University of Alberta

COMPARISON OF DISCRETE DATA METHODS FOR REPEATED MEASURES DATA WITH SMALL SAMPLES

by

Xuechen Zhang

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

in

Biostatistics

Department of Mathematical and Statistical Sciences

© Xuechen Zhang

Fall 2013

Edmonton, Alberta

Abstract

Various analytical methods are available to analyze repeated measures data for both continuous and discrete data. In the case of discrete data, most methods are based on the assumption of asymptotic normality, requiring large samples. Naturally, their small sample performance may not match the expectation satisfactorily. Two main methods, the non-linear mixed effects (NLME) model and the generalized estimating equations (GEE) method, are investigated for their small sample performance on repeated binary data. We generated binary data, considering two levels of correlation at ρ=0.3 and 0.7, with three cases of repeated measures with T=2, 4, or 6 and sample sizes ranging from 40 to 200.The two analysis methods are applied to each data set in 5000 simulations, and the resulting empirical size and power are compared. We conclude that the GEE performs quite well in small samples with satisfactory empirical size and statistical power and is therefore recommended.

Acknowledgements

I would like to express my gratitude to my supervisor, Dr. K. C. Carriere, for all her guidance during the process of research. I am also grateful to my committee members for their advice and all the professors in the department of Mathematical and Statistical Sciences as well as the School of Public Health for their invaluable courses.

Finally, I would like to thank my parents and friends for their encouragement and support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Variables can be classified into two categories, i.e. continuous variable and discrete variable, based on their values. Continuous data, such as temperature, height of a child and blood pressure, can take any value possible within the range. Discrete data, such as age group and gender, however, can take only on a countable number of values or categories (Hogg et al., 2005). The main difference between these two kinds of variables leads to different models that can be applied to deal with each of them. In simple linear regression

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \qquad (1.1)$$

where $i = 1, 2, \ldots, n$, $\{y_i\}$ are continuous data and for given $(x_1, \ldots, x_p)$ follow a normal distribution because the error term ε_i follows a normal distribution. Therefore this model can only be applied to continuous response variables.

Models such as logistic and poisson regressions have been well established for discrete data now.

Repeated measures data is collected from subjects repeatedly under same or different conditions (Lindstrom and Bates, 1990). Such data occurs commonly in health-related research, especially in longitudinal study, which is usually used to study trends over time. For example, the Framingham Heart Study starting from 1948 is one of the famous longitudinal studies. A total of 5,209 people living in the town of Framingham, Massachusetts were included in the study in 1948 and their sons and daughters were also included in the following study. The results of this study revealed those well-known risk factors concerning heart disease (O'Donnell and Elosua, 2008).

Because of the special way the data are collected, observations from one subject may be more alike than observations from other subjects, i.e. are correlated. It means that the independence assumption in some models may be violated due to the significant correlation among the data. It would be inappropriate to ignore such facts. To analyze such data, some particular approaches have been developed. Among those approaches, this thesis will focus on two common ones, which are the Generalized Estimating Equations (GEE) method and the Non-linear Mixed Effects (NLME) model.

The two methods have different targets of inference and should be selected based on the objective of a study. The GEE approach is referred to as a

marginal model, which focuses on measuring the effect of covariates on the whole population. Therefore, when the question to be addressed in a study is to assess the factors' influence on the population level, the GEE method would be an appropriate choice. However, if the focus of a study is on an individual's response or a cluster's response, the mixed effects models should be used because the random terms in the model will capture the variability among individuals or clusters (Davis, 2002).

Besides model selection, which requires much attention, sample size is another important aspect in applied research. Statistical methods perform quite well in large samples because of the large amount of information they can obtain. But large samples may not be always possible in many situations because of a lack of subjects or limited funding. In these cases, small samples that will simplify the data collection process and decrease the cost of a study will be preferred. The weak point of choosing small samples is that it will "rob a study of power to detect significant effects when they truly exist"(Rochon, 1998). Hence when designing experiments, one should choose the most appropriate sample size to reduce the cost and also guarantee the desired power. When the sample size is limited in a study, one should select the method that gives the most power and make the most use of information.

The related issue with repeated measures data has not been fully resolved. The object of this thesis is to investigate this issue by comparing performance

of the two methods mentioned above, i.e. the GEE and the NLME model, in small samples and arriving at appropriate recommendations on the choice of methods in small samples.

Chapter 2

Literature Review

2.1 Overview

Many methods can be used to analyze repeated measurement data. In this chapter, we mainly focus on the Generalized Estimating Equations method and the Non-linear Mixed Effects model.

2.2 Generalized Linear Model

2.2.1 Three Components of GLM

Generalized linear models (GLM), formulated by John Nelder and Robert Wedderburn in 1972, are extensions of ordinary regression models. They have proven to be a valuable methodology and has been widely used by statisticians

(McCullagh and Nelder 1989). A generalized linear model consists of three key components: a random component, a systematic component and a link function (Agresti, 2002).

Considering the model as proposed by Agresti(2002), a random component refers to the response variables $\{y_i\}$ and their distributions. In GLM, $Y = (Y_1, Y_2, \ldots, Y_n)'$. $Y_i's$ are independent variables and belong to a natural exponential family such as Poisson distribution and binomial distribution.

A systematic component refers to a linear predictor

$$\eta_i = \sum_j \beta_j x_{ij} = X_i'\beta, i = 1, 2, \ldots, n \tag{2.1}$$

where $x_{ij}, j = 1, 2, \ldots, n_i,$ denotes the value of the jth explanatory variable on the ith subject. So the linear predictor is in fact a linear combination of explanatory variables.

A link function is a differentiable function that connects the random and systematic components. Suppose the mean of the response variable Y_i is $\mathbb{E}(Y_i) = \mu_i$. Then $\mu_i's$ are linked to the linear predictor $\eta_i = X_i'\beta$ by a known one-to-one function $h(\cdot)$.

$$\mu_i = h(\eta_i) = h(X_i'\beta) \tag{2.2}$$

$$\eta_i = X_i'\beta = g(\mu_i) \qquad\qquad (2.3)$$

where $g(\cdot) = h^{-1}(\cdot)$ is called the link function. In other words, a link function transforms the mean of a response variable.

To specify a link function, one can use an identical link or a canonical link. A link function $g(\mu_i) = \mu_i$ is called an identical link. $g(\mu_i) = Q(\mu_i) = X_i'\beta$ is called a canonical link.

2.2.2 GLM for Binary Data

Binary responses, for instance the diagnosis of a certain disease (present or absent) or the status of smoking (smoker or nonsmoker), can be expressed with binary variables, denoted by 0 and 1. Then $\mathbb{E}(Y_i) = P(Y_i = 1)$

2.3 Non-linear Mixed Effects Model

Mixed effects models, which contain both fixed effects and random effects, have been widely used in applied sciences such as epidemiology and biology. This method is particularly useful in repeated measures data in longitudinal studies. Widely accepted, the linear mixed effects model is a powerful tool to analyze repeated measures data. But in some cases, nonlinear patterns inherently exist in the relationship between independent variables and dependent

variables and, therefore, the use of a linear model is no longer appropriate (Lee and Xu, 2004). Then the nonlinear mixed effects model proposed in recent years would be a better way to deal with such situations. Much work has been done on the form of the model, the estimation of the parameters and also the application in different areas. Lindstrom and Bates (1990) introduced the nonlinear mixed effects model as well as the parameter estimators. Vonesh and Carter (1992) considered a "generalized mixed effects nonlinear regression model" that can be used for "incomplete or unbalanced data". Davidian and Gallant (1993) specified a "general nonlinear mixed effects model, which makes no parametric assumption about the form of the random effects distribution". And Walker (1996) also introduced an EM algorithm for fitting the nonlinear random effects model.

Considering the nonlinear mixed effects model as proposed by Lindstrom and Bates (1990), the jth observation on the ith individual is modeled as

$$Y_{ij} = f(\boldsymbol{\phi_i}, x_{ij}) + e_{ij} \tag{2.4}$$

where Y_{ij} is the response j from individual i, x_{ij} is the explanatory variable vector for response j from individual i, $f(\cdot)$ is a nonlinear function and $e_{ij} \sim \mathcal{N}(0, \sigma^2)$ is an random error term with $\text{Cov}(e_{ij}, e_{ik}) = 0, 1 \leq j \neq k \leq n_i$. Responses from different individuals are independent from each other. In this

model, ϕ_i is a $r \times 1$ parameter vector with the following form:

$$\boldsymbol{\phi_i} = \boldsymbol{A_i}\boldsymbol{\beta} + \boldsymbol{B_i}\boldsymbol{b_i}, \quad \boldsymbol{b_i} \sim \mathcal{N}(0, \sigma_b^2 \boldsymbol{D})$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta p)^T$ is a $p \times 1$ vector of fixed parameters that is common to all individuals, $\boldsymbol{b_i} = (b_{i1}, b_{i2}, \ldots, b_{iq})^T$ is a $q \times 1$ vector of random effects that varies among individuals, $\boldsymbol{A_i}$, $\boldsymbol{B_i}$ are $r \times p$ and $r \times q$ design matrices and $\sigma_b^2 \boldsymbol{D}$ is a covariance matrix of random effects.

With highly developed computer science, parameter estimation can be efficiently done by softwares. For instance, the package PROC NLMIXED in SAS is an efficient tool to complete this task. It is capable of analyzing data from different distributions such as normal, binomial or Poisson and performs the maximum likelihood estimation (SAS/STAT® 9.22 User's Guide, 2010).

For non-normal responses Y_{ij}, the non-linear mixed effects model becomes $\mu_{ij} = E(y_{ij}|b_i) = f(\boldsymbol{\phi_i}, x_{ij})$. A special case of the non-linear mixed effects model, the generalized linear mixed effects model has the form

$$\mu_{ij} = E(y_{ij}|b_i) = h(\eta_{ij}) = h(\boldsymbol{X'_{ij}}\boldsymbol{\beta} + \boldsymbol{\omega_{ij}}\boldsymbol{b_i}), \quad \boldsymbol{b_i} \sim \mathcal{N}(0, \sigma_b^2 \boldsymbol{D}) \qquad (2.5)$$

Consider the special case with a univariate random effect and $\omega_{ij} = 1$.

$$\mu_{ij} = E(y_{ij}|b_i) = h(\eta_{ij}) = h(\boldsymbol{X'_{ij}}\boldsymbol{\beta} + b_i), \quad b_i \sim \mathcal{N}(0, \sigma_b^2) \qquad (2.6)$$

It has the form of an ordinary GLM with a random intercept.

