009	0.157	0.156	0.008	0.157	0.156
		12–15	-0.015	0.158	0.164	-0.015	0.158	0.164	-0.015	0.158	0.164
20%–50%	1.0	15	0.001	0.125	0.124	0.001	0.124	0.123	0.001	0.124	0.123
		14–15	-0.002	0.127	0.123	-0.002	0.125	0.123	-0.002	0.126	0.123
		12–15	0.000	0.130	0.130	0.000	0.128	0.130	0.000	0.128	0.130
	1.5	15	-0.004	0.117	0.115	-0.005	0.116	0.115	-0.005	0.116	0.115
		14–15	-0.007	0.117	0.116	-0.007	0.117	0.116	-0.007	0.117	0.116
		12–15	-0.030	0.120	0.112	-0.030	0.119	0.112	-0.031	0.119	0.112

* Averaged model-based standard error.

† Averaged robust standard error.

‡ Empirical standard error, i.e. sample standard deviation of simulation replicates.

5.3. Simulation II: Misspecification of Random Effects

Table 5.4: – Continued

(c) Relative efficiency[†] to maximum likelihood for estimating β_1 , when the time-dependent random effect follows an AR-1 process.

Autocorrelation, $\rho = 0.1$

Number of subjects, $M = 250$

$\beta_0 = \log(1/15)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GEE-EXCH	GEE-IND
30%	1.0	15	1.002	1.003
		14–15	1.001	1.003
		12–15	1.002	1.001
	1.5	15	1.000	1.001
		14–15	1.001	1.003
		12–15	1.002	1.001
20%–50%	1.0	15	1.002	1.004
		14–15	1.002	1.003
		12–15	1.002	1.001
	1.5	15	1.000	1.001
		14–15	1.002	1.001
		12–15	1.000	1.001

Autocorrelation, $\rho = 0.5$

Prop. $x_{ij} = 0$	RR	n_i	GEE-EXCH	GEE-IND
30%	1.0	15	1.002	1.002
		14–15	1.000	1.002
		12–15	1.003	1.002
	1.5	15	1.002	1.000
		14–15	1.001	1.000
		12–15	1.001	0.999
20%–50%	1.0	15	1.002	1.001
		14–15	1.002	1.001
		12–15	1.003	1.004
	1.5	15	1.000	1.002
		14–15	1.004	1.001
		12–15	1.002	1.001

[†]Relative efficiency defined as reciprocal of the ratio of empirical variances.

5.3. Simulation II: Misspecification of Random Effects

Table 5.4: – Continued

(d) Estimation of variance component in GLMM and computation issues of `glmer()`, when the time-dependent random effect follows an AR-1 process.

Autocorrelation, $\rho = 0.1$

Number of subjects, $M = 250$

$\beta_0 = \log(1/15)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	Average $\hat{\sigma}_b$	Prop. $\hat{\sigma}_b < 0.01$	Replicates replaced
30%	1.0	15	0.127	54.7%	9
		14–15	0.130	50.7%	11
		12–15	0.134	52.0%	34
	1.5	15	0.100	53.0%	47
		14–15	0.096	56.4%	72
		12–15	0.084	62.2%	425
20%–50%	1.0	15	0.117	55.3%	14
		14–15	0.122	55.2%	4
		12–15	0.134	54.2%	14
	1.5	15	0.096	55.1%	79
		14–15	0.099	56.0%	151
		12–15	0.079	65.5%	712

Autocorrelation, $\rho = 0.5$

Prop. $x_{ij} = 0$	RR	n_i	Average $\hat{\sigma}_b$	Prop. $\hat{\sigma}_b < 0.01$	Replicates replaced
30%	1.0	15	0.145	48.3%	13
		14–15	0.151	48.5%	19
		12–15	0.134	52.4%	37
	1.5	15	0.115	47.2%	47
		14–15	0.120	47.5%	96
		12–15	0.100	54.7%	470
20%–50%	1.0	15	0.144	48.0%	10
		14–15	0.144	48.1%	12
		12–15	0.150	49.3%	21
	1.5	15	0.118	47.7%	89
		14–15	0.111	50.7%	206
		12–15	0.104	55.3%	763

Table 5.5: Results of Simulation II-2: More Clusters

(a) Comparison of estimated log relative risk, β_1 , when the time-dependent random effect follows an AR-1 process, $\rho = 0.1$.

Autocorrelation, $\rho = 0.1$

Number of subjects, $M = 1000$
 $\beta_0 = \log(1/15)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GLMM			GEE-EXCH			GEE-IND		
			Bias	SE $^*_{(M)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$
30%	1.0	15	0.005	0.081	0.078	0.005	0.080	0.078	0.005	0.080	0.078
		14–15	0.003	0.081	0.084	0.003	0.081	0.084	0.003	0.081	0.084
		12–15	0.004	0.082	0.080	0.004	0.081	0.080	0.004	0.081	0.080
	1.5	15	0.001	0.078	0.081	0.001	0.078	0.081	0.001	0.078	0.081
		14–15	-0.001	0.078	0.081	-0.001	0.078	0.081	-0.001	0.078	0.081
		12–15	-0.015	0.079	0.078	-0.015	0.078	0.078	-0.015	0.078	0.078
20%–50%	1.0	15	0.005	0.062	0.064	0.005	0.062	0.064	0.005	0.062	0.064
		14–15	0.001	0.063	0.062	0.001	0.063	0.062	0.001	0.063	0.062
		12–15	-0.002	0.064	0.065	-0.002	0.064	0.065	-0.002	0.064	0.065
	1.5	15	-0.004	0.058	0.056	-0.004	0.058	0.056	-0.004	0.058	0.056
		14–15	-0.006	0.058	0.057	-0.006	0.058	0.057	-0.006	0.058	0.057
		12–15	-0.028	0.059	0.056	-0.028	0.059	0.056	-0.028	0.059	0.056

Continued on Next Page...

Table 5.5: – Continued

(b) Comparison of estimated log relative risk, β_1 , when the time-dependent random effect follows an AR-1 process, $\rho = 0.5$.

Autocorrelation, $\rho = 0.5$

Prop. $x_{ij} = 0$	RR	n_i	GLMM			GEE-EXCH			GEE-IND		
			Bias	SE $^*_{(M)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$
30%	1.0	15	0.001	0.081	0.080	0.001	0.080	0.080	0.001	0.080	0.080
		14–15	0.000	0.081	0.078	0.000	0.081	0.078	0.000	0.081	0.078
		12–15	0.004	0.082	0.079	0.004	0.081	0.079	0.004	0.081	0.079
	1.5	15	-0.003	0.078	0.075	-0.003	0.078	0.075	-0.003	0.078	0.075
		14–15	-0.004	0.078	0.075	-0.005	0.078	0.075	-0.005	0.078	0.075
		12–15	-0.021	0.078	0.079	-0.021	0.078	0.079	-0.021	0.078	0.079
20%–50%	1.0	15	0.001	0.062	0.064	0.001	0.062	0.064	0.001	0.062	0.064
		14–15	-0.003	0.063	0.061	-0.003	0.063	0.061	-0.003	0.063	0.061
		12–15	-0.001	0.065	0.065	-0.001	0.064	0.065	-0.001	0.064	0.065
	1.5	15	-0.003	0.058	0.057	-0.004	0.058	0.057	-0.004	0.058	0.057
		14–15	-0.005	0.059	0.058	-0.006	0.058	0.058	-0.006	0.058	0.058
		12–15	-0.026	0.060	0.058	-0.027	0.060	0.058	-0.027	0.060	0.058

* Averaged model-based standard error.

† Averaged robust standard error.

‡ Empirical standard error, i.e. sample standard deviation of simulation replicates.

5.3. Simulation II: Misspecification of Random Effects

Table 5.5: – Continued

(c) Relative efficiency[†] to maximum likelihood for estimating β_1 , when the time-dependent random effect follows an AR-1 process.

