Modelling Longitudinal Binary Outcomes for Analysing Covariate Dependence

Doctoral Thesis

Submitted by
Jahida Gulshan
Reg. No. 140
Session: 2014-15

Supervisors
Azmeri Khan
M Ataharul Islam

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

Institute of Statistical Research and Training (ISRT)
University of Dhaka
Dhaka 1000, Bangladesh.

January 2021

Dedicated to

The Martyrs of 1971

Supervisor's Declaration

Dated: January 21, 2021

The undersigned hereby certify that this thesis titled, **Modelling Longitudinal Binary Outcomes for Analysing Covariate Dependence** submitted as a requirement for the degree of Ph.D. is the result of Jahida Gulshan's (Reg. No: 140, Session: 2014-2015) research work under our supervision and that this study in whole or in part has not been submitted for any award, including a higher degree, to any other University or Institution.

M. Ataharul Islam, **Ph.D**
Q. M. Husain Professor
Institute of Statistical Research
and Training (ISRT)
University of Dhaka
Dhaka, Bangladesh.

Azmeri Khan, **Ph.D.**
Professor
Institute of Statistical Research
and Training (ISRT)
University of Dhaka
Dhaka, Bangladesh.

Acknowledgements

All praises are due to the Almighty Allah, who gave me the ability to make this dissertation possible and achieve my goal.

I would like to express my sincere gratitude to my supervisor Professor Azmeri Khan,PhD at the Institute of Statistical Research and Training (ISRT), University of Dhaka for her continuous support, encouragement and motivation to complete this study. A very special word of gratitude is due to my co-supervisor, QM Husain Professor, M Ataharul Islam,PhD, who inspired me to work on this research topic and guided me with motivation, encouragement, immense knowledge and strong patience.

My sincere thanks goes to the administration of University of Dhaka and Institute of Statistical Research and Training (ISRT) for giving me the opportunity to join the institution as a PhD candidate and giving me access to the library and other research facilities as and when required. Without these precious supports, it would not be possible to conduct this research so smoothly.

I would like to thank all of the faculty members and staff at the Institute of Statistical Research and Training (ISRT) for their encouragement and support throughout my Ph.D. research. My sincere thanks to Professor Syed Shahadat Hossain for his continuous encouragement and helpful suggestions to improve this work. Special thanks to my colleagues, Mr. Nabil Awan, Mr. Shamsul Alam, Dr. Anower Hossain, and my former students, Mr. Jahidur Rahman Khan and Mr. ASM Ferdous Hossain for their support. Thanks to Dr. Rafiquel Islam Chowdhury, for his supportive role.

Last but not the least, I would like to thank my family members, my husband, my only child, and my parents for their patience, support and encouragement that made it possible to complete this journey.

Table of Contents

List of Tables

Abbreviations

GLM	**G***eneralzied* **L***inear* **M***odel*
GEE	**G***eneralized* **E***stimating* **E***quation*
ALR	**A***lternating* **L***ogistic* **R***egression*
GLMM	**G***eneralized* **L***inear* **M***ixed* **M***odel*
QL	**Q***uasi* **L***ikelihood*
QLR	**Q***uasi* **L***ikelihood* **R***atio*
MCM	**M***arginal* **C***onditional* **M***odel*

Abstract

Correlated data which are very common in longitudinal studies should be analysed with models and methods that take the correlation into account. Most of the longitudinal models are based on marginal approaches, assuming an induced correlation between successive individuals, often lacking the proper specification of the dependence of binary outcomes. As a result, such models may fail to provide efficient estimation of parameters. Conditional models which is another commonly used approach for the above situation use a transition probability model to capture the dependence of outcome variables. While the selection of a model depends on the question under study, there is no clear directives in the existing literature about when to choose which model. Keeping in mind the limitations of the existing popular methods for analysing longitudinal data, the objectives of this study were set.

The objectives of this study are (i) to examine how well the dependence of repeated response are addressed in selected methods including GEE and ALR, (ii) to propose joint models based on a marginal conditional approach enabling to incorporate the true dependence relationship, using likelihood methods, (iii) to propose a joint model based on a marginal conditional approach under a quasi-likelihood setup appropriate for situation where the distribution of the outcome variables is unknown, (iv) to develop and demonstrate the inferential theories associated with all the proposed models, (v) make comparisons of the proposed models with the existing models and (vi) to illustrate the proposed models with applications to real life data. The proposed models demonstrated under objective (ii) link the marginal and sequence of conditional models to provide the joint model needed for predicting the covariate effects on dependent variable at different time points. In case of more than three repeated measurements,

the regressive model approach was proposed that can be extended for any order of dependence without complicating the theory and keeping the number of parameters of the model for repeated measures minimum. This model has the flexibility such that one can easily add interaction terms among previous outcomes and predictors in the proposed framework if and when required. A number of simulation studies resulted that the proposed methods perform better than GEE and ALR in terms of bias and 95% coverage probability.

The marginal conditional model developed under a quasi-likelihood setup captures the correlations among repeated observations in a built-in nature and unlike GEE or ALR, does not need to have a correlation parameter in the model. This model can be extended for any number of repeated measures without complicating the theory and keeping the number of parameters to a minimum. The simulation studies showed that, when the data are correlated or the distribution of the outcome variables are not identical at different time points, the estimates of this method has less bias than GEE or ALR.

The marginal conditional feature of the proposed models make the models very useful for analysing big data, one can use the existing software for model fitting and risk prediction of a sequence of events. The application using Health and Retirement Study data illustrate the performance of the proposed models and prove the usefulness of such models for longitudinal data.

Chapter 1

Introduction

1.1 Introduction

A longitudinal data set comprises of repeated measurements on each subject under a study and the term 'repeated measurements' refers to a dataset in which the response of each experimental unit or subject under a study is observed on multiple occasions or under multiple conditions over a period of time [18, 23, 38]. Although the response variable in a longitudinal study can be either univariate or multivariate, we restricted our consideration to univariate response variables measured at multiple occasions for each subject.

In a longitudinal study, data collected from the same subject are usually correlated and while modelling and/or analysing such data, this correlation should be taken into account [18, 38, 79]. The association among the repeated measures in a longitudinal data makes the analysis of such data distinctive and challenging and needs to adjust for the association.

Most of the models used for longitudinal data are based on marginal approaches (for example, Generalized Estimating Equations (GEE) by Zeger and Liang [79] or Alternating Logistic Regression (ALR) by Carey et al. [9]) in which the correlation considered among repeated measures on the same individuals are induced correlations. As a result, these models may fail to provide an efficient estimation of parameters due to lack of proper specification of the dependence. Relatively fewer approaches for longitudinal data analysis are based on transition probability based conditional models (for example, Bonney [7], Islam and Chowdhury [35, 36], Muenz and Rubinstein [59], etc.). Conditional models have the advantage that one may find whether or not the changes in a dependent variable are independent of previous observations as well as the independent variable. But conditional models themselves are not adequate to model the dependence.

An extensive literature review proves the existing controversy about the use of marginal and conditional models, particularly in the analysis of longitudinal data. We discussed the advantages and limitations of the marginal models and the conditional models. We regard the conditional model as fundamental and from conditional models, marginal predictions can be made [51].

A thorough review of literature confirms that analysis of a longitudinal data considering the correlation between repeated observations of an individual, it is more logical to consider a joint model where both the marginal and conditional probabilities can be expressed as a function of explanatory variables. So the motivation of this study is to propose appropriate models for analysing longitudinal binary data with time dependent covariates, which takes into account of both marginal and conditional probabilities of correlated binary events such that the joint function can be specified fully by unifying marginal and conditional probabilities. Marginal and conditional probability based joint models are not new [7, 34–36]. But such models were proposed mainly to focus on the char-

acterization problems and have not been employed to focus on the covariate dependent models with dependence in the outcomes, which is one of the main focus of this research.

This study proposed joint models as alternatives of the popular GEE or GEE based approaches for outcome variables with known and unknown distributions of outcome variables in a longitudinal study. The joint models are proposed along with the estimation procedure to study the relationship among the repeated measures of the dependent variable as well as relationship among dependent and independent variables. Related tests were suggested and a number of simulation studies have been performed to compare the proposed method with existing popular methods such as GEE or ALR. The proposed method has been illustrated using real life data and the findings are compared with popular existing methods.

1.2 Background of the Study

Modelling of binary outcomes are common in lifetime data analysis and logistic regression models are common choices in such analyses [14]. For correlated binary data in a longitudinal study, a joint multivariate distribution can be described in terms of the marginal means and correlations as shown by Bahadur [4]. For paired data, the Bahadur [4]'s model is relatively simple [55] and coincides with the joint distribution that is completely determined by the marginal means and pairwise correlations. But the full Bahadur [4]'s model is not routinely implemented for data analysis because of a large number of association parameters that increase the number of measurements per subject; in addition, the association parameters are subject to complex constraints that depend on the marginal means. Using multiple measurements per subject, Lipsitz et al. [55] implemented maximum likelihood method for analysis of the parameters of Bahadur [4]'s model.

Important initial work on longitudinal data and binary longitudinal data were done by Liang and Zeger [52] and Zeger and Liang [79]. Liang and Zeger [52] and Zeger and Liang [79] proposed the GEE models based on probability of the event and correlations or the first and the second moments. Lipsitz et al. [56], Liang et al. [53] and Carey et al. [9] employed the marginal odds ratios instead of correlations between pairs of binary responses. Le Cessie and Van Houwelingen [50] proposed use of different measures of dependence in modelling for logistic regression for correlated binary data. It has been observed, however, that the marginal measures may fail to provide the measure of dependence of binary outcomes due to lack of proper specification of the underlying model.

GEE [52, 79] is an iterative approach that alternates between solving the GEE for regression parameters and updating the estimate of the correlation parameter. GEE has been criticized for its lack of an objective function which complicates the development of objective measure for goodness of fit for this approach. However, the followers of GEE defined the major advantage of GEE that it yields a consistent estimator of the regression parameter even if the working correlation structure that models the pattern of association in the data is not specified correctly with a potential loss in precision of estimation if the assumed and true patterns of association are not close [18, 24]. Several authors identified that selection of an appropriate correlation structure in the GEE models as the limitation of GEE and tried to address the problems [10, 12, 25, 33] by suggesting different ways for selecting an appropriate correlation structure, although, the problem lies elsewhere.

ALR is the algorithm that alternates between a logistic regression using first order GEE to estimate regression coefficients and a logistic regression to estimate the odds ratios [9]. ALR uses a logistic model for an outcome conditional upon another outcome instead of specifying the models by additional marginal

description of the pairwise association.

We note that the consistency in the parameter estimates of GEE or related models, in spite of misspecified correlation structure, essentially indicates that the correlation used in GEE is a nuisance correlation and focus on selecting an appropriate correlation structure will not solve the problem and the problem of analysing the longitudinal data by properly addressing the correlation among the repeated measures needs a different approach than GEE or any marginal model.

Maximum Likelihood method or ML method for time independent covariates was proposed by Zeger et al. [80]. ML approach is suggested as the gold-standard approach in a few studies because of its attractive features, which include consideration of an objective function (the log-likelihood function) that can be used to assess the fit of competing models and construct likelihood ratio tests. However, the ML approach may be less robust to model misspecification than GEE. Guerra et al. [29] suggested an extension of this ML method for time dependent covariates. The method, named MARK1ML method, can also be considered as an extension of multiple measurements per subject of the approaches of McDonald [58] and Lipsitz et al. [55] who considered correlated binary data with two measurements per subject. The likelihood model for this method was based on the assumption of a Markovian model of first order (MARK1 model) so that the value of an outcome on a subject at a particular measurement occasion only depends on the value at the previous measurement occasion. The MARK1 model allows to consider the usual logistic model for a GEE analysis of binary data which has the benefit of easy interpretation of the regression parameters. This approach can be extended for at least second order because Markovian assumption of higher order is not addressed yet for such problems. MARK1ML method assumes that correlation between adjacent measurements on a subject depend on their separation in time and considers

the correlation between adjacent measurements on each subject where different assumptions regarding the adjacent correlations induces different patterns of correlation in the data. Furthermore, this approach considers one additional specification for the adjacent correlations and unstructured form that does not impose a particular pattern on the adjacent correlations within subjects. Like GEE, this approach is also robust to misspecification of the true correlation structure of the data but not to the same degree as is GEE.

Relatively fewer studies have been conducted in the conditional approaches as compared to the marginal models approach. Some of the studies on conditional approach include Muenz and Rubinstein [59], Bonney [7, 8], etc.. However, neither the marginal models nor the conditional models alone are adequate to model longitudinal data.

In a series of works, Islam et al. [34], Islam and Chowdhury [35, 36, 37, 38], Islam et al. [40, 41, 43] employed the covariate dependent conditional logistic models and regressive logistic models under the Markov assumptions to construct a joint model. The works of Bonney [7, 8], Islam and Chowdhury [35, 37], Islam et al. [43] was generalized to include both binary outcomes in previous times as well as covariates in the conditional models proposed by Islam et al. [34, 42]. A mixed effect model (GLMM) could also be an alternative choice to capture the picture since if a random component is taken in the model, the change in an individual over time can be considered in the model, but again the change would be added in the intercept part and the fixed effect part will be unchanged. So if a relationship between the repeated measures of the dependent variable as well as relationship between dependent and independent variables are required to be studied then a mixed model is not an appropriate choice.

1.3 Statement of the Problem

Let us consider an experiment for a specified time period for a sample of size N. In the experiment, we have data from several follow-ups on each of the N units. The repeated measurements on N units in the sample at times $\boldsymbol{T_i} = (t_{i1}, ..., t_{in_i})'$ $\forall\ i = 1, ..., N$ produce repeated outcomes data on the dependent variable with associated covariates $\boldsymbol{X}_{ij} = (X_{ij1}, ..., X_{ijp})'$. Longitudinal data are naturally correlated and the main challenge in analysing such data is to address the correlation or association among the repeated outcomes.

Most of the methods found in literature for analysing repeated measures data are based on marginal models and has a major limitation in the correlation structure considered. The marginal approaches consider an induced correlation structure to take into account the correlation among the repeated responses of each individual. Induced correlation, considered in many popular marginal model based approaches does not fit to the estimation procedure. Moreover, while a marginal model is taken, the correlation among the repeated responses are not properly addressed and hence it may fail to provide efficient estimation of parameters due to lack of proper specification of the dependence of binary outcomes in the model. Although most of these methods are routinely used and known to well represent the population averaged methods for the analysis of longitudinal binary data, as in almost all the cases, the marginal models consider induced correlations between successive individuals, it does not fit to the estimation procedure.

ALR alternates between a logistic regression using first order GEE to estimate regression coefficients and a logistic regression to estimate the odds ratios [9] with the use of a logistic model for an outcome conditional upon another outcome instead of specifying the models by additional marginal description of the pairwise association. Modification of ALR is also found in literature as an alter-

native of GEE using marginal conditional approach which was expected to be a help to get rid of limitations of GEE that uses a marginal model. But while a marginal model is taken, the correlation among repeated measures are not taken care of in a proper sense and hence an efficient estimation of parameters is not always possible due to lack of proper specification of the dependence of binary outcomes in the model.

An extensive literature review unravels the fact that the use of popular GEE, ALR or related methods for analysing longitudinal data is not necessarily logical. Although several researchers tried to address the problem with concentration on selecting an appropriate correlation structure, we would like to point out that the correlation structure used in GEE is a nuisance correlation and the estimates are robust to it. Hence the limitation of GEE does not lie in its correlation structure but the underlying model and an alternative to analyze longitudinal data should be looked for. The problem of addressing correlation among repeated measures in a proper sense was partly addressed by Darlington and Farewell [16] and Guerra et al. [29]. Both explained a joint model by conditional marginal model. Darlington and Farewell [16] pointed out that the relationship between outcome and explanatory variables may also depend on the dependence in outcomes and explanatory variables. According to the two approaches proposed by Darlington and Farewell [16], the models are designed to focus on the dependence of correlation structure on explanatory variables. Guerra et al. [29]'s work has scope to be extended further by using a model that takes into account the dependence patterns more appropriately which could be a model that takes into account the dependence between outcomes in repeated measurements as well as dependence between outcomes and explanatory variables under the framework of a quasi likelihood approach. Darlington and Farewell [16]'s method appear to be close enough to the actual correlation, but their method to address the correlation problem was not

complete. In true sense, Guerra et al. [29] eventually ended up with the same marginal model as the prior researchers in this field.

This study focuses on the fact that to feature the correlation between repeated outcomes of an individual, it is more logical to consider a joint model than a marginal or a conditional model. We propose the use of the joint model based on a marginal conditional approach for repeated binary outcomes that was addressed in a series of works by Islam et al. [34], Islam and Chowdhury [35, 36, 37, 38], Islam et al. [40, 41, 43]. These studies proposed covariate dependent conditional logistic regression models under the Markov assumptions and regressive logistic models to construct a joint model based on both the marginal and conditional probabilities of the correlated binary events such that the joint function can be specified fully by unifying the marginal and the conditional probabilities.

1.4 Objectives of the Study

A major feature of the longitudinal data is the repeated responses among each individual are expected to be correlated that makes the analysis of such data challenging. Keeping in mind the limitations of existing popular methods in addressing the correlation among the repeated measures, the objectives of this study were determined. The specific and detailed objectives of this study are listed below.

- To examine selected popular methods including GEE and ALR in order to figure out how well these methods addressed the dependence among the repeated outcomes;
- to propose joint models based on a marginal conditional approach enabling to incorporate the true dependence relationship using likelihood based methods;

- to develop a joint model based on a marginal conditional approach under a quasi-likelihood setup for longitudinal binary data under the assumption that the distribution of the repeated outcomes are unknown;
- to develop and demonstrate the inferential theories associated with all the proposed models;
- to compare the proposed models based on marginal conditional approach with popular marginal models including GEE and ALR; and finally
- to illustrate the proposed models with applications to real life longitudinal data.

In this study, we propose some joint models based on marginal conditional approaches as alternatives of GEE or related approaches for correlated binary outcomes. A comparison of these joint models with popularly used GEE or ALR is performed. These joint models for bivariate data are extended under the assumption of known distribution of the repeated outcomes of any number of follow ups. The estimation technique of parameters is shown based on likelihood based approaches. A joint model when the distribution of the repeated outcomes are not known is proposed under the set up of a quasi-likelihood method.

We used R software for the data analysis. The R packages, 'bindata', 'geepack' and 'alr' are used for estimating the parameters of GEE and ALR models discussed in this study. The other codes used for the simulation studies and applications are given in the appendix.

1.5 Organization of the Study

In this thesis, Chapter 1 is the introductory chapter, that includes a brief discussion on the background of this study with a literature review followed by an introduction to structure of repeated measures data to detail the statement of the problem along with the objectives and organization of the study. Chapter 2 contains a detailed literature review which describes the platform of this research work. The literature is described in a sequence in which the previous models for analysing longitudinal data were developed. Chapter 3 provides a detailed description of selected marginal models and quasi-likelihood and working likelihood approaches for analysing those models. In Chapter 4, the discussion in Chapter 3 is extended to a point-to-point identification of problems to describe the shortcomings of the earlier methods. In Chapter 5, the idea of conditional models are reviewed and the proposed joint models using a marginal conditional approach are described with related inferential procedures. A generalization of the proposed model is also shown. A number of simulation studies were performed followed by an example with a real life data. The results are presented in Chapter 6 . In Chapter 7, we proposed a new model that we developed under the setup of a quasi-likelihood approach. The inferential procedures are also described. The results of a set of simulation studies are reported to show the performance of the new model as compared to that of GEE and ALR. The new model was explained using a real life data. Finally in Chapter 8, a summary of findings of this dissertation as well as some discussions on directions for future work are placed. Last but not the least, Appendix A contains the full description of all codes required to run the models used in this study.

Chapter 2

Literature Review on Generalized Linear Models

2.1 Introduction

The development of methods for analysing longitudinal data or repeated measures categorical data has received substantial attention in the last few decades and has become an important and active area of research. A Generalized linear model (GLM) is a common choice for modeling repeated measures data with categorical response variables. In this chapter, the structure of data with repeated outcomes is described in the form of a table. A GLM for analysing repeated measures data is described in details stating the model assumptions. The GLM and the associated parameters and estimates of the parameters of a GLM using quasi-likelihood and likelihood based approaches are also discussed with a detailed literature review on analysis of repeated measures data.

2.2 Structure of Data with Repeated Outcomes

In the study of repeated measures data, several notations are required for the description of the data and related methodologies. Naturally, the notations used by different authors are not exactly the same. So the basic notations used in this study are defined here and further notations that would be required to explain the methodologies in following chapters will be defined accordingly.

Let us consider an experiment for a specified time period for a sample of size N. In the experiment, we have data from several follow-ups on each of N units. Suppose each of N units are observed on n_i occasions. Then the N units in the sample observed at n_i occasions produce data on the dependent or outcome variable. Let us define the outcome vector $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, ..., Y_{in_i})'$ and associated covariates $\boldsymbol{X}_{ij} = (X_{ij1}, ..., X_{ijp})'$, observed at times $\boldsymbol{T_i} = (t_{i1}, ..., t_{in_i})' \; \forall \; i = 1, ..., N$. The data structure can be shown as given in Table 2.1.

Table 2.1: Structure of Longitudinal Data

i	T_i	$\boldsymbol{Y}$	$\boldsymbol{X}_1$	$\boldsymbol{X}_2$	...	$\boldsymbol{X_p}$
	t_{11}	Y_{11}	X_{111}	X_{112}	...	X_{11p}
1	...	...	...	...	...	...
	t_{1n_1}	Y_{1n_1}	X_{1n_11}	X_{1n_12}	...	X_{1n_1p}
	t_{21}	Y_{21}	X_{211}	X_{212}	...	X_{21p}
2	...	...	...	...	...	
	t_{2n_2}	Y_{2n_2}	X_{2n_21}	X_{2n_22}	...	X_{2n_2p}
	...	...	...	...	...	
...	...	...	...	...	...	
	...	...	...	...	...	
	t_{N1}	Y_{N1}	X_{N11}	X_{N12}	.	X_{N1p}
N	...	...	...	...	...	...
	t_{Nn_N}	Y_{Nn_N}	X_{Nn_N1}	X_{Nn_N2}	...	X_{Nn_Np}

2.3 GLMs for analysing Repeated Measures Data

A GLM investigates the relationship between a response variable and one or more predictors.

2.3.1 Components of a GLM

A GLM has three components

- Random Component: The random component of a GLM is the probability distribution of the response variable Y_{ij}; For example, in a linear regression, Y_{ij} has a Normal distribution, or Y_{ij} has a binomial distribution in a binary logistic regression. The random component of a GLM is also known as a noise model or an error model.

- Systematic Component: A systematic component is the linear combination of the independent variables $(X_{ij1}, X_{ij2}, ..., X_{ijp})'$ in the model in creating the linear predictor. For example, for a linear regression with two explanatory variables X_{ij1} and X_{ij2} and vector of unknown parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ where β_0 is the intercept and β_1 and β_2 are coefficients of X_{ij1} and X_{ij2} respectively, a systematic component can be expressed as $\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2}$.

- Link Function: The function that describes the relationship between the expected value of the response variable and the systematic component is called the link function of GLM. A link function is often denoted by η or $g(\mu)$ and can be expressed as $\eta_{ij} = g(\mu_{ij}) = g(E(Y_{ij}))$. For linear regression, $\eta_{ij} = E(Y_{ij})$, for logistic regression, $\eta_{ij} = logit(\pi_{ij})$ (where πij is the probability of success of Y_{ij}), etc.

2.3.2 Model

As defined in section 2.2, the vector of outcome for the ith subject can be defined as $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, ..., Y_{n_i})'$ and $E(\mathbf{Y}_i) = \boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, ..., \mu_{in_i})'$. For the ith subject at jth follow-up, Y_{ij}, the GLM can be expressed as

$$f(y_{ij}, \theta_{ij}, \phi_{ij}) = exp\left[\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi_{ij})} + c(y_{ij}, \phi_{ij})\right], \tag{2.1}$$

where θ_{ij} is the natural parameter, $b(\theta_{ij})$ is a function of θ and $a(\phi_{ij})$ is the dispersion parameter. We may define

$$\begin{aligned} \theta_{ij} &= g(\mu_{ij}) = \boldsymbol{X}_{ij}\boldsymbol{\beta} = \beta_0 + \beta_1 X_{ij1} + ... + \beta_p X_{ijp}, \\ b'(\theta_{ij}) &= E(Y_{ij}) = \mu_{ij}, \\ b''(\theta_{ij}) &= V(\mu_{ij}) \text{ and} \\ V(Y_{ij}) &= a(\phi_{ij})b''(\theta_{ij}) = a(\phi_{ij})V(\mu_{ij}). \end{aligned}$$

2.3.3 Likelihood Function and Maximum Likelihood Estimates

The contribution of Y_{ij} to the likelihood function can be expressed as

$$L_{ij}(\theta_{ij}, \phi_{ij}, y_{ij}) = exp\left[\left(\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi_{ij})} + c(y_{ij}, \theta_{ij})\right)\right]. \tag{2.2}$$

The contribution of Y_{ij} to the log-likelihood function can be shown as

$$l_{ij} = \left[\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi_{ij})} + c(y_{ij}, \theta_{ij})\right]. \tag{2.3}$$

The log-likelihood function is

$$l(\theta, \phi, y) = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi_{ij})} + c(y_{ij}, \theta_{ij})\right]. \tag{2.4}$$

Using the Chain Rule, the estimating equations are

$$\begin{aligned} \frac{\delta l}{\delta \beta_j} = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left(\frac{\delta l_{ij}}{\delta \theta_{ij}}\frac{\delta \theta_{ij}}{\delta \mu_{ij}}\right).\left(\frac{\delta \mu_{ij}}{\delta \eta_{ij}}\frac{\delta \eta_{ij}}{\delta \beta_j}\right) &= 0 \\ \text{or, } \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left(\frac{y_{ij} - \mu_{ij}}{a(\phi_{ij}).V(\mu_{ij})}.\frac{\delta \mu_{ij}}{\delta \beta_j}\right) &= 0. \end{aligned} \tag{2.5}$$

The solution to these equations gives the maximum likelihood estimates of $\boldsymbol{\beta}$. It is well known that, maximum likelihood methods are used for fitting models under the assumptions that there is a known probability model for the data.

Knowledge of the physical process that lead to the data or substantial experience with similar data from previous studies is required for that.

Sometimes there is insufficient information about the data to specify a model, but some of the features of the data can be specified. For example, the type of the outcome variable (continuous or discrete), dependence among the response variables, etc. can be known. In such cases, maximum likelihood methods cannot be used and analytical methods are based on approximation to the likelihoods.

2.3.4 Quasi-likelihood Methods

Quasi-likelihood method was introduced by Wedderburn [77] as an approximation to the likelihoods in cases where maximum likelihood methods cannot be used. Let us consider a vector of responses,

$$\boldsymbol{Y} = (Y_{11}, ..., Y_{1n_1}, ..., Y_{N1}, ..., Y_{N_{n_N}})',$$

which are independent with mean

$$\boldsymbol{\mu} = (\mu_{11}, ..., \mu_{1n_1}, ..., \mu_{N1}, ..., \mu_{Nn_n})'.$$

The response vector for subject i can be shown as $\boldsymbol{Y_i} = (Y_{i1}, Y_{i2}, ..., Y_{in_i})'$ with mean vector $\boldsymbol{\mu_i} = (\mu_{i1}, \mu_{i2}, ..., \mu_{in_i})'$. We assume that μ_{ij} $i = 1, 2, ..., N$, $j = 1, 2, ..., n_i$ is a function of covariates, $\boldsymbol{X}_{ij}$, and some regression parameters $\boldsymbol{\beta}$ and covariance matrix of $\boldsymbol{Y}_i$ is $\sigma^2 V(\boldsymbol{\mu_i})$. We also assume that the form of the random components are not known, but μ_{ij} and $V(\mu_{ij})$ are known. Since the components of Y_{ij} are assumed to be independent, $V(\boldsymbol{\mu}_i)$ must be diagonal.