2.4 Generalized Estimating Equations Method

2.4.1 Quasi-likelihood

Quasi-likelihood estimation is an alternative approach to maximum likelihood estimation because of the similar properties they have (Agresti, 2002). It was proposed by Wedderburn in 1974. Unlike likelihood estimation, quasi-likelihood estimation only assumes the mean-variance relationship

$$\text{Var}(Y_i) = v(\mu_i)$$

instead of assuming the specific distribution of Y_i, i.e. we only specify a variance function $v(\cdot)$. Quasi-likelihood estimates are solutions of an equation called the quasi-score function, which has the same form as the score function in likelihood estimation except that the variance part is replaced by $v(\mu_i)$ here (Agresti, 2002).

2.4.2   Generalized Estimating Equations

With repeated responses, the correlation among observations for a given subject must be taken into consideration. Liang and Zeger (1986 and 1988) proposed an extension of GLM to analyze longitudinal data, i.e. the Generalized Estimating Estimations approach, based on quasi-likelihood.

Let $Y_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})^T, i = 1, 2, \ldots, N$ be the $n_i \times 1$ vector of outcome values representing the measurement j on the subject i. The corresponding mean vector is $\mu_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{in_i})^T$ and the covariance matrix is V_i. Let $X_i = (x_{i1}, x_{i2}, \ldots, x_{in_i})^T$ be the $n_i \times p$ matrix of covariates for the subject i. $g(\mu_i) = X_i^T \beta$, where $g(\cdot)$ is the link function as defined in GLM.

The generalized estimating equations method for estimating parameters can be applied to correlated data and has the following form:

$$S(\beta) = \sum_{i=1}^{N} D_i^T V_i^{-1}(Y_i - \mu_i(\beta)) = 0$$

where

$$D_i^T = \frac{\partial \mu_i^T}{\partial \beta} = \begin{pmatrix} \frac{x_{i11}}{g(\mu_{i1})^T} & \cdots & \frac{x_{in_i 1}}{g(\mu_{in_i})^T} \\ \vdots & \ddots & \vdots \\ \frac{x_{i1p}}{g(\mu_{i1})^T} & \cdots & \frac{x_{in_i p}}{g(\mu_{in_i})^T} \end{pmatrix}$$

$$V_i = \phi A_i^{\frac{1}{2}} W_i^{-\frac{1}{2}} R(\alpha) W_i^{-\frac{1}{2}} A_i^{\frac{1}{2}}$$

Let $R(\alpha)$ be a $n \times n$ symmetric matrix with the feature of a correlation matrix. We refer to $R(\alpha)$ as a working correlation matrix. A_i is a $n_i \times n_i$ diagonal matrix and W_i is a $n_i \times n_i$ diagonal weight matrix. V_i will be the true covariance matrix of Y_i if $R(\alpha)$ is correctly specified (SAS/STAT® 9.22 User's Guide, 2010).

Usually unknown, the working correlation matrix must be estimated in the iterative fitting process. Structures of the working correlation include independent, exchangeable, unstructured and autoregressive, and can be specified in the SAS procedure.

Table 2.1 is a summary of all the common choices of working correlation matrix. The independent working correlation matrix is, in fact, the identity matrix:

$$Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 \text{ if } j = k, \\ 0 \text{ if } j \neq k, \end{cases}$$

However, this may not be realistic in repeated measures data. The exchangeable working correlation matrix assumes that $Corr(Y_{ij}, Y_{ik})$ remains the same for all pairs of (j, k). It may be even more realistic to allow the correlation between each pair of (Y_{ij}, Y_{ik}) to be different. But the drawback is that it may need too many parameters especially when there are a large number of repeated measurements. Researchers can choose an appropriate working correlation matrix based on their experience. When the correlations are modest,

using different forms of a working correlation matrix would produce similar

results (Agresti, 2002).

Table 2.1: Working Correlation Matrix

Structure	Working Correlation Matrix
Fixed	$Corr(Y_{ij}, Y_{ik}) = r_{jk}$, where r_{jk} is the jk^{th} element of the correlation matrix specified by the user
Independent	$Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 \text{ if } j = k, \\ 0 \text{ if } j \neq k, \end{cases}$
m-Independent	$Corr(Y_{i,j}, Y_{i,j+t}) = \begin{cases} 1 \text{ if } t = 0, \\ \alpha_t \text{ if } t = 1, 2, \ldots, m \\ 0 \text{ if } t > m, \end{cases}$
Exchangeable	$Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 \text{ if } j = k, \\ \alpha \text{ if } j \neq k, \end{cases}$
Unstructured	$Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1 \text{ if } j = k, \\ \alpha_{jk} \text{ if } j \neq k, \end{cases}$
Autoregressive AR(1)	$Corr(Y_{i,j}, Y_{i,j+t}) = \alpha^t$, for $t = 0, 1, \ldots, n_i - j$

After specifying all the key elements in the model, the fitting algorithm in

PROC GENMOD is as follows[1].

1. *Compute an initial estimate of β with an ordinary generalized linear model assuming independence.*

2. *Compute the working correlations R based on the standardized residuals, the current β and the assumed structure R.*

3. *Compute an estimate of the covariance: $V_i = \phi A_i^{\frac{1}{2}} W_i^{-\frac{1}{2}} R(\hat{\alpha}) W_i^{-\frac{1}{2}} A_i^{\frac{1}{2}}$*

4. *Update β: $\beta_{r+1} = \beta_r + [\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} \frac{\partial \mu_i}{\partial \beta}]^{-1} \sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} (Y_i - \mu_i)$*

---

[1]The fitting algorithm is from SAS/STAT® 9.22 User's Guide, 2010

13

5. *Repeat step 2-4 until convergence (SAS/STAT$^{\circledR}$ 9.22 User's Guide, 2010).*

2.4.3 Sample Size Calculation in the GEE Approach

Designing an experiment with an adequate sample size is important in the study planning stage in applied research. Rochon(1998) introduced a method of calculating sample size for the GEE method in repeated measures experiments. A WLS estimation was conducted and the covariance matrix of estimators was provided. Once the significance level α and the power were specified, the minimum sample size was calculated based on a non-central version of the Wald χ^2 test statistic. This method can provide guidance in obtaining the minimum sample size. However, a drawback of this method is that it may be not valid if the sample size derived from this method is too small because the method is based on an asymptotic theory that requires a large sample in the first place.

2.5 The Monte Carlo Method

In this section, the generation of data from a specific distribution will be discussed. This is called a Monte Carlo method. The Monte Carlo technique was introduced in the 1940s by John von Neumann and Stanislaw Ulam who were working on a secret nuclear fission weapon project (Eckhardt, 1987). It

allows researchers to solve complicated problems that they cannot solve analytically. Since developed, Monte Carlo techniques have become increasingly popular in such disciplines as business, finance and science(Thomopoulos, 2013).

A random number generator is a device that can generate a stream of random numbers, i.e. numbers that have no patterns. Some well-known physical methods, such as rolling dice and flipping coins, have been used for years but it is not practical to generate a large steam of random numbers. This problem has been solved by the highly developed computer technology. Mathematical algorithms have been efficient tools that allow users to generate random numbers based on their needs. Except for the early popular computer languages such as C++ and JAVA, a number of computer languages have emerged recently such as SAS, SPSS, R and Splus (Thomopoulos, 2013). To construct the random number generators and to test their accuracy, great work has been done in this area (Hogg et al., 2005).

In a Monte Carlo process, a generator of random uniform observations is the key step and can be easily realized by the algorithms developed by mathematicians. For instance, the command $runif(n)$ in R software can be used to generate n random numbers from the uniform distribution $U(0,1)$ (Hogg et al., 2005).

In most cases, generation of variables can be conducted based on random

uniform variables no matter whether continuous or discrete. To generate numbers from a discrete distribution $F(x_i), i = 1, \ldots, n$, one can first generate a random uniform variable u and then obtain the value of x by comparing u with the cumulative distribution $F(x_i)$. For example, to generate a random number from a Bernoulli distribution with $p(X = 1) = 0.80$, we can

1.Generate a random variable $u \sim U(0, 1)$.

2.If $u < 0.80$, then x=1; Otherwise, x=0 (Thomopoulos, 2013).

For continuous data, the Inverse Transform Method can be used to generate realizations (Zio, 2013). "Suppose the random variable u has a uniform $U(0, 1)$ distribution. Let F be a continuous distribution function. Then the random variable $X = F^{-1}(u)$ has distribution function F" (Hogg et al., 2005).

The Monte Carlo method is a powerful way to emulate the real system. By generating a series of random numbers as input variables and transforming these random numbers through a certain model, the resulting observations or samples will simulate the real data with respect to its characteristics such as the distribution. The better the simulation model simulates the real system, the more reliable the sample is (Thomopoulos, 2013).

2.6   Normality Test methods

Normal distribution assumption is of extreme importance in statistics. Many statistical procedures such as t-tests and regression are based on an

underlying assumption of normality. When the normality assumption is violated, the analysis results may not be reliable and cannot be used for further interpretation. Graphical methods, such as Q-Q plot, are the most straightforward and effective tools for checking the normality of data. However, they are not able to provide objective evidence that normality assumption holds. To support graphic methods, formal normality tests should be performed to make a more reasonable decision.

2.6.1   Graphical Methods

Q-Q plot   The normal quantile-quantile plot (Q-Q plot) is the most commonly used graphical method. As we know, a Q-Q plot compares the quantiles of two distributions. When it is used in normality tests, the Q-Q plot compares the empirical distribution of a data set to the normal distribution. If the data follows the normal distribution, the points on the plot would fall approximately on a straight line.

Histogram   A histogram is also a useful tool to show the distribution of data. It displays the frequency of data with vertical bars. The area represented by each bar is proportional to the frequency of data in the corresponding interval. If the data follows the normal distribution, the histogram of the data would be symmetric and bell-shaped.

2.6.2 Statistical Test for Normality

Normality tests are classified by Dufour et al.(1998) as empirical distribution function tests, correlation tests and moment tests. In this section, we mainly focus on empirical distribution function tests and correlation tests. Empirical distribution function (EDF) tests test normality of the data by comparing the empirical distributions with the hypothetical distributions, i.e. the normal distribution. The most well-known tests such as the Kolmogorov-Smirnov test, the Lilliefors test and the Anderson-Darling test are all EDF tests. In correlation tests, the ratio of the WLS estimate under normality assuption and the sample variance is calculated. The most famous correlation test is the Shapiro-Wilk test (Dufour et al., 1998). Razali and Wah (2011) provided a comprehensive summary of these normality tests and compared the power of them.