Autocorrelation, $\rho = 0.1$

Number of subjects, $M = 1000$

$\beta_0 = \log(1/15)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GEE-EXCH	GEE-IND
30%	1.0	15	1.000	1.001
		14–15	1.001	1.003
		12–15	1.002	1.001
	1.5	15	1.001	1.000
		14–15	1.000	1.002
		12–15	1.000	1.000
20%–50%	1.0	15	1.002	1.002
		14–15	1.001	1.001
		12–15	1.001	1.003
	1.5	15	1.001	1.002
		14–15	1.000	1.000
		12–15	1.000	1.000

Autocorrelation, $\rho = 0.5$

Prop. $x_{ij} = 0$	RR	n_i	GEE-EXCH	GEE-IND
30%	1.0	15	1.001	1.002
		14–15	1.002	1.001
		12–15	1.002	1.002
	1.5	15	1.002	1.002
		14–15	1.001	1.000
		12–15	1.000	1.000
20%–50%	1.0	15	1.002	1.002
		14–15	1.001	1.001
		12–15	1.002	1.003
	1.5	15	1.001	1.000
		14–15	1.001	1.002
		12–15	1.001	1.001

[†]Relative efficiency defined as reciprocal of the ratio of empirical variances.

5.3. Simulation II: Misspecification of Random Effects

Table 5.5: – Continued

(d) Estimation of variance component in GLMM and computation issues of `glmer()`, when the time-dependent random effect follows an AR-1 process.

Autocorrelation, $\rho = 0.1$

Number of subjects, $M = 1000$

$\beta_0 = \log(1/15)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	Average $\hat{\sigma}_b$	Prop. $\hat{\sigma}_b < 0.01$	Replicates replaced
30%	1.0	15	0.098	51.7%	2
		14–15	0.108	49.9%	0
		12–15	0.101	53.9%	4
	1.5	15	0.075	52.7%	106
		14–15	0.072	52.9%	294
		12–15	0.056	64.1%	2429
20%–50%	1.0	15	0.109	48.1%	0
		14–15	0.108	49.6%	0
		12–15	0.109	51.2%	5
	1.5	15	0.078	53.3%	224
		14–15	0.067	59.0%	552
		12–15	0.049	69.4%	4581

Autocorrelation, $\rho = 0.5$

Prop. $x_{ij} = 0$	RR	n_i	Average $\hat{\sigma}_b$	Prop. $\hat{\sigma}_b < 0.01$	Replicates replaced
30%	1.0	15	0.129	39.3%	0
		14–15	0.128	42.0%	2
		12–15	0.138	41.3%	4
	1.5	15	0.107	37.5%	144
		14–15	0.107	38.6%	297
		12–15	0.079	51.5%	2904
20%–50%	1.0	15	0.131	41.1%	0
		14–15	0.125	43.4%	0
		12–15	0.140	40.1%	3
	1.5	15	0.111	39.0%	219
		14–15	0.097	44.4%	630
		12–15	0.077	56.1%	5469

Table 5.6: Results of Simulation II–3: Larger Cluster Size

(a) Comparison of estimated log relative risk, β_1 , when the time-dependent random effect follows an AR-1 process.

Autocorrelation, $\rho = 0.1$

Number of subjects, $M = 250$
 $\beta_0 = \log(5/100)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GLMM			GEE-EXCH			GEE-IND		
			Bias	SE $^*_{(M)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$
30%	1.0	100	0.002	0.075	0.075	0.002	0.074	0.075	0.002	0.074	0.075
	1.3	100	0.002	0.073	0.073	0.002	0.073	0.073	0.002	0.073	0.073
20%–50%	1.0	100	0.002	0.057	0.056	0.002	0.056	0.056	0.002	0.056	0.056
	1.3	100	0.003	0.054	0.054	0.003	0.054	0.054	0.003	0.054	0.054

Autocorrelation, $\rho = 0.5$

Prop. $x_{ij} = 0$	RR	n_i	GLMM			GEE-EXCH			GEE-IND		
			Bias	SE $^*_{(M)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$
30%	1.0	100	-0.001	0.075	0.078	-0.001	0.075	0.078	-0.001	0.075	0.078
	1.3	100	0.000	0.073	0.078	0.000	0.073	0.077	0.000	0.073	0.077
20%–50%	1.0	100	0.001	0.057	0.057	0.001	0.057	0.057	0.001	0.057	0.057
	1.3	100	0.001	0.054	0.054	0.001	0.054	0.054	0.001	0.054	0.054

* Averaged model-based standard error.

† Averaged robust standard error.

‡ Empirical standard error, i.e. sample standard deviation of simulation replicates.

Table 5.6: – Continued

(b) Relative efficiency[†] to maximum likelihood for estimating β_1 , when the time-dependent random effect follows an AR-1 process.

Autocorrelation, $\rho = 0.1$

Number of subjects, $M = 250$

$\beta_0 = \log(5/100)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GEE-EXCH	GEE-IND
30%	1.0	100	1.001	1.001
	1.3	100	1.001	1.000
20%–50%	1.0	100	1.000	1.001
	1.3	100	1.001	1.001

Autocorrelation, $\rho = 0.5$

Prop. $x_{ij} = 0$	RR	n_i	GEE-EXCH	GEE-IND
30%	1.0	100	1.001	1.002
	1.3	100	1.001	1.001
20%–50%	1.0	100	0.999	1.001
	1.3	100	1.000	1.000

[†]Relative efficiency defined as reciprocal of the ratio of empirical variances.

5.3. Simulation II: Misspecification of Random Effects

Table 5.6: – Continued

(c) Estimation of variance component in GLMM and computation issues of `glmer()`, when the time-dependent random effect follows an AR-1 process.

Autocorrelation, $\rho = 0.1$

Number of subjects, $M = 250$
 $\beta_0 = \log(5/100)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	Average $\hat{\sigma}_b$	Prop. $\hat{\sigma}_b < 0.01$	Replicates replaced
30%	1.0	100	0.049	55.4%	7
	1.3	100	0.044	55.8%	1
20%–50%	1.0	100	0.053	52.0%	2
	1.3	100	0.048	51.6%	2

Autocorrelation, $\rho = 0.5$

Prop. $x_{ij} = 0$	RR	n_i	Average $\hat{\sigma}_b$	Prop. $\hat{\sigma}_b < 0.01$	Replicates replaced
30%	1.0	100	0.057	48.8%	3
	1.3	100	0.054	45.8%	2
20%–50%	1.0	100	0.060	45.4%	2
	1.3	100	0.055	45.4%	0

5.4 Simulation III: Misspecification of the Random Effect Distribution

In the third scenario, we study the impact of a form of misspecification of the random effect distribution. All the features of the simulation study in Section 5.2 are retained in this section except for this departure of the random effect distribution.

Beta Distribution for the Baseline Conditional Probability

Firstly, a beta-binomial setup is considered. The baseline conditional probability, $P(Y_{ij} = 1 | b_i, x_{ij} = 0) = \exp(\beta_0 + b_i)$, as a function of b_i , is assumed to follow a beta distribution with mean $\pi_0 = \exp\left(\beta_0 + \frac{\sigma_b^2}{2}\right)$ and variance $\sigma_\pi^2 = (\exp(\sigma_b^2) - 1) \cdot \pi_0^2$. The equivalent parameterization in the usual shape parameters α and β of the beta distribution is

$$\alpha = \pi_0 \left(\frac{\pi_0(1 - \pi_0)}{\sigma_\pi^2} - 1 \right)$$

and

$$\beta = (1 - \pi_0) \left(\frac{\pi_0(1 - \pi_0)}{\sigma_\pi^2} - 1 \right).$$

For $\beta_0 = \log(5/100)$ and $\sigma_b = 0.3$, π_0 lies between 0.026 and 0.088 with probability 0.95 and between 0.020 to 0.010 with probability 0.99. The beta distribution for π_0 is analogous to implementing a slightly asymmetric distribution for b_i , with approximately, mean -0.003 , standard deviation 0.316 and skewness -0.35 . The probabilities corresponding to $x = 1$ are obtained by multiplying the baseline probabilities by the relative risk $\exp(\beta_1)$.

Figure 5.1 shows the probability density function of the random intercept when the baseline conditional probability follows the specified beta distri-

bution and compares it to a normal distribution to illustrate the extent it deviates from normal. The same first two moments of the normal distribution are matched to have mean -0.003 and standard deviation 0.316 . Under this beta-binomial setting, the random intercept distribution does not deviate much from a normal distribution and is only slightly left-skewed.

Results

Table 5.7a shows the estimated biases, the average standard error and the empirical standard error for the three estimation approaches. Results obtained by the three approaches are very similar, with numerical values within a margin of 0.001 . The estimated bias is small, less than 0.010 in absolute value in all cases. The average standard error and the empirical standard error closely resemble each other in general. The ratio of the grand average of the standard error to the grand average of the empirical standard error is 101.3% for GLMM and 100.0% for GEE-EXCH and GEE-IND.