To construct the quasi-likelihood, one may start with a single component Y_{ij} of $\boldsymbol{Y}$. The function,

$$Q_{ij} = Q(\mu_{ij}|y_{ij}) = \int_{y_{ij}}^{\mu_{ij}} \frac{y_{ij} - t}{V(t)} dt, \tag{2.6}$$

behaves like a log-likelihood function and this is referred to as a quasi-likelihood function [77]. Under the conditions given at the beginning of this section, the first derivative of the function Q_{ij} has several properties in common with the log-likelihood derivative (i.e. the score). In particular,

- $E(Q_{ij}) = 0$
- $V(Q_{ij}) = \frac{1}{\sigma^2 V(\mu_{ij})}$
- $-E(\frac{\delta^2 Q_{ij}}{\delta \mu_{ij}^2}) = \frac{1}{\sigma^2 V(\mu_{ij})}$

Most of the first order asymptotic theory concerned with the likelihood are founded on these properties [77]. Since the components of $\boldsymbol{Y}$ are independent by assumption, we may define the quasi-likelihood for the complete data as the sum of the individual contributions

$$Q = \sum_{i=1}^{N} \sum_{j=1}^{n_i} Q_{ij} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} Q(\mu_{ij}|y_{ij}) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \int_{y_{ij}}^{\mu_{ij}} \frac{y_{ij} - t}{V(t)} dt. \tag{2.7}$$

Quasi-estimating Equations

Then the quasi-estimating equations are obtained by differentiating the Quasi-likelihood function (2.7) with respect to the respective parameters and equating to zero as

$$S_\beta = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{\delta Q(\mu_{ij}|y_{ij})}{\delta \beta_k} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{y_{ij} - \mu_{ij}}{\sigma^2 V(\mu_{ij})} . \frac{\delta \mu_{ij}}{\delta \beta_k} = 0, k = 1, 2, ..., p, \tag{2.8}$$

which is an extension of GLM when μ_{ij} and $V(\mu_{ij})$ are needed to be known for obtaining estimating equations for any link function $g(\mu_{ij}) = \boldsymbol{X}_{ij}\boldsymbol{\beta}$.

2.4 A Review on Models for Longitudinal Data

GLMs were extended in different ways to model longitudinal data including marginal or population averaged models and transition or response conditional models. Three broad classes of regression models for longitudinal data include

(i) marginal or population averaged models, (ii) random-effects or mixed effects or subject-specific models, (iii) transition or response conditional models. These models differ in how the correlation among the repeated measures is accounted for. Besides, these models have regression parameters with distinctly different interpretations reflecting the different targets of inference of such models. We present a detailed review on marginal and conditional models in the following sections.

2.4.1 Marginal Models

The term marginal emphasizes that the model for the mean response at each occasion depends only on the covariates of interest and not on any random effects or previous responses. A marginal model is a straightforward way to extend GLMs to longitudinal data to model the mean response at each occasion using an appropriate link function [22]. The focus of a marginal model is on marginal mean and its dependence on the covariates. As a result, marginal models require only a regression model for the mean response and the full distributional assumptions for the vector of repeated responses is not necessarily known. We focus much of this review on marginal models for longitudinal binary data.
A marginal model for longitudinal data has three-parts

- The mean of each response $E(Y_{ij}|\boldsymbol{X}_{ij}) = \mu_{ij}$ is assumed to depend on the covariates through a known link function $g(\mu_{ij}) = \boldsymbol{X}'_{ij}\boldsymbol{\beta}$.
- The variance of each Y_{ij}, given the covariates is assumed to depend on the mean according to $V(Y_{ij}|\boldsymbol{X}_{ij}) = a(\phi_{ij})V(\mu_{ij})$, where $V(\mu_{ij})$ is a known variance function and $a(\phi_{ij})$ is a scale parameter that may be known or need to be estimated.

- The conditional within-subject association among the vector of repeated responses, given the covariates, is assumed to be a function of an additional set of association parameters , α and may also depend upon the means, μ_{ij}.

The extension of GLMs to longitudinal data is clear through this three-part specification of a marginal model. The first two parts of the marginal model correspond to the standard GLM, although with no explicit distributional assumptions about the responses. The main extension of GLMs to longitudinal data is represented by the third component, the incorporation of a model for the within-subject association among the repeated responses from the same individual.

In a marginal model, the mean response and within-subject association are modeled separately and this separation of the modelling of the mean response and the association among responses has important implications for interpretation of the regression parameters β. In marginal models, the regression parameters have population-averaged interpretations.

The development of marginal models for discrete longitudinal data has its origins in likelihood-based approaches. For a $n_i \times 1$ vector of responses for ith individual, $\mathbf{Y}_i = (Y_{i1}, ..., Y_{in_i})'$, the three-part specification given above is extended by making full distributional assumptions about the response vector. At least three main research threads can be distinguished in the development of likelihood based marginal models for discrete longitudinal data [24].

One of the earliest likelihood based approach is a latent variable model proposed by Gumbel [30] for multivariate binary data. This approach considers a vector of unobserved latent variables, say, $L_{i1}, ..., L_{in_i}$, where each of these is related to the observed binary responses via $Y_{ij} = 1$ when $L_{ij} \leq X'_{ij}\beta$, $Y_{ij} = 0$ when $L_{ij} > X_{ij}\beta$, β is the $p \times 1$ vector of unknown parameters. Assuming a multivariate joint distribution for $L_{i1}, ..., L_{in_i}$, identifies the joint distri-

bution for $Y_{i1},...,Y_{in_i}$ with $P(Y_{i1}=1,Y_{i2}=1,...,Y_{in_i}=1)=P(L_{i1}\leq X'_{i1}\beta,L_{i2}\leq X'_{i2}\beta,...,L_{in_i}\leq X'_{in_i}\beta)=F(X'_{i1}\beta,X'_{i2}\beta,...,X_{in_i}\beta)$ where $F(.)$ denotes the joint cumulative distribution function of the latent variables. Any dependence among the L_{ij} induces dependence among the Y_{ij}'s.

Another likelihood based approach was proposed by Bahadur [4] where an expansion for an arbitrary probability mass function for a vector of responses $Y_{i1},...,Y_{in_i}$ was proposed. The expansion for repeated binary responses is of the form

$$\begin{aligned} f(Y_{i1},Y_{i2},...,y_{in_i}) &= \left(\prod_{j=1}^{n_i}\pi_{ij}^{y_{ij}}(1-\pi_{ij})^{1-y_{ij}}\right)\times\left(1+\sum_{j<k}\rho_{ijk}z_{ij}z_{ik}\right. \\ &\quad \left.+\sum_{j<k<l}\rho_{ijkl}z_{ij}z_{ik}z_{il}+...+\rho_{i1...n_i}z_{i1}z_{ik}...z_{in_i}\right), \\ \text{where} \qquad Z_{ij} &= \frac{Y_{ij}-\pi_{ij}}{\sqrt{\pi_{ij}(1-\pi_{ij})}} \\ \pi_{ij} &= E(Y_{ij}) \\ \text{and} \qquad \rho_{ijk} &= E(Z_{ij}.Z_{ik}),...,\rho_{i1...n_i}=E(Z_{i1}.Z_{i2}...Z_{in_i}). \end{aligned}$$

Here ρ_{ijk} is the pairwise or second order correlation and additional parameters are related to the third and higher order correlations among the responses. The Bahadur [4] expansion is reproducible or upwardly compatible for any subset of the vector of responses, the same model holds. In addition, given the model parameters, the multinomial probabilities for the vector of binary responses are relatively straightforward to obtain. Altham [1], Kupper and Haseman [49] discussed applications of this model, with very simple pairwise correlation structure with an assumption that higher-order terms are zero.

The major limitation of the Bahadur [4] expansion is its parameterization of the higher-order associations in terms of correlation parameters that has limited its application to longitudinal data. For discrete data, the Bahadur [4] model requires a complicated set of inequality constraints on the model parameters that

make maximization of the likelihood very difficult. As a result, the model has not been widely applied to longitudinal data except in very simple settings with a small number of repeated measures. Because of the restrictions on the correlations, alternative multinomial models for the joint distribution of the vector of discrete responses have recently been proposed where the within-subject association is parameterized in terms of other metrics of association, for example, marginal odds ratio [27, 53, 55, 57]. In virtually all of the later advances, the application of the methodology has been hampered by at least three main factors [22]. First, there are no simple expressions for the joint probabilities in terms of the model parameters that makes maximization of the likelihood difficult to some extent. Second, these models are difficult to fit except when the number of repeated measures is relatively small and finally, many of these models are not robust to misspecification of the higher-order moments. That is, many of the likelihood-based methods require that the entire joint distribution be correctly specified. As a result, if the marginal model for the mean responses has been correctly specified but the model for any of the higher-order moments has not, then the maximum likelihood estimators of the marginal mean parameters will fail to converge in probability to the true mean parameters.

analysing discrete longitudinal data using marginal models had a remarkable advancement with the introduction of the generalized estimating equations (GEE) approach by Liang and Zeger [52] and Zeger and Liang [79]. The GEE approach is a natural extension of the quasi-likelihood approach [77] for GLM to the multivariate response setting, with an additional set of nuisance parameters incorporated to accommodate the within-subject association. The GEE methodology generated a lot of theoretical and applied research on the use of this methodology for analysing longitudinal data, e.g., to improve upon efficiency, Prentice [66] proposed joint estimating equations for both the main regression parameters, β, and the nuisance association parameters, α.

There is overwhelming use of correlated binary data since work on repeated measures data by Liang and Zeger [52], Zeger and Liang [79] has been published on GEE. However, GEE models are proposed based on probability of the event and correlations or the first and second moments by Liang and Zeger [52], Liang et al. [53], Lipsitz et al. [56], Prentice [66]. Although the GEE approach yields a consistent estimator of β under misspecification of the within subject associations, GEE, by nature, provides very limited information about the longitudinal associations within the repeated outcomes themselves. This is disorganized, since previous outcome is frequently an important predictor of future outcome for many studies. Carey et al. [9] developed models based on marginal odds ratios instead of correlations between pairs of binary responses to overcome the limitations of GEE.

GEE and related methods are based on marginal models using a Quasi-likelihood approach and the marginal models may fail to provide the measure of dependence of binary outcomes due to lack of proper specification of the underlying model. Although GEE and other approaches based on GEE consider correlation between the repeated outcomes, the correlation considered are induced correlations and anomalies caused by the induced correlation between repeated outcomes is beyond any explanation [38].

Many of the earlier studies tried to address correlation in a marginal model approach in a variety of ways (e.g. Darlington and Farewell [16], Guerra et al. [29], Zeger et al. [80], etc.) using Markov based transition probabilities. Zeger et al. [80] examined a robust likelihood estimation method for estimating the relationship between covariates and the transition matrix in the discrete time setting. Darlington and Farewell [16] examined modelling the correlation between the binary outcomes as a function of time dependent covariates also in the discrete time settings. The binary models of Zeger et al. [80] as well as Darlington and Farewell [16], Guerra et al. [29] also focused on marginal rela-

tionships using Markov transition probability.

The marginal measures may fail to provide the measure of dependence of binary outcomes due to lack of proper specification of the underlying model. Although GEE and other approaches based on GEE considers correlation between the repeated outcomes, the quasi-likelihood approaches are based on independence assumption and an induced correlation among the repeated measures may cause anomalies which is beyond any explanation. More over, the induced correlation which is, basically, a nuisance correlation, even if misspecified does not affect the parameter estimates. In Chapter 3 and Chapter 4, we tried to detail these features of GEE and ALR and some other marginal models which used transition probabilities but ended up with marginal parameter estimations.

2.4.2 Generalized Linear Mixed Models

A generalized linear mixed model incorporates the within subject variation in the linear model where random effects are attributable to within subject variation are incorporated. The generalized linear model for repeated measures can be expressed as

$$g(\mu_{ij}) = \boldsymbol{X}_{ij}\boldsymbol{\beta},\ i = 1,2,...,N;\ \ j = 1,2,...,n_i,$$

with $E(Y_{ij}|\boldsymbol{X}_{ij}) = \mu_{ij}$ and $V(Y_{ij}) = a(\phi)V(\mu_{ij})$ [38]. Then considering a random effect, u_i, for the repeated outcomes of the ith subject, the extended model can be written as

$$g(\mu_{ij}) = \boldsymbol{X}_{ij}\boldsymbol{\beta} + Z_i u_i, i = 1,2,...,N;\ \ j = 1,2,...,n_i,$$

where u_i follows multivariate normal distribution with mean 0, and variance covariance matrix Σ. Instead of normality assumptions, other assumptions may be considered depending on the type of the data.

2.4.3 Conditional Models

The transition models are the models, in which the conditional distribution of each response is expressed as an explicit function of the earlier responses and the covariates [22]. These models can be considered as conditional models as they model the conditional distribution of the response at any time point given the responses at previous time points and the covariates. The conditional models assume that the dependence among the repeated measures are due to influence of the past values of the response on the present observations. These models are attractive because of the sequential nature of longitudinal data and GLMs can be extended to model longitudinal data by modeling the mean and time dependence simultaneously through conditioning an outcome on previous outcome or on a subset of all previous outcomes by a transition or Markov model [22].

Number of previous studies show the use of Markov chains to model equally spaced discrete longitudinal data with a finite number of states or categories [2, 5, 13]. Cox [15], Korn and Whittemore [48], Zeger et al. [80] discuss applicability of transition models to longitudinal data.

An example of the simplest conditional model for longitudinal data can be first-order Markov chain in which the transition probabilities are assumed to be the same for each time interval. The most appealing aspect of the transition models is that the joint distribution of the vector of responses can be expressed as the product of a sequence of conditional distributions, that is,

$$f(y_{i1},...,y_{ij};\boldsymbol{\beta}) = \prod_{j=1}^{n_i} f(y_{ij}|y_{i1},...,y_{i,n_i};\boldsymbol{\beta}).$$

Markov models for binary longitudinal data have been explored by Darlington and Farewell [16], Guerra et al. [29], Tuma et al. [71], Zeger et al. [80], Zeger and Qaqish [81] as well as several others. Tuma et al. [71] examined modelling the two category instantaneous transition matrix corresponding to the

binary outcome as a simple function of covariates, in a manner similar to that of Kalbfleisch and Lawless [44], the first unified approach for applying Markov models to continuous time categorical panel data with time independent covariates. Zeger et al. [80] examined a robust likelihood estimation method for estimating the relationship between covariates and the transition matrix in the discrete time setting. Zeger and Qaqish [81] explored the use of higher order Markov process for modelling the relationship between a binary outcome and its covariates and prior outcome history in the discrete time setting. Darlington and Farewell [16] examined modelling the correlation between the binary outcomes as a function of time dependent covariates also in the discrete time settings. The binary models of Zeger et al. [80] as well as Guerra et al. [29] focused on marginal relationships while the models of Tuma et al. [71] and the models of Zeger and Qaqish [81] focused more on predictive relationships. Azzalini [3], Bonney [7, 8], Muenz and Rubinstein [59] employed the conditional logistic models and regressive models under the Markov assumptions.

In a series of works, Islam et al. [34], Islam and Chowdhury [35, 36, 37, 38], Islam et al. [40, 41, 43] employed the conditional covariate dependent conditional logistic models and regressive logistic models under the Markov assumptions. The works of Bonney [7, 8], Islam and Chowdhury [35, 37], Islam et al. [43] was generalized to include both binary outcomes in previous times as well as covariates in the conditional models [34, 42]. In Chapter 5 we discussed the joint model for binary outcomes based on marginal conditional model [43] and proposed it as a better alternative to GEE or related models for longitudinal data in many situations. We also proposed an extension of the joint model for bivariate outcomes to joint models for outcomes from exponential families with any number of follow-ups. A generalized form of the joint model using the regressive model approach [7] is also proposed for cases where there are more

than three follow-ups. A quasi-likelihood approach based marginal conditional model is developed in Chapter 7 for response variable with unknown distributions.

2.5 Conclusion

In this chapter we described the structure of repeated measures data and generalized linear models for analysing such data. A thorough literature review on earlier works on analysis of repeated measures data are discussed with concentration on marginal and conditional models. We also introduced the generalized linear mixed models for longitudinal data but avoided details on it as this model is beyond the concern of this study. In the next chapter, we discussed some methods based on marginal models for longitudinal data.

Chapter 3

Marginal Models for Repeated Measures Data

3.1 Introduction

For analysis of longitudinal binary data, most of the previous works on correlated outcome variables were based on the marginal models or marginal response probabilities. In this chapter, we discuss in details, the marginal approaches, GEE, ALR and methods proposed by Zeger et al. [80], Darlington and Farewell [16], Guerra et al. [29]. The assumptions, likelihood functions, score and information matrix for each of the models are discussed so that the methods can be examined thoroughly.

3.2 Structure of the Longitudinal Data

As described in section 2.2 of Chapter 2, consider a longitudinal study for a specified time period for a sample of size N. In the study, we have data from several follow-ups on each of N units. The N units in the sample produce data on the outcome variable, Y (Table 2.1) for the data structure. Let individual i is observed on n_i occasions taking value either 0 or 1 so that Y_{ij} be a binary outcome variable observed for individual i at time j, $i = 1,2,...,N$ and $j = 1,2,...,n_i$ and also let $\boldsymbol{X}_{ij}$ be the $p \times 1$ vector of covariates for Y_{ij}. Then the outcome vector for ith subject is $\boldsymbol{Y}_i = (Y_{i1}\ Y_{i2}\ ...\ Y_{in_i})'$.

3.3 Models for Repeated Binary Data

The repeated measures data are naturally correlated and the major challenge of the methods for analysing repeated measures categorical data is to model the probable correlations among the repeated outcomes on the same subject. The development of methods for analysis of repeated measures categorical data has received substantial attention and has become an important and active area of research in the last few decades and GLMs are commonly chosen for modelling repeated measures data with categorical response variables (Chapter 2). In the following sections, we examine selected methods based on quasi-likelihood approaches followed by selected working likelihood approaches for marginal models.

3.4 Quasi-likelihood Approaches

Among the popular quasi-likelihood based approaches, GEE and ALR models are discussed here.

3.4.1 Generalized Estimating Equations (GEE)

Generalized Estimating Equations (GEE) [52, 79] is a popular and one of the most widely used method for analysis of longitudinal data which is a quasi-likelihood approach that uses a marginal or population averaged model [9, 19, 24, 31, 66]. Generalized Estimating Equations(GEE) extend generalized linear models to accommodate correlated Y_{ij}'s obtained through longitudinal studies or from repeated measures on same individuals [52, 79]. GEE do not require to meet the classical assumptions of independence and normality, which are too restrictive for many problems [66] and needs only the correct specification of the form of the mean function μ_i, of the vector of responses for each individual.

Assumptions of GEE

- The responses, $Y_{i1}, Y_{i2}, ..., Y_{in_i}$ are correlated or clustered, i.e., cases are not independent.
- The homogeneity of variance does not need to be satisfied.
- Errors are correlated.
- It uses quasi-likelihood estimation rather than maximum likelihood estimation(MLE) or ordinary least squares(OLS) to estimate the parameters, but at times these will coincide.
- Covariance specification: There are typically four or more correlation structures that we assume apriori.

Mean and Variance of the Response Vector

Let Y_{ij} be a binary response observed for individual i at time j, $i = 1, 2, ..., N$ and $j = 1, 2, ..., n_i$ and also let $\boldsymbol{X}_i$ be the vector of parameters for individual i. Then for the ith individual we have a $n_i \times 1$ random vector of Bernoulli responses $\boldsymbol{Y}_i =$

$(Y_{i1}, Y_{i2}, ..., Y_{in_i})'$, and associated covariates of Y_{ij} are $\boldsymbol{X}_{ij} = (X_{ij1}, X_{ij2}, ..., X_{ijk})'$, $i = 1, 2, ..., N$, $j = 1, 2, ..., n_i$. The mean vector for ith individual is

$$\boldsymbol{\mu}_i = (\mu_{i1}\ \mu_{i2}\ ...\ \mu_{in_i})' = [E(Y_{i1})\ E(Y_{i2})\ ...\ E(Y_{in_i})]' = (p_{i1}\ p_{i2}\ ...\ p_{in_i})',$$

where, the probability of observing the event for ith individual at jth occasion, $p_{ij} = Pr(Y_{ij} = 1 | \boldsymbol{X}_{ij} = \boldsymbol{x}_{ij})$; $i = 1, 2, ..., N$; $j = 1, 2, ..., n_i$. For Binary outcome variables,

$$\mu_{ij} = p_{ij} = Pr(Y_{ij} = 1 | \boldsymbol{X}_{ij} = \boldsymbol{x}_{ij}) = \frac{\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})}. \tag{3.1}$$

The probability of not observing the event for ith individual at jth occasion is $q_{ij} = 1 - p_{ij}$, $i = 1, 2, ..., N$; $j = 1, 2, ..., n_i$ and the variance of Y_{ij} is $p_{ij}(1 - p_{ij})$. The variance covariance matrix of Y_i is given by

$$V_i = V(Y_i) = \begin{pmatrix} V(Y_{i1}) & cov(Y_{i1}, Y_{i2}) & ... & cov(Y_{i1}, Y_{in_i}) \\ cov(Y_{i2}, Y_{i1}) & V(Y_{i2}) & ... & cov(Y_{i2}, Y_{in_i}) \\ ... & ... & ... & ... \\ cov(Y_{in_i}, Y_{i1}) & cov(Y_{in_i}, Y_{i2}) & ... & V(Y_{in_i}) \end{pmatrix}.$$

Working Correlation and the Estimating Equations

In addition to the mean and covariance of the vector of responses, Liang and Zeger [52] suggested to take a $n_i \times n_i$ working correlation matrix for each $\boldsymbol{Y}_i$. The correlation matrix (denoted by $R_i(\alpha)$) is working in the sense that it is an approximation to the actual correlation matrix of $\boldsymbol{Y}_i$. It is assumed that $\boldsymbol{R}_i(\alpha)$ is fully specified by the vectors of unknown parameters α that is same for all subjects.

Following the quasi-likelihood approach, with a mean model, $\boldsymbol{\mu}_i$, and variance structure, $V(\boldsymbol{Y}_i)$, Liang and Zeger [52] expressed the GEE for $\boldsymbol{\beta}$ of the form

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{N} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = 0, \tag{3.2}$$

where, $\boldsymbol{D}_i = \frac{\delta \boldsymbol{\mu}_i}{\delta \boldsymbol{\beta}'}$ and $\boldsymbol{V}_i$ is a working or approximate covariance matrix of $\boldsymbol{Y}_i$, chosen by the investigator. In addition to the mean and covariance of the vector of responses, Liang and Zeger [52] suggested to take a $n_i \times n_i$ working correlation matrix for each $\boldsymbol{Y}_i$. The correlation matrix [denoted by $\boldsymbol{R}_i(\alpha)$] is working in the sense that it is an approximation to the actual correlation matrix of $\boldsymbol{Y}_i$. It is assumed that $\boldsymbol{R}_i(\alpha)$ is fully specified by the vectors of unknown parameters α that is same for all subjects.

The parameter estimates are obtained by solving $U(\boldsymbol{\beta}) = 0$ and are typically obtained via the Newton Raphson algorithm. The variance structure is chosen to improve the efficiency of the parameter estimates. The Hessian of the solution to the GEEs in the parameter space can be used to calculate robust standard error estimates. The term "variance structure" refers to the algebraic form of the covariance matrix between outcomes of ith individual, $\boldsymbol{Y}_{ij}, j = 1, 2, \cdots n_i$, in the sample. Examples of variance structure specifications include independence, exchangeable, autoregressive, stationary m-dependent, and unstructured. This working covariance matrix can be expressed in the form

$$\boldsymbol{V}_i = \boldsymbol{A}_i^{\frac{1}{2}} \boldsymbol{R}_i(\alpha) \boldsymbol{A}_i^{\frac{1}{2}},$$

where $\boldsymbol{A}_i = diag[V(Y_{i1}), ...V(Y_{in_i})]$, is a $n_i \times n_i$ diagonal matrix with $V(Y_{ij}) = \phi V(\mu_{ij})$, ϕ is a dispersion parameter, $\boldsymbol{R}_i(\alpha) = corr(\boldsymbol{Y}_i)$ is a $n_i \times n_i$ working correlation matrix and α represents a vector of parameters associated with a specified model for $corr(\boldsymbol{Y}_i)$. Here the form of the estimating equation is similar to the quasi-likelihood estimating equations described in McCullagh and Nelder [57]. With a binary response vector, these equations simply generalize the ordinary logistic regression estimating equations by introducing a working or approximate correlation matrix, $R_i(\alpha)$. This leads to estimating equations of the form

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{A}_i \boldsymbol{V}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i). \tag{3.3}$$

The followings are some common specifications for $corr(Y_i)$ in GEE approach.

a. Identity matrix

$\boldsymbol{R}_i(\alpha) = \boldsymbol{I}$, where $\boldsymbol{I}$ is simply a $n_i \times n_i$ identity matrix. This corresponds to the working independence assumption, and gives estimating equations identical to ML method.

b. Exchangeable Correlation or Compound Symmetry

In many longitudinal studies, the correlation within the responses of an individual remains same for change in time. Mathematically, for *i*th individual, $corr(Y_{ij}, Y_{ik}) = \alpha; j \neq k$. So the correlation matrix for *i*th individual is defined as $R_i(\alpha) = corr(Y_{ij}, Y_{ik}) = \alpha; j \neq k$.

c. Autoregressive correlation

Sometimes, in longitudinal studies, the correlation within the responses of an individual follows autoregressive property. That is for *i*th individual, the correlation within the responses can generally be defined as $corr_(Y_{ij}, Y_{ik}) = \alpha^{|j-k|}; j \neq k$. Here α is a correlation value and thus a fraction. So in this types of correlation, we consider that for all $k > t$, $\alpha^{|j-k|} > \alpha^{|j-t|}$. Then the correlation matrix can be defined as $(\boldsymbol{R}_i(\alpha))_{jk} = corr(Y_{ij}, Y_{ik}) = \alpha^{|j-k|}; j \neq k$.

d. Unstructured or Pairwise correlation

If the correlation matrix is totally unspecified then α forms a vector of order $\frac{n_i(n_i-1)}{2}$ that considers all the pairwise correlations within the repeated responses of the same individual. Here we can define the pairwise correlations as $corr(Y_{ij}, Y_{ik}) = \alpha_{jk}$; $j \neq k$. The correlation matrix can be written as $(\boldsymbol{R}_i(\alpha))_{jk} = corr(Y_{ij}, Y_{ik}) = \alpha_{jk}; j \neq k$ where, $\alpha_{j,j+1} = \alpha_{j+1,j}; j = 1, 2, .., n_i$; Thus α being a

$\frac{n_i(n_i-1)}{2} \times 1$ vector contains all the pairwise correlations.

Many other correlation structures can be considered and α can also depend on subject-specific covariates. Thus the specification of $R_i(\alpha)$ can be expressed more generally as $h(R_i) = Z_i\alpha$, where Z_i is a set of subject specific covariates and $h(R_i)$ is some suitable link function (e.g., inverse hyperbolic). Alternatively, Z_i might represent a common design matrix from the time-dependence.

Lipsitz et al. [56] and Liang et al. [53] suggested modelling the association by the pairwise marginal odds-ratios. With binary responses, the marginal odds-ratios are a natural measure of association, and $ln(OR)$ can be modeled as a linear function of covariates. Furthermore, given (μ_i, OR), we can always construct R_i since, given the means, the pairwise correlations are a one-to-one function of the pairwise marginal odds-ratios.

Limitations of GEE

The advantages and limitations of GEE are discussed in literature [12, 60]. Some researchers asserted that the major limitation of GEE is that it lacks a likelihood function (since the GEE does not specify completely the joint distribution). Likelihood-based methods are not available for testing fit, comparing models, and conducting inferences about parameters. Empirical based standard errors underestimate the true ones, unless very large sample size Chen and Lazar [10].

Although GEE is a popular approach for its flexibility, the very bottom line of GEE is to simply model the mean response and instead of attempting to model the within subject covariance structure, to treat it as a nuisance. Several studies pointed out that in GEE a consistent estimator for the regression parameter can be achieved without correctly specifying the correlation structure of the repeatedly measured outcomes. Yet, misspecification of working correlation could not only lead to loss of efficiency, but more seriously, could lead to non-

feasibility of the GEE solutions [74, 75]. Working correlation structure of GEE has been concern of many studies mainly focusing on examining the existing selection criteria and/or proposing new selection criteria for correlations structures [25, 33, 62, 64, 68, 76].

Unfortunately, although several studies pointed at the problems with the correlation structure of GEE and since the beginning of GEE, researchers are trying to improve and/or develop a method to capture the true correlation among the response variables, none of these studies addressed the root of the problem that the GEE are based on marginal models and considers independence assumption to make use of quasi-likelihood approach and then consider the correlation between successive measures of an individual to be induced correlation. As a result, with an arbitrary working assumption about the correlation among repeated measurements, the GEE estimator for the regression coefficients is always consistent. Hence, the solution of the problem lies elsewhere. In next chapter (Chapter 4), we described how the correlation among repeated observations remain not addressed in GEE. Due to lack of proper specification of the underlying model, marginal models such as GEE or ALR may fail to provide the measure of dependence of binary outcomes.