Kolmogorov-Smirnov Test and Lilliefors test　The Kolmogorov-Smirnov test is based on the "largest vertical distance" between the empirical distribution and the normal distribution (Razali and Wah, 2011). It is defined by Conover (1999) as $T = sup_x |F^*(x) - F_n(x)|$.

The Lilliefors test is an extension of the Kolmogorov-Smirnov Test and it is the most popular EDF normality test. The test statistic is computed as $D = \max_x |F^*(X) - S_n X|$ (Lilliefors, 1967).

In R, command "$lillie.test(\cdot)$" will perform the Lilliefors test.

Anderson-Darling Test "The Anderson-Darling test is a modification of the Cramer-von Mises test"(Razali and Wah, 2011). We will not present the details of Cramer-von Mises test here. For sample size n, Anderson and Darling (1954) gave the definition of this test statistic as $W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 \psi(F^*(x)) \, dF^*(x)$, where $F^*(x)$ is the cumulative probability function of a normal distribution and $\psi(F^*(x))$ is the nonnegative weight function. Usually, $\psi(F^*(x)) = [F^*(x)(1 - F^*(x))]^{-1}$.

In R, $ad.test(\cdot)$ can be used to perform the Anderson-Darling test.

Shapiro-Wilk Test The Shapiro-Wilk test, formulated by Shapiro and Wilk (1965), is one of the most famous normality tests. Let $y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n)}$ be an ordered random sample. The Shapiro-Wilk test statistic has the following form:

$$W = \frac{(\sum_{i=1}^{k} a_{n-i+1}(y_{(n-i+1)} - y_{(i)}))^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Small values of W indicate that the normality assumption is violated (Shapiro and Wilk, 1965). Originally, the Shapiro-Wilk test was only valid when the sample size is less than 50. After new algorithms was developed, the restriction of sample size has been greatly relaxed to 5000 (Razali and Wah, 2011). In R, command "$shapiro.test(\cdot)$" performs the Shapiro-Wilk test.

19

Pearson Chi-square Test   The Pearson Chi-square test is usually used as goodness-of-fit test, which evaluates whether the observed frequency distribution differs from a hypothetical distribution. As a special case, normality then can be assessed using the Pearson Chi-square test. The Pearson Chi-square test statistic has the following form: $P = \sum \frac{(C_i - E_i)^2}{E_i}$, where C_i is the observed value and E_i is the expected value. The test statistic follows the Chi-square distribution. In R, $"pearson.test(\cdot)"$ is the command to perform the Pearson Chi-square test.

2.6.3   Comments

Even though a graphical display is a straight-forward way to illustrate empirical distributions, it is highly recommended that graphic methods be supported by formal normality tests in order to provide conclusive evidence of whether or not the normality assumption is valid. Also, formal normality test may not be reliable either under a small sample size or when the sample size is too large. In other words, the formal normality tests may perform poorly in small samples because of limited information extracted from the data and may not perform as well as we expect when the sample size is too large because the slight deviation from the normal distribution may mislead researchers into rejecting the normality assumption. Careful interpretation of the results of normality tests is important.

Chapter 3

Simulation Analysis

3.1   Data Generation

In our study, the Monte Carlo method was used to evaluate the performance of the non-linear mixed effects model and the GEE method in small samples. Simulated data sets of clinical trial data comparing two treatments for a respiratory disorder are created with sample size $n =$40, 60, 80, 100 and 200. Half of the subjects in each data set are assigned to the active treatment group(treatment=1) and the other half are assigned to the placebo group(treatment=0). The selection process is randomized. The binary response y_{ij} represents the respiratory status (coded as 0=poor, 1=good) of subject i at the jth visit, $i = 1, 2, \ldots, n; j = 1, 2, \ldots, T$. $T =$2, 4 and 6 repeat measurements are generated for each subject based on $\pi_{ij} = P(Y_{ij} = 1)$. The

underlying model is

$$logit(\pi_i) = log(\frac{\pi_i}{1 - \pi_i}) = \beta_0 + \beta_1 * treatment_i, i = 1, 2, \ldots, n$$

For the subjects in the placebo group, we assume that $treatment = 0$ and $\pi_i = P(Y_i = 1) = 0.5$, i.e. the chance of having good respiratory status is 0.5. Thus $\beta_0 = logit(\pi_i) = 0$. It is also assumed that the active treatment is actually effective and it increases the probability of having good responses, i.e. $\pi_i = P(Y_i = 1) > 0.5$. So for the subjects in the active treatment group($treatment = 1$), we assume that $\pi_i = P(Y_i = 1)) = 0.6, 0.7, 0.8$ and 0.9. Since $\beta_0 = 0$ and $treatment = 1$, $\beta_0 + \beta_1 * 1 = logit(\pi_i) = log(\frac{\pi_i}{1 - \pi_i})$ then $\beta_1 = 0.41, 0.85, 1.39$ and 2.20, respectively. As we will see, the power of the two methods are quite high already when $\beta_1 = 1.39$. Therefore, power for $\beta_1 = 0.41, 0.85, 1.39$ will be observed for our study.

The R package "bindata" and command "rmvbin" were used to generate correlated binary responses. Command "rmvbin" creates correlated multivariate binary variables by "thresholding a normal distribution"(Leisch et al., 1998). When generating correlated binary variables, one can specify the joint probability, the correlations matrix of the binary distribution or the covariance matrix of the normal distribution that it uses to generate outcomes (Leisch et al., 1998). Exchangeable correlation matrix were generated with correlation coefficient $\rho = 0.3$ and 0.7 representing low and high correlation among

responses.

3.2   Analysis of Simulation

3.2.1   Software Package Introduction

In this section, the analysis of the simulated data is conducted using SAS software (Version 9.2, SAS Institute Inc. Cary, NC, USA). PROC NLMIXED and PROC GENMOD procedures are performed to fit the non-linear mixed effects model and the generalized estimating equations method, respectively.

PROC NLMIXED enables us to fit non-linear mixed models where the conditional distribution of the data is the normal, gamma, binary, Poisson distribution or a general function we code. Two major approximation methods, the daptive Gaussian quadrature and a first-order Taylor series approximation, are available in this procedure to provide the integral approximation to the likelihood. By maximizing the approximation to the likelihood, the estimates, standard errors and their p-values can be obtained (SAS 9.22 Users' Guide, 2010).

3.2.2 Non-linear Mixed Effects Model

As discussed in Chapter 2 and previous sections, the underlying model is given by

$$logit(\tilde{\pi}_{ij}) = log(\frac{\tilde{\pi}_{ij}}{1 - \tilde{\pi}_{ij}}) = \beta_0 + \beta_1 * treatment_i + e_i, i = 1, 2, \ldots, n; j = 1, 2, \ldots, T$$

The corresponding code is as following.

proc nlmixed data=simdata update=dfp;

parms beta0=0 beta1=0;

eta=beta0+beta1*treatment+z;

expeta=exp(eta);

p=expeta/(1+expeta);

model outcome binomial(1,p);

random z normal(0,sd*sd) subject=id;

predict eta out=eta;

ods output ParameterEstimates=NLMIXED;

run;

The PROC NLMIXED statement "invokes" the procedure and inputs the data set. The PARMS statement identifies the unknown parameters and their starting values. After specifying the form of the model, the MODEL statement specifies the distribution of the response variable(i.e. binomial distribution

here). Here z is a random term that follows the normal distribution. The SUBJECT statement indicates the cluster variable (i.e. individuals in our study)(SAS 9.22 Users' Guide, 2010).

3.2.3 The GEE Method

The following is the statements of PROC GENMOD:

proc genmod data=resp descend;

class id treatment(ref="0") / param=ref;

model outcome=treatment / dist=bin;

repeated subject=id / corr=exch corrw;

ods output GEEEmpPEst=GEE;

run;

By specifying the REPEATED statement, PROC GENMOD applies the GEE method. The DESCEND option means that outcome=1 will be modeled. Otherwise, outcome =0 will be modeled. The CLASS statement defines the categorical variables and their reference group as well. The MODEL statement specifies the form of the model and also the distribution of the response variable. Similar to PROC NLMIXED, the SUBJECT statement indicates that a cluster is defined by ID variable. The CORR option allows us to specify the structure of the working correlation matrix. For example, "EXCH" means that a exchangeable working correlation structure is used (SAS 9.22 Users'

Guide, 2010).

3.2.4 Testing the Normality Assumption

Normality assumption is crucial in statistical procedures. Unlike the asymptotic properties in large samples, the normality assumption may not be valid in the data sets of a small sample size and therefore the test of normality is necessary. In this section, we use Quantile-Quantile (Q-Q) plots and histograms to test the normality of parameter estimates. Also formal normality tests are performed to implement the graphical methods.

Histograms and Q-Q plots are created for $\hat{\beta}_1$ obtained from the non-linear mixed effects model and the GEE method when sample size is 40, 60, 80, 100 and 200, repeated measurement T=2, 4 and 6, ρ=0.3 and 0.7, and β_1 is 0, 0.41, 0.85 and 1.39. From the graphs, histograms are all approximately symmetric and bell-shaped. In all the Q-Q plots, most of the dots are placed on or near the straight line. For example, the following Figure 3.1 and Figure 3.2 are the histograms and Q-Q plots for $\hat{\beta}_1$ from the two methods when sample size is 40, T=6, $\rho = 0.3$ and $\beta_1 = 0$. The shape of the histograms approximates the normal distribution. Most of the dots are falling on or near the straight line except some outliers in the tail end. Based on the graphical methods, we can say that the normality assumption is not seriously violated.

As we mentioned previously, graphical methods are not strong enough to

provide conclusive evidence. Then formal normal tests are conducted. The following Table 3.1 is the formal normality test conducted when sample size is 40, T=4, $\rho = 0.3$ and $\beta_1 = 0$. In contrast to the graphic methods, the small p-values obtained from formal normality tests showed that the normality assumption of GEE method was seriously violated under 0.05 level. One explanation for this contradiction is that formal tests are extremely sensitive to slight deviations from the normal distribution such as outliers when the sample size is too large(here the sample size is 5000). Then we may conclude that the normal assumption is not valid as the histogram and Q-Q plot indicate but the deviation may not be as serious as we observed from the p-values.