Table 5.7b shows the estimated efficiencies of GEE-EXCH and GEE-IND relative to GLMM. Among the 18 cases we considered, GEE-EXCH is as efficient as GLMM with relative efficiency ranging from 99.9% to 100.4% . GEE-IND is in general less efficient in this setup as the estimated relative efficiencies mostly range from 98.2% to 100.0% with three cases where the relative efficiencies are larger than 100.0% (100.3% , 100.3% and 100.8%).

The average estimated variance components reported in Table 5.7c are close to true value, $\sigma_b = 0.3$. None of the estimated variance components is numerically 0. A considerable number of replicates, ranging from 48 to 311, are replaced.

This very mild misspecification of the random effect distribution does not appear to influence the bias and accuracy of the model-based standard

error of GLMM. In this particular setup, GEE-EXCH and GLMM have very similar efficiency, while GEE-IND has a slightly lower efficiency.

Triangular Distribution for the Random Intercept

A second setup considers a triangular distribution for the random effect. The random intercept b_i is generated from a distribution with density function

$$f_b(x) = \frac{2 \cdot (2\delta - x)}{(3\delta)^2}$$

for $x \in [-\delta, 2\delta]$. The cumulative distribution function (CDF) is

$$F_b(x) = \frac{(x + \delta) \cdot (5\delta - x)}{(3\delta)^2}$$

for $x \in [-\delta, 2\delta]$, and the inverse CDF is

$$F_b^{-1}(u) = (2 - 3\sqrt{1 - u}) \cdot \delta$$

for $u \in [0, 1]$. This triangular distribution has mean 0 and variance $\frac{\delta^2}{2}$. In this subsection, We choose a scale parameter $\delta = 0.3\sqrt{2}$ so that the standard deviation of b_i is the same as in our earlier simulations ($\sigma_b = 0.3$). Figure 5.2 compares the density of this triangular distribution with $\delta = 0.3\sqrt{2}$ to a normal distribution with zero mean and variance $\sigma_b^2 = 0.3^2$. This density function is perpendicular triangular in shape with skewness approximately 0.57. It imitates a situation when the majority of patients have relatively low event rates while a few exceptional patients have comparatively high event rates. The 18 basic settings listed in Section 5.2 are considered, with cluster sizes uniformly distribution from 80 to 100 and $\beta_0 = \log(5/100)$. This implies a baseline probability in the range from 0.03 to 0.12. Results for these settings are directly comparable to those in Tables 5.3a to 5.3c.

Results

Table 5.8a shows the estimated biases, average standard errors and the empirical standard errors. The estimated biases of all three approaches are small, less than 0.005 in absolute value in all cases. The average standard errors are in good agreement with the empirical standard errors in general. For empirical standard errors around 0.055 to 0.075, the discrepancies of the average standard errors are within a margin of 0.002. The ratio of the grand average of the standard errors to the grand average of the empirical standard errors reported in Table 5.8a is 101.4% for GLMM, 100.0% for GEE-EXCH and 100.2% for GEE-IND. For GLMM, the ratios of the standard errors are consistently larger than 100% ranging from about 100% to 104%. Exceptions include three cases with the ratios about 99.0% to 99.5%. In general, for all three approaches, there is essentially no difference between the robust standard errors or the model-based standard errors from their corresponding empirical standard errors.

Table 5.8b presents the estimated relative efficiencies. GEE-EXCH is basically as efficient as GLMM in this setting as well, with estimated relative efficiencies ranging from 99.9% to 100.4%; whereas GEE-IND tends to be slightly less efficient, with estimated relative efficiencies ranging from 98.3% to 100.1%.

Table 5.8c reports on the estimation of the variance component and number of simulated replicates replaced. The variance component is overestimated by 4.0% to 6.3%. The number of simulated replicates replaced is comparable to the previous setup with the beta distribution.

Even with this more extreme distribution for the random effect, regarding estimation of the regression coefficients, GLMM is still quite robust to this form of model misspecification. The variance component is however overes-

estimated on average leading to potentially slight inflation of the model-based standard error, although the effect is very limited in our cases. The bias and validity of the sandwich estimator of variance of the GEE approaches are not affected by the distribution of the random effect, as far as we observed. For the two setups considered in this section, GEE-EXCH has the same efficiency as GLMM while GEE-IND is slightly less efficient by 1% to 2%.

Summary

Our limited simulation studies in this section show that GLMM estimation of the regression coefficients is quite robust to moderate departures of the true random effect distribution from the normal distribution when the latter is conventionally assumed. Comparing the two settings of model misspecification in this section to the correct model specification setting in Section 5.2, similar biases, average standard errors and empirical standard errors are demonstrated. The number of replicates replaced are consistently smaller in the two model misspecification settings than in the correct model specification case. The number of replacements is however similar between these two settings.

For the average estimates of the variance component, those in the correct model specification setting are always the smallest, while the triangular random effect distribution setting always results in the largest variance component estimates among the three simulation settings. The efficiencies of GEE-EXCH and GEE-IND relative to GLMM are also similar in all three settings.

5.5 Conclusion

In the simulation studies for clustered binary responses and a binary covariate in this chapter, we have only considered circumstances with very low

marginal probability of events. Simulated response vectors usually consist of long strings of zeros and very rare events. In general, the pairwise correlation is bound to be low for small marginal probability, which is presumably the reason no loss of efficiency for the GEE approaches is observed in the simulations even for the independence “working” correlation structure.

The GLMM estimates of the regression coefficients are robust to departures of the random effect distribution from the conventional normal assumption even if the true underlying distribution is moderately skewed. Misspecification to the underlying stochastic process of the random effect of the type investigated also has no impact on the quality and efficiency for estimating the regression coefficients. However, it is worth noting that the GLMM log binomial model seems to be a difficult model to fit when the cluster size is small ($n = 15$) compared to fitting the corresponding marginal model by the GEE approach.

Figure 5.1: Probability density functions of random intercept b_i with mean -0.003 and variance $\sigma_b^2 = 0.316^2$ when 1) normally distributed and 2) base-line conditional probability follows beta a distribution.