The GEE or GEE based models, being constructed to describe the population averaged or marginal distribution of repeated measurements, may sometimes be appropriate for descriptive observational studies but should be used carefully in causal experiments [54]. Moreover, the correlation considered in these GEE based methods are induced correlations and anomalies caused by the induced correlation between repeated outcomes is beyond any explanation. Many studies tried to address this problem by modifying the approaches based on marginal models using Markov based transition probabilities as alternatives to GEE for fitting population averaged logistic models (see for example, Zeger et al., 1985; Darlington and Farewell, 1992; Guerra et al., 2012).

3.4.2 Alternating Logistic Regression

Carey et al. [9] introduced Alternating Logistic Regression (ALR) models based on marginal odds ratios combining the first order GEE for regression coefficients with new logistic regression equation for estimating the correlation parameter.

Let each subject, i, be observed for n_i occasions and let Y_{ij} be a binary outcome variable observed for individual i at time j, $i = 1,2,...,N$ and $j = 1,2,...,n_i$ and also let $\boldsymbol{X}_{ij}$ be the covariate vector for individual i at follow up j.

Then for the ith individual we have a $n_i \times 1$ random vector of Bernoulli responses $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, ..., Y_{in_i})'$, and the associated covariates of Y_{ij} are $\boldsymbol{X}_{ij} = (X_{ij1}, X_{ij2}, ..., X_{ijp})'$ where $i = 1,2,...,N$, $j = 1,2,...,n_i$.

The mean vector for ith individual is

$$\boldsymbol{\mu}_i = (\mu_{i1}\ \mu_{i2}\ ...\ \mu_{in_i})' = [E(Y_{i1})\ E(Y_{i2})\ ...\ E(Y_{in_i})]' = (p_{i1}\ p_{i2}\ ...\ p_{in_i})',$$

where, the probability of observing the event for ith individual at jth occasion, $p_{ij} = Pr(Y_{ij} = 1 | X_{ij} = x_{ij})$; $i = 1,2,...,N$; $j = 1,2,...,n_i$. The probability of not observing the event for ith individual at jth occasion is $q_{ij} = 1 - p_{ij}$, $i = 1,2,...,N$; $j = 1,2,...,n_i$ and the variance of Y_{ij} is $p_{ij}(1 - p_{ij})$. The variance covariance matrix of $\boldsymbol{Y}_i$ is given by

$$V(Y_{i1}) = \begin{pmatrix} V(Y_{i1}) & cov(Y_{i1}, Y_{i2}) & ... & cov(Y_{i1}, Y_{in_i}) \\ cov(Y_{i2}, Y_{i1}) & V(Y_{i2}) & ... & cov(Y_{i2}, Y_{in_i}) \\ ... & ... & ... & ... \\ cov(Y_{in_i}, Y_{i1}) & cov(Y_{in_i}, Y_{i2}) & ... & V(Y_{in_i}) \end{pmatrix}.$$

For binary data, the correlation between the jth and kth response is, by definition,

$$corr(Y_{ij}, Y_{ik}) = \frac{Pr(Y_{ij} = 1, Y_{ik} = y_{ik}) - \mu_{ij}\mu_{ik}}{\sqrt{\mu_{ij}(1 - \mu_{ij})\mu_{ik}(1 - \mu_{ik})}}. \tag{3.4}$$

The joint probability in the numerator satisfies the following bounds, by elementary properties of probability, because $\mu_{ij} = Pr(Y_{ij} = 1)$, $\max(0, \mu_{ij} + \mu_{ik} - 1) \leq Pr(Y_{ij} = 1, Y_{ik} = 1) \leq \max(\mu_{ij}, \mu_{ik})$.

Therefore, the correlation is constrained to be within limits that depend in a complicated way on the means of the data. For modelling binary data, the odds ratio OR_{ijk} between the jth and kth responses for the ith subject can be expressed as

$$OR(Y_{ij}, Y_{ik}) = \frac{Pr(Y_{ij} = 1, Y_{ik} = 1)Pr(Y_{ij} = 0, Y_{ik} = 0)}{Pr(Y_{ij} = 1, Y_{ik} = 0)Pr(Y_{ij} = 0, Y_{ik} = 1)}, \tag{3.5}$$

which is not constrained by the means and is preferred, in some cases, to correlations for binary data. In ALR approach, the associations between pairs of outcome measures are modeled with odds ratios. As stated in Carey et al. [9], the strategy of ALR was developed following the suggestions by Firth [21] and Diggle [17] in the discussion of Liang et al. [53]. The ALR procedure combines the first order GEE for $\boldsymbol{\beta}$ with new logistic regression equations for estimating the correlation parameter α. The first order approach for $\boldsymbol{\beta}$ is retained because it gives robust and reasonably efficient estimates when the assumed form of $cov(\boldsymbol{Y}_i)$ is close to the true covariance matrix. The new equations for α are designed to avoid the computational burden of second-order equations that results from evaluating and inerting the matrix, $cov(\boldsymbol{Y}_i, \boldsymbol{W}_i)$, where $\boldsymbol{W}_i$ denotes the $\binom{n_i}{2}$ cross products of the Y_{ij}'s, $(Y_{i1}Y_{i2}, Y_{i1}Y_{i3}, ..., Y_{i1}Y_{in_i}, ..., Y_{i,n_i-1}Y_{in_i})'$. The strategy was to estimate α using the $\binom{n_i}{2}$ conditional events Y_{ij} given $Y_{ik} = y_{ik}$. In the simple case with $log\psi_{j,k} \equiv \alpha$; α is estimated by regressing Y_{ij} on Y_{ik}, for $1 \leq j \leq k \leq n_i$ with an appropriate offset. The prior hypothesis is that weighting the conditional elements as if independent of one another and of Y_{ij} will yield reasonably efficient estimates of α in many problems.

Let γ_{ijk} be the log odds ratio between outcomes Y_{ij} and Y_{ik}, let $\mu_{ij} = Pr(Y_{ij} = 1)$

and $\nu_{ijk} = E(Y_{ij}Y_{ik}) = Pr(Y_{ij} = 1, Y_{ik} = 1)$. Then following Diggle [17],

$$logitPr(Y_{ij} = 1|Y_{ik} = y_{ik}) = \gamma_{ijk}y_{ik} + log\left(\frac{\mu_{ij} - \nu_{ijk}}{1 - \mu_{ij} - \mu_{ik} + \nu_{ijk}}\right).$$

If a particular cluster has n_i measures, then there is an $\begin{pmatrix} n_i \\ 2 \end{pmatrix}$ vector ν_i with elements $\nu_{ijk} = E(Y_{ij}Y_{ik})$ for $1 \leq j < k \leq n_i$.

Let ζ_i be the $\begin{pmatrix} n_i \\ 2 \end{pmatrix}$ vector with elements

$$\begin{aligned} \zeta_{ijk} &= E(Y_{ij}|Y_{ik} = y_{ik}) \\ &= logit^{-1}\left(\gamma_{ijk}y_{ik} + log\left(\frac{\mu_{ij} - \nu_{ijk}}{1 - \mu_{ij} - \mu ik + \nu_{ijk}}\right)\right) \end{aligned}$$

and e_i be the vector of residuals with elements

$$e_{ijk} = Y_{ij} - E(Y_{ij}|Y_{ik} = y_{ik}) = Y_{ij} - \zeta_{ijk}.$$

Estimating Equations

Mathematically, the ALR estimates for β and α are the solutions to the following unbiased estimating equations

$$U(\beta) = \sum_{i=1}^{N} \boldsymbol{D}_i' \boldsymbol{A}_i^{-\frac{1}{2}} \boldsymbol{R}_i^{-1}(\hat{\alpha}) \boldsymbol{A}_i^{-\frac{1}{2}} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = 0 \tag{3.6}$$

$$\text{and} \quad U(\alpha) = \sum_{i=1}^{m} \frac{\delta \zeta_i}{\delta \alpha}|_{\beta=\hat{\beta}} H_i^{-1}(\hat{\zeta_{ijk}})(\boldsymbol{Y}_i - \hat{\boldsymbol{\zeta}}_i) = 0, \tag{3.7}$$

where $\boldsymbol{D}_i = \frac{\delta \mu_i}{\delta \beta}$, $cov(Y_i) = \boldsymbol{A}_i^{\frac{1}{2}} \boldsymbol{R}_i(\hat{\alpha}) \boldsymbol{A}_i^{\frac{1}{2}}$, $H(\zeta_{ijk}) = \zeta_{ijk}(1 - \zeta_{ijk})$ and $\boldsymbol{C}_i = \frac{\delta \zeta_i}{\delta \alpha^*}|_{\beta=\hat{\beta}}$.

Recall that in a GEE model, the correlation parameters (α) are estimated using estimates of the regression parameters (β). The regression parameter estimates are, in turn, updated using estimates of the correlation parameters. The computational process alternately updates the estimates of the alphas and then the

betas until convergence is achieved. The ALR approach works in a similar manner, except that the alpha parameters are log odds ratio parameters rather than correlation parameters. Moreover, for the same data, an odds ratio between the jth and kth responses that is greater than 1 using an ALR model corresponds to a positive correlation between the jth and kth responses using a GEE model. Similarly, an odds ratio less than 1 using an ALR model corresponds to a negative correlation between responses. But, the correspondence is not one-to-one, and examples can be constructed in which the same odds ratio corresponds to different correlations.

3.5 Working Likelihood Methods for Marginal Models

When the distribution of the outcome variables are known, likelihood based approaches can be used to estimate parameters of the models for repeated measures data. Some researchers analyzed longitudinal data based on marginal models where the approach was to approximate the actual likelihood with working likelihoods. We discuss the working likelihood based approaches by Darlington and Farewell [16], Guerra et al. [29], Zeger et al. [80] in the following subsections.

3.5.1 Zeger et al.'s Approach

Zeger et al. [80] proposed models in which marginal probabilities were expressed as logistic functions of the covariates instead of conditional probabilities. The approach was to approximate the actual likelihood with working likelihoods that lead to consistent estimates of β under weak assumptions. Time series models were considered to account for time dependence.

Model I by Zeger et al.

The first model by Zeger et al. [80] is based on a working likelihood approach assuming independence among the repeated outcomes.

Assumption

The repeated outcomes for each subject are independent.

Likelihood and Log-likelihood Function

Under the assumption of Zeger et al. [80], the marginal distribution of Y_{ij} is

$$logit[Pr(Y_{ij}=1|\boldsymbol{X}_i=\boldsymbol{x}_i)] \quad = \quad \boldsymbol{\beta}'\boldsymbol{x}_i$$

and the correlation between the successive outcomes can be shown as

$$corr(Y_{ij},Y_{ij-1}|\boldsymbol{X}_i=\boldsymbol{x}_i)=0 \text{ for } j=2,3,...,n_i.$$

Under the assumption of independent repeated outcomes, the standard likelihood analysis of the logistic regression model [14] is appropriate. The working likelihood function was defined as

$$L(\boldsymbol{\beta})=\prod_{i=1}^{N}\prod_{j=1}^{n_i} {p_{ij}}^{y_{ij}}(1-p_{ij})^{1-y_{ij}}, \tag{3.8}$$

$$\text{where } p_{ij}=Pr(Y_{ij}=1|\boldsymbol{X}_i=\boldsymbol{x}_i)=\frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta})}.$$

The log-likelihood function can be expressed as

$$l(\boldsymbol{\beta}) \quad = \quad \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left(y_{ij}\ln p_{ij}+(1-y_{ij})ln(1-p_{ij})\right). \tag{3.9}$$

Score Equations and Information Matrix

Differentiating $l(\boldsymbol{\beta})$ in equation (3.9) with respect to β_k and equating to zero, the score equations are obtained as

$$S_k = \frac{\delta l}{\delta \beta_k} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} x_{ijk}(y_{ij} - p_{ij}) = 0,\ k = 0, 1, ..., p. \tag{3.10}$$

Solving the score equations (3.10), the estimates of $\boldsymbol{\beta}$ can be obtained. Differentiating the log likelihood equations with respect to β_k, and $\beta_{k'}$ we have,

$$\frac{\delta^2 l}{\delta \beta_k \delta \beta_{k'}} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} -x_{ijk}^2 p_{ij}(1 - p_{ij}). \tag{3.11}$$

The working Fisher information matrix is obtained as

$$I_0(\boldsymbol{\beta}) = -\frac{\delta^2 l}{\delta \beta \delta \beta'}, \tag{3.12}$$

with diagonal elements $-\sum_{i=1}^{N} \sum_{j=1}^{n_i} -x_{ij}^2 p_{ij}(1 - p_{ij})$.
Under the assumption of independent repeated outcomes, $\hat{\boldsymbol{\beta}}$, the estimator of $\boldsymbol{\beta}$ proposed by Zeger et al. [80], is consistent and asymptotically Gaussian as $N \to \infty$ for any set of stationary processes such that $logit(p_i) = \boldsymbol{x}_i' \boldsymbol{\beta}$. Let $Y_{ij}, j = 1, 2, ..., n_i, i = 1, 2, ..., N$, be a stationary binary series such that $logit E[Y_{ij} | \boldsymbol{X}_i = \boldsymbol{x}_i] = \boldsymbol{x}_i' \boldsymbol{\beta}$. Then $N^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically multivariate Normal with expectation $\mathbf{0}$ and covariance matrix $\boldsymbol{I}_0^{-1} \boldsymbol{I}_0^* \boldsymbol{I}_0^{-1}$ where $\boldsymbol{I}_0^{-1} = E(SS')$.

Model II by Zeger et al.

The second model proposed by Zeger et al. [80] for repeated binary data considered a stationary Markov chain of order one among the repeated outcomes.

Assumption

Each series of the repeated outcomes for a subject is a stationary Markov chain of order one.

Likelihood and Log-likelihood Function

Under the assumption that each binary series is a realization of a stationary Markov chain, Zeger et al. [80] expressed as

$$Pr(Y_{ij} = 1|\boldsymbol{X}_i = \boldsymbol{x}_i) = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta})}.$$

The correlation between jth and $(j+1)$th outcome of ith response is

$$corr(Y_{ij}, Y_{ij-1}|\boldsymbol{X}_i = \boldsymbol{x}_i) = \rho = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\tau})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\tau})}, \tag{3.13}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ are the vectors of unknown parameters to be estimated. Under the assumption of Markov dependence among the repeated outcomes with a common lag one autocorrelation, the working likelihood function is defined as

$$L(\boldsymbol{\beta}, \boldsymbol{\tau}) = \prod_{i=1}^{N} p_i^{y_{i1}}(1-p_i)^{1-y_{i1}} \prod_{j=2}^{n_i} p_{ij}^{*\,y_{ij}}(1-p_{ij}^{*})^{1-y_{ij}}, \tag{3.14}$$

where the marginal probability p_i, the conditional probability p_{ij}^{*} and the association parameter ρ can be defined as

$$\begin{aligned}
p_i &= Pr(Y_{ij} = 1|\boldsymbol{X}_i = \boldsymbol{x}_i) \\
&= \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta})}, \\
p_{ij}^{*} &= Pr(Y_{ij} = 1|Y_{ij-1}, \boldsymbol{X}_i = \boldsymbol{x}_i) \\
&= E(Y_{ij}|Y_{ij-1}, \boldsymbol{X}_i = \boldsymbol{x}_i) \\
&= p_i + \rho(Y_{ij-1} - p_i), \\
\rho &= \frac{\exp(\boldsymbol{x}_i\boldsymbol{\tau})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\tau})}.
\end{aligned}$$

The log-likelihood function can be expressed as

$$\begin{aligned} l(\boldsymbol{\beta},\boldsymbol{\tau}) = & \sum_{i=1}^{N} \left(y_{i1}\ln(p_{i1}) + (1-y_{i1})\ln\left(1-p_{i1}\right)\right) \\ & + \sum_{i=1}^{N}\sum_{j=1}^{n_i} y_{ij}\ln\left(p_{ij}+\rho_i\left(y_{ij}-p_{ij}\right)\right) \\ & + \sum_{i=1}^{N}\sum_{j=1}^{n_i} (1-y_{ij})\ln\left(1-p_{ij}-\rho_i\left(y_{ij}-p_{ij}\right)\right). \end{aligned}$$

Score Equations and Information Matrix

Differentiating $l(\beta,\tau)$ with respect to β_k and ρ, and equating to zero, we have the score equations. Solving them simultaneously, we have the estimates of $\boldsymbol{\beta}$ and ρ.

Differentiating the score equations with respect to β_k' and ρ, respectively, we have, the information matrix as follow

$$\boldsymbol{I}_1 = \begin{pmatrix} \boldsymbol{I}_{\beta\beta} & \boldsymbol{I}_{\beta\rho} \\ \boldsymbol{I}_{\rho\beta} & \boldsymbol{I}_{\rho\rho} \end{pmatrix}.$$

$N^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ is, as $N\to\infty$, asymptotically Normal with expectation 0 and covariance matrix $\boldsymbol{V} = \boldsymbol{I}_1^{-1}\boldsymbol{I}_1^{*}\boldsymbol{I}_1^{-1}$ where $\boldsymbol{I}_1$ is the working Fisher information matrix as obtained above and $\boldsymbol{I}_1^{*} = E(\boldsymbol{S}\boldsymbol{S}')$.

3.5.2 Darlington and Farewell's Method

Darlington and Farewell [16] extended the model of Zeger et al. [80]. They showed that the dependence among repeated outcomes may depend on the explanatory variables when the focus of the study is to identify the relationship among the binary response and a set of independent variables. A robust estimate of the variance-covariance matrix of coefficient estimates were suggested to provide estimates of standard errors. This method considers conditional prob-

ability and one may argue that the work is more a conditional model than a marginal one. But since it ends up estimating a marginal $\boldsymbol{\beta}$, we considered reviewing this method under marginal models.

Assumptions

For a $k \times 1$ vector of unknown parameters $\boldsymbol{\eta}$, Darlington and Farewell [16] defined the conditional probability, $Pr(Y_{ij} = 1|Y_{ij-1}, \boldsymbol{X}_i) = \boldsymbol{\eta'} \boldsymbol{X}_i$.

Likelihood and Log-likelihood Functions

For this model, the working likelihood function can be written as

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} {p_i}^{y_{i1}} (1-p_i)^{1-y_{i1}} \prod_{j=2}^{n_i} {p^*_{ij}}^{y_{ij}} (1-p^*_{ij})^{1-y_{ij}}, \tag{3.15}$$

where

$$\begin{aligned} p_i &= Pr(Y_{ij} = 1|\boldsymbol{X}_i = \boldsymbol{x}_i) \\ &= \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta})}, \\ p^*_{ij} &= Pr(Y_{ij} = 1|Y_{ij-1}, \boldsymbol{X}_i = \boldsymbol{x}_i) \\ &= E(Y_{ij}|Y_{ij-1}, \boldsymbol{X}_i = \boldsymbol{x}_i) \\ &= p_i + \rho_i(Y_{ij-1} - p_i) \end{aligned}$$

and

$$\rho_i = \frac{\exp(\boldsymbol{\eta'}\boldsymbol{x}_i) - \exp(\boldsymbol{\beta'}\boldsymbol{x}_i)}{1+\exp(\boldsymbol{x}_i\boldsymbol{\eta})}.$$

The log-likelihood function can be written as

$$\begin{aligned} l(\boldsymbol{\beta}, \boldsymbol{\eta}) &= \sum_{i=1}^{N} \{y_{i1} \ln(p_1) + (1-y_{i1}) \ln(1-p_1)\} + \\ &\quad \sum_{i=1}^{N} \sum_{j=1}^{n_i} \{y_{ij} \ln\left(p_1 + \rho_i\left(y_{ij} - p_1\right)\right)\} + \\ &\quad \sum_{i=1}^{N} \sum_{j=1}^{n_i} \{\left(1-y_{ij}\right) \ln\left(1 - p_1 - \rho_i\left(y_{ij} - p_1\right)\right)\}. \end{aligned} \tag{3.16}$$

Score Equations and Fishers Information Matrix

First derivative of the log likelihood with respect to β_k and η equating to zero gives the score equations. Solving the score equations simultaneously give the estimates of $\hat{\beta}$ and ρ.

The Fisher's Information Matrix for the model proposed by Darlington and Farewell [16] has the form

$$I = \begin{pmatrix} I_{\beta\beta} & I_{\beta\eta} \\ I_{\eta\beta} & I_{\eta\eta} \end{pmatrix}.$$

Second Derivative of the log likelihood with respect to β and η gives the components of Fisher's Information Matrix. The element of uth row and kth colum of $\boldsymbol{I}_{\beta\beta}$ is $-\frac{\delta^2 l}{\delta\beta_u\delta\beta_k}$. The element of kth row and uth column of $\boldsymbol{I}_{\beta_k\eta_u}$ are obtained from $-\frac{\delta^2 l}{\delta\beta_k\delta\eta_u}$. and elements of uth row and lth column of $\boldsymbol{I}_{\eta\eta}$ are obtained from $-\frac{\delta^2 l}{\delta\eta_u\eta_l}$.

3.5.3 MARK1ML Approach

The maximum likelihood (ML) approach by Guerra et al. [29] for time-varying covariates is also an extension of the method for time-independent covariates by Zeger et al. [80]. Since this method also ends up estimating parameters of a marginal nature, we considered this method under marginal approaches.

Assumption

A Markovian model of first order (MARK1 model) is assumed so that the value of an outcome on a subject at a particular measurement occasion only depends on the value at the previous measurement occasion. Mathematically,

$$Pr(Y_{ij+1} = y_{ij+1}|Y_{i1} = y_{i1},...,Y_{ij} = y_{ij}) = Pr(Y_{ij+1} = y_{ij+1}|Y_{ij} = y_{ij}).$$

Likelihood and Log-likelihood Function

The joint likelihood can be expressed as the product of the pairwise probabilities of consecutive outcomes on each subject, with a logistic model for the marginal means. Guerra et al. [29] showed that the joint probability $Pr(Y_{i1} = y_{i1}, ..., Yij = y_{ij})$ of any particular permutation $(y_{i1}, ..., y_{ij})$ of zeros and ones can be expressed by conditioning Y_{ij} on $Y_{i1}, ..., Y_{ij-1}$. The joint probability of the n_i outcomes on subject i is

$$\begin{aligned} &Pr(Y_{i1} = y_{i1}, ..., Y_{in_i} = y_{in_i}) \\ &= Pr(Y_{i1} = y_{i1}) \prod_{j=1}^{n_i-1} Pr(Y_{ij+1} = y_{ij+1} | Y_{i1} = y_{i1}, ..., Y_{ij} = y_{ij}). \end{aligned} \tag{3.17}$$

Substituting the Markov assumption of order one into the equation (3.17) yields the following model for the n_i outcomes on subject i

$$\begin{aligned} &Pr(Y_{i1} = y_{i1}, ..., Y_{in_i} = y_{in_i}) \\ &= Pr(Y_{i1} = y_{i1}) \prod_{j=1}^{n_i-1} Pr(Y_{ij+1} = y_{ij+1} | Y_{ij} = y_{ij}) \\ &= Pr(Y_{i1} = y_{i1}) \prod_{j=1}^{n_i-1} \frac{Pr(Y_{ij+1} = y_{ij+1}, Y_{ij} = y_{ij})}{Pr(Y_{ij} = y_{ij})}. \end{aligned} \tag{3.18}$$

The pairwise probabilities in equation (3.18) can be expressed as

$$\begin{aligned} &Pr(Y_{ij+1} = y_{ij+1}, Y_{ij} = y_{ij}) \\ &= p_{ij+1}^{y_{ij+1}} q_{ij+1}^{1-y_{ij+1}} p_{ij}^{y_{ij}} q_{ij}^{1-y_{ij}} \left(1 + C_{ijj+1}(\alpha) \frac{(y_{ij+1} - p_{ij+1})(y_{ij} - p_{ij})}{(p_{ij+1} \cdot q_{ij+1} \cdot p_{ij} \cdot q_{ij})^{\frac{1}{2}}} \right), \end{aligned}$$

where $C_{ijj+1}(\alpha) = Corr(Y_{ij+1}, Y_{ij})$ and $q_{ij} = 1 - p_{ij}$ [66]. Then

$$\begin{aligned} &Pr(Y_{i1} = y_{i1}, ..., Y_{in_i} = y_{in_i}) \\ &= p_{i1}^{y_{i1}} q_{i1}^{1-y_{i1}} \prod_{j=1}^{n_i-1} p_{ij+1}^{y_{ij+1}} q_{ij+1}^{1-y_{ij+1}} \left(1 + C_{ijj+1}(\alpha) \frac{(y_{ij+1} - p_{ij+1})(y_{ij} - p_{ij})}{(p_{ij+1} \cdot q_{ij+1} \cdot p_{ij} \cdot q_{ij})^{\frac{1}{2}}} \right). \end{aligned} \tag{3.19}$$

The log-likelihood is obtained as

$$\begin{aligned} l(\beta,\alpha) &= \sum_{i=1}^{N}\sum_{j=1}^{n_i}\{y_{ij}log(p_{ij})+(1-y_{ij})log(1-p_{ij})\} \\ &\quad +\sum_{i=1}^{N}\sum_{j=1}^{n_i} log\left(1+C_{ijj+1}(\alpha)\frac{(y_{ij+1}-p_{ij+1})(y_{ij}-p_{ij})}{(p_{ij+1}.q_{ij+1}.p_{ij}.q_{ij})^{\frac{1}{2}}}\right). \end{aligned} \tag{3.20}$$

The correlation between Y_{ij} and Y_{ik} for $j<k$ can be expressed as

$$Corr(Y_{ij},Y_{ik}) = C_{ijj+1}(\alpha)...C_{i(k-1)k}(\alpha) = \prod_{s=j}^{k-1} C_{iss+1}(\alpha).$$

Different functional forms were specified for $C_{ijj+1}(\alpha)$ to obtain the correlation structures such as autoregressive, Markov or unstructured correlation.

Score Equations and Information Matrix

The score equation for β can be obtained by differentiating the log-likelihood function with respect to β and equating to zero. The elements of the information matrix are obtained by differentiating the score equation $U(\beta)$ with respect to the respective parameters.

Note that, although the model proposed by Guerra et al. [29] is based on the assumption of a first order Markovian model, the joint likelihood of their model was expressed as a product of the pairwise probabilities of successive outcomes on each subject, with a logistic model for the marginal means instead of a model for the conditional probabilities. Hence, the MARK1ML model proposed by Guerra et al. [29] basically considers the usual logistic model similar to a GEE model for binary data.

3.6 Conclusion

In this chapter, we described selected marginal models in details including the very popular GEE and ALR methods. We discussed that these models, because of their marginal features, fail to make use of the main feature of a longitudinal data. We showed that, although the Models proposed by Zeger et al. [80], Darlington and Farewell [16] and Guerra et al. [29] incorporated Markov based transition probabilities to define conditional probabilities, the methods ended up with estimation of marginal effect of covariates for each follow ups. In the next chapter, we extend the discussion on these models to examine the correlation structures between repeated outcomes considered in these models.

Chapter 4

Analysis of Correlation Structures used in Marginal Models

4.1 Introduction

Generalized linear models were extended in different ways to model longitudinal data including marginal or population averaged models and transition or conditional probability models (Chapter 3). We discussed the GEE approach as a natural extension of the quasi-likelihood approach [77] for GLM to the multivariate response setting, with an additional set of nuisance parameters incorporated to take into account the within-subject association.

Although the GEE or GEE based approaches yield consistent estimators of $\boldsymbol{\beta}$, even under misspecification of the within subject associations, GEE or related methods, by nature, provide very limited information about the associations within the repeated outcomes themselves. This is disorganized, since previ-

ous outcome is frequently an important predictor of future outcomes for many studies. Many of the earlier studies identified this problem and tried to address this problem in a variety of ways and proposed alternatives to GEE based on marginal models. Among those, we discussed the working likelihood based approaches by Zeger et al. [80], Darlington and Farewell [16], Guerra et al. [29] using Markov based transition probabilities (Chapter 3). We showed that in spite of using Markov transition probability, the binary models of Zeger et al. [80] as well as Darlington and Farewell [16], Guerra et al. [29] focused on marginal models.

In this chapter, we continue the discussion and compare GEE, ALR and models proposed by Zeger et al. [80], Darlington and Farewell [16], Guerra et al. [29] in terms of the correlation pattern considered among the repeated outcomes on same individuals.

4.2 Statement of the Problem in Repeated Measures Data

As defined earlier in Chapter 2, let us consider a longitudinal study for a specified time period for a sample of size N. We have data from several follow-ups on each of N units. The N units in the sample produce data on the outcome variable, Y. Let Y_{ij} be a binary response observed for individual i at time j. Let each subject, i, be observed for n_i occasions so that $i = 1,2,...,N$ and $j = 1,2,...,n_i$. Also let $\boldsymbol{X}_{ij}$ be the vector of parameters for jth response of ith individual.