Figure 3.1: Histogram of $\hat{\beta}_1$ when $\beta_1 = 0$, T=6, $\rho = 0.3$ and n=40 (Left: NLME; Right: GEE)
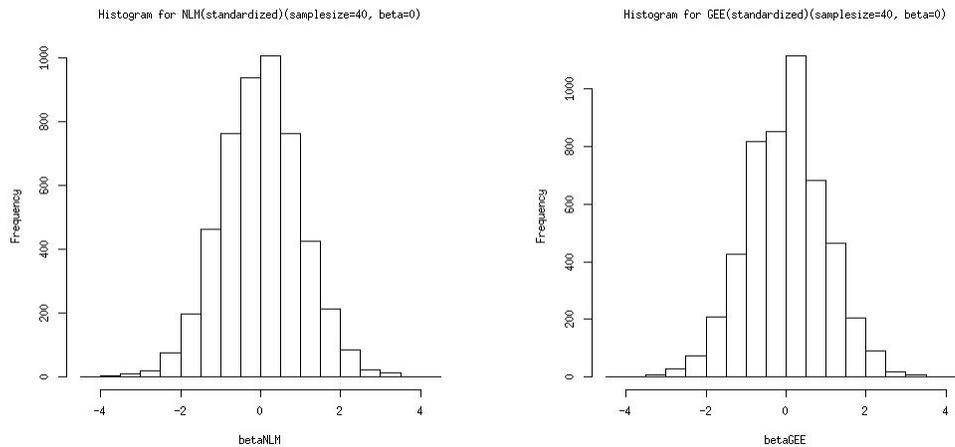
Figure 3.2: Q-Q plot of $\hat{\beta}_1$ when $\beta_1 = 0$, T=6, $\rho = 0.3$ and n=40 (Left: NLME; Right: GEE)

Normal Q-Q Plot for NLM(standardized)(samplesize=40,beta=0)

Normal Q-Q Plot for GEE(standardized)(samplesize=40,beta=0)
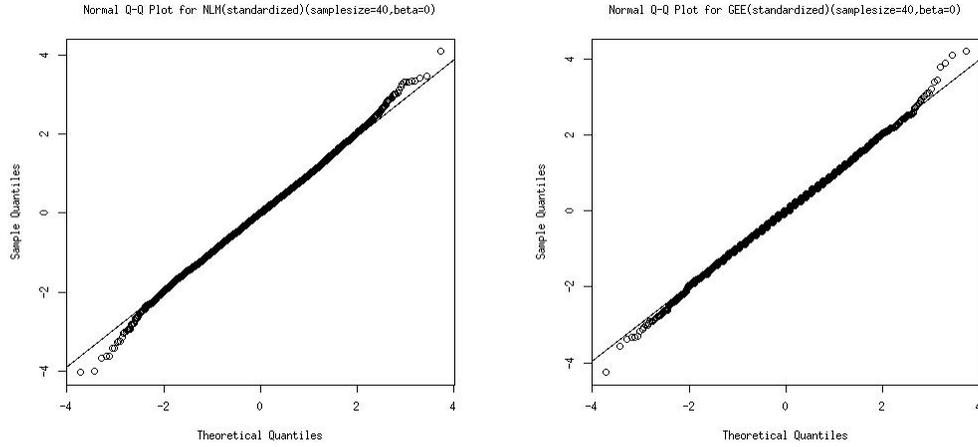
Table 3.1: Results of Normality Test for $\hat{\beta}_1$ when n=40, $\beta_1 = 0$, T=4, $\rho = 0.3$

Method	Tests of normality	P-value
NLMIXED	Shapiro-Wilk test	0.0021
	Anderson-Darling test	0.0535
	Pearson Chi-square test	< 0.0001
	Lilliefors test	0.2801
GEE	Shapiro-Wilk test	< 0.0001
	Anderson-Darling test	< 0.0001
	Pearson Chi-square test	< 0.0001
	Lilliefors test	< 0.0001

28

3.2.5 Empirical Size and Power Comparison

Size and Power

In hypothesis testing, two types of error can occur. A Type I error occurs if H_0 is rejected when H_0 is true. A Type II error occurs if H_0 is accepted when H_1 is true. Usually the Type I error is considered as the worse kind of the two types of error. Therefore we desire to minimize the Type I error by selecting a critical region that bounds the Type I error (Hogg et al., 2005). We say a critical region C is of size α if

$$\alpha = \max_{\theta \in \omega_0} P_\theta[(X_1, X_2, \ldots, X_n) \in C]$$

Thus, α=P(reject H_0 | H_0 is true).

For all the critical regions of size α, we also desire to minimize Type II Error=P(do not reject H_0 | H_1 is true), or equivalently maximizing 1-Type II Error=P(do not reject H_1 | H_1 is true) which is the probability of making the right decision. The right side of the equation is called the power of a test. The greater the power is, the more powerful the test is and the better the procedure is.

Empirical Size and Power

In this section, we compare the empirical size and power of the two methods for different sample size. The empirical size of a test, an estimate of the size, is calculated as the ratio of the number of rejected tests and the total number of tests when the null hypothesis is true. Empirical power is, however, calculated as the ratio of number of rejected tests and the total number of tests under the condition that the alternative hypothesis is true (Lin, 2003). In our simulation, the denominator is 5000.

In large samples, the empirical size, the estimate of size α_0, follows the normal distribution $\mathcal{N}(\alpha_0, \sigma^2)$, where $\sigma = \sqrt{\frac{\alpha_0(1-\alpha_0)}{5000}}$. Therefore the value of empirical size is expected to fall into the interval $(\alpha_0 \pm 1.96\sigma)$. The corresponding intervals are $(0.007, 0.013)$, $(0.044, 0.056)$ and $(0.092, 0.108)$ for $\alpha = 0.01, 0.05, 0.10$ respectively. Empirical Size of the two models under the level of 0.01, 0.05 and 0.10 are listed in Table 3.2, Table 3.3 and Table 3.4

In general, the non-linear mixed effects model is too conservative and the empirical size increases as the sample size increases. However, the GEE method is liberal and the empirical size decreases as the sample size increases. Under the level of 0.01, the GEE method is acceptable when n is larger than 60 while the non-linear mixed effects model can only be used after n=100. Under the level of 0.05, the GEE method can be used when n is larger than 40 while the non-linear mixed effects model can be used when n is larger than 80.

30

Table 3.2: Empirical Size for the GEE method and the NLME model in 5000 simulations under the level of 0.01

ρ	T	Method	$n = 40$	$n = 60$	$n = 80$	$n = 100$	$n = 200$
$\rho =0.3$	T=2	GEE	0.0114	**0.0154**	0.0108	0.0108	0.0126
		NLMIXED	**0.0034**	0.0072	**0.0064**	**0.0060**	0.0096
	T=4	GEE	**0.0174**	0.0100	0.0112	0.0098	0.0110
		NLMIXED	0.0098	0.0082	0.0092	0.0080	0.0102
	T=6	GEE	**0.0164**	**0.0152**	0.0118	0.0086	0.0090
		NLMIXED	0.0100	0.0128	0.0096	0.0076	0.0080
$\rho =0.7$	T=2	GEE	0.0110	0.0102	0.0114	0.0106	0.0100
		NLMIXED	**0.0396**	**0.0418**	**0.0330**	**0.0292**	0.0100
	T=4	GEE	0.0118	**0.0136**	0.0098	0.0130	0.1020
		NLMIXED	**0.0050**	0.0070	**0.0052**	0.0070	0.0078
	T=6	GEE	0.0122	0.0124	0.0120	0.0106	0.0118
		NLMIXED	**0.0052**	**0.0064**	0.0076	**0.0068**	0.0112

Note: the entries are the proportion of rejections out of 5000 simulations.
For Type I Error=0.01, the 95% CI is (0.007, 0.013).

Under the level of 0.10, the GEE method is good when n is 60 or larger while the non-linear mixed effects model is good when n is larger than 100. In summary, the GEE method has better empirical size than the non-linear mixed effects model and thus is recommended. When sample size is 60 or lower, GEE is too liberal and the non-linear mixed effects model is too conservative and therefore are not recommended to use. When sample size is 100 or up, both of the methods will have satisfactory empirical size.

The power of the non-linear mixed effects model and GEE method are listed in Table 3.5, Table 3.6 and Table 3.7.

The results above show that the GEE method always has a higher or at

Table 3.3: Empirical Size for the GEE method and the NLME model in 5000 simulations under the level of 0.05

ρ	T	Method	$n=40$	$n=60$	$n=80$	$n=100$	$n=200$
ρ =0.3	T=2	GEE	**0.0562**	0.0552	0.0554	0.0520	0.0542
		NLMIXED	**0.0366**	**0.0438**	0.0474	0.0454	0.0512
	T=4	GEE	**0.0602**	0.046	0.0536	**0.0598**	**0.0566**
		NLMIXED	0.0500	**0.0402**	0.0488	0.0548	0.0546
	T=6	GEE	**0.0598**	0.0550	0.0524	0.0470	0.0544
		NLMIXED	0.0496	0.0490	0.0488	0.0446	0.0522
ρ =0.7	T=2	GEE	**0.0636**	0.0520	0.0536	0.0506	0.0536
		NLMIXED	**0.0784**	**0.0688**	**0.0594**	0.0492	0.0424
	T=4	GEE	0.0520	**0.0598**	0.0560	0.0532	0.0516
		NLMIXED	**0.0338**	**0.0392**	**0.0378**	0.0444	0.0474
	T=6	GEE	**0.0572**	0.0524	0.0552	0.0500	0.0512
		NLMIXED	**0.0396**	**0.0380**	**0.0432**	**0.0434**	0.0538

Note: the entries are the proportion of rejections out of 5000 simulations. For Type I Error=0.05, the 95% CI is (0.044, 0.056).

least equal power compared to the non-linear mixed effects model. However, the power of a test is only comparable when the empirical size is accurate. For example, the GEE method may seem to be more powerful because it tends to reject more and the non-linear mixed effects model may have lower power due to its conservativeness. Therefore, based on the empirical size we obtained, the power of the two methods should be adjusted accordingly.