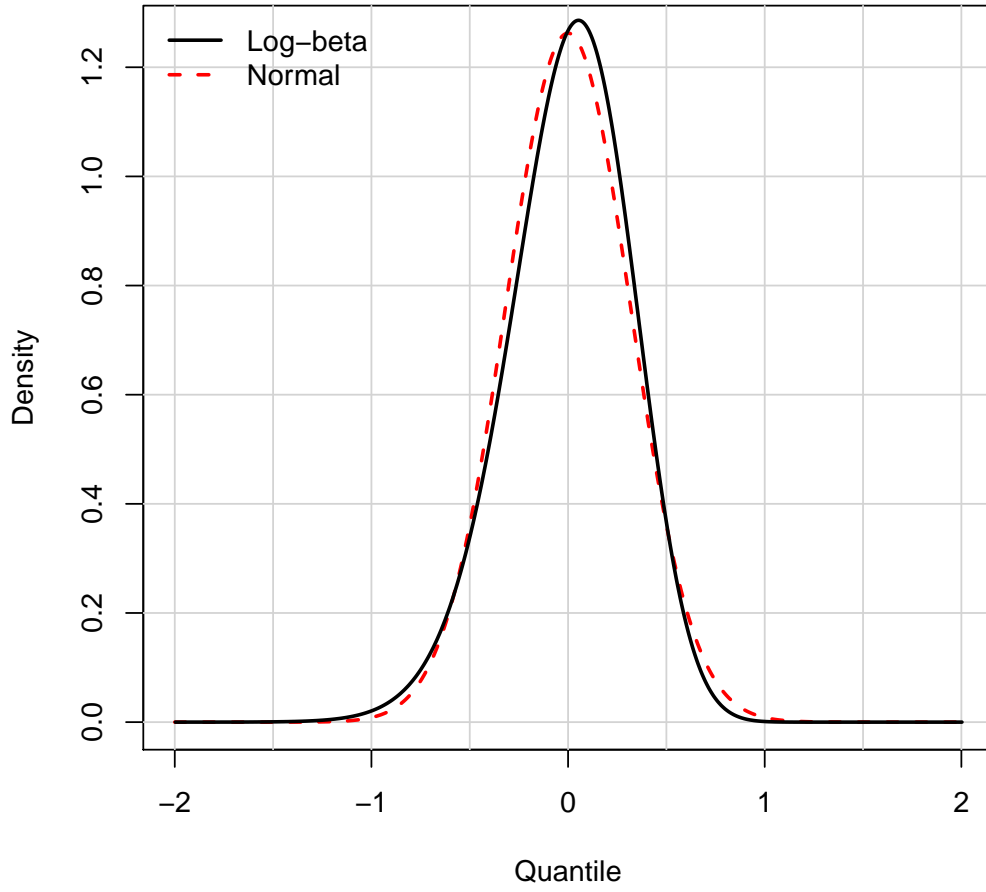


Table 5.7: Results of Simulation III-1: Beta-Binomial Setup

(a) Comparison of estimated log relative risk, β_1 when the baseline probability follows a beta distribution.Number of subjects, $M = 250$ $\beta_0 = \log(5/100)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GLMM			GEE-EXCH			GEE-IND		
			Bias	$SE^*_{(M)}$	$SE^\dagger_{(E)}$	Bias	$SE^\dagger_{(R)}$	$SE^\dagger_{(E)}$	Bias	$SE^\dagger_{(R)}$	$SE^\dagger_{(E)}$
30%	1.0	100	0.006	0.076	0.077	0.006	0.075	0.077	0.006	0.075	0.078
		90-100	0.006	0.076	0.075	0.006	0.075	0.075	0.006	0.076	0.076
		80-100	0.008	0.077	0.074	0.008	0.076	0.074	0.007	0.076	0.074
	1.3	100	-0.001	0.074	0.075	-0.002	0.073	0.074	-0.001	0.074	0.075
		90-100	0.002	0.075	0.075	0.002	0.074	0.074	0.002	0.074	0.075
		80-100	0.001	0.075	0.075	0.001	0.074	0.075	0.000	0.074	0.074
	1.5	100	0.001	0.074	0.073	0.000	0.073	0.073	0.001	0.073	0.073
		80-100	0.004	0.074	0.075	0.004	0.073	0.075	0.003	0.074	0.076
		90-100	0.003	0.074	0.072	0.002	0.073	0.072	0.002	0.074	0.072
20%-50%	1.0	100	0.002	0.057	0.057	0.002	0.056	0.057	0.002	0.057	0.057
		90-100	0.002	0.058	0.055	0.002	0.057	0.055	0.002	0.057	0.055
		80-100	-0.005	0.059	0.057	-0.005	0.058	0.057	-0.005	0.058	0.057
	1.3	100	0.000	0.055	0.055	0.000	0.054	0.055	0.000	0.054	0.055
		90-100	0.003	0.055	0.054	0.003	0.055	0.054	0.003	0.055	0.054
		80-100	0.002	0.056	0.054	0.001	0.055	0.054	0.001	0.056	0.055
	1.5	100	0.001	0.054	0.054	0.001	0.053	0.054	0.001	0.053	0.054
		90-100	0.001	0.054	0.054	0.001	0.053	0.053	0.001	0.054	0.054
		80-100	0.002	0.055	0.054	0.001	0.054	0.054	0.001	0.055	0.054

* Averaged model-based standard error.

 † Averaged robust standard error. ‡ Empirical standard error, i.e. sample standard deviation of simulation replicates.

5.5. Conclusion

Table 5.7: – Continued

(b) Relative efficiency[†] to maximum likelihood for estimating β_1 when the baseline probability follows a beta distribution.

Number of subjects, $M = 250$
 $\beta_0 = \log(5/100)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GEE-EXCH	GEE-IND
30%	1.0	100	1.003	0.982
		90–100	1.002	0.986
		80–100	1.001	0.999
	1.3	100	1.003	0.986
		90–100	1.001	0.986
		80–100	0.999	1.003
	1.5	100	1.004	0.990
		80–100	1.004	0.983
		90–100	1.000	0.995
20%–50%	1.0	100	1.002	0.990
		90–100	1.002	0.985
		80–100	1.003	0.995
	1.3	100	1.004	0.997
		90–100	1.004	1.008
		80–100	1.001	0.987
	1.5	100	1.002	1.003
		90–100	1.004	0.994
		80–100	1.002	1.000

[†]Relative efficiency defined as reciprocal of the ratio of empirical variances.

5.5. Conclusion

Table 5.7: – Continued

(c) Estimation of variance component in GLMM and computation issues of `glmer()` when the baseline probability follows a beta distribution.

Number of subjects, $M = 250$
 $\beta_0 = \log(5/100)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	Average $\hat{\sigma}_b$	Prop. $\hat{\sigma}_b < 0.01$	Replicates replaced
30%	1.0	100	0.302	0%	164
		90–100	0.302	0%	235
		80–100	0.301	0%	311
	1.3	100	0.301	0%	80
		90–100	0.301	0%	115
		80–100	0.302	0%	199
	1.5	100	0.298	0%	83
		90–100	0.299	0%	128
		80–100	0.300	0%	185
20%–50%	1.0	100	0.297	0%	48
		90–100	0.300	0%	83
		80–100	0.299	0%	102
	1.3	100	0.301	0%	48
		90–100	0.300	0%	75
		80–100	0.299	0%	97
	1.5	100	0.297	0%	53
		90–100	0.299	0%	76
		80–100	0.299	0%	103

Figure 5.2: Probability density functions of random intercept b_i with mean 0 and variance $\sigma_b^2 = 0.3^2$ when 1) normally distributed and 2) triangularly distributed.