4.2.1 Structure of the Data on Repeated Outcomes

Let us consider the simplest case, individual i is observed on two occasions taking value either 0 or 1. The value of the outcome variable at time point 1, Y_{i1} can take values 0 or 1. Then for $Y_{i1} = 0$, the value of the outcome variable at time point 2, Y_{i2}, can be either 0 or 1. Again for $Y_{i1} = 1$, the outcome variable

at time point 2, Y_{i2} can take either of the two values, 0 or 1.

In a longitudinal data, the repeated outcomes on the same subject are expected to be correlated. Consider a study with two repeated binary outcome variables Y_{i1} and Y_{i2} for each subject i where Y_{i1} and Y_{i2} can take values 0 or 1.

The outcome vector for ith subject can be expressed as $\boldsymbol{Y}_i = (Y_{i1}\ Y_{i2})'$. The mean vector for ith subject is $\boldsymbol{\mu}_i = (\mu_{i1}\ \mu_{i2})' = [E(Y_{i1})\ E(Y_{i2})]' = (p_{i1}\ p_{i2})'$, where, $p_{ij} = Pr(Y_{ij} = 1|\boldsymbol{X}_{ij} = \boldsymbol{x}_{ij})$; $i = 1,2,...,N$; $j = 1,2$. The probability of not observing the event for ith individual at jth occasion is $q_{ij} = 1 - p_{ij}$. The variance covariance matrix of $\boldsymbol{Y}_i$ is

$$cov(\boldsymbol{Y}_i) = \begin{pmatrix} V(Y_{i1}) & cov(Y_{i1}, Y_{i2}) \\ cov(Y_{i2}, Y_{i1}) & V(Y_{i2}) \end{pmatrix},$$

where $V(Y_{ij}) = p_{ij}(1 - p_{ij})$.

We may define $\boldsymbol{\beta}_1$ as the $(p+1) \times 1$ vector of parameters of the marginal model $P(Y_{i1} = 1|\boldsymbol{X}_i = \boldsymbol{x}_i)$ and $\boldsymbol{\beta}_2$ as the $(p+1) \times 1$ vector of parameters of the marginal model $P(Y_{i2} = 1|\boldsymbol{X}_i = \boldsymbol{x}_i)$. Also we may define, $\boldsymbol{\beta}_{01}$ as the $(p+1) \times 1$ vector of parameters of the conditional model $P(Y_{i2} = 1|Y_{i1} = 0, \boldsymbol{X}_i = \boldsymbol{x}_i)$ and $\boldsymbol{\beta}_{11}$ as the $(p+1) \times 1$ vector of parameters of the conditional model $P(Y_{i2} = 1|Y_{i1} = 1, \boldsymbol{X}_i = \boldsymbol{x}_i)$. A generalized linear model, which is a common choice for analysing longitudinal data, can be expressed as

$$f(y_{ij}, \theta_{ij}, \phi_{ij}) = exp[(y_{ij}\theta_{ij} - b(\theta_{ij}))a^{-1}(\phi_{ij}) + c(y_{ij}, \phi_{ij})],$$

where the canonical parameter $\theta_{ij} = g(\mu_{ij}) = \boldsymbol{X}_{ij}\boldsymbol{\beta}$ and $\boldsymbol{\beta}$ is the coefficient of $\boldsymbol{X}_{ij}$ (equation (2.1), Chapter 2).

The first derivative of $b(\theta_{ij})$ is $b'(\theta_{ij}) = E(Y_{ij}) = \mu_{ij}$; the second derivative of $b(\theta_{ij})$ is $b''(\theta_{ij}) = V(\mu_{ij})$ and $V(Y_{ij}) = a(\phi_{ij})b''(\theta_{ij}) = a(\phi_{ij})V(\mu_{ij})$, where $a(\phi_{ij})$ is the dispersion parameter.

For two Bernoulli populations, we have the logit link functions

$$\begin{aligned}
\theta_{i1} = g(\mu_{i1}) &= ln\frac{\mu_{i1}}{1-\mu_{i1}} = \boldsymbol{X}_{i1}\boldsymbol{\beta}_1, \\
\mu_{i1} = E(Y_{i1}) &= p_{i1} = \frac{\exp(\boldsymbol{X}_{i1}\boldsymbol{\beta}_1)}{1+\exp(\boldsymbol{X}_{i1}\boldsymbol{\beta}_1)}, \\
\theta_{i2} = g(\mu_{i2}) &= ln\frac{\mu_{i2}}{1-\mu_{i2}} = \boldsymbol{X}_{i2}\boldsymbol{\beta}_2, \\
\mu_{i2} = E(Y_{i2}) &= p_{i2} = \frac{\exp(\boldsymbol{X}_{i2}\boldsymbol{\beta}_2)}{1+\exp(\boldsymbol{X}_{i2}\boldsymbol{\beta}_2)},
\end{aligned}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively, are the parameters of the marginal models of Y_{i1} and Y_{i2}.

4.2.2 Correlation among the Repeated Outcomes

Let us consider two binary outcomes Y_{i1} and Y_{i2} for ith individual. Consider binary outcomes Y_{i1} and Y_{i2} for ith individual. If Y_{i1} and Y_{i2} are not independent, then the conditional probability of Y_{i2} given Y_{i1} can be expressed as [16, 65]

$$\begin{aligned}
P(Y_{i2}=1|y_{i1},\boldsymbol{x}_{i2}) &= P(Y_{i2}=1|\boldsymbol{X}_{i2}=\boldsymbol{x}_{i2}) \\
&\quad +\rho_i\left(Y_{i1}-P(Y_{i1}|\boldsymbol{X}_{i1}=\boldsymbol{x}_{i1})\right), \qquad (4.1)
\end{aligned}$$

where ρ is the correlation between Y_{i1} and Y_{i2}. For $Y_{i1}=0$, equation (4.1) can be expressed as,

$$\begin{aligned}
P(Y_{i2}=1|Y_{i1}=0,\boldsymbol{X}_{i2}=\boldsymbol{x}_{i2}) &= P(Y_{i2}=1|\boldsymbol{x}_{i2})+\rho_i\left(0-P(Y_{i1}|\boldsymbol{x}_{i1})\right) \\
\text{or,}\quad \frac{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01})}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01})} &= \frac{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_2)}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_2)}-\rho_i.\frac{\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}{1+\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}. \qquad (4.2)
\end{aligned}$$

For $Y_{i1}=1$, equation (4.1) can be expressed as

$$\begin{aligned}
&P(Y_{i2}=1|Y_{i1}=1,\boldsymbol{x}_{i2}) = P(Y_{i2}=1|\boldsymbol{x}_{i2})+\rho_i\left(1-P(Y_{i1}|\boldsymbol{x}_{i1})\right) \\
\text{or,}\quad &\frac{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{11})}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{11})} = \frac{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_2)}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_2)}+\rho_i\left(1-\frac{\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}{1+\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}\right). \qquad (4.3)
\end{aligned}$$

Clearly ρ_i is a function of β_1, β_2 and $\beta_{2.1}$ where β_1 and β_2 are the parameters of the marginal models, $P(Y_{i1}=1)$ and $P(Y_{i2}=1)$, respectively. For Bernoulli outcomes, $\beta_{2.1}$, denote the vector of parameters of the two conditional models $P(Y_{i1}=1|Y_{i1}=0)$ and $P(Y_{i1}=1|Y_{i1}=1)$. If Y_{i1} and Y_{i2} are not correlated, $\rho_i=0$ and

$$P(Y_{i2}=1|Y_{i1},\boldsymbol{X}_{i2}=\boldsymbol{x}_{i2})=P(Y_{i2}=1|\boldsymbol{X}_{i2}=\boldsymbol{x}_{i2})=\frac{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_2)}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_2)}. \quad (4.4)$$

Theoretically, the observed correlation between two repeated outcome variables, Y_{i1} and Y_{i2}, can be shown as

$$\rho_i=\frac{cov(Y_{i1},Y_{i2})}{\sqrt{V(Y_{i1})}\sqrt{V(Y_{i2})}}=\frac{E(Y_{i1}Y_{i2})-E(Y_{i1})E(Y_{i2})}{\sqrt{\mu_{i1}(1-\mu_{i1})}\sqrt{\mu_{i2}(1-\mu_{i2})}}, \quad (4.5)$$

$$\begin{aligned}\text{where}\quad E(Y_{i1}Y_{i2}) &= \sum_{y_{i1},y_{i2}=0}^{1} y_{i1}y_{i2}P(Y_{i1}=y_{i1},Y_{i2}=y_{i2})\\ &= P(Y_{i1}=1,Y_{i2}=1)\\ &= P(Y_{i2}=1|Y_{i1}=1)P(Y_{i1}=1)\\ &= \frac{\exp(x_{i2}\boldsymbol{\beta}_{11})}{1+\exp(x_{i2}\boldsymbol{\beta}_{11})}.\frac{\exp(x_{i1}\boldsymbol{\beta}_1)}{1+\exp(x_{i1}\boldsymbol{\beta}_1)}. \quad (4.6)\end{aligned}$$

If $\boldsymbol{X}_{ij}$ is time invariant then the correlation between Y_{i1} and Y_{i2}, can be shown as

$$\rho_i=\exp\left(\frac{1}{2}\boldsymbol{x}_i(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)\right)\frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta}_{11})-\exp(\boldsymbol{x}_i\boldsymbol{\beta}_2)}{(1+\exp(\boldsymbol{x}_i\boldsymbol{\beta}_{11}))}. \quad (4.7)$$

Equation (4.7) shows that correlation between Y_{i1} and Y_{i2} is equal to zero when $\boldsymbol{\beta}_{11}=\boldsymbol{\beta}_2$. However, this condition does not completely define no association between Y_{i1} and Y_{i2}. The equations (4.2) and (4.3) show that for the independence of Y_{i1} and Y_{i2}, it is necessary that both $\boldsymbol{\beta}_{01}$ and $\boldsymbol{\beta}_{11}$ are equal and equal to $\boldsymbol{\beta}_2$. If $\boldsymbol{\beta}_{01}\neq\boldsymbol{\beta}_{11}$ then Y_{i1} and Y_{i2} are associated. Islam et al. [40] showed that the dependence in bivariate Bernoulli outcome variables can be tested by

testing the equality of these two conditional models. Y_{i1} and Y_{i2} are independent if

$$\begin{aligned} P(Y_{i2}=1|Y_{i1}=y_{i1},\boldsymbol{X}_{i2}=\boldsymbol{x}_{i2}) &= P(Y_{i2}=y_{i2}|Y_{i1}=0,\boldsymbol{X}_{i2}=\boldsymbol{x}_{i2}) \\ &= P(Y_{i2}=y_{i2}|Y_{i1}=1,\boldsymbol{X}_{i2}=\boldsymbol{x}_{i2}) \\ &= P(Y_{i2}=y_{i2}|\boldsymbol{X}_{i2}=\boldsymbol{x}_{i2}), \end{aligned}$$

i.e. $\boldsymbol{\beta}_{2.1}=\boldsymbol{\beta}_{01}=\boldsymbol{\beta}_{11}=\boldsymbol{\beta}_2$.

It should also be noted that even if the distribution of Y_{i1}, Y_{i2}, ..., Y_{ij} are independent, i.e., $\boldsymbol{\beta}_{j.12\ldots j-1}=\boldsymbol{\beta}_j$, $\forall j=2,3,\ldots,n_i$, this does not necessarily mean that the distribution of Y_{ij}'s are identical. Distribution of Y_{i1}, Y_{i2}, ..., Y_{ij} are identical only if $\boldsymbol{\beta}_1=\boldsymbol{\beta}_{2.1}=\ldots=\boldsymbol{\beta}_{j.12\ldots j-1}=\boldsymbol{\beta}$.

4.3 Correlation under GEE

Following the quasi-likelihood approach [77], with a mean model, μ_{ij}, and variance structure, V_{ij}, Liang and Zeger [52] expressed the GEE for $\boldsymbol{\beta}$ of the form (as shown in equation (3.2))

$$\begin{aligned} U(\boldsymbol{\beta}) &= \sum_{i=1}^{N}\boldsymbol{D}_i'\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i-\boldsymbol{\mu}_i)=0, \\ \text{with } \boldsymbol{D}_i &= \frac{\delta\boldsymbol{\mu}_i}{\delta\boldsymbol{\beta}}, \\ \boldsymbol{V}_i &= \boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{R}_i(\alpha)\boldsymbol{A}_i^{\frac{1}{2}}, \text{ a working or covariance matrix of } Y_i, \\ \boldsymbol{A}_i &= diag[V(Y_{i1}),\ldots,V(Y_{in_i})], \text{ a } n_i\times n_i \text{ diagonal matrix}, \\ \boldsymbol{R}_i(\alpha) &= corr(Y_{ij},Y_{ik}),\ j\neq k \text{ is a working correlation matrix}. \end{aligned}$$

In a longitudinal data, the repeated observations on same individual are expected to be correlated. The correlation between Y_{i1} and Y_{i2} were shown in equation (4.5) in section 4.2.2 as

$$\rho=\exp\left(\frac{1}{2}\boldsymbol{X}_i(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)\right)\frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta}_{11})-\exp(\boldsymbol{X}_i\boldsymbol{\beta}_2)}{1+\exp(\boldsymbol{X}_i\boldsymbol{\beta}_{11})},$$

where β_1 and β_2 are parameters of the marginal models of Y_1 and Y_2 and β_{11} is the parameter of the conditional model of Y_2 given $Y_1 = 1$. Note that the parameters to be estimated in GEE model are the parameters of a marginal or population averaged model (same in each follow-up), i.e. $\beta_1 = \beta_2 = \beta(say)$. Under this assumption, the equation (4.5) becomes

$$\rho = \frac{\exp(\boldsymbol{X}_i\beta_{11}) - \exp(\boldsymbol{X}_i\beta)}{1 + \exp(\boldsymbol{X}_i\beta_{11})}.$$

Also to be noted that GEE does not take into account any transition model (such as $P(Y_{ij} = 1|Y_{ij-1})$) and as a population averaged model, it considers the same distribution for all $Y_{ij}, j = 1,2,...,n_i$. In other words, even for the correlated outcomes, GEE estimates the same regression parameter β for the model $P(Y_{ij} = 1|Y_{ij-1})$. That inevitably proves that, under the assumption of GEE, $\beta_{01} = \beta_{11} = \beta$, and with this assumption, the equation (4.5) becomes $\rho = 0$.
One might argue that, in GEE, the correlation among the repeated outcomes are attempted to be addressed by incorporating some nuisance correlation parameters, assuming correlation structures such as independence, autoregressive, exchangeable, etc.
But it must be remembered that being a marginal or population averaged model, GEE considers $\beta_1 = \beta_{2.1} = ... = \beta_{n_i.12...n_i-1} = \beta$, which logically fits only when $\boldsymbol{Y}_i$'s are identically and independently distributed. Hence inducing a (nuisance) correlation structure contradicts with the basic assumptions of GEE unless the correlation structure considered is independent correlation. So inducing a correlation structure might contradict with the true correlations among the repeated measures unless the correlation structure considered is an independent correlation, such that, the repeated outcomes are independent.

4.4 Correlation in Alternating Logistic Regression

The alternating logistic regression procedure or ALR proposed by Carey et al. [9] combines the first order GEE for β with new logistic regression equations for estimating α, the correlation parameter. The first order approach for β is retained because it gives robust and reasonably efficient estimates when the assumed form of $cov(Y_i)$ is close to the true covariance matrix. The new equations for α are designed to avoid the computational burden of second-order equations that results from evaluating and inverting the covariance matrix, $cov(Y_i, W_i)$. The strategy was to estimate α using the $\binom{n_i}{2}$ conditional events Y_{ij} given $Y_{ik} = y_{ik}$. In the simple case α is estimated by regressing Y_{ij} on Y_{ik}, for $1 \leq j \leq k \leq n_i$ with an appropriate offset. The prior hypothesis is that weighting the conditional elements as if independent of one another and of Y_{ij} will yield reasonably efficient estimates of α in many problems.

The ALR strategy follows from the suggestions by Firth [21] and Diggle [17] in the discussion of Liang et al. [53].

Let γ_{ijk} be the log odds ratio between out comes Y_{ij} and Y_{ik}, $\mu_{ij} = Pr(Y_{ij} = 1)$ and $\nu_{ijk} = Pr(Y_{ij} = 1, Y_{ik} = 1)$. Then following Diggle [17], $\text{logitPr}(\text{Y}_{\text{ij}} = 1|\text{Y}_{\text{ik}} = \text{y}_{\text{ik}}) = \gamma_{\text{ijk}}\text{y}_{\text{ik}} + \log\left(\frac{\mu_{\text{ij}} - \nu_{\text{ijk}}}{1 - \mu_{\text{ij}} - \mu\text{ik} + \nu_{\text{ijk}}}\right)$.

Unlike tha correlation parameters in GEE, the association parameters in ALR are the log odds ratio parameters, estimated using estimates of the regression parameters, β. The regression parameter estimates are, in turn, updated using estimates of the correlation parameters. The computational process alternately updates the estimates of the log odds ratios and then the β's until convergence is achieved. Furthermore, for the same data, an odds ratio between the jth and kth responses that is greater than 1 using an ALR model corresponds to a positive correlation between the jth and kth responses using a GEE model. Similarly, an odds ratio less than 1 using an ALR model corresponds to a nega-

tive correlation between responses. But, the correspondence is not one-to-one, and examples can be constructed in which the same odds ratio corresponds to different correlations.

4.5 Correlation in Zeger et al.'s Model

If Y_{ij} is a binary outcome variable observed for individual i at time j, $i = 1,2,...,N$ and $j = 1,2,...,n_i$ and $\boldsymbol{X}_i$ is the $p \times 1$ vector of covariates for individual i at each follow up j, then under the assumptions of Zeger et al. [80], the marginal distribution of Y_{ij} and correlation between Y_{ij} and Y_{ij+1} can be expressed as (section 3.5.1 in Chapter 3)

$$Pr(Y_{ij} = 1|\boldsymbol{x}_i) = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta})} \quad \text{and} \quad \rho = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\tau})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\tau})}.$$

Under the assumption of Markov dependence among the repeated outcomes, the working likelihood function is (equation (3.14))

$$L(\beta,\tau) = \prod_{i=1}^{m} p_i^{y_{i1}}(1-p_i)^{1-y_{i1}} \prod_{j=2}^{n_i} p_{ij}^{*\,y_{ij}}(1-p_{ij}^{*})^{1-y_{ij}},$$

$$\begin{aligned}
\text{where } p_i &= Pr(Y_{ij} = 1|\boldsymbol{X}_i = \boldsymbol{x}_i) = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta})}, \\
p_{ij}^{*} &= Pr(Y_{ij} = 1|Y_{ij-1}, \boldsymbol{X}_i) = E(Y_{ij}|Y_{ij-1}, \boldsymbol{X}_i) \\
&= p_i + \rho(Y_{it-1} - p_i) \\
\text{and } \rho &= \frac{\exp(\boldsymbol{x}_i\boldsymbol{\tau})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\tau})}.
\end{aligned}$$

The method proposed by Zeger et al. [80] uses a marginal model, where a marginal $\boldsymbol{\beta}$ is considered for each follow up. Hence, in this method, $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}$. With this assumption in equation (4.5), we have

$$\rho = \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta}_{11}) - \exp(\boldsymbol{X}_i\boldsymbol{\beta})}{1+\exp(\boldsymbol{X}_i\boldsymbol{\beta}_{11})}.$$

In addition, as Zeger et al. [80]'s method considers a marginal $\boldsymbol{\beta}$, the method considers $\boldsymbol{\beta}_{01} = \boldsymbol{\beta}_{11} = \boldsymbol{\beta}$, which results in $\rho = 0$. Clearly, the correlation defined by Zeger et al. [80], $\rho = \frac{\exp(\boldsymbol{X}_i\boldsymbol{\tau})}{1+\exp(\boldsymbol{X}_i\boldsymbol{\tau})}$, is an induced correlation not representing the true association scenario and it is not logical to consider such correlation among repeated measures.

4.6 Correlation by Darlington and Farewell

Darlington and Farewell [16] assumed Markov dependence among the repeated outcomes and redefined the working likelihood function of Zeger et al. [80] as (equation (3.15), section 3.5.2, Chapter 3)

$$L(\boldsymbol{\beta},\boldsymbol{\eta}) = \prod_{i=1}^{N} p_i^{y_{i1}}(1-p_i)^{1-y_{i1}} \prod_{j=2}^{n_i} p_{ij}^{*\,y_{ij}}(1-p_{ij}^{*})^{1-y_{ij}},$$

$$\begin{aligned} \text{where} \quad p_i &= Pr(Y_{ij}=1|\boldsymbol{X}_i=\boldsymbol{x}_i) = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta})}, \\ p_{ij}^{*} &= Pr(Y_{ij}=1|Y_{ij-1},\boldsymbol{X}_i=\boldsymbol{x}_i) = E(Y_{ij}|Y_{ij-1},\boldsymbol{X}_i=\boldsymbol{x}_i) \\ &= p_i + \rho_i(Y_{ij-1}-p_i) \end{aligned}$$

$$\text{and} \quad \rho_i = \frac{\exp(\boldsymbol{\eta}'\boldsymbol{x}_i)-\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)}{1+\exp(\boldsymbol{x}_i\boldsymbol{\eta})}.$$

Under the assumption of Darlington and Farewell [16], $\beta_1 = \beta_2 = \beta$ and under this assumption, the equation (4.5) becomes

$$\rho_i = \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta}_{11})-\exp(\boldsymbol{X}_i\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta}_{11})},$$

which is the correlation defined by Darlington and Farewell [16],

$$\rho_i = \frac{\exp(\boldsymbol{\eta}'\boldsymbol{x}_i)-\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)}{1+\exp(\boldsymbol{x}_i\boldsymbol{\eta})}, \text{ if } \boldsymbol{\beta}_{11}=\boldsymbol{\eta}.$$

According to Darlington and Farewell [16], complete independence is present when $\rho = 0$ i.e. $\boldsymbol{\beta} = \boldsymbol{\eta}$. But we showed in section 4.2.2 that $\rho = 0$, when $P(Y_{ij} = 1|Y_{ij-1,}\boldsymbol{X}_i) = P(Y_{ij} = 1|\boldsymbol{X}_i)$. The two conditional models for Y_{i2} given $Y_{i1} = 1$ and $Y_{i1} = 0$, are

$$\begin{aligned} p^*_{ij} &= P(Y_{ij} = 1|Y_{ij-1} = 1) = \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta}_{11})}{1+\exp(\boldsymbol{X}_i\boldsymbol{\beta}_{11})};\ if\ Y_{ij-1} = 1 \\ &= P(Y_{ij} = 1|Y_{ij-1} = 0) = \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta}_{01})}{1+\exp(\boldsymbol{X}_i\boldsymbol{\beta}_{01})};\ if\ Y_{ij-1} = 0. \end{aligned}$$

When both the p^*_{ij} are equal to the marginal probability $p_i = P(Y_{ij} = 1|\boldsymbol{X}_i)$, the outcomes are independent, or in other words, if the two outcomes are independent, $p^*_{ij} = p_i$. This implies that Y_{i1} and Y_{i2} are independent when both of the following are true

$$\frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta}_{11})}{1+\exp(\boldsymbol{X}_i\boldsymbol{\beta}_{11})} = \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta})}{1+\exp(\boldsymbol{X}_i\boldsymbol{\beta})} \tag{4.8}$$

$$\text{and} \quad \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta}_{01})}{1+\exp(\boldsymbol{X}_i\boldsymbol{\beta}_{01})} = \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta})}{1+\exp(\boldsymbol{X}_i\boldsymbol{\beta})}. \tag{4.9}$$

We can say that both the tests $H_{01} : \boldsymbol{\beta}_{01} = \boldsymbol{\beta}_2$ and $H_{02} : \boldsymbol{\beta}_{11} = \boldsymbol{\beta}_2$ are needed to be performed for independence of the outcome variables on repeated occasions or follow ups [39]. Clearly, Darlington and Farewell [16] closely attempted to address the correlation among the repeated measures but could only, partially, portray it.

4.7 Correlation in MARK1 Model

Guerra et al. [29] made some improvements over the model proposed by Zeger et al. [80] based on a Markovian model of the first order. Guerra et al. [29] showed that the consecutive pairwise probabilities can be expressed as [66]

$$
\begin{aligned}
&Pr(Y_{ij+1} = y_{ij+1}, Y_{ij} = y_{ij}) \\
&= p_{ij+1}^{y_{ij+1}} q_{ij+1}^{1-y_{ij+1}} p_{ij}^{y_{ij}} q_{ij}^{1-y_{ij}} \left(1 + C_{ijj+1}(\alpha) \frac{(y_{ij+1} - p_{ij+1})(y_{ij} - p_{ij})}{(p_{ij+1} . q_{ij+1} . p_{ij} . q_{ij})^{\frac{1}{2}}} \right),
\end{aligned}
$$

where $C_{ijj+1}(\alpha) = Corr(Y_{ij+1}, Y_{ij})$. Different correlation structures, for example, independence, exchangeable, autoregressive, etc. can be induced among the repeated responses. Although Guerra et al. [29] introduced a Markovian approach in their model, they specified a logistic model for the marginal means under the set up of marginal probabilities instead of a model for the conditional probabilities (section 3.5.3, Chapter 3). As a result, this approach, too, ultimately ended up with a similar problem in identifying the association pattern between repeated measures as the earlier marginal approaches of Zeger et al. [80] and Darlington and Farewell [16].

4.8 Conclusion

In this chapter, we showed that both GEE and ALR are based on marginal models and are inadequate to provide the measure of dependence of binary outcomes due to its marginal or population averaged feature. One may argue that GEE and ALR considers correlation between the repeated outcomes, but we explained in this chapter that the correlation considered are induced and nuisance correlations and anomalies caused by the induced correlation between repeated outcomes is beyond any explanation. As a result, an alternative model is required for longitudinal data

Chapter 5

Proposition of Marginal Conditional Models

5.1 Introduction

Most of the longitudinal models are based on marginal approaches. In almost all the cases, the marginal models consider an induced correlation between successive individuals. We discussed in Chapter 3 and 4 that the marginal models have limitations in analysing the repeated measures data, mainly, because of their marginal features and due to the correlation structures considered. Use of correlation in a marginal model lacks in proper specification of the dependence of binary outcomes in the model. Furthermore, an Induced correlation does not fit to the estimation procedure while a marginal model is taken. Hence it may fail to provide efficient estimation of parameters of the model considered.

Markov models for binary longitudinal data have been explored by many re-

searchers. Cox [14] proposed an extension of logistic regression in which the conditional rather than the marginal probabilities of a Markov chain are expressed as logistic functions of the covariates. This method have been applied to the analysis of many binary series [48, 59]. Cox [14]'s model leads to propensity estimates which depend strongly on the specification of the time dependence, for example, choice of order of Markov chain. When the covariates are categorical, an alternative approach is introduced by Grizzle et al. [28], Koch et al. [47] who proposed repeated measures models to account for time dependence in dichotomous data.

Azzalini [3], Bonney [7, 8] also used logistic regression for autocorrelated data with repeated measures using a conditional model approach. Islam et al. [34], Islam and Chowdhury [35], Islam et al. [39, 40, 41, 42] carried out a series of research works using Markov based conditional models based on Markov transition probability for repeated binary data. The conditional regressive logistic model of Bonney (1986, 1987) were generalized by Islam et al. [39, 42] to include both binary outcomes in previous times as well as covariates in the conditional models. In this Chapter, we discuss the Markov based transition probability models proposed by Islam and Chowdhury [35] along with necessary tests and propose joint models based on marginal conditional approaches for analysing Repeated Binary data as alternatives to GEE based approaches.

Although the idea of joint model using a marginal conditional approach is not a new one, the earlier works on such model mainly focused on estimating the transition probability or testing dependence among repeated measures. Islam and Chowdhury [38] described the estimation procedure of the parameters and the test procedures for overall model. However, pertinence of such a model for analysing longitudinal binary data is completely neglected in literature. We examined the performance of the model in terms of bias of the estimates and coverage probability, compared the estimates of parameters of the joint model

with the estimates of parameters of GEE and ALR and proposed the application of joint model based on marginal conditional approach for analysis of longitudinal binary data as a better alternative to GEE or ALR. In this Chapter, for analysis of longitudinal data we propose two different strategies:

- Joint models based on marginal-conditional approach
- Joint model based on extended Regressive Model approach.

The proposed methods can be used for outcome variables with known distribution using likelihood based methods. The proposed joint models based on marginal conditional approach for repeated binary outcomes are generalized for exponential family. The limitation of the conditional probability based model for more than 3 repeated measures is discussed and the extended regressive model is shown as a better alternative for such cases.