3.2.6 Adjusted Size and Power Comparison

The original critical values, which are calculated based on the asymptotic theory, may not be valid in this case because the sample size is small and the

Table 3.4: Empirical Size for the GEE method and the NLME model in 5000 simulations under the level of 0.10

ρ	T	Method	$n=40$	$n=60$	$n=80$	$n=100$	$n=200$
$\rho=0.3$	T=2	GEE	**0.1088**	0.1022	0.1058	0.1040	0.1014
		NLMIXED	**0.0894**	**0.0906**	0.0968	0.0942	0.0970
	T=4	GEE	**0.1128**	0.0968	0.1066	0.1078	**0.1090**
		NLMIXED	0.1000	**0.0888**	0.1002	0.1052	0.1070
	T=6	GEE	**0.1096**	0.1078	0.1036	0.0934	0.1002
		NLMIXED	0.1014	0.1026	0.0982	**0.0910**	0.0968
$\rho=0.7$	T=2	GEE	**0.1142**	0.1054	0.0994	0.0998	0.1010
		NLMIXED	**0.1154**	0.1034	0.0932	**0.0904**	**0.0876**
	T=4	GEE	**0.1092**	**0.1116**	**0.1092**	0.0992	0.1064
		NLMIXED	**0.0760**	**0.0850**	**0.0910**	**0.0918**	0.1064
	T=6	GEE	**0.1096**	0.1048	0.1074	0.1026	0.1078
		NLMIXED	**0.0820**	**0.0864**	0.0982	0.0994	**0.1126**

Note: the entries are the proportion of rejections out of 5000 simulations.
For Type I Error=0.10, the 95% CI is (0.092, 0.108).

normality assumption is violated. Then to ensure that we can obtain the true power of the models, adjusted critical values should be calculated to make the empirical size to be exactly 0.01, 0.05 and 0.10. Then adjusted power will be calculated to reflect the true power of the tests. In PROC NLMIXED, a t-test is performed to evaluate the significance of factor effects and in PROC GENMOD, a Z-test is performed. Then the original and adjusted critical values can be obtained. For example, the original and adjusted critical values for α=0.01, 0.05 and 0.10 when ρ=0.3, T=2, 4, 6, n=40, 60, 80, 100 and 200 are displayed in Table 3.8, Table 3.9 and Table 3.10. The corresponding adjusted power values are displayed in Table 3.11, Table 3.12 and Table 3.13,

Table 3.5: Power for the GEE method and the NLME model in 5000 simulations under the level of 0.01

ρ	T	β_1	Method	$n=40$	$n=60$	$n=80$	$n=100$	$n=200$
ρ =0.3	T=2	0.41	GEE	0.0228	0.0316	0.0380	0.0448	0.0790
			NLMIXED	0.0106	0.0160	0.0236	0.0324	0.0692
		0.85	GEE	0.0628	0.1088	0.1580	0.1832	0.4144
			NLMIXED	0.0324	0.0606	0.1068	0.1410	0.3942
		1.39	GEE	0.1362	0.2660	0.3864	0.4982	0.8442
			NLMIXED	0.0898	0.1836	0.3036	0.4346	0.8188
	T=4	0.41	GEE	0.0236	0.0252	0.0266	0.0380	0.0554
			NLMIXED	0.0170	0.0188	0.0206	0.0328	0.0518
		0.85	GEE	0.0532	0.0846	0.1076	0.1322	0.3040
			NLMIXED	0.0522	0.0674	0.0890	0.1162	0.2884
		1.39	GEE	0.1114	0.1864	0.2890	0.3618	0.6926
			NLMIXED	0.1202	0.1734	0.2600	0.3402	0.6772
	T=6	0.41	GEE	0.0204	0.0226	0.0314	0.0358	0.0548
			NLMIXED	0.0192	0.0176	0.0262	0.0304	0.0516
		0.85	GEE	0.0464	0.0656	0.0934	0.1106	0.2466
			NLMIXED	0.0524	0.0560	0.0858	0.1012	0.2368
		1.39	GEE	0.0906	0.1756	0.2366	0.3052	0.6254
			NLMIXED	0.1162	0.1680	0.2224	0.2862	0.6130
ρ =0.7	T=2	0.41	GEE	0.0072	0.0342	0.0446	0.0568	0.1092
			NLMIXED	0.0072	0.0264	0.0338	0.0384	0.0834
		0.85	GEE	0.0334	0.1504	0.1968	0.2456	0.5320
			NLMIXED	0.0170	0.0584	0.1210	0.1694	0.4512
		1.39	GEE	0.1784	0.3468	0.4958	0.5940	0.9282
			NLMIXED	0.0662	0.1218	0.3282	0.4526	0.8768
	T=4	0.41	GEE	0.0272	0.0322	0.0416	0.0514	0.0926
			NLMIXED	0.0092	0.0172	0.0208	0.0302	0.0780
		0.85	GEE	0.0806	0.1270	0.1796	0.2358	0.4978
			NLMIXED	0.0432	0.0734	0.1194	0.1602	0.4338
		1.39	GEE	0.1812	0.3370	0.4642	0.5740	0.9042
			NLMIXED	0.1020	0.2216	0.3378	0.4570	0.8600
	T=6	0.41	GEE	0.0308	0.0342	0.0446	0.0460	0.0924
			NLMIXED	0.0162	0.0196	0.0252	0.0314	0.0856
		0.85	GEE	0.0692	0.1184	0.1754	0.2242	0.4648
			NLMIXED	0.0370	0.0740	0.1248	0.1620	0.4256
		1.39	GEE	0.1756	0.3220	0.4428	0.5612	0.8848
			NLMIXED	0.1094	0.2256	0.3444	0.4648	0.8590

Table 3.6: Power for the GEE method and the NLME model in 5000 simulations under the level of 0.05

ρ	T	β_1	Method	$n = 40$	$n = 60$	$n = 80$	$n = 100$	$n = 200$
$\rho =0.3$	T=2	0.41	GEE	0.0868	0.1040	0.1196	0.1376	0.2086
			NLMIXED	0.0586	0.0818	0.1046	0.1232	0.2008
		0.85	GEE	0.1828	0.2642	0.3238	0.3866	0.6398
			NLMIXED	0.1418	0.2252	0.2884	0.3586	0.6242
		1.39	GEE	0.3398	0.4860	0.6182	0.7224	0.9428
			NLMIXED	0.2938	0.4526	0.5846	0.6984	0.9412
	T=4	0.41	GEE	0.0854	0.0936	0.0940	0.1148	0.1698
			NLMIXED	0.0758	0.0824	0.0878	0.1094	0.1676
		0.85	GEE	0.1558	0.2124	0.2628	0.2976	0.5212
			NLMIXED	0.1494	0.2000	0.2498	0.2834	0.5138
		1.39	GEE	0.2772	0.3918	0.5056	0.5856	0.8686
			NLMIXED	0.2868	0.3800	0.4940	0.5760	0.8640
	T=6	0.41	GEE	0.0822	0.0860	0.0996	0.1064	0.1598
			NLMIXED	0.0742	0.0782	0.0942	0.1016	0.1554
		0.85	GEE	0.1422	0.1782	0.2302	0.2644	0.4650
			NLMIXED	0.1400	0.1666	0.2220	0.2550	0.4580
		1.39	GEE	0.2478	0.3640	0.4534	0.5214	0.8150
			NLMIXED	0.2550	0.3590	0.4416	0.5116	0.8086
$\rho =0.7$	T=2	0.41	GEE	0.0206	0.1164	0.1414	0.1544	0.2640
			NLMIXED	0.0226	0.1022	0.1134	0.1256	0.2410
		0.85	GEE	0.0900	0.3254	0.3978	0.4564	0.7590
			NLMIXED	0.0618	0.2504	0.3154	0.3668	0.7172
		1.39	GEE	0.3994	0.5858	0.7198	0.8018	0.9804
			NLMIXED	0.2402	0.4628	0.6072	0.7010	0.9638
	T=4	0.41	GEE	0.0994	0.1170	0.1372	0.1436	0.2390
			NLMIXED	0.0604	0.0814	0.1024	0.1182	0.2284
		0.85	GEE	0.2208	0.2976	0.3682	0.4486	0.7196
			NLMIXED	0.1562	0.2230	0.3026	0.3920	0.6986
		1.39	GEE	0.3936	0.5644	0.6864	0.7836	0.9718
			NLMIXED	0.2994	0.4834	0.6140	0.7260	0.9654
	T=6	0.41	GEE	0.0934	0.1126	0.1298	0.1404	0.2326
			NLMIXED	0.0660	0.0836	0.1072	0.1178	0.2300
		0.85	GEE	0.2022	0.2830	0.3674	0.4312	0.6952
			NLMIXED	0.1432	0.2278	0.3182	0.3924	0.6828
		1.39	GEE	0.3900	0.5560	0.6708	0.7656	0.9664
			NLMIXED	0.3176	0.4910	0.6176	0.7312	0.9618

Table 3.7: Power for the GEE method and the NLME model in 5000 simulations under the level of 0.10