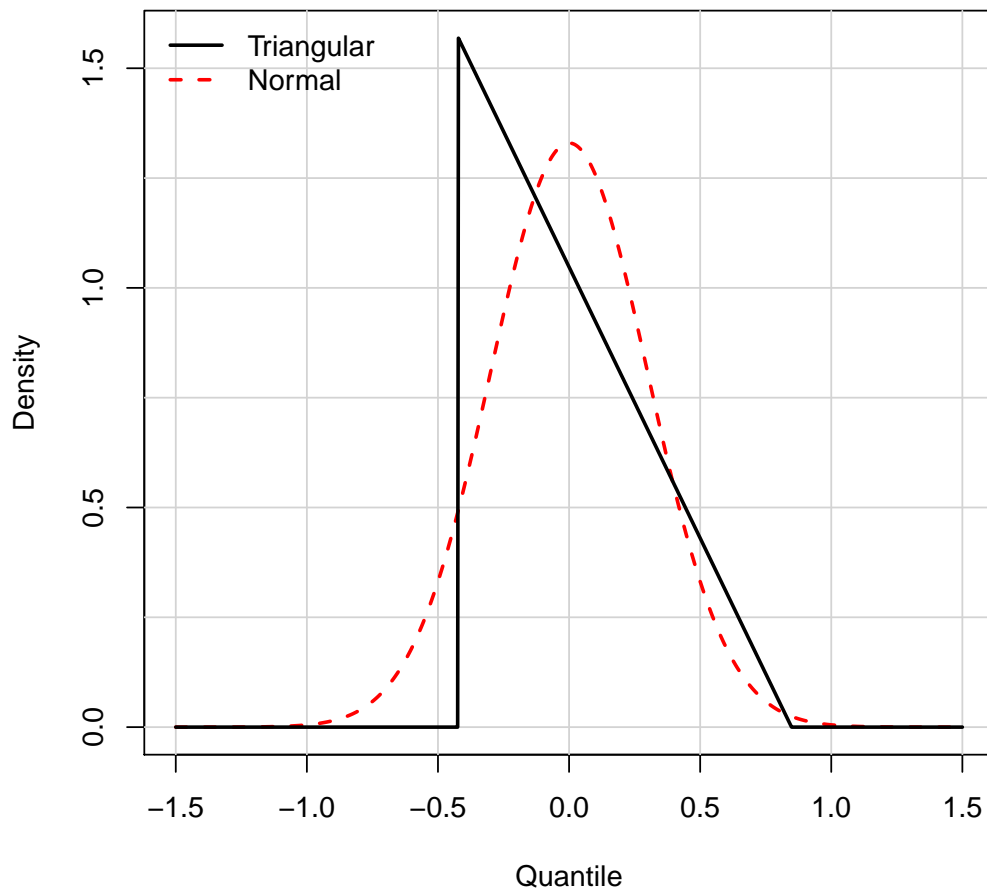


Table 5.8: Results of Simulation III-2: Triangular Distribution

(a) Comparison of estimated log relative risk, β_1 when the random effect follows a triangular distribution.Number of subjects, $M = 250$ $\beta_0 = \log(5/100)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GLMM			GEE-EXCH			GEE-IND		
			Bias	SE $^*_{(M)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$	Bias	SE $^\dagger_{(R)}$	SE $^\dagger_{(E)}$
30%	1.0	100	0.002	0.076	0.075	0.002	0.075	0.075	0.002	0.075	0.076
		90–100	0.001	0.076	0.076	0.001	0.075	0.075	0.001	0.076	0.076
		80–100	0.004	0.077	0.075	0.004	0.076	0.075	0.003	0.076	0.075
	1.3	100	0.000	0.074	0.074	0.000	0.073	0.074	0.000	0.074	0.074
		90–100	0.002	0.075	0.074	0.002	0.074	0.074	0.002	0.074	0.074
		80–100	0.004	0.075	0.073	0.003	0.074	0.073	0.003	0.074	0.073
	1.5	100	0.002	0.074	0.072	0.002	0.073	0.072	0.002	0.073	0.073
		80–100	0.004	0.074	0.074	0.004	0.073	0.074	0.004	0.073	0.074
		90–100	0.004	0.074	0.073	0.003	0.073	0.073	0.003	0.074	0.073
20%–50%	1.0	100	0.001	0.057	0.056	0.001	0.056	0.056	0.000	0.057	0.056
		90–100	0.002	0.058	0.058	-0.002	0.057	0.058	-0.002	0.057	0.059
		80–100	0.003	0.059	0.060	0.003	0.058	0.060	0.003	0.058	0.060
	1.3	100	0.001	0.055	0.054	0.000	0.054	0.054	0.000	0.054	0.055
		90–100	0.000	0.055	0.054	0.000	0.054	0.054	0.000	0.055	0.054
		80–100	0.002	0.056	0.055	0.002	0.055	0.055	0.002	0.056	0.055
	1.5	100	0.001	0.053	0.054	0.000	0.053	0.054	0.000	0.053	0.054
		90–100	0.003	0.054	0.053	0.002	0.053	0.053	0.001	0.054	0.053
		80–100	0.000	0.055	0.053	-0.001	0.054	0.053	-0.001	0.054	0.053

* Averaged model-based standard error.

 † Averaged robust standard error. ‡ Empirical standard error, i.e. sample standard deviation of simulation replicates.

5.5. Conclusion

Table 5.8: – Continued

(b) Relative efficiency[†] to maximum likelihood for estimating β_1 , when the random effect follows a triangular distribution.

Number of subjects, $M = 250$
 $\beta_0 = \log(5/100)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	GEE-EXCH	GEE-IND
30%	1.0	100	1.002	0.995
		90–100	1.004	1.000
		80–100	1.002	0.995
	1.3	100	1.000	1.001
		90–100	1.000	0.996
		80–100	1.002	0.996
	1.5	100	1.002	0.987
		80–100	1.003	0.984
		90–100	0.999	1.000
20%–50%	1.0	100	1.003	0.998
		90–100	1.002	0.996
		80–100	1.002	1.000
	1.3	100	1.002	0.984
		90–100	1.003	0.989
		80–100	1.002	1.001
	1.5	100	1.002	0.990
		90–100	1.001	0.983
		80–100	1.001	0.987

[†]Relative efficiency defined as reciprocal of the ratio of empirical variances.

5.5. Conclusion

Table 5.8: – Continued

(c) Estimation of variance component in GLMM and computation issues of `glmer()`, when the random effect follows a triangular distribution.

Number of subjects, $M = 250$
 $\beta_0 = \log(5/100)$, $\sigma_b = 0.3$

Prop. $x_{ij} = 0$	RR	n_i	Average $\hat{\sigma}_b$	Prop. $\hat{\sigma}_b < 0.01$	Replicates replaced
30%	1.0	100	0.316	0%	161
		90–100	0.319	0%	235
		80–100	0.318	0%	334
	1.3	100	0.313	0%	80
		90–100	0.314	0%	102
		80–100	0.318	0%	185
	1.5	100	0.313	0%	70
		90–100	0.315	0%	94
		80–100	0.315	0%	169
20%–50%	1.0	100	0.317	0%	66
		90–100	0.314	0%	83
		80–100	0.316	0%	108
	1.3	100	0.315	0%	38
		90–100	0.314	0%	70
		80–100	0.315	0%	104
	1.5	100	0.312	0%	60
		90–100	0.314	0%	88
		80–100	0.314	0%	127

Chapter 6

Conclusions and Discussion

A log-linear random intercept model for estimating relative risk for clustered or longitudinal data is described in this report. This model implies a common mean structure at the population average level and the subject specific level, which leads to two approaches for estimation of the same regression coefficient. In the case of a binary response, the corresponding regression coefficient represents the log relative risk. The first model fitting approach, maximum likelihood, is based on the integrated likelihood of the GLMM model; the second, the GEE approach, is an estimating equations approach based on the marginalized mean structure of this GLMM model.

This regression model was applied to the BEYOND trial data to assess the effects of neutralizing antibodies on the efficacy of interferon beta-1b in relapsing-remitting multiple sclerosis. Analyses based on daily relapse data showed that the model fits obtained by GLMM and GEE-EXCH are very similar. The estimated correlations by GEE-EXCH were very low, about 0.001. The BEYOND trial data suggested some NAB effect for medium and high level NAB titer values in the 500 mcg group, but this pattern was barely significant for GEE-EXCH and not significant for GLMM. For the 250 mcg group, no evidence of NAB effect was suggested based on NAB status. When using the NAB titers directly, no evidence of NAB effects was suggested, again, in the 250 mcg group. On the other hand, in the 500 mcg group, a positive association between the titer value and the risk of relapse onset was indicated for both approaches.

Another version of the data was considered by collapsing the binary indicators of daily relapse onset to the number of relapse onsets within intervals between two consecutive serum sample collections. Analyses based on the collapsed data resulted in generally the same conclusions on NAB effect. The estimated relative risks obtained by GEE-EXCH were slightly larger than those obtained by GLMM. The correlation estimated by GEE-EXCH was about 0.05 for the 250 mcg group and 0.11 for the 500 mcg group. Comparing the standard errors of the estimated log relative risk based on the daily relapse data and the collapsed data, there was little loss of efficiency for GLMM due to collapsing. For GEE-EXCH, mild to moderate loss of efficiency was observed. The variance of $\hat{\beta}$ corresponding to NAB status was inflated by up to 20% in the 500 mcg group.

To assess the efficiency of GEE-EXCH and GLMM, the variances of $\hat{\beta}$ were compared. Based on the daily relapse data, GEE-EXCH was slightly more efficient except in the 500 mcg group with direct use of NAB titer when GLMM was more efficient. Based on the collapsed data, in the 250 mcg group, the efficiency of the two approaches were similar. In the 500 mcg group, since a substantial loss of efficiency was observed for GEE-EXCH due to collapsing, it was less efficient compared to GLMM with both NAB status and NAB titers.

When the relapse onsets are very sparse among the binary responses, the within-subject correlation is usually very low. Our simulations showed, not surprisingly, that GEE-EXCH was as efficient as GLMM when the events were sparse and the within cluster correlation was smaller than 0.1. We also looked at the impact of model misspecification under the scenario of low event probability and low correlation. When the random effects follow a latent AR-1 process, our simulations showed that for estimation of the fixed effect, GLMM was robust to this form of model misspecification, even

though the variance component was poorly estimated. Similar robustness of GLMM to departures from the conventionally assumed normal form of the random effect distribution was also observed. In the situation when the random effects follow a triangular distribution with moderate skewness, the estimated regression coefficients were essentially unaffected and their model-based standard errors are accurate, although the variance component was slightly overestimated. In these cases of misspecification of the random effect distribution, GEE-EXCH has the same efficiency as GLMM while GEE-IND is almost as efficient with a loss of less than 2%.

Various studies of the efficiency of GEE estimators can be found in the literature. Neuhaus[21] compared a mixed-effects logistic model, a marginal logistic model and Rosner's model for clustered binary data. Although the regression coefficients in the three models do not have the same interpretation, they estimate the same null covariate effect when the true regression coefficients are zero. The mixed-effects model and Rosner's model can be fitted by maximum likelihood whereas the marginal logistic model can be fitted by GEE. He considered a setting where a single covariate has exchangeable correlation structure within cluster. Assuming the mixed-effects model holds with no actual covariate effect, approximate analytic results relating the regression coefficients as well as the standard errors obtained by the three approaches were derived asymptotically. Based on those results, the asymptotic relative efficiencies (AREs) of the derived Wald-type tests of the three approaches were compared. It was concluded that when the covariate only varies at the cluster level, all three approaches have similar power in detecting covariate effects. However, when the covariate varies within cluster, the mixed-effects model and GEE-EXCH are similar and more efficient than GEE-IND and Rosner's model. The gain in efficiency increases with the within cluster correlation of the response.