5.2 Conditional Models for Repeated Binary Data

Let Y_{ij} be a time dependent binary outcome variable for subject i at time j, $i = 1,2,...,N$ and $j = 1,2,...,n_i$. The outcome vector for subject i can be defined as $\mathbf{Y}_i = (Y_{i1}\ Y_{i2},...,Y_{in_i})'$ with mean

$$\boldsymbol{\mu}_i = E(\mathbf{Y}_i) = (E(Y_{i1})\ E(Y_{i2})\ ...\ E(Y_{in_i}))' = (\mu_{i1}\ \mu_{i2}\ ...\ \mu_{in_i})'.$$

Suppose, $P(Y_{i1} = y_{i1})$ denote the marginal distribution of the outcome variable Y_{i1} and $P(Y_{ij} = y_{ij}|\boldsymbol{X}_i = \boldsymbol{x}_i, Y_{i1} = y_{i1}, Y_{i2} = y_{i2},...,Y_{ij-1} = y_{ij-1})$ denote the conditional distribution of Y_{ij} given $Y_{i1} = y_{i1},...,Y_{i2} = y_{i2}, Y_{ij-1} = y_{ij-1}$, $j = 2,3,...,n_i$. Also let $\boldsymbol{X}_{ij}$ be the $p \times 1$ vector of covariates for subject i at jth occasion. If $\mathbf{Y}_i$ is covariate dependent, then the marginal distribution of the outcome variable Y_{i1} can be defined as $P(Y_{i1} = y_{i1}|\boldsymbol{X}_{i1} = \boldsymbol{x}_{i1})$ and the conditional distribution of Y_{ij} given $Y_{i1} = y_{i1},...,Y_{i2} = y_{i2}, Y_{ij-1} = y_{ij-1},\ j = 2,3,...,n_i$, is denoted by $P(Y_{ij} = y_{ij}|\boldsymbol{X}_{ij} = \boldsymbol{x}_{ij}, Y_{i1} = y_{i1}, Y_{i2} = y_{i2},...,Y_{ij-1} = y_{ij-1})$.

Let $\boldsymbol{\theta} = (\theta_1, \theta_{2.1}, ..., \theta_{n_i.1,2,...,n_i-1})'$ be the vector of unknown parameters where $\theta_j = g(\mu_{ij}) = X_{ij}\boldsymbol{\beta}_j$, $\theta_{j.12...j-1} = g(\mu_{ij.12...j-1}) = \boldsymbol{X}_{ij}\boldsymbol{\beta}_{j.12...j-1}$ and g is an appropriate link function.

For example, for Bernoulli outcome variables Y_{ij}, $i = 1,2,...,N$ and $j = 1,2,...,n_i$, if the Y_{ij}'s are time dependent, the marginal probability of Y_{ij} observing an event can be expressed as

$$p_{ij} = Pr(Y_{ij} = 1|\boldsymbol{x}_{ij}) = \frac{\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}_j)}{1+\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}_j)};\ i = 1,2,...,N; j = 1,2,...,n_i, \quad (5.1)$$

where $\boldsymbol{\beta}_j = (\beta_{j0}, ..., \beta_{jp})'$ is a $(p+1)\times 1$ vector of parameters of the marginal model of Y_{ij}. Consequently, the marginal probability of not observing an event can be expressed as

$$1 - p_{ij} = 1 - Pr(Y_{ij} = 1|\boldsymbol{x}_{ij}) = \frac{1}{1+\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}_j)}.$$

The conditional probability of Y_{ij} observing an event, given $Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, ..., Y_{ij-1} = y_{ij-1}$, can be defined as

$$\begin{aligned} p^*_{ij} &= Pr(Y_{ij} = 1|y_{i1}, ..., y_{ij-1}, \boldsymbol{x}_{ij}) \\ &= \frac{\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}_{j.12...j-1})}{1+\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}_{j.12...j-1})};\ i = 1,2,...,N;\ j = 2,...,n_i, \end{aligned} \quad (5.2)$$

where $\boldsymbol{\beta}_{j.12...j-1}$ is the vector of parameters of the conditional model of $P(Y_{ij} = 1|Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, ..., Y_{ij-1} = y_{ij-1})$; $j = 2,...,n_i - 1$. For each combination of $y_{i1}, ..., y_{in_i-1}$, we get one conditional model. For example, when $n_i = 2$, $\boldsymbol{\beta}_{2.1}$ is the vector of parameters of the two conditional models $P(Y_{i2} = 1|X_{i2} = x_{i2}, Y_{i1} = y_{i1})$, $y_{i1} = 0, 1$. We may denote the vector of parameters of the conditional model, $P(Y_{i2} = 1|X_{i2} = x_{i2}, Y_{i1} = 0)$, as $\boldsymbol{\beta}_{01} = (\beta_{010}, ..., \beta_{01p})'$, and the vector of parameters of the conditional model, $P(Y_{i2} = 1|X_{i2} = x_{i2}, Y_{i1} = 1)$, as $\boldsymbol{\beta}_{11} = (\beta_{110}, ..., \beta_{11p})'$. To distinguish between the vector of parameters of a marginal model and the vector of parameters of a marginal conditional model, from this point onward, we will use two different notations, $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta}$ respec-

tively.

The repeated measures data are naturally correlated and the major challenge of the methods for analysing repeated measures categorical data is to take care of the probable correlations among the repeated observations on the same subject. We start from the model considered by Darlington and Farewell [16] with repeated binary outcomes with the working likelihood function

$$L(\boldsymbol{\beta}^{*},\boldsymbol{\beta}_{11}) = \prod_{i=1}^{N} {p_i}^{y_{i1}}(1-p_i)^{1-y_{i1}} \prod_{j=2}^{n_i} {p^*_{ij}}^{y_{ij}}(1-p^*_{ij})^{1-y_{ij}},$$

$$\begin{aligned} \text{where} \quad p_i &= Pr(Y_{ij}=1|\boldsymbol{X}_i=\boldsymbol{x}_i) = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta}^{*})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta}^{*})}, \qquad (5.3)\\ p^*_{ij} &= Pr(Y_{ij}=1|Y_{ij-1},\boldsymbol{X}_i=\boldsymbol{x}_i) = E(Y_{ij}|Y_{ij-1},\boldsymbol{X}_i=\boldsymbol{x}_i)\\ &= p_i+\rho_i(Y_{ij-1}-p_i) \qquad (5.4) \end{aligned}$$

and

$$\rho_i = \frac{\exp(\boldsymbol{\beta}'_{11}x_i)-\exp(\boldsymbol{x}_i\boldsymbol{\beta}^{*})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta}_{11})}. \qquad (5.5)$$

$\max(-\frac{p_i}{1-p_i},-\frac{1-p_i}{p_i}) < \rho_i < 1$ because the likelihood must be maximized at $0 < p_i < 1$ and $0 < p_{ij} < 1$. The limitations of this model is that it does not consider the transition probability from $Y_{ij-1} = 0$ to $Y_{ij} = 1$ and considered $p^*_{ij} = Pr(Y_{ij} = 1|Y_{ij-1} = 1, \boldsymbol{X}_i)$. Although, the transition probability $P(Y_{ij} = 1|Y_{ij-1} = 0)$ was not considered in determining the correlation, while defining the range of ρ_i, the transition from $Y_{ij-1} = 0$ was considered which contradicts with the definition of ρ_i. A straight forward and simple way to improve the model discussed by Darlington and Farewell [16] by including both the transition probabilities $P(Y_{ij} = 1|Y_{ij-1} = 0)$ and $P(Y_{ij} = 1|Y_{ij-1} = 1)$ in the working likelihood function.

We named our proposed model 1, proposed model 2 and proposed model 3 as marginal conditional model 1 $(MCM1)$, marginal conditional model 2 $(MCM2)$ and marginal conditional model 3 $(MCM3)$, respectively.

5.3 Proposed Marginal Conditional Model 1 (MCM1)

To overcome the limitations of the working likelihood function proposed by Darlington and Farewell [16], an extension of the model by Darlington and Farewell [16] is proposed which is the marginal conditional model 1 (MCM1). For simplicity, let us consider, first order Markov model for the two consecutive binary outcomes in a follow-up study.

5.3.1 Likelihood Function

The working Markov likelihood can be expressed as

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} p_i^{y_{i1}} (1-p_i)^{1-y_{i1}} \prod_{i=1}^{N} \prod_{j=2}^{n_i} p_{ij}^{*\, y_{ij}} (1-p_{ij}^{*})^{1-y_{ij}}, \tag{5.6}$$

where the marginal probability is

$$p_i = Pr(Y_{ij} = 1 | \boldsymbol{x}_{ij}) = \frac{\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}^*)}{1+\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}^*)}, \tag{5.7}$$

where $\boldsymbol{\beta}^*$ is the vector of parameters of the marginal model $P(Y_{ij} = 1|\boldsymbol{x}_{ij})$ and the conditional probabilities are

$$p_{ij}^{*} = \begin{cases} p_{ij01} = & Pr(Y_{ij} = 1 | Y_{ij-1} = 0, \boldsymbol{x}_{ij}) = p_i + \rho_{i1}(Y_{ij-1} - p_i) \\ p_{ij11} = & Pr(Y_{ij} = 1 | Y_{ij-1} = 1, \boldsymbol{x}_{ij}) = p_i + \rho_{i2}(Y_{ij-1} - p_i), \end{cases} \tag{5.8}$$

$$\text{with } \rho_{i1} = \frac{\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}_{01}) - \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}^*)}{1+\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}_{01})}$$

$$\text{and } \rho_{i2} = \frac{\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}_{11}) - \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}^*)}{1+\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}_{11})},$$

where $\boldsymbol{\beta}_{01}$ and $\boldsymbol{\beta}_{11}$ are the vector of parameters of the conditional models, $P(Y_{ij} = 1|Y_{ij-1} = 0)$ and $P(Y_{ij} = 1|Y_{ij-1} = 1)$, respectively; $(-\frac{1-p_i}{p_i}) < \rho_{1i} < 1$, $p_i \geq 0.5$ and $(-\frac{p_i}{1-p_i}) < \rho_{i2} < 1$, $p_i \leq 0.5$.

5.3.2 Score Equations and Information Matrix

The score equations for the likelihood function can be obtained by differentiating the working log likelihood in equation (5.6) with respect to β_k^*, β_{01k} and β_{11k} respectively and equating to zero as

$$\begin{aligned}
\frac{\delta l}{\delta \beta_k^*} &= \sum_{i=1}^{N} x_{i1k}(y_{i1}-p_i) + \sum_{i=1}^{N}\sum_{j=2}^{n_i} \frac{x_{ijk}(y_{ij}-p_{ij}^*)}{p_{ij}^*(1-p_{ij}^*)}\{(1-\rho_{i1})p_i(1-p_i)\} \\
&\quad - \sum_{i=1}^{N}\sum_{j=2}^{n_i} \frac{x_{ijk}(y_{ij}-p_{ij}^*)}{p_{ij}^*(1-p_{ij}^*)}\left\{\frac{\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}^*)}{1+\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}_{01})}(y_{ij-1}-p_i)\right\} = 0, \\
\frac{\delta l}{\delta \beta_{01k}} &= \sum_{i=1}^{N} \frac{x_{i2k}\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01k})}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01k})}(1-\rho_{i1}) \sum_{j=1}^{n_i-1} \frac{(y_{ij}-p_{ij})(y_{ij+1}-p_{ij+1}^*)}{p_{ij+1}^*(1-p_{ij+1}^*)} = 0, \\
\frac{\delta l}{\delta \beta_{11k}} &= \sum_{i=1}^{N} \frac{x_{i2k}\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{11k})}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{11k})}(1-\rho_{i2}) \sum_{j=1}^{n_i-1} \frac{(y_{ij}-p_{ij})(y_{ij+1}-p_{ij+1}^*)}{p_{ij+1}^*(1-p_{ij+1}^*)} = 0.
\end{aligned}$$

However, the MCM1, which is an extension of Darlington and Farewell [16], although considers both the transition probabilities $P(Y_{ij}=1|Y_{ij-1}=1)$ and $P(Y_{ij}=1|Y_{ij-1}=0)$, ignores the fact that the marginal distribution of Y_{ij} may vary with respect to time and hence the model might not be able to estimate the outcome-covariate relationship at different time points except in some special cases (i.e. $\boldsymbol{\beta}_1=\boldsymbol{\beta}_2=\boldsymbol{\beta}$).

5.3.3 Test of Hypothesis

Since estimation of parameters of proposed model 1 (MCM1) is based on likelihood based procedures, a likelihood ratio test can be used for testing the significance of the model parameters. The individual parameters can be tested using a Wald test.

5.4 Proposed Marginal Conditional Model 2 (MCM2)

For any order of Markov chain with covariate dependence, the proposed marginal conditional model 2 (MCM2) is a further generalization of proposed marginal conditional model 1 (MCM1).

Suppose Y_{ij} be a time dependent outcome variable for subject i at time j, $i = 1,2,...,N$ and $j = 1,2,...,n_i$. The outcome vector for subject i can be defined as $\boldsymbol{Y}_i = (Y_{i1}\ Y_{i2},...,Y_{in_i})'$ with mean

$$\boldsymbol{\mu}_i = E(\boldsymbol{Y}_i) = (E(Y_{i1})\ E(Y_{i2})\ ...\ E(Y_{in_i}))' = (\mu_{i1}\ \mu_{i2}\ ...\ \mu_{in_i})'.$$

Also let $\boldsymbol{X}_{ij}$ be the $p \times 1$ vector of covariates for subject i at jth occasion. Suppose, $P(Y_{i1} = y_{i1})$ denote the marginal distribution of the outcome variable Y_{i1} and $P(Y_{ij} = y_{ij}|Y_{i1} = y_{i1}, Y_{i2} = y_{i2},..., Y_{ij-1} = y_{ij-1})$ denote the conditional distribution of Y_{ij} given $Y_{i1} = y_{i1},..., Y_{i2} = y_{i2}, Y_{ij-1} = y_{ij-1}$, $j = 2,3,...,n_i$. Let $\boldsymbol{\theta} = (\theta_1, \theta_{2.1},..., \theta_{n_i.1,2,...,n_i-1})$ be the vector of unknown parameters where $\theta_1 = g(\mu_{i1}) = X_{i1}\boldsymbol{\beta}_1$, and $\theta_{j.12...j-1} = g(\mu_{ij.12...j-1}) = \boldsymbol{X}_{ij}\boldsymbol{\beta}_{j.12...j-1}$ and g is an appropriate link function.

If $\boldsymbol{Y}_i$ is covariate dependent, the joint distribution of Y_{i1}, Y_{i2}, ..., Y_{in_i} can be expressed as

$$\begin{aligned} &P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2},..., Y_{in_i} = y_{in_i}|\boldsymbol{X}_i = \boldsymbol{x}_i) \\ &= P(Y_{i1} = y_{i1}|\boldsymbol{x}_{i1}).P(Y_{i2} = y_{i2}|y_{i1}, \boldsymbol{x}_{i2}) \\ &\ ...\ P(Y_{in_i} = y_{in_i}|y_{i1}, y_{i2},..., y_{in_i-1}, \boldsymbol{x}_{in_i}). \end{aligned} \tag{5.9}$$

5.4.1 Likelihood and Log-likelihood Function

The likelihood function can be expressed as

$$\begin{aligned} L(\boldsymbol{\beta}) \ &= \ \prod_{i=1}^{N} f(y_{i1}|\boldsymbol{x}_{i1}, \boldsymbol{\beta}_1) f(y_{i2.1}|\boldsymbol{x}_{i2}, \boldsymbol{\beta}_{2.1}) \\ &\quad ...\ f(y_{n_i.1,2,\ ...,\ n_i-1}|\boldsymbol{x}_{in_i}, \boldsymbol{\beta}_{n_i.1,2,\ ...,\ n_i-1}), \end{aligned} \tag{5.10}$$

where, $f(y_{i1}|\boldsymbol{x}_{i1},\boldsymbol{\beta}_1)$ is the marginal distribution of y_{i1} for given $\boldsymbol{X}_{i1}=\boldsymbol{x}_{i1}$ and the conditional probabilities of y_{ij}, given $Y_{i1}=y_{i1},...,Y_{ij-1}=y_{ij-1}$, and $\boldsymbol{X}_{ij}=\boldsymbol{x}_{ij}$ are $f(y_{ij.1...j-1})=f(y_{ij}|\boldsymbol{x}_{ij},y_{i1},...,y_{ij-1},\boldsymbol{\beta}_{j.1,...,j-1})$, $j=2,3,...,n_i$ and $\boldsymbol{\beta}=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_{2.1},...,\boldsymbol{\beta}_{n_i.12...n_i-1})$. Let l_{ij} be the contribution of ijth term to the log likelihood function. Then the log likelihood is $l=\sum_{i=1}^{N}\sum_{j=1}^{n_i}l_{ij}$.

5.4.2 Score Equations and Information Matrix

Differentiating the log-likelihood, $l=\sum_{i=1}^{N}\sum_{j=1}^{n_i}l_{ij}$, with respect to corresponding parameters, and equating to zero, the estimating equations are

$$\frac{\delta l}{\delta\beta_k}=\sum_{i=1}^{N}\sum_{j=1}^{n_i}\frac{\delta l_{ij}}{\delta\theta_j}\frac{\delta\theta_j}{\delta\mu_{ij}}.\frac{\delta\mu_{ij}}{\delta\beta_k}=0. \tag{5.11}$$

The estimates of $\boldsymbol{\beta}$ can be obtained by maximum likelihood method. The variance of the estimates, $V(\hat{\boldsymbol{\beta}})$, is obtained from the inverse of the information matrix I, where I is a $(2^{n_i}-1)(p+1)\times(2^{n_i}-1)(p+1)$ matrix with kk'th elements $-\frac{\delta^2 l}{\delta\beta_k\delta\beta_k'};k,k'=0,1,...,p$.

Example 1: Proposed MCM2 for Binary Outcome Variables, $n_i=2$

Let Y_{ij} be a time dependent binary outcome variable for subject i at time j, $i=1,2,...,N$ and $j=1,2$. Then the outcome vector for subject i can be defined as $\boldsymbol{Y}_i=(Y_{i1}\ Y_{i2})'$ with mean vector $\boldsymbol{\mu}_i=E(\boldsymbol{Y}_i)=(E(Y_{i1})\ E(Y_{i2}))'=(\mu_{i1}\ \mu_{i2})'=(p_{i1}\ p_{i2})'$. Also let $\boldsymbol{X}_{ij}$ be the $p\times 1$ vector of covariates for subject i at jth occasion. The canonical parameters θ_1 and $\theta_{2.1}$ are the logit link functions, where

$$\theta_1=g(\mu_{i1})=ln\frac{\mu_{i1}}{1-\mu_{i1}}=\boldsymbol{X}_{i1}\boldsymbol{\beta}_1$$
$$\text{and}\ \ \theta_{2.1}=g(\mu_{i2.1})=ln\frac{\mu_{i2.1}}{1-\mu_{i2.1}}=\boldsymbol{X}_{i2}\boldsymbol{\beta}_{2.1}.$$

The joint distribution of Y_{i1} and Y_{i2},as defined in equation (5.9) is

$$\begin{aligned}&P(Y_{i1}=y_{i1},Y_{i2}=y_{i2}|\boldsymbol{X}_i=\boldsymbol{x}_i)\\&=P(Y_{i1}=y_{i1}|\boldsymbol{X}_{i1}=\boldsymbol{x}_{i1}).P(Y_{i2}=y_{i2}|Y_{i1}=y_{i1},\boldsymbol{X}_{i2}=\boldsymbol{x}_{i2}).\end{aligned}$$

The marginal probabilities can be defined as

$$\begin{aligned}P(Y_{i1}=1|\boldsymbol{X}_{i1}=\boldsymbol{x}_{i1}) &= \frac{\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}{1+\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}\\ \text{and } P(Y_{i1}=0|\boldsymbol{X}_{i1}=\boldsymbol{x}_{i1}) &= \frac{1}{1+\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)},\end{aligned}$$

where $\boldsymbol{\beta}_1=(\beta_{10},\beta_{11},...,\beta_{1p})'$.

The conditional probabilities considering covariate vector $\boldsymbol{X}_{ij}$, can be expressed in terms of logit link functions as

$$\begin{aligned}P(Y_{i2}=1|Y_{i1}=0,\boldsymbol{X}_{i2}=\boldsymbol{x}_{i2}) &= \frac{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01})}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01})},\\ P(Y_{i2}=1|Y_{i1}=1,\boldsymbol{X}_{i2}=\boldsymbol{x}_{i2}) &= \frac{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{11})}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{11})},\\ P(Y_{i2}=0|Y_{i1}=0,\boldsymbol{X}_{i2}=\boldsymbol{x}_{i2}) &= \frac{1}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{00})}\\ \text{and } P(Y_{i2}=0|Y_{i1}=1,\boldsymbol{X}_{i2}=\boldsymbol{x}_{i2}) &= \frac{1}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01})},\end{aligned}$$

where $\boldsymbol{\beta}_{01}=(\beta_{010},\beta_{011},...,\beta_{01p})'$ denote the vector of the regression parameters of the conditional model with transition from $Y_{i1}=0$ to $Y_{i2}=1$ and $\boldsymbol{\beta}_{11}=(\beta_{110},\beta_{111},...,\beta_{11p})'$ denote the vector of regression parameters of the conditional model with transition from $Y_{i1}=1$ to $Y_{i2}=1$. The joint probabilities are

$$\begin{aligned}P(Y_{i2}=1|Y_{i1}=0,\boldsymbol{x}_{i2})P(Y_{i1}=0|\boldsymbol{x}_{i1}) &= \tfrac{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01})}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01})}\cdot\tfrac{1}{1+\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)},\\ P(Y_{i2}=1|Y_{i1}=1,\boldsymbol{x}_{i2})P(Y_{i1}=1|\boldsymbol{x}_{i1}) &= \tfrac{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{11})}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{11})}\cdot\tfrac{\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}{1+\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)},\\ P(Y_{i2}=0|Y_{i1}=0,\boldsymbol{x}_{i2})P(Y_{i1}=0|\boldsymbol{x}_{i1}) &= \tfrac{1}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{00})}\cdot\tfrac{1}{1+\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}\\ \text{and}\quad P(Y_{i2}=0|Y_{i1}=1,\boldsymbol{x}_{i2})P(Y_{i1}=1|\boldsymbol{x}_{i1}) &= \tfrac{1}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01})}\cdot\tfrac{\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}{1+\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}.\end{aligned}$$

Likelihood and Log-likelihood Function

Let us denote, $Y_{i2} = y_{i00}$ when $Y_{i1} = 0$ and $Y_{i2} = 0$, $Y_{i2} = y_{i01}$ when $Y_{i1} = 0$ and $Y_{i2} = 1$, $Y_{i2} = y_{i10}$ when $Y_{i1} = 1$ and $Y_{i2} = 0$ and $Y_{i2} = y_{i11}$ when $Y_{i1} = 1$ and $Y_{i2} = 1$. Then the likelihood function for the parameters $\boldsymbol{\beta}$ can be expressed as

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^{N} P(Y_{i1}|\boldsymbol{X}_{i1} = \boldsymbol{x}_{i1}) P(Y_{i2}|Y_{i1}, \boldsymbol{X}_{i2} = \boldsymbol{X}_{i2}) \\ &= \prod_{i=1}^{N} \left(\frac{\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}{1+\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}\right)^{y_{i1}} \left(\frac{1}{1+\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}\right)^{1-y_{i1}} \\ &\quad \left(\frac{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01})}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01})}\right)^{y_{i01}} \left(\frac{1}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01})}\right)^{y_{i00}} \\ &\quad \left(\frac{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{11})}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{11})}\right)^{y_{i11}} \left(\frac{1}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{11})}\right)^{y_{i10}}. \end{aligned} \tag{5.12}$$

The log likelihood function takes the following form

$$\begin{aligned} l &= \sum_{i=1}^{N} [\{y_{i1}\boldsymbol{x}_{i1}\boldsymbol{\beta}_1 - ln(1+\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1))\} \\ &\quad +\{y_{i01}\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01} - (y_{i00}+y_{i01})ln(1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01}))\} \\ &\quad +\{y_{i11}\boldsymbol{x}_{i2}\boldsymbol{\beta}_{11} - (y_{i10}+y_{i11})ln(1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{11}))\}]. \end{aligned} \tag{5.13}$$

Score Equations and Information Matrix

The score equations can be obtained by differentiating the log likelihood function (5.13) with respect to the respective parameters. The score equations for the marginal model can be expressed as

$$\frac{\delta l}{\delta \beta_{1k}} = \sum_{i=1}^{N} x_{i1k}\left[y_{i1} - \frac{\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}{1+\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}\right] = 0, \;\; k = 0, 1, ..., p. \tag{5.14}$$

The score equations for the conditional models can be expressed as

$$\begin{aligned} \frac{\delta l}{\delta \beta_{u1k}} = \sum_{i=1}^{N}\left[x_{i2k}y_{iu1} - (y_{iu0}+y_{iu1})\frac{x_{i2k}\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{u1})}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{u1})}\right] = \;\; 0; \\ u = 0, 1; \;\; k = 0, 1, ..., \;\; p. \end{aligned} \tag{5.15}$$

Solving the score equations (5.14) and (5.15) iteratively, the estimates of $\beta_{1k}, k = 0,1,...,p$ and $\beta_{u1k},\ u = 0,1;\ k = 0,1,...,p$, can be obtained.

Elements of the variance covariance matrix can be obtained from the inverse of the observed information matrix using the second derivatives

$$\begin{aligned} -\frac{\delta^2 l}{\delta\beta_{1k}\delta\beta_{1k'}} &= \sum_{i=1}^{N} x_{i1k}x_{i1k'}\frac{\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}{[1+\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)]^2};\ k = 1,2,...,p, \\ \text{and}\quad -\frac{\delta^2 l}{\delta\beta_{u1k}\delta\beta_{u1k'}} &= \sum_{i=1}^{N}(y_{iu0}+y_{iu1})x_{i2k}x_{i2k'}\cdot\frac{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{u1})}{[1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{u1})]^2}; \\ & \qquad k = 1,2,...,p;\ \ u = 0,1. \quad (5.16) \end{aligned}$$

Diagonal elements of the information matrix are obtained when $k = k'$.