ρ	T	β_1	Method	$n=40$	$n=60$	$n=80$	$n=100$	$n=200$
$\rho=0.3$	T=2	0.41	GEE	0.1568	0.1724	0.1934	0.2270	0.3176
			NLMIXED	0.1314	0.1562	0.1794	0.2140	0.3132
		0.85	GEE	0.2758	0.3690	0.4452	0.5074	0.7538
			NLMIXED	0.2462	0.3444	0.4234	0.4892	0.7460
		1.39	GEE	0.4678	0.6074	0.7288	0.8214	0.9704
			NLMIXED	0.4398	0.5906	0.7150	0.8092	0.9684
	T=4	0.41	GEE	0.1452	0.1548	0.1640	0.1942	0.2642
			NLMIXED	0.1376	0.1460	0.1592	0.1838	0.2614
		0.85	GEE	0.2408	0.3088	0.3724	0.4116	0.6384
			NLMIXED	0.2390	0.2990	0.3612	0.4054	0.6348
		1.39	GEE	0.3820	0.5134	0.6240	0.6964	0.9252
			NLMIXED	0.3938	0.5086	0.6158	0.6930	0.9248
	T=6	0.41	GEE	0.1368	0.1516	0.1654	0.1772	0.2502
			NLMIXED	0.1342	0.1448	0.1600	0.1712	0.2498
		0.85	GEE	0.2270	0.2712	0.3332	0.3762	0.5830
			NLMIXED	0.2278	0.2628	0.3228	0.3690	0.5860
		1.39	GEE	0.3596	0.4872	0.5744	0.6416	0.8858
			NLMIXED	0.3738	0.4786	0.5728	0.6382	0.8834
$\rho=0.7$	T=2	0.41	GEE	0.0438	0.1938	0.2310	0.2488	0.3850
			NLMIXED	0.0414	0.1666	0.1952	0.2162	0.3642
		0.85	GEE	0.1418	0.4450	0.5302	0.5808	0.8476
			NLMIXED	0.1128	0.3702	0.4456	0.5050	0.8276
		1.39	GEE	0.5496	0.7070	0.8176	0.8762	0.9914
			NLMIXED	0.4000	0.6032	0.7210	0.8030	0.9822
	T=4	0.41	GEE	0.1648	0.1918	0.2176	0.2334	0.3416
			NLMIXED	0.1220	0.1544	0.1886	0.2080	0.3406
		0.85	GEE	0.3170	0.4076	0.4896	0.5724	0.8116
			NLMIXED	0.2568	0.3468	0.4366	0.5352	0.8018
		1.39	GEE	0.5192	0.6776	0.7818	0.8582	0.9856
			NLMIXED	0.4440	0.6218	0.7428	0.8264	0.9852
	T=6	0.41	GEE	0.1572	0.1904	0.2100	0.2256	0.3430
			NLMIXED	0.1282	0.1656	0.1932	0.2092	0.3468
		0.85	GEE	0.2994	0.3964	0.4826	0.5476	0.7990
			NLMIXED	0.2488	0.3602	0.4544	0.5256	0.7902
		1.39	GEE	0.5162	0.6686	0.7742	0.8442	0.9832
			NLMIXED	0.4644	0.6338	0.7514	0.8340	0.9832

respectively.

Table 3.8: Adjusted critical values for the GEE method and the NLME model in 5000 simulations under the level of 0.01

ρ	T	Method		$n = 40$	$n = 60$	$n = 80$	$n = 100$	$n = 200$
$\rho = 0.3$	T=2	GEE	Original	2.5758	2.5758	2.5758	2.5758	2.5758
			Adjusted	2.6184	2.7729	2.5916	2.5854	2.6194
		NLMIXED	Original	2.7079	2.6618	2.6395	2.6264	2.6008
			Adjusted	2.4120	2.5606	2.4868	2.4940	2.5872
	T=4	GEE	Original	2.5758	2.5758	2.5758	2.5758	2.5758
			Adjusted	2.7569	2.5737	2.6017	2.5538	2.6119
		NLMIXED	Original	2.7079	2.6618	2.6395	2.6264	2.6008
			Adjusted	2.6987	2.5638	2.5797	2.5689	2.6015
	T=6	GEE	Original	2.5758	2.5758	2.5758	2.5758	2.5758
			Adjusted	2.7527	2.7542	2.6697	2.5152	2.5135
		NLMIXED	Original	2.7079	2.6618	2.6395	2.6264	2.6008
			Adjusted	2.7036	2.7218	2.6033	2.5119	2.5083

From the tables above, the adjusted critical values are relatively smaller than the original critical value for the non-linear mixed effects model since it is conservative in small samples. For the GEE method, the adjusted critical value will become larger than the original ones because the GEE method is liberal and tends to reject more. They are consistent with the tables of empirical size.

Based on the adjusted critical value, we can obtain the adjusted power.

The tables above show that the power of the two methods after adjustment are much more closer than before, which means that the two methods are equally powerful after adjustment.

We should note that even though the two methods are equally powerful after adjustment, researchers may not be able to adjust for the power as we do

Table 3.9: Adjusted critical values for the GEE method and the NLME model in 5000 simulations under the level of 0.05

ρ	T	Method		$n = 40$	$n = 60$	$n = 80$	$n = 100$	$n = 200$
ρ =0.3	T=2	GEE	Original	1.9600	1.9600	1.9600	1.9600	1.9600
			Adjusted	2.0042	2.0073	2.0198	1.9711	2.0029
		NLMIXED	Original	2.0227	2.0010	1.9905	1.9842	1.9720
			Adjusted	1.8977	1.9361	1.9618	1.9285	1.9927
	T=4	GEE	Original	1.9600	1.9600	1.9600	1.9600	1.9600
			Adjusted	2.0523	1.9382	1.9990	2.0336	2.0122
		NLMIXED	Original	2.0227	2.0010	1.9905	1.9842	1.9720
			Adjusted	2.0211	1.9266	1.9751	2.0204	2.0102
	T=6	GEE	Original	1.9600	1.9600	1.9600	1.9600	1.9600
			Adjusted	2.0412	1.9965	1.9751	1.9424	1.9995
		NLMIXED	Original	2.0227	2.0010	1.9905	1.9842	1.9720
			Adjusted	2.0172	1.9946	1.9762	1.9323	1.9887

Table 3.10: Adjusted critical values for the GEE method and the NLME model in 5000 simulations under the level of 0.10

ρ	T	Method		$n = 40$	$n = 60$	$n = 80$	$n = 100$	$n = 200$
ρ =0.3	T=2	GEE	Original	1.6449	1.6449	1.6449	1.6449	1.6449
			Adjusted	1.7168	1.6535	1.6734	1.6576	1.6488
		NLMIXED	Original	1.6849	1.6711	1.6644	1.6604	1.6525
			Adjusted	1.6413	1.6120	1.6494	1.6365	1.6442
	T=4	GEE	Original	1.6449	1.6449	1.6449	1.6449	1.6449
			Adjusted	1.7089	1.6278	1.6712	1.6910	1.6822
		NLMIXED	Original	1.6849	1.6711	1.6644	1.6604	1.6525
			Adjusted	1.6841	1.6159	1.6642	1.6805	1.6842
	T=6	GEE	Original	1.6449	1.6449	1.6449	1.6449	1.6449
			Adjusted	1.6903	1.6907	1.6688	1.6086	1.6447
		NLMIXED	Original	1.6849	1.6711	1.6644	1.6604	1.6525
			Adjusted	1.6915	1.6818	1.6568	1.6195	1.6314

Table 3.11: Adjusted power for the GEE method and the NLME model in 5000 simulations under the level of 0.01

ρ	T	β_1	Method	$n = 40$	$n = 60$	$n = 80$	$n = 100$	$n = 200$
ρ =0.3	T=2	0.41	GEE	0.0204	0.0210	0.0366	0.0434	0.0724
			NLMIXED	0.0204	0.0226	0.0362	0.0343	0.0708
		0.85	GEE	0.0592	0.0756	0.1524	0.1804	0.3972
			NLMIXED	0.0618	0.0772	0.1434	0.1706	0.3872
		1.39	GEE	0.1298	0.2044	0.3800	0.4948	0.8304
			NLMIXED	0.1494	0.2162	0.3748	0.4862	0.8224
	T=4	0.41	GEE	0.0136	0.0252	0.0250	0.0396	0.0520
			NLMIXED	0.0174	0.0236	0.0240	0.0374	0.0518
		0.85	GEE	0.0388	0.0846	0.1020	0.1354	0.2912
			NLMIXED	0.0524	0.0808	0.0994	0.1264	0.2876
		1.39	GEE	0.0776	0.1866	0.2798	0.3688	0.6766
			NLMIXED	0.1206	0.1952	0.2802	0.3574	0.6768
	T=6	0.41	GEE	0.0150	0.0148	0.0250	0.0404	0.0632
			NLMIXED	0.0192	0.0146	0.0284	0.0390	0.0620
		0.85	GEE	0.0294	0.0470	0.0830	0.1220	0.2678
			NLMIXED	0.0524	0.0492	0.0894	0.1220	0.2632
		1.39	GEE	0.0690	0.1314	0.2068	0.3252	0.6452
			NLMIXED	0.1164	0.1548	0.2324	0.3184	0.6402

Table 3.12: Adjusted power for the GEE method and the NLME model in 5000 simulations under the level of 0.05

ρ	T	β_1	Method	$n=40$	$n=60$	$n=80$	$n=100$	$n=200$
$\rho=0.3$	T=2	0.41	GEE	0.0802	0.0944	0.1084	0.1348	0.1958
			NLMIXED	0.0814	0.0932	0.1106	0.1366	0.1948
		0.85	GEE	0.1726	0.2496	0.3016	0.3826	0.6234
			NLMIXED	0.1736	0.2464	0.2984	0.3818	0.6178
		1.39	GEE	0.3244	0.4666	0.5946	0.7188	0.9388
			NLMIXED	0.3436	0.4754	0.5980	0.7184	0.9376
	T=4	0.41	GEE	0.0708	0.0978	0.0878	0.1024	0.1572
			NLMIXED	0.0758	0.0948	0.0898	0.1022	0.1588
		0.85	GEE	0.1362	0.2172	0.2512	0.2706	0.4990
			NLMIXED	0.1498	0.2194	0.2548	0.2704	0.5000
		1.39	GEE	0.2478	0.3978	0.4908	0.5592	0.8554
			NLMIXED	0.2868	0.4082	0.5016	0.5636	0.8568
	T=6	0.41	GEE	0.0694	0.0794	0.0978	0.1108	0.1504
			NLMIXED	0.075	0.0804	0.0968	0.1104	0.1524
		0.85	GEE	0.1234	0.1670	0.2262	0.2688	0.4484
			NLMIXED	0.1414	0.1680	0.2250	0.2714	0.4504
		1.39	GEE	0.2174	0.3492	0.4454	0.5292	0.8036
			NLMIXED	0.2564	0.3610	0.4478	0.5310	0.8056

Table 3.13: Adjusted power for the GEE method and the NLME model in 5000 simulations under the level of 0.10

ρ	T	β_1	Method	$n = 40$	$n = 60$	$n = 80$	$n = 100$	$n = 200$
$\rho =0.3$	T=2	0.41	GEE	0.1392	0.17000	0.1838	0.2220	0.3160
			NLMIXED	0.1396	0.1716	0.1842	0.2240	0.3166
		0.85	GEE	0.2520	0.3662	0.4360	0.5026	0.7528
			NLMIXED	0.2612	0.3660	0.4294	0.5010	0.7492
		1.39	GEE	0.4368	0.6058	0.7176	0.8174	0.9700
			NLMIXED	0.4594	0.6126	0.7206	0.8168	0.9694
	T=4	0.41	GEE	0.1326	0.1604	0.1570	0.1746	0.2538
			NLMIXED	0.1376	0.1580	0.1592	0.1774	0.2518
		0.85	GEE	0.2242	0.3128	0.3620	0.3958	0.6238
			NLMIXED	0.2390	0.3188	0.3612	0.3972	0.6232
		1.39	GEE	0.3592	0.5202	0.6134	0.6810	0.9192
			NLMIXED	0.3940	0.5316	0.6158	0.6872	0.9184
	T=6	0.41	GEE	0.1298	0.1410	0.1602	0.1884	0.2504
			NLMIXED	0.1324	0.1418	0.1622	0.1818	0.2586
		0.85	GEE	0.2136	0.2544	0.3228	0.3906	0.5830
			NLMIXED	0.2260	0.2584	0.3254	0.3832	0.5882
		1.39	GEE	0.3424	0.4682	0.5666	0.6544	0.8858
			NLMIXED	0.3712	0.4768	0.5754	0.6526	0.8890

in this project. Although at n=40 NLME is actually more powerful, the sample size of 40 is not sufficiently large enough to apply these methods. Therefore, it is important to keep in mind that there is no real power advantage between the two analysis methods.