Fitzmaurice[22] compared the efficiency of GEE-EXCH and GEE-IND to maximum likelihood when the binary responses follow a general multivariate distribution with mean structure related to covariates through a logit link. The longitudinal data are measured at the same time points across clusters. A binary covariate, either cluster-level or within-cluster, was considered. The AREs of the estimated time effects were close to 1 so the focus was on the binary covariate. For cluster-level covariates, the efficiency of GEE-EXCH and GEE-IND are almost the same. Their AREs decrease with increasing correlation of the responses but are quite high (> 0.9) provided the correlation is modest (< 0.3). For within-cluster covariates, GEE-EXCH retains a similar pattern of ARE. However, GEE-IND results in a considerable loss of efficiency, as its ARE drops more sharply than that of GEE-EXCH as the correlation increases. GEE-IND was also compared to GEE-EXCH under the null covariate effect when both the response and the covariate has an exchangeable correlation structure within cluster. Conclusions similar to those of Neuhaus[21] were reached: loss of efficiency is associated with correlation of both the covariates and the responses and the two approaches are equally efficient when the covariate is constant within cluster. Fitzmaurice's results[22] explained the paradoxical findings in some previous studies on relative efficiency of GEE-IND to GEE-EXCH where either a cluster-level or a within-cluster covariate was considered, but not both.

Mancl and Leroux[23] investigated extensively the loss of efficiency due to the use of an independence “working” correlation structure when the true correlation structure is exchangeable. Asymptotic efficiency comparing GEE-IND to GEE-EXCH was based on the analytic form of the covariance matrices of the estimated regression coefficients for the case of linear models. The ARE was shown to depend on the covariate distribution, the cluster size, the within cluster correlations of the responses and the regression coefficients. The distribution of the covariates was described in terms of the proportion

of covariate variation attributable to between-cluster variation. Mancl and Leroux[23] identified that GEE-IND is fully efficient relative to GEE-EXCH in two situations: 1) when the covariate is cluster-level (i.e. covariate variation is entirely between-cluster); or 2) when cluster size is constant and the within-cluster covariate is mean-balanced across clusters (i.e. covariate variation is entirely within-cluster). The latter situation had been overlooked in previous studies but actually had been demonstrated by Fitzmaurice’s time covariate, which is a special case of a mean-balanced covariate as it is identical across clusters. The ARE decreases as the covariate distribution deviates from these two extremes. Larger cluster size was shown to aggravate the loss of efficiency. When the cluster size is large (e.g. 100), even when the response correlation is low (e.g. $\rho_y = 0.1$), the ARE can be as low as about 30%. They noted that the results may give good approximations for more general situations with other link functions and variance functions, especially if the resulting weights do not vary to a great extent. In an example for binary response and binary covariate with $\rho_y = 0.2$ and using a logit link, the ARE is about 0.9 for a constant cluster size of 4 and drops to about 0.5 for a constant cluster size of 20. The ARE is apparently unrelated to the baseline probability and increases only slightly with the magnitude of the covariate effect. Compared to our simulation study where a binary covariate varying within cluster was considered, we did not see any loss of efficiency due to variations in covariates, cluster size or covariate effects. This was most likely due to the low correlation in our simulated responses.

Chaganty and Joe[12] pointed out that the usual GEE “working” correlation matrix, assumed to be constant over possible values of covariates, cannot be the true correlation structure in general for non-normal, especially binary responses. For binary responses, the correlations are constrained by Fréchet bounds which are determined by the marginal means. The bounds could be quite narrow when the range of the covariate is wide. Also, strong

positive dependence is possible without high correlation. They suggested to regard the “working” correlation matrix merely as a weight matrix. A set of guidelines for selecting the optimal weight matrix was described. Since the optimal weight matrix is fixed, no association parameters are involved so the weight matrix can be ensured to be positive definite. By simulation, the efficiency of GEE is shown to be high relative to maximum likelihood when data is generated from a multivariate probit model. In this case, the optimal weight matrix turns out to be a rough approximate to the average true correlation matrix.

Chaganty and Joe[12] also argued that some studies of the relative efficiency of GEE in the literature are invalid. For example, they pointed out that Fitzmaurice’s[22] comparison of the GEE approach to maximum likelihood was invalid due to the questionable likelihood used. The correlations between the binary responses were assumed constant across possible values of the covariate, which was generally impossible. In contrast, our comparisons of the GEE approach to GLMM were based on a correct likelihood in which the correlations between the responses actually depended on the covariates².

To conclude, when the marginal probability of the binary response is low, the within cluster correlation is usually low. In this case, the GEE approach is almost as efficient as GLMM even when an independence “working” correlation matrix is used. Our simulations also indicated that GLMM was robust to misspecification of the random effect distribution as well as when the random intercept followed a latent AR-1 process within cluster. The estimation of regression coefficients was not sensitive to the normality assumption of the random intercept in general. Overall, the GEE approach demonstrated high efficiency. Fitting the marginal model was also computationally easier

²Although one may argue that a normal random effect distribution is theoretically impossible when a log link is applied to binary responses, we think this is not an issue provided the variance component is small enough.

and less demanding than fitting the GLMM. All in all, the GEE approach is recommended over GLMM for estimation of relative risk when the relative frequency of the event is very low for the correlated binary response.

In our analyses of the BEYOND data, the time trend of relapse risk was modelled as linear based on the conclusion of the likelihood ratio tests. Confirmatory analyses fitting a second degree spline or a quadratic time trend can be performed for both daily and collapsed data to see if more comparable NAB effects are estimated for both IFNB treatment groups. Our study only compared the efficiency of the GEE approach to GLMM when the response probability is low and the within cluster correlation is low. Further research could be done to study the relative efficiency for higher marginal probability and within cluster correlation. Also, the robustness of GLMM under more extreme departure from the normal random effect distribution could be examined. Since with the log link, the mean is not bounded above by 1, problems may arise when fitting the log binomial models to data with higher marginal probability. Some of these problems are documented in the literature in the context of cross-sectional studies. These issues of fitting the log-binomial model in the context of longitudinal studies and possible alternatives to circumvent the difficulties could be topics for future studies.

Bibliography

- [1] IFNB Multiple Sclerosis Group (1993). Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomized double-blind, placebo-controlled trial. *Neurology*, **43**, 655–661.
- [2] Paty, D. W., Li, D. K. B., UBC MS/MRI Study Group, and IFNB Multiple Sclerosis Study Group (1993). Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. II. MRI analysis results of a multicenter, randomized double-blind, placebo-controlled trial. *Neurology*, **43**, 662–667.
- [3] IFNB Multiple Sclerosis Study Group and UBC MS/MRI Analysis Group (1995). Interferon beta-1b in the treatment of multiple sclerosis: Final outcome of the randomized controlled trial. *Neurology*, **45**, 1277–1285.
- [4] IFNB Multiple Sclerosis Study Group and UBC MS/MRI Analysis Group (1996). Neutralizing antibodies during treatment of multiple sclerosis with interferon beta-1b: Experience during the first three years. *Neurology*, **47**, 889–894.
- [5] Petkau, A. J., White, R. A., Ebers, G. C., Reder, A. T., Sibley, W. A., Lublin, F. D., Paty, D. W., and the IFNB Multiple Sclerosis Study Group (2004). Longitudinal analyses of the effects of neutralizing antibodies on interferon beta-1b in relapsing-remitting multiple sclerosis. *Multiple Sclerosis*, **10**, 126–138.

- [6] O’Conner, P., Filippi, M., Arnason, B., Comi, G., Cook, S., Goodin, D., Hartung, H.-P., Jeffery, D., Kappos, L., Boateng, F., Filippov, V., Groth, M., Knappertz, V., Kraus, C., Sandbrink, R., Pohl, C., Bogumilt, T. (2009). 250 μg or 500 μg interferon beta-1b versus 20 mg glatiramer acetate in relapsing-remitting multiple sclerosis: a prospective, randomised, multicentre study. *Lancet Neurology*, **8**, 889–897.
- [7] Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, **135**, 370–384.
- [8] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall/CRC Press, 2nd edition.
- [9] Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439–447.
- [10] Bahadur, R. R. (1961). A representation of the joint distribution of responses to n dichotomous items. Solomon, H. (ed.), *Studies on item analysis and prediction*, pp. 158–68, Stanford Mathematical Studies in the Social Science VI, Stanford University Press, Stanford, California.
- [11] Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press, 2nd edition.
- [12] Chaganty, N. R. and Joe, H. (2004). Efficiency of generalized estimating equations for binary responses. *Journal of the Royal Statistical Society, Series B*, **66**, 851–860.
- [13] Liang, K.-Y. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.
- [14] Zeger, S. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121–130.