Example 2: Proposed MCM2 for Binary Outcome Variables, $n_i > 2$

Consider possibly correlated Binary outcome variables $Y_{i1},...,Y_{in_i}$, with probability of success $p_{i1}, p^*_{i2},..., p^*_{in_i}$, where p_{i1} is the marginal probability $P(Y_{ij} = 1|\boldsymbol{x}_{ij})$ p^*_{ij} denotes the conditional probability, $P(Y_{ij} = 1|y_{i1},...,y_{ij-1},\boldsymbol{x}_{ij})$, $j = 2,...,n_i$. The marginal distribution of Y_{i1} is

$$f(y_{i1}|\boldsymbol{x}_{i1},\boldsymbol{\beta}_1) = p_{i1}^{y_{i1}}(1-p_{i1})^{1-y_{i1}}$$

and the conditional distribution of Y_{ij} given $Y_{i1} = y_{i1},...,Y_{ij-1} = y_{ij-1}$ is

$$f(y_{ij.12...j-1}|\boldsymbol{x}_{ij},\boldsymbol{\beta}_{j.12...j-1}) = {p^*}_{ij}^{y_{ij}}(1-{p^*}_{ij})^{1-y_{ij}},\ j = 2,...,n_i.$$

We may write,

$$\begin{aligned} f(y_{i1}|\boldsymbol{x}_{i1},\theta_1,\phi_1) &= p_{i1}^{y_{i1}}(1-p_{i1}^{1-y_{i1}}) \\ &= \exp(y_{i1}lnp_{i1}+(1-y_{i1})ln(1-p_{i1})) \\ &= \exp(y_{i1}ln\frac{p_{i1}}{1-p_{i1}}-(-ln(1-p_{i1}))) \\ &= \exp\left(\frac{[y_{i1}\theta_1-b(\theta_1)]}{a(\phi_1)}+c(y_{i1},\phi_1)\right), \end{aligned}$$

$$
\begin{aligned}
\text{where} \qquad \theta_1 &= ln\frac{p_{i1}}{1-p_{i1}} = \beta_{10} + \beta_{11}x_{i11} + ... + \beta_{1p}x_{i1p}, \\
p_{i1} &= \frac{\exp(\theta_1)}{1+\exp(\theta_1)}, \\
b(\theta_1) &= -ln(1-p_{i1}) = ln(1+\exp(\theta_1)) \\
a(\phi_1) &= 1, \\
c(y_{i1},\phi_1) &= 0, \\
E(Y_{i1}) &= b'(\theta_1) = \frac{\exp(\theta_1)}{1+\exp(\theta_1)} = p_{i1}, \\
\text{and } V(Y_{i1}) &= a(\phi_{i1})b''(\theta_1) = p_{i1}(1-p_{i1}).
\end{aligned}
$$

For $j = 2, ..., n_i - 1$, the conditional distribution of Y_{ij} is expressed as

$$
\begin{aligned}
& f(y_{ij}|\boldsymbol{X}_{ij}, y_{i1}, y_{i2}, ..., y_{ij}, \theta_{j.1,2,...,j-1}, \phi_j) \\
&= p_{ij}^{y_{ij}}(1-p_{ij}^{1-y_{ij}}) \\
&= \exp[y_{ij} ln p_{ij} + (1-y_{ij}) ln(1-p_{ij})] \\
&= \exp[y_{ij} ln\frac{p_{ij}}{1-p_{ij}} - (-ln(1-p_{ij}))] \\
&= \exp\frac{y_{ij}\theta_{j.1,2,...,j-1} - b(\theta_{j.1,2,...,j-1})}{a(\phi_j)} + c(y_j\phi_j).
\end{aligned}
$$

We may define, $\forall j = 2, 3, ..., n_i$,

$$
\begin{aligned}
\theta_{j.1,2,...,j-1} &= ln\frac{p_{ij}}{1-p_{ij}}, \\
p_{ij} &= \frac{\exp(\theta_{j.1,2,...,j})}{1+\exp(\theta_{j.1,2,...,j-1})}, \\
b(\theta_{j.1,2,...,j-1}) &= -ln(1-p_{ij}) \\
&= ln(1+\exp(\theta_{j.1,2,...,j-1})), \\
a(\phi_j) &= 1, \\
\text{and } c(y_{ij},\phi_j) &= 0.
\end{aligned}
$$

Mean and variance of Y_{ij}, respectively, can be expressed as

$$\begin{aligned} E(Y_{ij}|y_{i1},y_{i2},...,y_{ij-1}) &= b'(\theta_{j.1,2,...,j-1}) = \frac{\exp(\theta_{j.1,2,...,j-1})}{1+\exp(\theta_{j.1,2,...,j-1})} \\ \text{and} \quad Var(Y_{ij}|y_{i1},y_{i2},...,y_{ij-1}) &= a(\phi_j)b''(\theta_{j.1,2,...,j-1}) = p_{ij}(1-p_{ij}). \end{aligned}$$

Likelihood and Log-likelihood Function

The likelihood function of $\boldsymbol{\theta}$ can be expressed as

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{N} f(y_{i1},y_{i2},...,y_{in_i}|\theta_1,\theta_{2.1},...,\theta_{n_{i.1,2,...,n_i-1}},\phi_1,\phi_2,...,\phi_{n_i}) \\ &= \prod_{i=1}^{N} f(y_{i1}|\boldsymbol{X}_{i1},\theta_1,\phi_1)\prod_{j=2}^{n_i} f(y_{ij}|\boldsymbol{X}_{ij},y_{i1},...,y_{ij-1},\theta_{j.1,2,...,j-1},\phi_j) \\ &= \prod_{i=1}^{N} \exp\{y_{i1} ln\frac{p_{i1}}{1-p_{i1}} + y_{i2} ln\frac{p^*_{i2}}{1-p^*_{i2}} + ... + y_{in_i} ln\frac{p^*_{in_i}}{1-p^*_{in_i}}\}. \end{aligned} \tag{5.17}$$

The log likelihood function takes the following form

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{N}\{y_{i1}\theta_1 - b(\theta_1)\} + \sum_{i=1}^{N}\sum_{j=2}^{n_i}\{y_{ij}\theta_{j.12...j-1} - b(\theta_{j.1,2,...,j-1})\} \\ &= \sum_{i=1}^{N}\left[y_{i1} ln\frac{p_{i1}}{1-p_{i1}} + y_{i2} ln\frac{p^*_{i2}}{1-p^*_{i2}} + ... + y_{in_i} ln\frac{p^*_{in_i}}{1-p^*_{in_i}}\right], \end{aligned} \tag{5.18}$$

where θ is a function of regression parameters $\boldsymbol{\beta}$. The estimates of the required regression parameters can be obtained by ML method from the above likelihood function. The score equations for $\boldsymbol{\beta}$ can be obtained by differentiating the log likelihood with respect to the respective parameters.

Score Equations and Information Matrix

Differentiating the log-likelihood in equation (5.18) with respect to the respective parameters and equating to zero, score equations for β are obtained as

$$\frac{\delta l}{\delta \beta_k} = \sum_{i=1}^{N} X_{i1k}(y_{i1} - p_{i1}) + \sum_{i=1}^{N}\sum_{j=2}^{n_i} X_{ijk}(y_{ij} - p^*_{ij}),\ k = 0, 1, ..., p. \qquad (5.19)$$

The information matrix I_β is a $(2^{n_i} - 1)(p+1) \times (2^{n_i} - 1)(p+1)$ matrix with elements $-\frac{\delta^2 l}{\delta \beta_k \delta \beta'_k};\ k, k' = 0, 1, ..., p.$

Example of Proposed Model MCM2: Response Variables belong to Exponential Family with $n_i = 2$

For simplicity, on each of N individuals, consider two possibly correlated outcome variables (Y_{i1}, Y_{i2}) (i.e. $i = 1, ..., N$ and $j = 1, 2$). Suppose Y_{i1} and Y_{i2} given $Y_{i1} = y_{i1}$ are known to belong to an exponential family. Then following equation (5.9), the joint density of the repeated outcomes of the ith subject can be expressed as

$$\begin{aligned} & f(y_{i1}, y_{i2} | \boldsymbol{X}_i = \boldsymbol{x}_i, \theta_1, \theta_{2.1}, \phi_1, \phi_2) \\ &= f(y_{i1} | \boldsymbol{x}_{i1}, \theta_1, \phi_1) f(y_{i2} | \boldsymbol{x}_{i2}, y_{i1}, \theta_{2.1}, \phi_2) \\ &= f(y_{i1}) f(y_{i2.1}) \text{ (say)}, \end{aligned} \qquad (5.20)$$

where

$$\begin{aligned} f(y_{i1}) &= f(y_{i1} | x_{i1}, \theta_1, \phi_1), \\ f(y_{i2.1}) &= f(y_{i2} | x_{i2}, y_{i1}, \theta_{2.1}, \phi_2), \\ \theta_1 &= g(\mu_1), \\ \theta_{2.1} &= g(\mu_{2.1}), \\ (\mu_1\ \mu_{2.1})' &= (E(Y_{i1})\ E(Y_{i2} | Y_{i1} = y_{i1}))', \\ \phi_1, \phi_2 &= \text{dispersion parameters}, \\ \text{and } g &= \text{an appropriate link function.} \end{aligned}$$

The Mean and the Variance of Marginal and Conditional Distributions

If Y_{i1} belongs to an exponential family with parameters θ_1 and ϕ_1, then the marginal distribution of y_{i1} can be expressed as

$$f(y_{i1}) = f(y_1|x_{i1},\theta_1,\phi_1) = \exp\Big[\frac{\{y_{i1}\theta_1 - b(\theta_1)\}}{a(\phi_1)} + c(y_{i1},\phi_1)\Big], \qquad (5.21)$$

$$\begin{aligned} \text{where} \qquad E(Y_{i1}) &= b'(\theta_1), \\ V(Y_{i1}) &= a(\phi_{i1})b''(\theta_1), \\ \theta_1 &= g(\mu_{i1}) = \beta_{10} + \beta_{11}x_{i11} + ... + \beta_{1p}x_{i1p}. \end{aligned}$$

Here β_{10} is the intercept and $\beta_{11},...,\beta_{1p}$ are the coefficients of the covariates $X_{i11},...,X_{i1p}$ respectively.

If Y_{i2} given $Y_{i1} = y_{i1}$ belongs to an exponential family with parameters $\theta_{2.1}$ and ϕ_2, then the distribution of y_{i2} given y_{i1} can be expressed as

$$\begin{aligned} f(y_{i2.1}) &= f(y_{i2}|x_{i2},y_1,\theta_{2.1},\phi_{i2}) \\ &= f(y_{i2.1},\theta_{2.1},\phi_{ij}) \\ &= \exp\Big[\frac{\{y_{i2.1}\theta_{2.1} - b(\theta_{2.1})\}}{a(\phi_{i2})} + c(y_{i2.1},\phi_{i2})\Big]. \qquad (5.22) \end{aligned}$$

We may define,

$$\begin{aligned} E(Y_{i2}|Y_{i1} &= y_{i1}) = E(Y_{i2.1}) = b'(\theta_{2.1}) \\ \text{and } Var(Y_{i2}|Y_{i1} &= y_{i1}) = Var(Y_{i2.1}) = a(\phi_{i2})b''(\theta_{2.1}), \end{aligned}$$

$$\text{where } \theta_{2.1} = g(\mu_{i2.1}) = \beta_{2.10} + \beta_{2.11}x_{i21} + ... + \beta_{2.1p}x_{i2p}.$$

So the mean and variance of $\mathbf{Y}_i$ can be shown, respectively, as

$$E(\mathbf{Y}_i) = (\mu_{i1} \;\; \mu_{i2.1})'$$

and

$$V(\mathbf{Y}_i) = \begin{pmatrix} a(\phi_{i1})V(\mu_{i1}) & 0 \\ 0 & a(\phi_{i2})V(\mu_{i2.1}) \end{pmatrix}.$$

We may define, $\theta_1 = \beta_{10} + \beta_{11}x_{i11} + ... + \beta_{1p}x_{i1p}$; and we may denote $\theta_{2.1} = \beta_{2.1,0} + \beta_{2.1,1}x_{i21} + ... + \beta_{2.1,p}x_{i2p}$. Clearly β_1 is the set of parameters of the marginal part of the model and $\beta_{2.1}$ is the set of parameters of the conditional part of the model.

Likelihood and Log-likelihood Function

The likelihood function can be expressed as

$$\begin{aligned} L &= \prod_{i=1}^{N} f(y_{i1}, y_{i2}|\theta_1, \theta_{2.1}, \phi_1, \phi_2) \\ &= \prod_{i=1}^{N} f(y_{i1}|X_{i1}, \theta_1, \phi_1) f(y_{i2.1}|X_{i2}, y_{i1}, \theta_{2.1}, \phi_{i2}) \\ &= exp \sum_{i=1}^{N} \Big[\frac{y_{i1}\theta_1 - b(\theta_1)}{a(\phi_1)} + c(y_{i1}\phi_1) + \frac{y_{i2}\theta_{2.1} - b(\theta_{2.1})}{a(\phi_2)} + c(y_{i2.1}\phi_2) \Big]. \end{aligned}$$

The log likelihood function takes the following form

$$l(\theta) = \sum_{i=1}^{N} \Big[\frac{y_{i1}\theta_1 - b(\theta_1)}{a(\phi_1)} + c(y_{i1}\phi_1) + \frac{y_{i2.1}\theta_{2.1} - b(\theta_{2.1})}{a(\phi_2)} + c(y_{i2.1}\phi_2) \Big]. \tag{5.23}$$

Score Equations and Information Matrix

Differentiating the log likelihood function (5.23) with respect to the respective parameters and equating to zero, we get the score equations. Solution of score equations gives the estimates of the parameters. The kk' th elements of the information matrix is $-\frac{\delta^2 l}{\delta\beta_k \delta\beta_k'}$.

Example of Proposed Model MCM2: Outcome Variable belongs to Exponential Family, $n_i > 2$

If Y_{i1}, Y_{i2} given $Y_{i1} = y_{i1}$, Y_{i3} given $Y_{i1} = y_{i1}, Y_{i2} = y_{i2}$, ..., $Y_{in_i} = y_{in_i}$ given $Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, ..., Y_{in_i-1} = y_{in_i-1}$ belong to an exponential family, then the joint density of repeated outcomes for subject i is

$$\begin{aligned}
&f(y_{i1}, y_{i2}, ..., y_{in_i}|X, \theta_1, \theta_{2.1}, ..., \theta_{n_i.1,2,...,n_i-1}, \phi_1, \phi_2, ..., \phi_{n_i}) \\
&= f(y_{i1}|x_{i1}, \theta_1, \phi_1) f(y_{i2}|x_{i2}, y_{i1}, \theta_{2.1}, \phi_2) ... f(y_{in_i}|x_{n_i}, y_{i1}, \\
&\qquad\qquad ..., y_{in_i-1}, \theta_{n_i.1,2,...,n_i-1}, \phi_{n_i}) \\
&= f(y_{i1}).f(y_{i2.1}).....f(y_{in_i.1,...,n_i-1}). \qquad (5.24)
\end{aligned}$$

Mean and Variance of the Marginal and the Conditional Distributions

The marginal distribution of y_{i1} is given by

$$f(y_{i1}) = f(y_1|x_{i1}, \theta_1, \phi_1) = \exp\Big[\frac{\{y_{i1}\theta_1 - b(\theta_1)\}}{a(\phi_1)} + c(y_{i1}, \phi_1)\Big]. \qquad (5.25)$$

$E(Y_{i1}) = b'(\theta_1)$, $Var(Y_{i1}) = a(\phi_{i1})b''(\theta_1)$ and and if β_{10} is the intercept and $\beta_{11}, ..., \beta_{1p}$ are the coefficients of the covariates $X_{i11}, ..., X_{i1p}$ respectively, then $\theta_1 = g(\mu_1) = \beta_{10} + \beta_{11}x_{i11} + ... + \beta_{1p}x_{i1p}$.

For $j = 2, ..., n_i$, the conditional distribution of Y_{ij} given $Y_{i1}, ..., Y_{ij-1}$ can be expressed as

$$\begin{aligned}
&f(y_{ij.1,...,j-1}) \\
&= f(y_{ij}|x_{ij}, y_{i1}, y_{i2}, ..., y_{ij-1}, \theta_{j.1,2,...,j-1}, \phi_{ij}) \\
&= f(y_{ij.1,2,...,j-1}, \theta_{j.1,...,j-1}, \phi_{ij}) \\
&= \exp\left(\frac{y_{ij.1,2,...,j-1}\theta_{j.1,2,...,j-1} - b(\theta_{j.1,2,...,j-1})}{a(\phi_{ij})} + c(y_{ij.1,2,...,j-1}, \phi_{ij})\right). \qquad (5.26)
\end{aligned}$$

We may define, $\forall\ j = 2, 3, ..., n_i$,

$$\begin{aligned}
\mu_{ij.12...j-1} &= E(Y_{ij}|y_{i1}, y_{i2}, ..., y_{ij-1}) = b'(\theta_{j.1,2,...,j-1}), \\
V(Y_{ij}|y_{i1}, y_{i2}, ..., y_{ij-1}) &= V(Y_{ij.1,2,...,j-1}) = a(\phi_{ij})b''(\theta_{j.1,2,...,j-1}), \\
\text{where } \theta_{j.1,2,...,j-1} &= g(\mu_{ij.1,2,...,j-1}).
\end{aligned}$$

Then $V(\boldsymbol{Y}_i)$ can be expressed as

$$V(\boldsymbol{Y}_i) = a \begin{pmatrix} \phi_{i1}V(\mu_{i1}) & 0 & ... & 0 \\ 0 & \phi_{i2}V(\mu_{i2.1}) & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \phi_{in_i}V(\mu_{in_i.1,2,...,n_i-1}) \end{pmatrix}.$$

Likelihood and Log-likelihood Function

The likelihood function can be expressed as

$$\begin{aligned}
L(\boldsymbol{\beta}) &= \prod_{i=1}^{N} f(y_{i1}, y_{i2}, ..., y_{in_i}|\theta_1, \theta_{2.1}, ..., \theta_{n_{i.1,2,...,n_i-1}}, \phi_1, \phi_2, ..., \phi_{n_i}) \\
&= \prod_{i=1}^{N} f(y_{i1}|X_{i1}, \theta_1, \phi_1) \prod_{j=2}^{n_i} f(y_{j.1,2,...,j-1}|X_i, \theta_{j|1,2,...,j-1}, \phi_{in_i}) \\
&= \exp \sum_{i=1}^{N} \left[\frac{y_{i1}\theta_1 - b(\theta_1)}{a(\phi_1)} + c(y_{i1}\phi_1) \right. \\
&\quad \left. + \sum_{j=2}^{n_i} \left(\frac{y_{ij.1,2,...,j-1}\theta_{j.1,2,...,j-1} - b(\theta_{j.1,...,j-1})}{a(\phi_j)} + c(y_{ij.1,...,j-1}\phi_j) \right) \right].
\end{aligned}$$

The log likelihood function takes the following form

$$\begin{aligned}
l(\theta) &= \sum_{i=1}^{N} \left[\frac{y_{i1}\theta_1 - b(\theta_1)}{a(\phi_1)} + c(y_{i1}\phi_1) + \frac{y_{i2}\theta_{2.1} - b(\theta_{2.1})}{a(\phi_2)} + c(y_{i2.1}\phi_2)... \right. \\
&\quad \left. + \frac{y_{in_i.1,...,n_i-1}\theta_{n_i.1,...,n_i-1} - b(\theta_{n_i.1,...,n_i-1})}{a(\phi_{n_i})} + c(y_{in_i.1,...,n_i-1}\phi_{n_i}) \right].
\end{aligned} \tag{5.27}$$

Score Equations and Information Matrix

The score equations are obtained by differentiating the log likelihood equation with respect to the respective parameters. Differentiating the score equations with respect to the parameters give the information matrix.

5.4.3 Tests for the Proposed Model MCM2

The test for the significance of the proposed model MCM2 is straightforward using a likelihood ratio test. The individual parameters can be tested using Wald test.

Test for Overall Model

To test the significance of the overall model, the null and alternative hypothesis can be expressed as

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta_0} \text{ vs } H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta_0},$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2.1}, ..., \boldsymbol{\beta}_{n_i.1,2,...n_i-1})$ and $\boldsymbol{\beta}_0$ is the value of $\boldsymbol{\beta}$ under null hypothesis of no covariate effect. The test statistic,

$$\Lambda = -2[\ln L(\boldsymbol{\beta}_0) - \ln L(\boldsymbol{\beta})],$$

has a chi-square distribution under H_0 with $(2^n - 1)p$ d.f. where $n = \max(n_i), i = 1, ..., N$. Here $\ln L(\boldsymbol{\beta})$ is the log likelihood of the full model and $\ln L(\boldsymbol{\beta_0})$ is the log likelihood of the reduced model for no covariate effects, i.e. the value of $\ln L(\boldsymbol{\beta})$ under H_0.

For example, for Bernoulli outcome variables with $n_i = 2$, the hypotheses for the overall model are

$$H_0 : \boldsymbol{\beta}^* = (\boldsymbol{\beta}_1^*, \boldsymbol{\beta}_{01}^*, \boldsymbol{\beta}_{11}^*) = 0, \quad \text{vs.} \quad H_1 : \boldsymbol{\beta}^* \neq 0.$$

Here $\boldsymbol{\beta}_1^* = (\beta_{11}, \beta_{12}, ..., \beta_{1p})$,

$\boldsymbol{\beta}_{01}^* = (\beta_{011}, \beta_{012}, ..., \beta_{01p})$, and

$\boldsymbol{\beta}_{11}^* = (\beta_{111}, \beta_{112}, ..., \beta_{11p})$.

Then the quantity, $-2\left[lnL(\beta_{10}, \beta_{010}, \beta_{110}) - lnL(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{01}, \boldsymbol{\beta}_{11})\right]$, can be shown to be asymptotically distributed as χ^2 with $3p$ degrees of freedom.

Test for Individual Parameters

For testing individual parameter in the marginal model, Wald test can be used for the following hypothesis

$H_0 : \beta_{1k} = 0$ vs. $H_1 : \beta_{1k} \neq 0$.

The Wald test statistic for the marginal model is $W = \frac{\hat{\beta}_{1k}}{\hat{se}(\beta_{1k})}$.

For testing parameter in the conditional models, the null and the alternative hypotheses can be defined as

$H_0 : \beta_{u1k} = 0$ vs. $H_1 : \beta_{u1k} \neq 0$, $u = 0, 1$.

The Wald test statistic for the marginal model is $W = \frac{\hat{\beta}_{u1k}}{\hat{se}(\beta_{u1k})}$.

Tests for Dependence among Repeated Outcomes

From the Bivariate Bernoulli Distribution described in section 5.4.2, equality of conditional models hold if $\beta_{01} = \beta_{11}$. If $\beta_{01} \neq \beta_{11}$, indicates dependence of Y_{i2} on Y_{i1}. The odds ratio is

$$OR = \frac{\frac{P_{11}(\boldsymbol{x}_{i2})}{1-P_{11}(\boldsymbol{x}_{i2})}}{\frac{P_{01}(\boldsymbol{x}_i)}{1-P_{01}(x_i)}} = \frac{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{11})}{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_{01})} = \exp\{\boldsymbol{x}_{i2}(\boldsymbol{\beta}_{11} - \boldsymbol{\beta}_{01})\}. \tag{5.28}$$

Islam et al. [40] showed that testing for $H_0 : \boldsymbol{\beta}_{01} = \boldsymbol{\beta}_{11}$ is equivalent to test for the association $OR = 1$ and $\ln(OR) = 0$ where both indicate independence of the two binary outcomes in the presence of covariates. Any departure from $OR = 1$ will measure the extent of dependence, where OR greater than 1 implies a positive association and $OR < 1$, a negative association.

For testing the null hypothesis $H_0 : \boldsymbol{\beta}_{01} = \boldsymbol{\beta}_{11}$, the following test statistics can be used as suggested by Islam et al. [40]

$$\chi^2 = (\hat{\boldsymbol{\beta}}_{01} - \hat{\boldsymbol{\beta}}_{11})'[\hat{Var}(\hat{\boldsymbol{\beta}}_{01} - \hat{\boldsymbol{\beta}}_{11})]^{-1}(\hat{\boldsymbol{\beta}}_{01} - \hat{\boldsymbol{\beta}}_{11}), \tag{5.29}$$

which is distributed asymptotically as chi-square with p degrees of freedom. Another alternative test can be performed from the relationship between the conditional and marginal probabilities for the outcome variable, Y_{i2}. The null and the alternative hypotheses are

$$H_{01} : \boldsymbol{\beta}_{01} = \boldsymbol{\beta}_2 \qquad \text{and} \qquad H_{02} : \boldsymbol{\beta}_{11} = \boldsymbol{\beta}_2.$$

The test statistics are, as suggested by Islam et al. [40]

$$\chi_1^2 = (\hat{\boldsymbol{\beta}}_{01} - \hat{\boldsymbol{\beta}}_2)'[\hat{V}(\hat{\boldsymbol{\beta}}_{01} - \hat{\boldsymbol{\beta}}_2)]^{-1}(\hat{\boldsymbol{\beta}}_{01} - \hat{\boldsymbol{\beta}}_2) \tag{5.30}$$

$$\text{and } \chi_2^2 = (\hat{\boldsymbol{\beta}}_{11} - \hat{\boldsymbol{\beta}}_2)'[\hat{V}(\hat{\boldsymbol{\beta}}_{11} - \hat{\boldsymbol{\beta}}_2)]^{-1}(\hat{\boldsymbol{\beta}}_{11} - \hat{\boldsymbol{\beta}}_2). \tag{5.31}$$

It is noteworthy to mention again that Darlington and Farewell [16] proposed a transition probability model based on the following logit functions with marginal specification

$$P(Y_{i2} = 1|Y_{i1} = 1, \boldsymbol{x}_i) = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta}_{11})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta}_{11})}$$
$$\text{and } P(Y_{i2} = 1|\boldsymbol{x}_i) = \frac{\exp\{\boldsymbol{x}_i\boldsymbol{\beta}\}}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta})}.$$

The Darlington and Farewell [16]'s method did not consider the transition probability, transition from $Y_{i1} = 0$ to $Y_{i2} = 1$, in their model and noted that due to asymmetry this may not be suitable for all applications. The measure of correlation proposed by Darlington and Farewell [16] is

$$\rho_i = corr(Y_{i1}, Y_{i2}|\boldsymbol{x}_i) = \frac{\exp(\boldsymbol{\beta}_{11}\boldsymbol{x}_i) - \exp(\boldsymbol{\beta}\boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta}_{11}\boldsymbol{x}_i)},$$

which can be tested by the corresponding chi-square test shown in the alternative tests as shown in 5.31. However, it is necessary for the independence that all the alternative tests (5.30 and 5.31) should be performed. This test can be

simply extended for more than two follow ups.

The major limitation of the proposed joint model based on Markov transition probability in equation (5.9) is the rapid increase in the number of parameters for increasing number of follow-ups. With n_i follow-ups, the number of parameters to be estimated is as big as $(2^{n_i}-1)(p+1)$ where $p+1$ is the number of covariates.

5.5 Proposed Model 3: A Marginal Conditional Model using Extended Regressive Approach

For Binary outcome variables, the number of parameters of the joint model with two follow-ups is $3\times(p+1)$. The number of parameters of the joint model with three follow-ups for binary outcome variables is $7\times(p+1)$. This, inevitably, shows that when the number of follow-ups increases, the number of parameters in the proposed joint model increases rapidly.

So we propose an alternative using a regressive model approach [7] for such cases where there are more than two follow-ups as an alternative to marginal approach in order to analyze repeated measures data.

The generalized form of the regressive model was proposed by Islam et al. [42]. For Bivariate data, Islam et al. [42] generalized the works of Bonney [7, 8], and Islam and Chowdhury [35] to include both binary outcomes in previous times as well as covariates in the conditional models. We denote this model as marginal conditional model 3 (MCM3).

5.5.1 Framework of the Regressive Model

Consider n_i possibly correlated outcome variables $(Y_{i1}, Y_{i2}, ..., Y_{in_i})$ on ith individual $(i = 1, ..., N$ and $j = 1, 2, ..., n_i)$. Considering the regression model for the conditional probabilities as proposed by Bonney [7], the canonical parame-

ter for Y_{ij}, $j = 2, ..., n_i$, can be defined as

$$\begin{aligned}\theta_{j.1,2,...,j-1} &= g(\mu_{ij.1,2,...,j-1}) \\ &= \beta_0 + \beta_1 x_{ij1} + ... + \beta_p x_{ijp} + \gamma_1 Y_{i1} + ... + \gamma_{j-1} Y_{ij-1},\end{aligned}$$

where β_0 is the intercept, $\beta_1, \beta_2, ..., \beta_p$ are the coefficients of the covariates, $X_{ij1}, X_{ij2}, ..., X_{ij1}$, respectively and $\gamma_1, \gamma_2, ..., \gamma_{j-1}$ are the coefficients of the previous outcomes $Y_{i1}, Y_{i2}, ..., Y_{ij-1}$, respectively.