Chapter 4

Data Analysis

In this chapter, a data set of clinical trial data comparing two treatments for a respiratory disorder from Stokes et al. (2000) is used. A total of 56 patients came from center1 and 55 patients came from center2. Patients in each of the two centers are randomly assigned to groups that receive the active treatment or a placebo. The response is the respiratory status. It is a binary variable (0=poor, 1=good) and is recorded for each of the four visits. Variables such as center, treatment, gender and baseline status are also binary and variable age is a continuous variable. According to our simulation study, either approach would satisfy the nominal levels with equivalent powers on a data set with N=111, and T=4. Therefore, we apply both methods on this data set.

The Y_{ij} is the respiratory status of patient i at the jth visit, $j = 1, 2, 3, 4$. The partial data set is listed below.

Table 4.1: Respiratory disorder clinical trial: Partial dataset including first 5 subjects

Center	ID	Trt	Sex	Age	Baseline	Visit1	Visit2	Visit3	Visit4
1	1	P	M	46	0	0	0	0	0
1	2	P	M	28	0	0	0	0	0
1	3	A	M	23	1	1	1	1	1
1	4	P	M	44	1	1	1	1	0
1	5	P	F	13	1	1	1	1	1

Suppose that x_{ij1} is the intercept, x_{ij2} is the treatment allocation(0=placebo, 1=active treatment), x_{ij3} is the center that patients came from (0=center1, 1=center2), x_{ij4} is the gender(0=male, 1=female), x_{ij5} is the age and x_{ij6} is the baseline status(0=poor, 1=good). Then the GEE model will be

$$logit(\mu_{ij}) = x'_{ij}\beta$$

where $\mu_{ij} = \mathbb{E}(Y_i)$ is the mean of the respiratory status. The non-linear mixed effects model will have the same form except that the intercept will be random.

The output of the GEE method is shown in Table 4.2 and the output of the non-linear mixed effects model is shown in Table 4.4. Based on the output of the GEE method, only treatment and baseline status are significant with the p-value of 0.003 and <0.0001, respectively. The same result can be observed in the output of the non-linear mixed effects model. In the GEE method, $\hat{\beta}_1 = 1.2442$, which means that the odds of a good respiratory status are 3.4702 (exp(1.2442), 95% CI=(1.7628, 6.8305)) times as high for the treatment group

as for the placebo group.

In the non-linear mixed effects model, $\hat{\beta}_1 = 2.0059$, which means that the estimated odds of a good respiratory status for a subject in the treatment group is 7.4328(exp(2.0059), 95% CI=(2.5847, 21.3767)) times of a subject with similar random effect values in the placebo group. We note that NLME offers additional information about the individual's heterogeneity, which has a variance of 1.93 and was highly significant even at the level of .0001.

Table 4.2: Respiratory disorder clinical trial: Parameter estimate from the GEE approach, N=111 subjects, T=4 repeated measurements

GEE Parameter Estimates							
Parameter		Estimate	SE	95% LCL	95 % UCL	Z	Pr> $|Z|$
Intercept		-0.8882	0.4568	-1.7835	0.0071	-1.94	0.0519
Treatment	1	1.2442	0.3455	0.5669	1.9214	3.60	0.0003
Center	1	0.6558	0.3512	-0.0326	1.3442	1.87	0.0619
Sex	1	0.1128	0.4408	-0.7512	0.9768	0.26	0.7981
Age		-0.0175	0.0129	-0.0427	0.0077	-1.36	0.1728
Baseline	1	1.8981	0.3441	1.2237	2.5725	5.52	< .0001

Note: SE:Standard Error, LCL: Lower Confidence Limit, UCL: Upper Confidence Limit

Table 4.3: Respiratory disorder clinical trial: The estimated working correlation matrix in the GEE approach, N=111 subjects, T=4 repeated measurements

1.0000	0.3351	0.2140	0.2953
0.3351	1.0000	0.4429	0.3581
0.2140	0.4429	1.0000	0.3964
0.2953	0.3581	0.3964	1.0000

Table 4.4: Respiratory disorder clinical trial: Parameter estimate from the NLME method, N=111 subjects, T=4 repeated measurements

NLMIXED Parameter Estimates						
Parameter	Estimate	SE	95% LCL	95 % UCL	t value	Pr> |t|
Intercept	-1.4661	0.7591	-2.9704	0.03815	-1.93	0.0560
Treatment	2.0059	0.5330	0.9496	3.0623	3.76	0.0003
Center	0.9701	0.5350	-0.09013	2.0303	1.81	0.0725
Sex	0.2399	0.6660	-1.0682	1.5479	0.36	0.7170
Age	-0.02659	0.01982	-0.06586	0.01268	-1.34	0.1824
Baseline	2.8849	0.5660	1.7632	4.0066	5.10	< .0001
sd	1.9348	0.3060	1.3283	2.5412	6.32	< .0001

Note: SE: Standard Error, LCL: Lower Confidence Limit, UCL: Upper Confidence Limit

On the other hand, the computational time of the GEE method is much shorter than that of the non-linear mixed effects model.

Based on the results of the full models, the nonsignificant terms can be omitted from the model. The fitting information of reduced models are shown in Table 4.5 and Table 4.7. The QIC is 512.3416 for the GEE full model and 510.2242 for the reduced model. However, the QICu for the GEE full model was 499.6081 and increased to 503.9955 for the reduced model. The AIC was 446.6 for the NLMIXED full model and 445.0 for the reduced model. According to p-values, there was a strong association between the baseline status and respiratory status. And treatment had a significant effect on the respiratory status of patients. In the GEE model, $\hat{\beta}_1 = 1.2394$, which means that the odds of a good respiratory status are 3.4535 (exp(1.2394), 95% CI=(1.8116, 6.5837)) times as high for the treatment group as for the placebo group.

In the non-linear mixed effects model, $\hat{\beta}_1 = 2.0058$, which means that estimated odds of a good respiratory status for a subject in the treatment group is 7.4320(exp(2.0058), 95% CI=(2.6000, 21.2467)) times of a subject with similar random effect values in the placebo group. We note that NLME offers additional information about the individual's heterogeneity, which has a variance of 2.00 and was highly significant even at the level of .0001.

Table 4.5: Respiratory disorder clinical trial: Parameter estimate from the GEE approach in the reduced model, N=111 subjects, T=4 repeated measurements

GEE Parameter Estimates (reduced)							
Parameter		Estimate	SE	95% LCL	95 % UCL	Z	Pr> $|Z|$
Intercept		-1.1823	0.3010	-1.7723	-0.5923	-3.93	< .0001
Treatment	1	1.2394	0.3292	0.5942	1.8846	3.77	0.0002
Baseline	1	2.0590	0.3212	1.4295	2.6884	6.41	< .0001

Note: SE: Standard Error, LCL: Lower Confidence Limit, UCL: Upper Confidence Limit

Table 4.6: Respiratory disorder clinical trial: The estimated working correlation matrix in the GEE approach in the reduced model, N=111 subjects, T=4 repeated measurements

1.0000	0.3314	0.2217	0.3255
0.3314	1.0000	0.4552	0.4197
0.2217	0.4552	1.0000	0.4124
0.3255	0.4197	0.4124	1.0000

We then performed diagnostic tests to assess the goodness-of-fit of the GEE approach and the non-linear mixed effects model. Classification tables are listed below to compare the predicted values and the observed values of

Table 4.7: Respiratory disorder clinical trial: Parameter estimate from the NLME method in the reduced model, N=111 subjects, T=4 repeated measurements

NLMIXED Parameter Estimates (Reduced)						
Parameter	Estimate	SE	95% LCL	95 % UCL	t value	Pr> $|t|$
Intercept	-1.9418	0.4529	-2.8394	-1.0442	-4.29	< .0001
Treatment	2.0058	0.5300	0.9555	3.0562	3.78	0.0003
Baseline	3.2018	0.5709	2.0704	4.3331	5.61	< .0001
sd	2.0020	0.3101	1.3874	2.6166	6.46	< .0001

Note: SE: Standard Error, LCL: Lower Confidence Limit, UCL: Upper Confidence Limit

the two models. Under the cutting point of 0.5, the predicted response is 1 if the predicted probability $\hat{\pi}_{ij} > 0.5$. Table 4.8 is the classification table of GEE model. Table 4.9 is the classification table of the non-linear mixed effects model where the predicted values were generated with only the fixed effect and Table 4.10 is the classification table of the non-linear mixed effects model where the predicted values were generated with both the fixed effect and the random intercept. Two principle statistics are used to assess the accuracy of a test: sensitivity and specificity. Sensitivity is the probability of a test giving positive diagnosis given that the true responses are positive. Specificity is the probability of a test giving negative diagnosis when the true responses are negative. In a perfect model, there will not be any false positive diagnosis nor false negative diagnosis. All cases will be on the diagonal of the classification table and the sensitivity and specificity will be 100% (Liang, 2002). After obtaining the sensitivity and specificity of a test, we can calculate

the misclassification rate.