- [15] Paik, M. C. (1988). Repeated measurement analysis for nonnormal data in small samples. *Communications in Statistics - Simulation and Computation*, **17**, 1155–1171.
- [16] Halekoh, U., Højsgaard, S., and Yan, J. (2006). The R package **geepack** for generalized estimating equations. *Journal of Statistical Software*, **15**, 1–11.
- [17] Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**, 605–610.
- [18] Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545–554.
- [19] McCulloch, C. E. (1994). Maximum likelihood variance components estimation for binary data. *Journal of American Statistical Association*, **89**, 330–335.
- [20] Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2004). *Applied Longitudinal Analysis*. John Wiley and Sons.
- [21] Neuhaus, J. M. (1993). Estimation efficiency and tests of covariate effects with clustered binary data. *Biometrics*, **49**, 989–996.
- [22] Fitzmaurice, G. M. (1995). A caveat concerning independence estimating equations with multivariate binary data. *Biometrics*, **51**, 309–317.
- [23] Mancl, L. A. and Leroux, B. G. (1996). Efficiency of regression estimates for clustered data. *Biometrics*, **52**, 500–511.

Appendix A

Details of Model Fits for the Daily Relapse Data

Table A.1: Model fit based on eventually NAB+ patients only: linear time trend, NAB status.

(a) The 250 mcg group

All switches considered; confirmed NAB+, confirmed NAB−

	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
(Intercept)	-6.720	0.139	-48.46	<0.01	-6.943	0.137	-50.52	<0.01
Time	-0.185	0.104	-1.78	0.08	-0.192	0.110	-1.74	0.08
NAB+ vs. NAB−	0.088	0.158	0.56	0.58	0.097	0.171	0.57	0.57
α (GEE-EXCH)	0.001	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	0.973	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.729	0.069	10.60	<0.01

(b) The 500 mcg group

All switches considered; confirmed NAB+, confirmed NAB−

	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
(Intercept)	-7.087	0.151	-47.02	<0.01	-7.354	0.156	-47.22	<0.01
Time	-0.052	0.113	-0.46	0.64	-0.050	0.116	-0.43	0.66
NAB+ vs. NAB−	0.257	0.172	1.49	0.14	0.252	0.183	1.38	0.17
α (GEE-EXCH)	0.001	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	0.970	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.802	0.073	10.94	<0.01

Table A.2: Model fit based on eventually NAB+ patients only: linear time trend, refined NAB status.

(a) The 250 mcg group

All Switches considered; confirmed NAB+, unconfirmed L/M/H and confirmed NAB-

	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
(Intercept)	-6.722	0.140	-48.11	<0.01	-6.946	0.138	-50.28	<0.01
Time	-0.182	0.104	-1.75	0.08	-0.187	0.113	-1.66	0.10
Low NAB+ vs. NAB-	0.073	0.168	0.43	0.67	0.090	0.182	0.49	0.62
Med NAB+ vs. NAB-	0.147	0.221	0.66	0.51	0.140	0.221	0.63	0.53
High NAB+ vs. NAB-	0.034	0.261	0.13	0.90	0.023	0.270	0.09	0.93
α (GEE-EXCH)	0.001	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	0.973	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.729	0.069	10.59	<0.01

(b) The 500 mcg group

All Switches considered; confirmed NAB+, unconfirmed L/M/H and confirmed NAB-

	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
(Intercept)	-7.062	0.150	-47.02	<0.01	-7.334	0.157	-46.80	<0.01
Time	-0.100	0.121	-0.83	0.41	-0.096	0.121	-0.80	0.43
Low NAB+ vs. NAB-	0.105	0.201	0.52	0.60	0.095	0.207	0.46	0.65
Med NAB+ vs. NAB-	0.432	0.220	1.97	0.05	0.436	0.223	1.96	0.05
High NAB+ vs. NAB-	0.471	0.237	1.99	0.05	0.466	0.241	1.93	0.05
α (GEE-EXCH)	0.001	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	0.971	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.809	0.074	11.00	<0.01

*Titers for NAB status — Low : [20, 100), Med : [100, 400), High : [400, ∞).

Table A.3: Model fits based on 250 mcg arm, eventually NAB+ patients only: linear time trend, log NAB titers.

	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
cutoff = ∞								
(Intercept)	-6.703	0.135	-49.59	<0.01	-6.924	0.132	-52.32	<0.01
Time	-0.151	0.092	-1.63	0.10	-0.148	0.102	-1.45	0.15
log (NAB Titer/20)	-0.002	0.054	-0.03	0.97	-0.009	0.056	-0.16	0.87
α (GEE-EXCH)	0.001	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	0.972	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.728	0.069	10.57	<0.01
cutoff = 1000 NU/mL								
(Intercept)	-6.704	0.135	-49.54	<0.01	-6.925	0.133	-52.26	<0.01
Time	-0.157	0.093	-1.69	0.09	-0.155	0.103	-1.51	0.13
log (min {NAB Titer, 1000} /20)	0.006	0.062	0.09	0.93	-0.001	0.062	-0.02	0.98
α (GEE-EXCH)	0.001	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	0.972	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.729	0.069	10.59	<0.01
Continued on Next Page...								

Table A.3 – Continued

	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
cutoff = 400 NU/mL								
(Intercept)	-6.703	0.135	-49.58	<0.01	-6.924	0.133	-52.22	<0.01
Time	-0.154	0.094	-1.64	0.10	-0.152	0.103	-1.47	0.14
$\log(\min\{\text{NAB Titer}, 400\}/20)$	0.002	0.071	0.03	0.97	-0.005	0.072	-0.07	0.94
$\alpha(\text{GEE-EXCH})$	0.001	—	—	—	—	—	—	—
$\phi(\text{GEE-EXCH})$	0.972	—	—	—	—	—	—	—
$\sigma_b(\text{GLMM})$	—	—	—	—	0.728	0.069	10.58	<0.01
cutoff = 100 NU/mL								
(Intercept)	-6.707	0.136	-49.18	<0.01	-6.916	0.133	-51.92	<0.01
Time	-0.166	0.095	-1.75	0.08	-0.134	0.102	-1.31	0.19
$\log(\min\{\text{NAB Titer}, 100\}/20)$	0.028	0.109	0.25	0.80	-0.050	0.111	-0.45	0.65
$\alpha(\text{GEE-EXCH})$	0.001	—	—	—	—	—	—	—
$\phi(\text{GEE-EXCH})$	0.972	—	—	—	—	—	—	—
$\sigma_b(\text{GLMM})$	—	—	—	—	0.727	0.069	10.56	<0.01

Table A.4: Model fits based on 500 mcg arm, eventually NAB+ patients only: linear time trend, log NAB titers.

	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
cutoff = ∞								
(Intercept)	-7.032	0.146	-48.18	<0.01	-7.309	0.148	-49.30	<0.01
Time	-0.086	0.117	-0.74	0.46	-0.085	0.113	-0.75	0.45
log (NAB Titer/20)	0.108	0.051	2.10	0.04	0.110	0.049	2.23	0.03
α (GEE-EXCH)	0.001	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	0.970	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.814	0.074	11.04	<0.01
cutoff = 1000 NU/mL								
(Intercept)	-7.044	0.148	-47.76	<0.01	-7.319	0.149	-49.14	<0.01
Time	-0.102	0.119	-0.86	0.39	-0.100	0.114	-0.87	0.38
log (min {NAB Titer, 1000} /20)	0.137	0.061	2.24	0.03	0.138	0.057	2.42	0.02
α (GEE-EXCH)	0.001	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	0.971	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.812	0.074	11.02	<0.01
Continued on Next Page...								

Table A.4 – Continued

	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
cutoff = 400 NU/mL								
(Intercept)	-7.056	0.148	-47.55	<0.01	-7.331	0.150	-48.96	<0.01
Time	-0.101	0.119	-0.85	0.40	-0.100	0.115	-0.87	0.38
log (min {NAB Titer, 400} /20)	0.162	0.073	2.23	0.03	0.163	0.068	2.39	0.02
α (GEE-EXCH)	0.001	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	0.971	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.811	0.074	11.01	<0.01
cutoff = 100 NU/mL								
(Intercept)	-7.079	0.150	-47.23	<0.01	-7.354	0.152	-48.28	<0.01
Time	-0.084	0.116	-0.72	0.47	-0.082	0.112	-0.73	0.47
log (min {NAB Titer, 100} /20)	0.240	0.118	2.04	0.04	0.244	0.113	2.17	0.03
α (GEE-EXCH)	0.001	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	0.970	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.797	0.073	10.91	<0.01

Appendix B

Details of Model Fits for the Collapsed Data

Table B.1: Model fit based on collapsed data, eventually NAB+ patients only: linear time trend, NAB status.