Following mostly the notations of Islam et al. [42], let us define

$$\begin{aligned}\boldsymbol{\lambda}'_j &= (\boldsymbol{\beta}', \boldsymbol{\gamma}'_{j-1}, \boldsymbol{\rho}'_{j-1}, \boldsymbol{\eta}'_{j-1}), \text{ and} \\ \boldsymbol{W}'_j &= (\boldsymbol{X}'_{ij}, \boldsymbol{Y}'_{j-1}, \boldsymbol{\nu}'_{j-1}, \boldsymbol{Z}'_{j-1}),\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{X}_{ij} &= (1, X_{ij1}, X_{ij2}, ..., X_{ijp}), \\ \boldsymbol{Y}_{j-1} &= (Y_{i1}, ..., Y_{ij-1})', \\ \boldsymbol{\nu}_{j-1} &= (\nu_{12}, \nu_{123}, ..., \nu_{12...j-1})', \text{interaction terms among } Y_{ij}s \\ &= (y_{i1}y_{i2}, y_{i1}y_{i2}y_{i3}, ..., y_{i1}y_{i2}...y_{ij-1})';\ , j = 1, ..., n_i, \\ \boldsymbol{Z}_{j-1} &= (z_{11}, ..., z_{1p}, ..., z_{j-11}, ..., z_{j-1p})' \\ &= \text{ interaction terms among } \boldsymbol{X}_{ij} \text{ and } \boldsymbol{Y}_i \\ &= (x_{i1}y_{i1}, ..., x_{ip}y_{i1}, ..., x_{i1}y_{ij-1}, ..., x_{ip}y_{ij-1})', \\ \boldsymbol{\beta}' &= (\beta_0, \beta_1, ..., \beta_p) \text{ are the coefficients of } \boldsymbol{X}_{ij}, \\ \boldsymbol{\gamma}'_{j-1} &= (\gamma_1, ..., \gamma_{j-1}), \\ &\quad \text{the parameters corresponding to } Y_{i1}, ..., Y_{ij-1}, \\ \boldsymbol{\rho}'_{j-1} &= (\rho_{12}, \rho_{13}, ..., \rho_{123}, ..., \rho_{12...j-1}), \\ &\quad \text{the } (2^j - j - 1) \times 1 \text{ vector of coefficients of } \boldsymbol{\nu}_{j-1}, \\ \text{and } \boldsymbol{\eta}'_{j-1} &= (\eta_{11}, ..., \eta_{1p}, ..., \eta_{j-11}, ...\eta_{j-1p}), \\ &= \text{ vector of parameters corresponding to } \boldsymbol{Z}_{j-1}.\end{aligned}$$

The regressive model for the jth follow-up is defined as

$$P(Y_{ij} = s|w_{j-1}) = \frac{\exp\{\boldsymbol{\lambda}_j' \boldsymbol{w}_j s\}}{1+\exp\{\boldsymbol{\lambda}_j' \boldsymbol{w}_{j-1}\}};\ s = 0,1,\ j = 2,...,n_i. \tag{5.32}$$

While considering possible interactions among $\boldsymbol{X}_i$ and $\boldsymbol{Y}_i$ as well as among repeated outcomes on $\boldsymbol{Y}_i$, the full parametric form of the model be can be expressed as [42]

$$\begin{aligned} &P(Y_{ij} = s|\boldsymbol{y}_{ij-1}, \boldsymbol{x}_{ij}, \boldsymbol{\nu}_{j-1}, \boldsymbol{z})_{j-1} \\ &= \frac{\exp\{(\boldsymbol{\beta}'\boldsymbol{x}_{ij} + \boldsymbol{\gamma}_{j-1}\boldsymbol{y}_{j-1} + \boldsymbol{\rho}_{j-1}\boldsymbol{v}_{j-1} + \boldsymbol{\eta}_{j-1}'\boldsymbol{z})s\}}{1+\exp\{(\boldsymbol{\beta}'\boldsymbol{x}_{ij} + \boldsymbol{\gamma}_{j-1}'\boldsymbol{y}_{j-1} + \boldsymbol{\rho}_{j-1}\boldsymbol{v}_{j-1} + \boldsymbol{\eta}_{j-1}'\boldsymbol{z}_{j-1})\}};\ s = 0,1. \end{aligned} \tag{5.33}$$

Considering the full regressive model for the conditional probabilities, for $j = 2,...,n_i$, the canonical parameter can be defined as

$$\theta_{j.1,2,...,j-1} = g(\mu_{ij.1,2,...,j-1}) = (\boldsymbol{\beta}'\boldsymbol{x}_{ij} + \boldsymbol{\gamma}_{j-1}'\boldsymbol{y}_{j-1} + \boldsymbol{\rho}_{j-1}\boldsymbol{v}_{j-1} + \boldsymbol{\eta}_{j-1}'\boldsymbol{z}_{j-1}).$$

5.5.2 Likelihood and Log-likelihood Functions

The likelihood function can be expressed as before

$$\begin{aligned} L &= \prod_{i=1}^{N} f(y_{i1}, y_{i2}, ..., y_{in_i}|\theta_1, \theta_{2.1}, ..., \theta_{n_{i.1,2,...,n_i-1}}, \phi_1, \phi_2, ..., \phi_{n_i}) \\ &= \prod_{i=1}^{N} f(y_{i1}|X_{i1}, \theta_1, \phi_1) f(y_{i2.1}|X_{i2}, y_{i1}, \theta_{2.1}, \phi_{i2})... \\ &\quad f(y_{n_i.1,2,...,n_i-1}|X_i, y_{i1}, y_{i2}, ..., y_{in_i-1}, \theta_{n_i.1,2,...,n_i-1}, \phi_{in_i}). \end{aligned} \tag{5.34}$$

For example, for outcomes from exponential family, the likelihood function can be expressed as

$$\begin{aligned} L &= \exp \sum_{i=1}^{N} \left[\frac{y_{i1}\theta_1 - b(\theta_1)}{a(\phi_1)} + c(y_{i1}\phi_1) + \frac{y_{i2}\theta_{2.1} - b(\theta_{2.1})}{a(\phi_2)} + \ldots \right. \\ &\quad \left. + \frac{y_{in_i.1,2,\ldots,n_i-1}\theta_{n_i.1,2,\ldots,n_i-1} - b(\theta_{n_i.1,2,\ldots,n_i-1})}{a(\phi_{n_i})} + c(y_{in_i.1,2,\ldots,n_i-1}\phi_{n_i}) \right]. \end{aligned} \tag{5.35}$$

The log likelihood function takes the following form

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{N} \left[\frac{y_{i1}\theta_1 - b(\theta_1)}{a(\phi_1)} + c(y_{i1}\phi_1) + \frac{y_{i2}\theta_{2.1} - b(\theta_{2.1})}{a(\phi_2)} + \ldots \right. \\ &\quad \left. + \frac{y_{in_i.1,\ldots,n_i-1}\theta_{n_i.1,\ldots,n_i-1} - b(\theta_{n_i.1,\ldots,n_i-1})}{a(\phi_{n_i})} + c(y_{in_i.1,\ldots,n_i-1}\phi_{n_i}) \right]. \end{aligned} \tag{5.36}$$

The parameters can be estimated by using ML method.

5.5.3 Score Equations and Information Matrix

Differentiating the log-likelihood function with respect to the corresponding parameters and using the Chain rule, we obtain the score equations. For example, for outcome variables from exponential family, (assuming no interaction terms, i.e. $\lambda'_{j-1} = (\beta, \gamma_{j-1})$) the score equations and information matrix can be obtained as follow

$$\frac{\delta l}{\delta \beta_k} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \left(\frac{y_{ij} - b'(\theta_j)}{a(\phi_{ij}).V(\mu_{ij})} . x_{ijk} \right) = 0;\ k = 0, 1, \ldots, p. \tag{5.37}$$

$$\frac{\delta l}{\delta \gamma_s} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \left(\frac{y_{ij} - b'(\theta_j)}{a(\phi_{ij}).V(\mu_{ij})} . y_{is} \right) = 0;\ s = 1, \ldots, j-1. \tag{5.38}$$

The information matrix can be expressed as

$$I = \begin{pmatrix} I_1 & 0 \\ 0 & I_2 \end{pmatrix},$$

where I_1 is a $(p+1)\times(p+1)$ matrix with elements, $I_1 = -\frac{\delta^2 l}{\delta\beta_k \delta\beta_k'}$.

$$-\frac{\delta^2 l}{\delta\beta_k \delta\beta_k'} = \sum_{i=1}^{N}\sum_{j=1}^{3} X_{ijk}X_{ijk'}a(\phi_j)b''(\theta_j),\ k = 0,1,...,p. \tag{5.39}$$

The diagonal elements of I_1 are obtained when $k = k'$.

Similarly, I_2 is a $(n_i-1)\times(n_i-1)$ matrix with elements $I_2 = -\frac{\delta^2 l}{\delta\gamma_s \delta\gamma_{s'}}$, $s, s' = 1,2,...,n_i-1$ where

$$-\frac{\delta^2 l}{\delta\gamma_s \delta\gamma_s'} = \sum_{i=1}^{N}\sum_{j=s+1}^{n_i} Y_{is}Y_{is'}a(\phi_j)b''(\theta_j). \tag{5.40}$$

The diagonal elements of I_2 are obtained when $s = s'$.

5.5.4 Test for the Proposed Regressive Model MCM3

To test the independence of the repeated outcomes, the hypotheses to be tested are $H_0 : \boldsymbol{\lambda}_{j-1}^* = 0$ against $H_1 : \boldsymbol{\lambda}_{j-1}^* \neq 0$ for model (5.33) where $\boldsymbol{\lambda}_{j-1}$ is the vector of parameters of the model with covariate effects and all interaction effects and $\lambda_{j-1}^* = (\boldsymbol{\gamma}_{j-1}, \boldsymbol{\rho}_{j-1}, \boldsymbol{\eta}_{j-1})$ is the vector of parameters need to be tested. The total number of parameters need to be tested is $j-1$ for $\boldsymbol{\gamma}_{j-1}$, $2^j - j - 1$ for $\boldsymbol{\rho}_{j-1}$ and $(j-1)\times p$ for $\boldsymbol{\eta}_{j-1}$. This test can be performed using the likelihood ratio test and the test statistic follows chi-square with $j-1+2^j-j-1+(j-1)\times p = 2^j-2+(j-1)\times p$ degrees of freedom.

Test of Dependence among Repeated Outcomes

$\boldsymbol{\gamma}$ is the vector of parameters associated with the outcome variables at earlier time points $\boldsymbol{Y}_{j-1}$ such that, $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ indicates a lack of dependence among Y_{ij}s. However, in several instances,the regressive model (5.33) may fail to rec-

ognize the true nature of relationship between Y_{ij}'s in the presence of covariates $X_{ij1}, X_{ij2}, ..., X_{ijp}$ in the model due to the fact that dependence among Y_{ij}s depends on the dependence between the outcome variables and the covariates as well [16]. However, if interested, one may use the tests, (5.29) or (5.30) and (5.31) or their extensions to take care of this problem.

5.6 Conclusion

In this chapter, we proposed three joint models based on a marginal conditional approach, $MCM1$, $MCM2$ and $MCM3$, as alternatives to GEE or related models based on marginal approaches. The proposed models take care of the correlation among the repeated measures in a built-in nature and can be extended for any order of dependence without complicating the theory. The proposed model 1 ($MCM1$) is an extension of Darlington and Farewell [16] that shows the likelihood for models based on Markovian assumption of first order more explicitly. The second model $MCM2$ (proposed model 2) is a further generalization based on marginal and conditional models for any order of a Markov chain with covariate dependence. For more than three repeated responses, $MCM2$, the proposed model 2, has restricted use due to overwhelming increase in the number of models and parameters to be estimated. At this backdrop, a further extension is considered by including previous outcomes as covariates. This model is denoted as proposed model 3 or $MCM3$. The number of parameters in $MCM3$ can be kept as minimum as possible for any order of the underlying Markov Chain. Hence for practical reasons, the proposed model 3 can be used to analyze longitudinal data effectively and conveniently when number of repeated measures is large. In the next chapter, we present the results of the simulation studies performed to check the competence of the proposed models in terms of bias and coverage probability of the estimates and to compare the proposed models with GEE and ALR with an application of the models to a real life data.

Chapter 6

A Comparison of Proposed Models, GEE and ALR

6.1 Introduction

We proposed three joint models in Chapter 5 for the outcome variables of a longitudinal data. The joint models (proposed models 1, 2 and 3 or $MCM1$, $MCM2$ and $MCM3$) shown in Chapter 5 take care of the correlation among the repeated measures in a built in nature and can be extended for any order of dependence without complicating the theory. The proposed model 1 is an extension of Darlington and Farewell [16] that shows the likelihood for models based on Markovian assumption of first order more explicitly. The second model (proposed model 2) is a further generalization based on marginal and conditional models for any order of a Markov chain with covariate dependence. For more than three repeated responses, the proposed model 2 has restricted

use due to overwhelming increase in the number of models and parameters to be estimated. At this backdrop, a joint model is considered which is based on a regressive approach and includes previous outcomes as covariates. This model is denoted as proposed model 3. The number of parameters in the proposed model 3 can be kept as minimum as possible for any order of the underlying Markov Chain. In this chapter we discuss the performance of the parameters of the proposed joint (marginal-conditional) models (5.6), (5.9), when $n_i = 2$ and (5.32) when $n_i > 3$. The estimates of the proposed models are compared with the estimates of GEE and ALR in terms of bias and coverage probability of the estimates. We start with the data generation steps in the next section and the results and findings followed by an application to HRS Data in the following sections.

6.2 Generation of Data

A simulation study was carried out to compare the properties of estimates of regression coefficients of the models discussed in the earlier sections. The repeated measures can be associated in a variety of ways and in this study, the cases considered are (i) Y_{ij}'s are identically and independently distributed, (ii) Y_{ij}'s are identically distributed and associated (iii) Y_{ij}'s are not identical and their distributions are independent. For simplicity of the study, we restrict the simulation study for the conditional marginal model to two follow ups, Y_{i1} and Y_{i2} on ith subject and only one explanatory variable, X_{i1} for each of the N individuals where X_{i1} is fixed and time invariant. We assumed that Y_{i1} and Y_{i2} are two binary random variables with $Y_{i1} \sim B(1, p_{i1})$ and $Y_{i2} \sim B(1, p_{i2})$. The corresponding GLMs are

$$g(\mu_{i1}) = \frac{\exp(\boldsymbol{X}_{i1}\boldsymbol{\beta}_1)}{1+\exp(\boldsymbol{X}_{i1}\boldsymbol{\beta}_1)} \text{ and } g(\mu_{i2}) = \frac{\exp(\boldsymbol{X}_{i2}\boldsymbol{\beta}_2)}{1+\exp(\boldsymbol{X}_{i2}\boldsymbol{\beta}_2)}.$$

The simulation followed the following steps

- Data generation
  - Step I: An explanatory variable X_{i1} was generated first from Bernoulli distribution with probability of success 0.5.
  - Step II: The probability of success for the outcome variable in the first follow-up, p_{i1} was calculated using the equation

$$P(Y_{i1} = 1|\boldsymbol{X}_{i1} = \boldsymbol{x}_{i1}) + \frac{\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}{1+\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1)}$$

    where $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11})$.
  - Step III: N values, a_i, were generated from uniform distribution within range $(0,1)$ and then the outcome variable at first time point, Y_{i1} was generated such that $Y_{i1} = 1$ if $a_i < P(Y_{i1} = 1|X_{i1})$ and 0 otherwise. i.e. Data on Y_{i1}, the outcome variable at first time point, was generated such that $Y_{i1} = 1$ if $runif(N,0,1) < P(Y_{i1} = 1|X_{i1})$ and 0 otherwise.
  - Step IV: To generate data on Y_{i2}, the value of the outcome variable at time point 2, first, the probability of success at time point 2, p_{i2} was calculated as

$$P(Y_{i2} = 1|\boldsymbol{X}_{i1} = \boldsymbol{x}_{i1}) = \frac{\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_2)}{1+\exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}_2)}$$

    where $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \gamma_1)$, β_{20} is the intercept term, β_{21} is the coefficient of X_{i2} and γ_1 is the coefficient of Y_{i1}.
  - Step V: Similar as step *III*, N values, b_i, were generated from uniform distribution within range $(0,1)$ and then the outcome variable at second time point, Y_{i2} was generated such that $Y_{i2} = 1$ if $b_i < P(Y_{i2} = 1|\boldsymbol{X}_{i2} = \boldsymbol{x}_{i2})$ and 0 otherwise.
  - The cases considered are as follow
    - Y_{i1} and Y_{i2} are **independent** and their **distributions are identical**, i.e $\beta_{10} = \beta_{20}$, $\beta_{11} = \beta_{21}$ and $\gamma_1 = 0$.

  - Y_{i1} and Y_{i2} are **not independent** and their **distributions are identical**, i.e $\beta_{10} = \beta_{20}$, $\beta_{11} = \beta_{21}$ and $\gamma_1 \neq 0$.
  - Y_{i1} and Y_{i2} are **independent** and their **distributions are not identical**, i.e $\beta_{10} \neq \beta_{20}$, $\beta_{11} \neq \beta_{21}$ and $\gamma_1 = 0$.
  - Y_{i1} and Y_{i2} are **not independent** and their **distributions are not identical**, i.e $\beta_{10} \neq \beta_{20}$, $\beta_{11} \neq \beta_{21}$ and $\gamma_1 \neq 0$.

- For illustration of the regressive model, Y_{i1}, Y_{i2}, Y_{i3} and Y_{i4} were generated in a similar way as Y_{i1} with $\boldsymbol{\beta}_1 = (\beta_0, \beta_1)'$, $\boldsymbol{\beta}_2 = (\beta_0, \beta_1, \gamma_1)'$, $\boldsymbol{\beta}_3 = (\beta_0, \beta_1, \gamma_1, \gamma_2)'$ and $\boldsymbol{\beta}_4 = (\beta_0, \beta_1, \gamma_1, \gamma_2, \gamma_3)'$, respectively. The cases considered are as follow
  - Y_{i1}, Y_{i2}, Y_{i3} and Y_{i4} are **independent** and their **distributions are identical**, i.e $\beta_{10} = \beta_{20} = \beta_{30} = \beta_{40} = \beta_0$, and $\beta_{11} = \beta_{21} = \beta_{31} = \beta_{41} = \beta_1$ and $(\gamma_1, \gamma_2, \gamma_3) = (0, 0, 0)$.
  - Y_{i1} and Y_{i2} are **not independent** and their **distributions are identical**, i.e $\beta_{10} = \beta_{20} = \beta_{30} = \beta_{40} = \beta_0$, and $\beta_{11} = \beta_{21} = \beta_{31} = \beta_{41} = \beta_1$ and $(\gamma_1, \gamma_2, \gamma_3) = (1, 1, 1)$.

- The bias, mean squared error and coverage probability of the 95% confidence interval were constructed over a range of scenarios for varying correlation among the responses.

- The sample size was 500 and number of samples was 1000 to construct each of the tables in the next section.

6.3 Results of the Simulation Study

The findings of the simulation study (estimates, bias, standard error and coverage probability) are summarized in Table 6.1 to Table 6.3. In all these tables, GEE(In), GEE(Ex), GEE(AR) and ALR(Ex) stand for GEE models under in-

dependent, exchangeable and autoregressive correlation, and ALR model under exchangeable correlation, respectively. The parameters of the joint model are β_{10}, β_{11}, β_{010}, β_{011}, β_{110} and β_{111}. Here, β_{10} and β_{11}, respectively, denote the intercept and the regression coefficient of the marginal model $P(Y_{i1} = 1|\boldsymbol{X}_i = \boldsymbol{x}_i)$; β_{010} and β_{011}, respectively, denote the intercept and the regression coefficient of the conditional model $P(Y_{i2} = 1|Y_{i1} = 0, \boldsymbol{X}_i = \boldsymbol{x}_i)$; and β_{110} and β_{111}, respectively, denote the intercept and regression coefficient of the conditional model $P(Y_{i2} = 1|Y_{i1} = 1, \boldsymbol{X}_i = \boldsymbol{x}_i)$. GEE or ALR, being approaches based on marginal models, estimate the parameters of such models as average of the parameters of two populations from where Y_{i1} and Y_{i2} were generated, and to distinguish the parameters of GEE and ALR from joint model, we used the notation $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1^*)'$ to denote the parameters of GEE and ALR. In Table 6.1,

$$P(Y_{i1} = 1|\boldsymbol{X}_i = \boldsymbol{x}_i) = \frac{\exp(\beta_{10} + \beta_{11}x_i)}{1 + \exp(\beta_{10} + \beta_{11}x_{i11})} = \frac{\exp(0.5 + 0.2x_{i11})}{1 + \exp(0.5 + 0.2x_{i11})}$$

$$\text{and } P(Y_{i2} = 1|Y_{i1} = y_{i1}, \boldsymbol{X}_i = \boldsymbol{x}_i) = \frac{\exp(\beta_{20} + \beta_{21}x_{i21} + \gamma_1 y_{i1})}{1 + \exp(\beta_{20} + \beta_{21}x_{i21} + \gamma_1 y_{i1})},$$

where, β_{20} and β_{21}, respectively, denote the intercept and regression coefficients of the marginal model $P(Y_{i2} = 1|\boldsymbol{X}_i = \boldsymbol{x}_i)$; So if $\gamma_1 = 0$, the true values of the parameters to be estimated for the joint model are $\beta_{10} = 0.5$, $\beta_{11} = 0.2$, $\beta_{010} = \beta_{20} + \gamma_1 \times (y_{i1} = 0) = 0.5 + 0 \times 0 = 0.5 = \beta_{20}$, $\beta_{011} = \beta_{21} = 0.2$, $\beta_{110} = \beta_{20} + \gamma_1 \times (y_{i1} = 1) = 0.5 + 0 \times 1 = 0.5$ and $\beta_{111} = \beta_{21} = 0.2$. When $\gamma_1 = 1$, the true values of the parameters to be estimated for the joint model are $\beta_{10} = 0.5$, $\beta_{11} = 0.2$, $\beta_{010} = \beta_{20} + \gamma_1 \times (y_{i1} = 0) = 0.5 + 1 \times 0 = 0.5$, $\beta_{011} = \beta_{21} = 0.2$, $\beta_{110} = \beta_{20} + \gamma_1 \times (y_{i1} = 1) = 0.5 + 1 \times 1 = 1.5$ and $\beta_{111} = \beta_{21} = 0.2$.

Table 6.1 shows that bias and the standard error of estimates of the proposed Model 1 (extension of Darlington and Farewell [16]), Proposed Model 2, GEE and ALR are competitive for longitudinal data when the repeated measures are independent ($\gamma_1 = 0.0$).

Table 6.1: Parameters (Par), estimates(Est), bias, standard error(SE) and coverage probability(CP) of estimates for independent ($\gamma_1 = 0.0$) and correlated outcomes ($\gamma_1 = 1.0$) with identical distributions of Y_{i1} and Y_{i2}, (No. of Simulation $= 1000$, $N = 500$, $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}) = (0.5, 0.2)$, $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}) = (0.5, 0.2)$)

	$\gamma_1 = 0$					$\gamma_1 = 1$				
	Par	Est	Bias	SE	CP	Par	Est	Bias	SE	CP
Model 1	β_0^*=0.5	0.5037	-0.0037	0.1456	0.9510	β_0^*=0.5	0.7756	-0.2756	0.1520	0.5670
	β_1^*=0.2	0.2054	-0.0030	0.2102	0.9570	β_1^*=0.2	0.2181	-0.0030	0.2212	0.9370
	β_{010}=0.5	0.5013	-0.0013	0.3420	0.9660	β_{010}=0.5	0.5013	-0.0013	0.3420	0.9660
	β_{011}=0.2	0.2046	-0.0046	0.5138	0.9500	β_{011}=0.2	0.2046	-0.0046	0.5138	0.9500
	β_{110}=0.5	0.5079	-0.0079	0.2628	0.9530	β_{110}=1.5	1.5365	-0.0365	0.3357	0.9640
	β_{111}=0.2	0.2169	-0.0169	0.3733	0.9550	β_{111}=0.2	0.2228	-0.0228	0.4896	0.9630
Model 2	β_{10}=0.5	0.5127	-0.0127	0.2066	0.9580	β_{10}=0.5	0.5127	-0.0127	0.2066	0.9580
	β_{11}=0.2	0.2030	-0.0030	0.2985	0.9550	β_{11}=0.2	0.2030	-0.0030	0.2985	0.9550
	β_{20}=0.5	0.4997	0.0003	0.2064	0.9510	β_{010}=0.5	0.5013	-0.0013	0.3420	0.9660
	β_{21}=0.2	0.2109	-0.0109	0.2982	0.9460	β_{011}=0.2	0.2046	-0.0046	0.5138	0.9500
						β_{110}=1.5	1.5365	-0.0365	0.3357	0.9640
						β_{111}=0.2	0.2228	-0.0228	0.4896	0.9630
GEE(In)	β_0^*=0.5	0.5037	-0.0037	0.1453	0.9470	β_0^*=0.5	0.7756	-0.2756	0.1659	0.6400
	β_1^*=0.5	0.2054	-0.0054	0.2101	0.9510	β_1^*=0.2	0.2181	-0.0181	0.2409	0.9520
GEE(Ex)	β_0^*=0.5	0.5037	-0.0037	0.1453	0.9470	β_0^*=0.5	0.7756	-0.2756	0.1659	0.6400
	β_1^*=0.5	0.2054	-0.0054	0.2101	0.9510	β_1^*=0.2	0.2181	-0.0181	0.2409	0.9520
ALR(Ex)	β_0^*=0.5	0.5037	-0.0037	0.1453	0.9470	β_0^*=0.5	0.7756	-0.2756	0.1659	0.6400
	β_1^*=0.5	0.2054	-0.0054	0.2101	0.9510	β_1^*=0.2	0.2181	-0.0181	0.2409	0.9520

Table 6.2: Estimates(Est), bias, standard error(SE) and coverage probability(CP) of estimates for independent outcomes with non-identical distributions, (No. of Simulation = 1000, $N = 500$, $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}) = (0.5, 0.2)$, $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \gamma_1) = (0.2, 0.7, 0.0)$).

Methods	Par	Est	Bias from $\boldsymbol{\beta}_1$	Bias from $\boldsymbol{\beta}_2$	SE	CP for $\boldsymbol{\beta}_1$	CP for $\boldsymbol{\beta}_2$
Model 1	β_0^*	0.3522	0.1478	-0.1522	0.1433	0.8070	0.8040
	β_1^*	0.4566	-0.2566	0.2434	0.2107	0.7720	0.7830
	β_{010}	0.1960		0.0040	0.3330		0.9590
	β_{011}	0.7231		-0.0231	0.5210		0.9570
	β_{110}	0.2047		-0.0047	0.2556		0.9640
	β_{111}	0.7248		-0.0248	0.3763		0.9530
Model 2	β_{10}	0.5127	-0.0127		0.2066	0.9580	
	β_{11}	0.2030	-0.0030		0.2985	0.9550	
	β_{20}	0.1992		0.0008	0.2010		0.9520
	β_{21}	0.7157		-0.0157	0.3010		0.9520
GEE(In)	β_0^*	0.3522	0.1478	-0.1522	0.1426	0.818	0.828
	β_1^*	0.4566	-0.2566	0.2434	0.2101	0.763	0.780
GEE(Ex)	β_0^*	0.3522	0.1478	-0.1522	0.1426	0.818	0.828
	β_1^*	0.4566	-0.2566	0.2434	0.2101	0.763	0.780
ALR(Ex)	β_0^*	0.3522	0.1478	-0.1522	0.1426	0.818	0.828
	β_1^*	0.4566	-0.2566	0.2434	0.2101	0.763	0.780

Inadequacy of GEE or ALR to portray the relationship between $\boldsymbol{X}$ and $\boldsymbol{Y}$ are visible with the presence of dependence relationship between Y_{i1} and Y_{i2} as shown in last 5 columns of Table 6.1 where the data were generated from two associated populations ($\gamma_1 = 1.0$). The marginal parameters in the Model 1 proposed as an extension of Darlington and Farewell [16] does not make much improvement in the performance of the parameters in terms of bias and standard error. The proposed joint model 2 (Model 2) gives better estimates than other models in this case.

The inadequacy of GEE or ALR to portray the relationship between $\boldsymbol{X}_i$ and $\boldsymbol{Y}_i$ are also observed in Table 6.2 where the data are generated from two independent but nonidentical populations. The estimates of parameters of GEE are not portraying the actual relationship between the covariates and the response variable because the relationship between $\boldsymbol{X}_i$ and $\boldsymbol{Y}_i$ are different at different time points. And the actual bias from population 1 (from where Y_{i1} were generated) and population 2 (from where Y_{i2} were generated) are shown in Table 6.2. Clearly, even if the repeated measures are not associated, while data come from two different populations, the GEE or ALR or other marginal models are not adequate to capture the relationship between the covariates and the response variable.

While there are three or more than three repeated measurements on same subject, the covariate dependent Markov Chain based joint models need to estimate too many parameters and we proposed the Model 3, a general form of the regressive model approach [42], as an alternative of GEE based approaches. The results of the simulation study (Table 6.3) show that when the outcomes are independent and identically distributed, the estimates of the parameters of a regressive model produce similar results as GEE or ALR in terms of bias and coverage probability. The regressive model performs better while the repeated responses are associated and GEE or ALR fails to portray the scenario.

Table 6.3: Parameters(Par), Estimates(Est), Bias, standard error(SE) and coverage probability(CP) of estimates of different models for independent and associated distribution (No. of Simulation $= 1000$, $N = 500$, $\beta_{10} = \beta_{20} = \beta_{30} = \beta_{40} = \beta_0 = 0.2$, $\beta_{11} = \beta_{21} = \beta_{31} = \beta_{41} = \beta_1 = 0.7$, $(\gamma_1, \gamma_2, \gamma_3) = (0,0,0)$ and $(1,1,1)$).