Table 4.8: Respiratory disorder clinical trial: Classification table of the GEE approach in the reduced model,, N=111 subjects, T=4 repeated measurements

	Truth +	Truth -	Total
Predict+	217	103	320
Predict-	31	93	124
Total	248	196	444

Table 4.9: Respiratory disorder clinical trial: Classification Table of the NLME Method in the reduced model with Only Fixed Effect, N=111 subjects, T=4 repeated measurements

	Truth +	Truth -	Total
Predict+	217	103	320
Predict-	31	93	124
Total	248	196	444

Table 4.10: Respiratory disorder clinical trial: Classification table of the NLME method in the reduced model with both Fixed Effect and Random Effect, N=111 subjects, T=4 repeated measurements

	Truth +	Truth -	Total
Predict+	226	38	264
Predict-	22	158	180
Total	248	196	444

Then the sensitivity of the GEE approach was 0.8750 and the specificity was 0.4745, which was rather low. The misclassification rate for the GEE approach was 0.3018. The same results were obtained for the non-linear mixed effects model predicted with only the fixed effect term as expected, because

the calculation only accounted for the marginal effects. The sensitivity of the non-linear mixed effects model predicted with both fixed and random effects was 0.9113 and the specificity was 0.8061. And the misclassification rate was 0.1351. The non-linear mixed effects model offered a better fit in terms of classification rates. It was also the method of choice for this data set, as the data exhibited a high level of heterogeneity.

Chapter 5

Conclusion and

Recommendation

This thesis investigated the performance of the two most popular techniques for analyzing discrete repeated measures data. The two techniques we focus on are the GEE method and a non-linear mixed effects model. Using a simulation study, we compared the two methods with respect to their small sample performance, such as the empirical size, statistical power, in an attempt to draw guidelines for the analysts.

Our study revealed that the non-linear mixed effects model is rather too conservative, while the GEE method tended to be too liberal. As the sample size and the number of repeated measures increase, the empirical size of the non-linear mixed effects model generally increases, while that of the GEE

method decreases, both approaches eventually reaching the nominal level. For example, at the nominal level of .05, the GEE method can be used with n is larger than 40, while the non-linear mixed effects model requires n to be at least larger than 100, especially when the correlation is high.

In terms of the level of correlation, the GEE method is not largely affected by it, and the empirical size remains stable and close to the nominal level as long as the sample size is 60 or larger. For the non-linear mixed effects model, the sample size should be larger than 60, but as the correlation rises, the sample size needs to be larger than 100.

In terms of the number of repeated measures per subject, there did not seem any discernable differences between the two approaches and the empirical sizes were not affected by how many repeated measures per subject there were.

We also examined their statistical power. Given that the empirical sizes are satisfactory, the GEE method was observed to have much higher power. However, since the empirical levels were in general higher for the GEE method, we considered adjusting the power by selecting a critical value at an exact nominal level so that the comparison was made on an equal footing. This investigation indicated that the higher statistical power of the GEE method was indeed due to inflated empirical level. In fact, when we compared them on an equal footing, the non-linear mixed effects model was more powerful.

Overall, the GEE approach appears to keep the empirical size at a rea-

sonable level when the sample size is larger than 40. The non-linear mixed effects model requires a larger sample size and we found that sample size larger than 100 appears to suffice the nominal level. In general, the GEE method appears to be more powerful, but the adjusted power indicates that both of the approaches are equally powerful. In practice, researchers may not be able to adjust the power for every analysis. Our study suggests that researchers are to be reminded that the power advantage of the GEE method should be taken with caution.

In the last few years, the Monte Carlo procedures have been widely used in statistical inference. Bootstrap is one of those procedures. Bootstrap procedures are computationally simple resampling methods that can be used to evaluate the accuracy of an estimator. Instead of assuming specific distributions of the estimator (e.g. the asymptotic normal distribution), they only depend on the empirical distributions. By Monte Carlo sampling, we can numerically evaluate the standard deviation of an estimator and obtained the bootstrap estimate of standard error as well as the bootstrap confidence interval. The algorithm has the following steps: (1) randomly draw a large number (in most of situations, a total number of 50-200 is adequate) of bootstrap samples from the data. These bootstrap samples are drawn with replacement and should have the same sample size with the original dataset; (2) calculate the estimate $\{ \hat{\theta}_1, \ldots, \hat{\theta}_B \}$ for each bootstrap sample. B is the total number

of bootstrap samples; (3) evaluate the standard deviation and the confidence intervals of the estimator based on those estimates { $\hat{\theta}_1, \ldots, \hat{\theta}_B$ } (Efron and Tibshirani, 1986).

Besides the GEE approach and the non-linear mixed effects model, the penalized quasi-likelihood (PQL) estimation is another option to perform the analysis. The PQL is a computationally simple technique that "uses a Laplace approximation to the integrated mixed model likelihood" (Dean et al., 2004) and has been commonly used in the fitting process of the generalized mixed effects model (GLMM). It is helpful when the number of the random effects is relatively large (Fitzmaurice et al., 2011). PQL method is implemented in software packages such as SAS, Splus and R.

Bibliography

[Agresti2002] Agresti, A. (2002). *Categorical Data Analysis*. John Wiley & Sons, Inc., Second edition.

[Anderson and Darling1954] Anderson, T. and Darling, D. (1954). A Test of Goodness of Fit. *Journal of American Statistical Association*, 49(268):765–769.

[Conover1999] Conover, W. (1999). *Practical Nonparametric Statistics*. John Wiley & Sons, Inc., Third edition.

[Davidian and Gallant1993] Davidian, M. and Gallant, A. (1993). The Nonlinear Mixed Effect Model with a Smooth Random Effects Density. *Biometrika*, 80(3):475–488.

[Davis2002] Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. Springer-Verlag New York, inc.

[Dean et al.2004] Dean, C., Ugarte, M., and Militino, A. (2004). Penalized Quasi-likelihood with Spatially Correlated Data. *Computaional Statistics & Data Analysis*, 45:235–248.

[Dufour et al.1998] Dufour, J. M., Farhat, A., Cardiol, L., and Khalaf, L. (1998). Simulation-based Finite Sample Normality Tests in Linear Regression. *Econometrics Journal*, 1:154–173.

[Eckhardt1987] Eckhardt, R. (1987). Stan Ulam, John von Neumann, and the Monte Carlo method". *Los Alamos Science*, Special Issue(15):131–137.

[Efron and Tibshirani1986] Efron, B. and Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, 1(1):54–75.

[Fitzmaurice et al.2011] Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2011). *Applied Longitudinal Analysis*. John Wiley & Sons, Inc., Second edition.

[Hogg et al.2005] Hogg, R., Mckean, J., and Craig, A. (2005). *Introduction to Mathematical Statistics.* Pearson Education, Inc., Sixth edition.

[Kiernan et al.2012] Kiernan, K., Tao, J., and Gibbs, P. (2012). Tips and Strategies for Mixed Modeling with SAS/STAT ®Procedures. *Statistics and Data Analysis, SAS Global Forum*, pages 332–2012.

[Lee and Xu2004] Lee, S.-Y. and Xu, L. (2004). Influence Analyses of Nonlinear Mixed Effects Models. *Computational Statistcs & Data Analysis*, 45:321–341.

[Leisch et al.1998] Leisch, F., Weingessel, A., and Hornik, K. (1998). On the Generation of Correlated Artificial Binary Data. Working Paper Series, SFB "Adaptive Information Systems and Modelling in Economics and Management Science", Vienna University of Economics, http://www.wu-wien.ac.at/am.

[Liang2002] Liang, Y. (2002). Analysis of Familial Data on Alcoholism. Master's thesis, University of Alberta.

[Lilliefors1967] Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of American Statistical Association*, 62(318):399–402.

[Lin2003] Lin, Y. (2003). Hierarchical Data and Analytic Methods: A Similation Study. Master's thesis, University of Alberta.

[Lindtrom and Bates1990] Lindtrom, M. and Bates, D. (1990). Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics*, 46(3):673–687.

[McCullagh and Nelder1997] McCullagh, P. and Nelder, J. (1997). *Generalized Linear Model.* Chapman & Hall, Second edition.

[O'Donnell and R.Elosua2008] O'Donnell, C. and R.Elosua (2008). Cardiovascular Risk Factors. Insights From Framingham Heart Study. *Rev Esp Cardiol*, 61(3):299–310.

[Park and Lee2004] Park, T. and Lee, S.-Y. (2004). Model Diagnostic Plots for Repeated Measures Data. *Biom. J*, 46:441–452.

[Pearson et al.1977] Pearson, E., D'Agostino, R., and Bowman, K. (1977). Tests for Departure from Normality: Comparison of Powers. *Biometrika*, 64(2):231–246.

[Razali and Wah2011] Razali, N. M. and Wah, Y. B. (2011). Power Comparison of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *Journal of Statistical Modeling and Analytics*, 2(1):21–33.

[Rochon1998] Rochon, J. (1998). Application of GEE Procedures for Sample Size Calculations in Repeated Measures Experiments. *Statistics in Medicine*, 17:1643–1658.

[SAS2010] SAS (2010). SAS Institute Inc. $SAS/STAT^{\textregistered}$ 9.22 User's Guide, Second Edition. Cary NC: SAS Institute Inc.

[Shapiro and Wilk1965] Shapiro, S. and Wilk, M. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4):591–611.

[Stokes et al.2000] Stokes, M., Davis, C., and Koch, G. (2000). *Categorical Data Analysis Using SAS System*. SAS Institute Inc. Cary, NC, USA, Second edition.

[Thomopoulos2013] Thomopoulos, N. T. (2013). *Essentials of Monte Carlo Simulation: Statistical Methods for Building Simulation Models*. Springer Science+Business Media New York.

[Vonesh and Carter1992] Vonesh, E. and Carter, R. (1992). Mixed-effects Nonlinear Regression for Unbalanced Repeated Measures. *Biometrics*, 48:1–18.

[Walker1996] Walker, S. (1996). An EM Algorithm for Nonlinear Random Effects Models. *Biometrics*, 52:934–944.

[Wedderburn1974] Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61:439–447.

[Yazici and Yolacan2007] Yazici, B. and Yolacan, S. (2007). A Comparison of Various Tests of Normality. *Journal of Statisitcal Computation and Simulation*, 77(2):175–183.

[Zeger and Liang1986a] Zeger, S. and Liang, K. (1986a). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrics*, 73(1):13–22.

[Zeger et al.1988] Zeger, S., Liang, K., and Albert, P. (1988). Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics*, 44(4):1049–1060.

[Zeger and Liang1986b] Zeger, S. L. and Liang, K. (1986b). Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, 42(1):121–130.

[Zio2013] Zio, E. (2013). *The Monte Carlo Simulation Method for System Reliability and Risk Analysis.* Springer-Verlag London.