(a) The 250 mcg group								
<i>All switches considered; confirmed NAB+, confirmed NAB−</i>								
	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
(Intercept)	-0.904	0.134	-6.76	<0.01	-1.177	0.131	-8.98	<0.01
Time	-0.245	0.115	-2.14	0.03	-0.235	0.114	-2.05	0.04
NAB+ vs. NAB−	0.142	0.167	0.85	0.40	0.138	0.174	0.79	0.43
α (GEE-EXCH)	0.050	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	2.242	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.818	0.071	11.51	<0.01

(b) The 500 mcg group								
<i>All switches considered; confirmed NAB+, confirmed NAB−</i>								
	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
(Intercept)	-1.304	0.154	-8.48	<0.01	-1.600	0.149	-10.72	<0.01
Time	-0.040	0.130	-0.31	0.76	-0.030	0.117	-0.25	0.80
NAB+ vs. NAB−	0.269	0.188	1.43	0.15	0.240	0.183	1.31	0.19
α (GEE-EXCH)	0.105	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	1.934	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.905	0.075	12.00	<0.01

Table B.2: Model fit based on collapsed data, eventually NAB+ patients only: linear time trend, refined NAB status.

(a) The 250 mcg group

All Switches considered; confirmed NAB+, unconfirmed L/M/H and confirmed NAB−

	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
(Intercept)	-0.905	0.134	-6.75	<0.01	-1.178	0.131	-8.97	<0.01
Time	-0.243	0.114	-2.13	0.03	-0.234	0.117	-2.00	0.05
Low NAB+ vs. NAB−	0.125	0.181	0.70	0.49	0.122	0.185	0.66	0.51
Med NAB+ vs. NAB−	0.201	0.232	0.87	0.39	0.195	0.224	0.87	0.38
High NAB+ vs. NAB−	0.094	0.274	0.34	0.73	0.097	0.275	0.35	0.72
α (GEE-EXCH)	0.050	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	2.239	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.819	0.071	11.51	<0.01

(b) The 500 mcg group

All Switches considered; confirmed NAB+, unconfirmed L/M/H and confirmed NAB−

	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
(Intercept)	-1.296	0.156	-8.31	<0.01	-1.593	0.150	-10.63	<0.01
Time	-0.095	0.141	-0.68	0.50	-0.082	0.123	-0.66	0.51
Low NAB+ vs. NAB−	0.103	0.218	0.48	0.63	0.084	0.207	0.41	0.68
Med NAB+ vs. NAB−	0.476	0.239	1.99	0.05	0.433	0.222	1.95	0.05
High NAB+ vs. NAB−	0.502	0.256	1.96	0.05	0.469	0.243	1.93	0.05
α (GEE-EXCH)	0.105	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	2.183	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.913	0.076	12.07	<0.01

*Titers for NAB status — Low : [20, 100), Med : [100, 400), High : [400, ∞).

Table B.3: Model fits based on collapsed data, 250 mcg arm, eventually NAB+ patients only: linear time trend, log NAB titers.

	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
cutoff = ∞								
(Intercept)	-0.868	0.126	-6.86	<0.01	-1.142	0.122	-9.37	<0.01
Time	-0.196	0.100	-1.95	0.05	-0.189	0.106	-1.79	0.07
log (NAB Titer/20)	0.005	0.057	0.09	0.93	0.008	0.057	0.14	0.89
α (GEE-EXCH)	0.054	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	2.054	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.819	0.071	11.49	<0.01
cutoff = 1000 NU/mL								
(Intercept)	-0.870	0.127	-6.85	<0.01	-1.145	0.122	-9.36	<0.01
Time	-0.200	0.102	-2.00	0.05	-0.196	0.106	-1.84	0.07
log (min {NAB Titer, 1000} /20)	0.012	0.065	0.22	0.82	0.018	0.063	0.28	0.78
α (GEE-EXCH)	0.054	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	2.068	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.819	0.071	11.51	<0.01
Continued on Next Page...								

Table B.3 – Continued

	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
cutoff = 400 NU/mL								
(Intercept)	-0.870	0.127	-6.85	<0.01	-1.145	0.123	-9.32	<0.01
Time	-0.200	0.103	-1.95	0.05	-0.193	0.107	-1.81	0.07
$\log(\min\{\text{NAB Titer}, 400\}/20)$	0.012	0.075	0.16	0.87	0.016	0.073	0.21	0.83
$\alpha(\text{GEE-EXCH})$	0.054	—	—	—	—	—	—	—
$\phi(\text{GEE-EXCH})$	2.068	—	—	—	—	—	—	—
$\sigma_b(\text{GLMM})$	—	—	—	—	0.819	0.071	11.50	<0.01
cutoff = 100 NU/mL								
(Intercept)	-0.877	0.129	-6.80	<0.01	-1.154	0.125	-9.25	<0.01
Time	-0.213	0.104	-2.04	0.04	-0.207	0.108	-1.91	0.06
$\log(\min\{\text{NAB Titer}, 100\}/20)$	0.044	0.115	0.38	0.70	0.051	0.114	0.45	0.66
$\alpha(\text{GEE-EXCH})$	0.053	—	—	—	—	—	—	—
$\phi(\text{GEE-EXCH})$	2.111	—	—	—	—	—	—	—
$\sigma_b(\text{GLMM})$	—	—	—	—	0.820	0.071	11.52	<0.01

Table B.4: Model fits based on collapsed data, 500 mcg arm, eventually NAB+ patients only: linear time trend, log NAB titers.

	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
cutoff = ∞								
(Intercept)	-1.257	0.145	-8.66	<0.01	-1.571	0.138	-11.42	<0.01
Time	-0.085	0.135	-0.63	0.53	-0.076	0.116	-0.65	0.51
log (NAB Titer/20)	0.118	0.054	2.17	0.03	0.114	0.050	2.30	0.02
α (GEE-EXCH)	0.106	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	2.078	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.920	0.076	12.11	<0.01
cutoff = 1000 NU/mL								
(Intercept)	-1.276	0.148	-8.64	<0.01	-1.584	0.139	-11.43	<0.01
Time	-0.102	0.138	-0.74	0.46	-0.089	0.117	-0.76	0.45
log (min {NAB Titer, 1000} /20)	0.151	0.066	2.29	0.02	0.142	0.057	2.47	0.01
α (GEE-EXCH)	0.106	—	—	—	—	—	—	—
ϕ (GEE-EXCH)	2.100	—	—	—	—	—	—	—
σ_b (GLMM)	—	—	—	—	0.918	0.076	12.09	<0.01
Continued on Next Page...								

Table B.4 – Continued

	GEE-EXCH				GLMM			
	estimate	SE	Z-score	p-value	estimate	SE	Z-score	p-value
cutoff = 400 NU/mL								
(Intercept)	-1.291	0.149	-8.64	<0.01	-1.596	0.140	-11.39	<0.01
Time	-0.103	0.139	-0.74	0.46	-0.087	0.117	-0.75	0.45
$\log(\min\{\text{NAB Titer}, 400\}/20)$	0.180	0.079	2.27	0.02	0.166	0.069	2.42	0.02
$\alpha(\text{GEE-EXCH})$	0.107	—	—	—	—	—	—	—
$\phi(\text{GEE-EXCH})$	2.086	—	—	—	—	—	—	—
$\sigma_b(\text{GLMM})$	—	—	—	—	0.916	0.076	12.08	<0.01
cutoff = 100 NU/mL								
(Intercept)	-1.312	0.153	-8.57	<0.01	-1.609	0.143	-11.22	<0.01
Time	-0.083	0.135	-0.62	0.54	-0.066	0.115	-0.58	0.56
$\log(\min\{\text{NAB Titer}, 100\}/20)$	0.268	0.131	2.05	0.04	0.240	0.114	2.11	0.03
$\alpha(\text{GEE-EXCH})$	0.107	—	—	—	—	—	—	—
$\phi(\text{GEE-EXCH})$	1.978	—	—	—	—	—	—	—
$\sigma_b(\text{GLMM})$	—	—	—	—	0.909	0.076	12.03	<0.01