		$(\gamma_1,\gamma_2,\gamma_3)=(0,0,0)$					$(\gamma_1,\gamma_2,\gamma_3)=(1,1,1)$			
Methods	Par	Est	Bias	SE	CP	Par	Est	Bias	SE	CP
	β_0^*=0.2	0.208	-0.008	0.226	0.946	β_0^*=0.2	0.252	-0.052	0.316	0.952
	β_1^*=0.7	0.698	0.002	0.199	0.953	β_1^*=0.7	0.748	-0.048	0.397	0.944
Model 3	γ_1=0.0	-0.012	0.012	0.197	0.940	γ_1=1.0	1.010	-0.010	0.370	0.942
	γ_2=0.0	-0.001	0.001	0.197	0.942	γ_2=1.0	0.987	0.013	0.352	0.949
	γ_3=0.0	0.003	-0.003	0.197	0.945	γ_3=1.0	1.001	-0.001	0.366	0.950
GEE	β_0^*=0.2	0.201	-0.001	0.062	0.950	β_0^*=0.2	0.927	-0.727	0.085	0.000
(In)	β_1^*=0.7	0.695	0.005	0.095	0.950	β_1^*=0.7	0.799	-0.099	0.136	0.889
GEE	β_0=0.2	0.201	-0.001	0.062	0.950	β_0^*=0.2	0.927	-0.727	0.085	0.000
(Ex)	β_1^*=0.7	0.695	0.005	0.095	0.950	β_1^*=0.7	0.799	-0.099	0.136	0.889
GEE	β_0^*=0.2	0.201	-0.001	0.062	0.948	β_0^*=0.2	0.921	-0.721	0.085	0.000
(AR)	β_1^*=0.7	0.695	0.005	0.095	0.951	β_1^*=0.7	0.788	-0.088	0.136	0.902
ALR	β_0^*=0.2	0.201	-0.001	0.062	0.950	β_0^*=0.2	0.927	-0.727	0.085	0.000
(Ex)	β_1^*=0.7	0.695	0.005	0.095	0.950	β_1^*=0.7	0.799	-0.099	0.136	0.889

Indubitably, GEE and ALR performed well only when repeated measures come from identical population and are not associated. The simulation study also finds that basically there is no difference in the estimates of GEE under different correlation structures (Table 6.1 to Table 6.3). Also ALR does not show any noticeable difference from GEE estimates in most cases. The proposed model 2 produces better estimates in terms of bias and coverage probability than GEE or ALR in the cases when responses are associated or the responses at different time points has different distributions. Proposed Model 3 is suggested for longitudinal data with more than 3 follow ups on same individual.

6.4 Application to HRS Data

The first three waves of the longitudinal data from the Health and Retirement Study (HRS) conducted by the University of Michigan [73] were used for comparison of the selected methods. The study started in 1992 on American individuals over the age of 50 years and their spouses and the subjects are observed every two years. In wave 1, the sample size was 9760 and the sample size was reduced to 9750 due to dropping of 10 cases with missing values of outcome variable at round 1. Finally the number of individuals were 8657 who reported that they were not hospitalized at wave 1. The panel data from the waves for 1992, 1994 and 1996 have been used in this study. Elderly population may suffer from repeated spells of depression which may change over time [20, 37] and result in other health problems and chronic illness [45]. The literature on depression among elderly helped filling many gaps in our understanding of the factors associated with depression and also the outcome of depression [6]. But understanding depression and its associated factors more explicitly is important. In many studies on clinical and non-clinical populations, *CESD* (Center for Epidemiologic Studies Depression) scale is employed to measure depressive symptoms [69].

The dependent variable for this study is Depression status with outcomes no depression (CESD score $= 0$) and depression (CESD score > 0). The independent variables are gender (male=1), marital status (married/partnered=1), education, ethnicity Black (Black=1), ethnicity White (White=1), drinking habit (drink=1) and number of health conditions. In Table 6.4 and Table 6.5, Mstat stands for marital status, White stands for white ethnicity, Black stands for Black ethnicity, Drink means drinking habit and No. of Cond. is the number of health conditions.

Table 6.4: Estimates of parameters of GEE and ALR on HRS Data

	GEE(In)			GEE(Ex)		
	Est	SE	p-value	Est	SE	p-value
Intercept	2.023	0.206	0.000	2.023	0.206	0.000
Gender	-0.059	0.059	0.321	-0.059	0.059	0.321
Mstat	-0.621	0.065	0.000	-0.621	0.065	0.000
Education	-0.153	0.010	0.000	-0.153	0.010	0.000
White	-0.363	0.166	0.029	-0.363	0.166	0.029
Black	-0.085	0.177	0.629	-0.085	0.177	0.629
Drink	-0.091	0.055	0.097	-0.091	0.055	0.097
No. of Cond.	0.389	0.024	0.000	0.389	0.024	0.000
	GEE(AR)			ALR(Ex)		
	Est	SE	p-value	Est	SE	p-value
Intercept	1.944	0.206	0.000	2.019	0.192	0.000
Gender	-0.056	0.060	0.351	-0.059	0.057	0.153
Mstat	-0.613	0.065	0.000	-0.624	0.063	0.000
Education	-0.151	0.010	0.000	-0.153	0.010	0.000
White	-0.338	0.166	0.042	-0.366	0.153	0.008
Black	-0.067	0.177	0.706	-0.085	0.163	0.301
Drink	-0.082	0.055	0.135	-0.091	0.053	0.045
No. of Cond.	0.391	0.024	0.000	0.391	0.023	0.000

In GEE models, we observe that marital status, education year, ethnicity White and number of health conditions are significantly associated with depression. The GEE model under the assumption of independence and exchangeable correlation produces the same results and finds that marital status, education, White ethnicity and number of health conditions have significant influence on depres-

sion. ALR under exchangeable correlation, in addition, finds drinking habit as a significant factor for depression. GEE model under the assumption of autoregressive correlation shows that marital status, education, white ethnicity and number of health conditions are significantly associated with the depression status. Gender is not significant in GEE based models.

Table 6.5 shows the estimates of the parameters of the proposed model 2, MCM2 that shows that the effects of the covariates are different on the depression status at different follow ups.

At the baseline, marital status, education, white ethnicity and number of conditions have significant effect on depression. Married people are less depressed as compared to their single counterparts, more are the respondents educated, less are they depressed, white people are less depressed, more physical conditions results in more risk of depression.

In second follow-up, the effects of the covariates were notably different depending on the depression status of the respondent in the previous follow ups. Depression status of patients (who were not depressed in baseline or first follow-up) were significantly associated with marital status, education and drinking habit.

Depression status of patients (who were not depressed in baseline but were depressed in first follow up) were significantly associated with education. Education had significant effect on depression status of patients in second follow up for those who were depressed in baseline but not depressed in first follow up. Respondents' depression status was significantly associated with marital status and education for those who were depressed in first as well as in second follow ups.

Clearly, the covariate effects on the depression status were different at different follow ups and using a marginal model like GEE or ALR may not be appropriate to estimate the covariate effects on depression status of the elderly people.

Table 6.5: Estimates of Parameters of the Proposed Marginal Conditional Model 2 (*MCM*2) for HRS Data

	β_1			β_{01}			β_{11}			β_{001}		
	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value
Intercept	1.230	0.179	0.000	1.837	0.254	0.000	2.856	0.319	0.000	-0.038	0.378	0.920
Gender	-0.012	0.054	0.826	-0.273	0.067	0.000	-0.047	0.093	0.614	-0.139	0.087	0.110
Mstat	-0.525	0.060	0.000	-0.334	0.081	0.000	-0.455	0.103	0.000	-0.213	0.111	0.056
Education	-0.111	0.009	0.000	-0.140	0.012	0.000	-0.140	0.016	0.000	-0.083	0.016	0.000
White	-0.454	0.144	0.002	-0.595	0.199	0.003	-0.288	0.250	0.249	-0.056	0.307	0.855
Black	-0.094	0.154	0.544	-0.312	0.214	0.146	-0.143	0.266	0.591	0.155	0.327	0.636
Drink	-0.079	0.054	0.147	-0.076	0.069	0.272	-0.127	0.095	0.182	0.256	0.094	0.007
No. of Cond.	0.354	0.024	0.000	0.285	0.034	0.000	0.282	0.041	0.000	0.175	0.048	0.000
		β_{011}			β_{101}			β_{111}				
		Est	SE	p-value	Est	SE	p-value	Est	SE	p-value		
Intercept		1.187	0.363	0.001	1.937	0.581	0.001	1.973	0.350	0.000		
Gender		-0.039	0.109	0.722	0.032	0.156	0.835	-0.117	0.121	0.334		
Mstat		-0.121	0.126	0.339	-0.335	0.181	0.064	-0.346	0.130	0.008		
Education		-0.092	0.018	0.000	-0.115	0.028	0.000	-0.076	0.019	0.000		
White		-0.092	0.284	0.746	-0.373	0.435	0.392	0.228	0.280	0.414		
Black		0.060	0.308	0.847	-0.029	0.467	0.950	0.136	0.296	0.647		
Drink		0.002	0.110	0.989	-0.188	0.161	0.245	-0.192	0.123	0.119		
No. of Cond.		0.320	0.052	0.000	0.052	0.070	0.453	0.317	0.053	0.000		

These findings confirm our assertion that the extensively used GEE based models fail to specify the covariate effects adequately for longitudinal data. The results demonstrate that a joint model based on marginal conditional approach explains the covariate effects more meaningfully.

6.5 Conclusion

In this chapter, we summarized the findings of the simulation studies to compare the proposed joint models using marginal conditional approach with GEE and ALR which are based on marginal approaches. It is evident from the simulation studies as well as the application of GEE, ALR and the proposed models, that the proposed model 2 is expected to provide more specified model in a more simplified set up. Proposed model 2 produces less bias and has better 95% coverage probability as compared to GEE or ALR. For more than 3 repeated outcomes, the proposed model 3 is the most convenient model that performs better than GEE or ALR. The results of the simulation study indicate that in terms of bias and coverage probability, the proposed models appears to be competitive or sometimes better than the alternatives, GEE, or ALR. Hence for practical reasons, the proposed models can be used to analyze longitudinal data effectively and conveniently. Both the theoretical and practical users are expected to find it more useful and interpretable using the proposed models in appropriate cases.

Chapter 7

Proposition of a Marginal Conditional Model using Quasi-likelihood Methods

7.1 Introduction

A marginal or population averaged model can not make use of the main advantage of a longitudinal data of visualizing the change in individual responses over time (Chapter 3 and Chapter 4). In Chapter 5 and 6, we explained how a joint model using a marginal-conditional approach enables studying the relationship among covariates and outcome variables at different time points. As a result, these models are expected to portray the outcome-covariate relationship in a more meaningful and explicit way as compared to marginal models. The models proposed in Chapter 5 use likelihood based methods for repeated out-

comes under the assumption that the distribution of the outcome variables are known.

A marginal-conditional model in a quasi-likelihood set up is not found in literature. In this chapter, we describe a new marginal-conditional model developed for repeated outcomes with unknown distributions. We started this chapter with a very short review of GEE and ALR and then described the proposed model along with its parameter estimation procedure and necessary tests. The new model is compared with GEE and ALR using a set of simulation studies under varying conditions. An application is shown using Health and Retirement Study (HRS) data [73].

7.2 Models for Repeated Binary Outcomes using Quasi-likelihood Approaches

Consider a binary outcome variable $Y_{ij},\ i = 1,2,...,N,\ j = 1,2,...,n_i$, observed for subject i at time j. Let $\boldsymbol{X}_{ij}$ be the $(p+1)\times 1$ vector of covariates for individual i at time j. The outcome vector for subject i is $\boldsymbol{Y}_i = (Y_{i1}\ Y_{i2}\ ...\ Y_{in_i})'$ with mean vector $\boldsymbol{\mu}_i = E(\boldsymbol{Y}_i) = (\mu_{i1}\ \mu_{i2}\ ...\ \mu_{in_i})' = (p_{i1}\ p_{i2}\ ...\ p_{in_i})'$. If Y_{ij}'s are time dependent, then the marginal probability that Y_{ij} observes an event is (as defined in equation (5.1), Chapter 5)

$$p_{ij} = Pr(Y_{ij} = 1|\boldsymbol{x}_{ij}) = \frac{\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}_j)}{1+\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}_j)},$$

where $\boldsymbol{\beta}_j$ is the $(p+1)\times 1$ vector of parameters of the marginal model $Pr(Y_{ij} = 1|\boldsymbol{x}_{ij})$. The marginal probability that Y_{ij} does not observe an event is $q_{ij} = 1 - p_{ij}$. The conditional probability that Y_{ij} observes an event given $(y_{i1},...,y_{ij-1})$ is

$$\begin{aligned} p^*_{ij.1...j-1} &= Pr(Y_{ij} = 1|y_{i1},...,y_{ij-1},\boldsymbol{x}_{ij}) \\ &= \frac{\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}_{j.12...j-1})}{1+\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}_{j.12...j-1})}; i = 1,...,N; j = 2,...,n_i, \end{aligned}$$

where $\boldsymbol{\beta}_{j.12...j-1}$ is the vector of parameters of the conditional model $Pr(Y_{ij} = 1|y_{i1},...,y_{ij-1},\boldsymbol{x}_{ij})$; $j = 2,...,n_i$ (equation (5.2), Chapter 5). Assume that the distributional form of Y_{ij} is unknown, but mean μ_{ij} is a known function of the set of regression parameters and variance of Y_{ij}, denoted by $V(Y_{ij})$, is a known function of μ_{ij}. Then for each observation Y_{ij}, we may define a quantity,

$$Q_{ij} = Q(\mu_{ij}|y_{ij}) = \int_{y_{ij}}^{\mu_{ij}} \frac{y_{ij}-t}{a(\phi_{ij})V(t)}dt.$$

7.2.1 Log Quasi-Likelihood Function

Under the assumptions above, the integral Q_{ij} behaves like a log-likelihood function and is referred to as a log quasi-likelihood function of the parameters μ_{ij} [61, 77] (section 2.3.4, Chapter 2). The log quasi-likelihood for the complete data is the sum of the individual contributions

$$\begin{aligned}\sum_{i=1}^{N}\sum_{j=1}^{n_i} Q(\mu_{ij}|y_{ij}) &= \sum_{i=1}^{N}\Big[Q_{i1}(\mu_{i1}|y_{i1}) + ... + Q_{in_i}(\mu_{in_i}|y_{in_i})\Big] \\ &= \sum_{i=1}^{N}\Big[\int_{y_{i1}}^{\mu_{i1}} \frac{y_{i1}-t}{a(\phi_{i1})V(t)}dt + ... + \int_{y_{in_i}}^{\mu_{in_i}} \frac{y_{in_i}-t}{a(\phi_{in_i})V(t)}dt\Big].\end{aligned} \tag{7.1}$$

7.2.2 Quasi-likelihood Estimating Equations

Differentiating equation (7.1) with respect to β_k, we have quasi-likelihood estimating equations or quasi-score equations for β_k

$$\sum_{i=1}^{N}\sum_{j=1}^{n_i} \frac{\delta Q(\mu_{ij}|y_{ij})}{\delta\beta_k} = \sum_{i=1}^{N}\sum_{j=1}^{n_i} \frac{y_{ij}-\mu_{ij}}{a(\phi_{ij})V(\mu_{ij})}.\frac{\delta\mu_{ij}}{\delta\beta_k} = 0. \tag{7.2}$$

We may write that the quasi-likelihood estimating equation, or the quasi score function as $U(\hat{\boldsymbol{\beta}}) = \mathbf{0}_p$, where

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{N} D_i' V_i^{-1} \frac{(y-\mu)}{a(\phi)} = 0$$

$$\text{with} \quad \boldsymbol{D}_i = \frac{\delta\boldsymbol{\mu}_i}{\delta\beta_k} \quad \text{and} \quad \boldsymbol{V}_i^{-1} = \frac{1}{a(\phi)V(\boldsymbol{\mu}_i)}.$$

While analysing longitudinal data with correlated response variables or response variables from independent but non-identical populations at different time points, fitting marginal models like GEE or ALR for Y_{ij}'s is logically not an appropriate choice as such models fail to utilize the major advantage of longitudinal data of observing the change in the outcome variable over time and models based on marginal conditional methods are preferred (Chapter 5 and Chapter 6).
In the following section, we propose a new method to obtain a joint model using a marginal-conditional approach under the framework of quasi-likelihood method for outcome variables with unknown distributions.

7.3 Proposed Joint Model

In section 7.2, we defined $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, ..., Y_{in_i})'$, $i = 1, 2, ..., N$. To discuss the proposed joint model using quasi-likelihood method, we start with the assumptions and basic notations used for the proposed model.

7.3.1 Assumptions and Basic Notations

Considering the probable dependence among the repeated outcomes, let us redefine the outcome vector for subject i as

$$\boldsymbol{Y}_i = (Y_{i1}, Y_{i2.1} ..., Y_{in_i.12...n_i-1})',$$

where, $Y_{ij.12...j-1}$ denotes the outcome variable of subject i at jth time point given $\{Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, ..., Y_{ij-1} = y_{ij-1}\}$. The mean vector can be redefined as $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{2.1}, ..., \mu_{in_i.12...n_i-1})'$, where $\mu_{ij.1...j-1}$ is the expected value of Y_{ij} given $\{Y_{i1} = y_{i1}, ..., Y_{ij-1} = y_{ij-1}\}$. Now the elements of $\boldsymbol{Y}_i$ are independent of each other with mean vector $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2.1}, ..., \mu_{in_i.12...n_i-1})'$ where $\mu_{i1} = E(Y_{i1}|\boldsymbol{x}_i)$ and $\mu_{ij.1,2,...,j-1} = E(Y_{ij.12...j-1}|\boldsymbol{x}_i)$. The covariance of $Y_{ij.12...j-1}$ is $a(\phi_{ij})V(\mu_{ij.12...j-1})$, where $a(\phi_{ij})$ is the scale parameter. If the distributional form of the outcome variable is unknown, then for each outcome variables,

$Y_{i1}, Y_{i2.1}, ..., Y_{in_i.12...n_i-1}$, that are independent of one another, we may define the quantities,

$$Q_{i1} = Q(\mu_{i1}|y_{i1}) = \int_{y_{i1}}^{\mu_{i1}} \frac{y_{i1}-t}{V(t)}dt, \tag{7.3}$$

$$\begin{aligned} Q_{ij} &= Q(\mu_{ij.12...j-1}, y_{ij.12..j-1}) \\ &= \int_{y_{ij.12...j-1}}^{\mu_{ij.12...j-1}} \frac{y_{ij.12...j-1}-t}{a(\phi_{ij})V(t)}dt;\ j=2,...,n_i. \end{aligned} \tag{7.4}$$

7.3.2 Log Quasi-likelihood Function

Given the components of $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2.1}, ..., Y_{in_i.12...n_i-1})'$ are independent, the log quasi-likelihood for the complete data is the sum of the individual contributions (as shown in equation (2.7) in Chapter 2)

$$\begin{aligned} Q_J = \sum_{i=1}^{N}\sum_{j=1}^{n_i} Q_{ij} &= \sum_{i=1}^{N}\Big[Q(\mu_{i1}|y_{i1}) + Q(\mu_{i2.1}|y_{i2.1}) + ... \\ &+ Q_(\mu_{in_i.1,2,...,n_i-1}|y_{in_i.1,2,...,n_i-1})\Big] \\ &= \sum_{i=1}^{N}\sum_{j=1}^{n_i.12...n_i-1}\int_{y_{ij}}^{\mu_{ij}} \frac{Y_{ij}-t}{a(\phi_{ij}).V(t)}dt. \end{aligned} \tag{7.5}$$

The sum of the integrals Q_J in equation (7.5), then, behaves like a log-likelihood function and following Nelder and Lee [61], Wedderburn [77], the equation (7.5) can be referred to as a log quasi-likelihood function of the parameters μ_{ij}. The $n_i.(p+1)\times 1$ vector of parameters to be estimated for the proposed model is denoted by $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2.1}, ..., \boldsymbol{\beta}_{n_i.1,2,...,n_i-1})'$, where $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, ..., \beta_{1p})'$, $\boldsymbol{\beta}_{2.1} = (\beta_{20}, \beta_{21}, ..., \beta_{2p})'$, ..., $\boldsymbol{\beta}_{n_i.1...n_i-1} = (\beta_{n_i0}, \beta_{n_i1}, ..., \beta_{n_ip})'$. Clearly, if the covariate effects on the dependent variables are similar at each follow-up, the proposed model reduces to the marginal model with $p+1$ parameters, $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)'$.

7.3.3 Score Equation and Variance Covariance Matrix

Differentiating equation (7.5) with respect to the respective parameter and equating to zero, the following quasi score equations can be obtained similarly as equation (2.8) as given in Chapter 2,

$$\begin{aligned}
\frac{\delta Q_J}{\delta \beta_{1k}} &= \sum_{i=1}^{N} \frac{\delta Q_J}{\delta \beta_{1k}} = \sum_{i=1}^{N} \left(\frac{y_{i1} - \mu_{i1}}{a(\phi_{i1}).V(\mu_{i1})} . \frac{\delta \mu_{i1}}{\delta \beta_{1k}} \right) = 0, \\
\frac{\delta Q_J}{\delta \beta_{2.1k}} &= \sum_{i=1}^{N} \frac{\delta Q_J}{\delta \beta_{2.1k}} = \sum_{i=1}^{N} \left(\frac{y_{i2.1} - \mu_{i2.1}}{a(\phi_{i2}).V(\mu_{i2.1})} . \frac{\delta \mu_{i2.1}}{\delta \beta_{2.1k}} \right) = 0, \\
&\ldots \\
&\ldots \\
\frac{\delta Q_J}{\delta \beta_{n_i.1\ldots n_i-1k}} &= \sum_{i=1}^{N} \left(\frac{y_{in_i.1\ldots n_i-1} - \mu_{in_i.1\ldots n_i-1}}{a(\phi_{in_i}).V(\mu_{in_i.1\ldots n_i-1})} . \frac{\delta \mu_{in_i.1\ldots n_i-1}}{\delta \beta_{n_i.1\ldots n_i-1k}} \right) = 0.
\end{aligned} \tag{7.6}$$

The quasi-likelihood estimating equations for $\boldsymbol{\beta}$ given in equation (7.6) and the likelihood equations for generalized linear models are equivalent and makes only second moment assumptions about the distribution of $\boldsymbol{Y}_i$ rather than full distributional assumptions required for generalized linear models.

The information matrix $\boldsymbol{I}$ is a $n_i.(p+1) \times n_i.(p+1)$ matrix with kk'th elements, $-\frac{\delta^2 Q_J}{\delta \beta_k \delta \beta_k'}$. The variance covariance matrix for the parameters of the proposed model can be expressed as

$$V(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} V(\hat{\boldsymbol{\beta}}_1) & 0 & \ldots & 0 \\ 0 & V(\hat{\boldsymbol{\beta}}_{2.1}) & \ldots & 0 \\ : & : & : & : \\ 0 & 0 & \ldots & V(\hat{\boldsymbol{\beta}}_{n_i.1,\ldots,n_i-1}) \end{pmatrix}. \tag{7.7}$$

7.3.4 Tests of Hypotheses

The test for the overall model and for individual parameters can be obtained easily as follow.

The Quasi Likelihood Ratio Test for Overall Model

We followed Wedderburn [77] for obtaining a test for the overall model. The quasi maximum likelihood estimates (QMLE) were discussed by Wedderburn [77] showing that the precision of QMLE may be estimated from the expected second derivatives of log quasi likelihood functions in the same way as the precision of maximum likelihood estimates may be estimated from the expected second derivatives of the log likelihood. Several researchers suggested the use of quasi likelihood ratio (QLR) statistic defined as the quasi-log-likelihood ratio computed from QMLE [77] under the null and the alternative hypothesis to test the significance of the overall model when the distribution is unknown (see for example [11]).
In our proposed setup, to compare the full model with a reduced model containing an intercept term only, the hypotheses can be defined as

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0^* \text{ vs. } H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0^*,$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2.1}, ..., \boldsymbol{\beta}_{n_i.1,2,...n_i-1})'$ and $\boldsymbol{\beta}_0^* = (\beta_{10}, \beta_{20}, ..., \beta_{n_i 0})'$ are the parameters of the full and reduced models, respectively. The test statistic can be defined as

$$QLR = -2[Q(\boldsymbol{\beta}_0^*) - Q(\boldsymbol{\beta})]$$

which follows a chi-square distribution with $n_i \times p$ degrees of freedom under H_0. Here $Q(\boldsymbol{\beta}_0^*)$ is the quasi likelihood of the reduced model for no covariate effects, i.e. the value of $Q(\boldsymbol{\beta})$ under H_0.

Tests for Individual Parameters: Score Test

Let us consider testing the null hypothesis

$$H_0 : \boldsymbol{C}\boldsymbol{\beta} = d \text{ vs. } H_1 : \boldsymbol{C}\boldsymbol{\beta} \neq d,$$
$$\text{where} \quad \boldsymbol{C} = \begin{bmatrix} \boldsymbol{O}_{(p-q)\times(p-q)} & \boldsymbol{O}_{(p-q)\times q} \\ \boldsymbol{O}_{q\times(p-q)} & \boldsymbol{I}_q \end{bmatrix}$$

is a $p \times q$ matrix with rank $q \leq p$ not depending on the data or $\boldsymbol{\beta}$. All elements of $\boldsymbol{O}_{(p-q)\times(p-q)}$, $\boldsymbol{O}_{(p-q)\times q}$ and $\boldsymbol{O}_{q\times(p-q)}$ are zero and $\boldsymbol{I}_q$ is the $q \times q$ identity matrix. In the absence of a likelihood function, an efficient score statistic [67] which do not involve existence of a likelihood function, can be adopted for quasi-likelihood based approaches and can be generalized directly to an estimating function setting.

Suppose, $\boldsymbol{V} = V(S_\beta) = E(S_\beta S'_\beta)$ is the variance of the quasi-score function S_β, $\hat{\boldsymbol{\beta}}$ is the unrestricted quasi-likelihood estimate of $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}_0^*$ is the quasi-likelihood estimate of $\boldsymbol{\beta}$ under H_0. Under $H_0 : \boldsymbol{C}\boldsymbol{\beta} = d$ and for ergodic case, the analog of the efficient score statistic [67],

$$\mu = S'_\beta [E(S_\beta S'_\beta)]^- S_\beta, \tag{7.8}$$

is envisioned in the circumstances under which

$$\boldsymbol{V}^{-\frac{1}{2}} S_\beta \overset{d}{\sim} \text{MVN}(\boldsymbol{0}, \boldsymbol{I}_p); \tag{7.9}$$

then μ is approximately distributed as χ^2_q under H_0.

It can be shown that $S_\beta \overset{d}{\sim} \boldsymbol{P} S_\beta$ where $\boldsymbol{P} = \boldsymbol{C}(\boldsymbol{C}'\boldsymbol{V}^{-1}\boldsymbol{C})^-\boldsymbol{C}'\boldsymbol{V}^{-1}$. Then $\boldsymbol{V}^{-\frac{1}{2}} S_\beta \overset{d}{\sim} \text{MVN}(\boldsymbol{0}, \boldsymbol{I}_p)$ and $S_\beta \overset{d}{\sim} \text{MVN}(\boldsymbol{0}, \boldsymbol{P}\boldsymbol{V}\boldsymbol{P}')$, where $V(\hat{\boldsymbol{\beta}}) \approx \boldsymbol{P}\boldsymbol{V}\boldsymbol{P}'$; hence

$$\begin{aligned} \mu &\overset{d}{\sim} (\boldsymbol{P} S_\beta)'(\boldsymbol{P}\boldsymbol{V}\boldsymbol{P}')^- \boldsymbol{P} S_\beta \\ &= (\boldsymbol{V}^{-\frac{1}{2}} S_\beta)' \boldsymbol{V}^{\frac{1}{2}} \boldsymbol{P}' (\boldsymbol{P}\boldsymbol{V}\boldsymbol{P}')^- \boldsymbol{P}\boldsymbol{V}^{\frac{1}{2}} (\boldsymbol{V}^{-\frac{1}{2}} S_\beta). \end{aligned} \tag{7.10}$$

In view of equation (7.9) and since $\boldsymbol{V}^{\frac{1}{2}} \boldsymbol{P}' (\boldsymbol{P}\boldsymbol{V}\boldsymbol{P}')^- \boldsymbol{P}\boldsymbol{V}^{\frac{1}{2}}$ is idempotent with rank q, μ is approximately distributed as χ^2 with q degrees of freedom. [32].

7.4 Simulation Study

To assess the properties of estimates (bias, standard error and 95% coverage probability) of the regression coefficients obtained by the proposed model, a simulation study was performed using R 3.4.3. For simplicity of the study, we restrict the simulation study to two follow ups of the outcome variable, Y_{i1} and Y_{i2} and only one explanatory variable, $X_{ij1}, j = 1,2$ for each of the N individuals. We assumed that Y_{i1} and Y_{i2} are two Bernoulli random variables with $Y_{i1} \sim B(1, p_{i1})$ and $Y_{i2} \sim B(1, p_{i2})$. The corresponding link functions are $g(\mu_{i1}) = \frac{e^{\boldsymbol{X}_{i1}\boldsymbol{\beta}_1}}{1+e^{\boldsymbol{X}_{i1}\boldsymbol{\beta}_1}}$ and $g(\mu_{i2}) = \frac{e^{\boldsymbol{X}_{i2}\boldsymbol{\beta}_{2.1}}}{1+e^{\boldsymbol{X}_{i2}\boldsymbol{\beta}_{2.1}}}$. At first, the explanatory variable $\boldsymbol{X}_{ij}$ was generated from $B(1, 0.5)$. For chosen values of β_1 and $\beta_{2.1}$, the probability of success for the outcome variable at the first and second time points, p_{i1} and p_{i2} respectively, were calculated as $p_{i1} = P(Y_{i1} = 1|\boldsymbol{X}_{i1} = \boldsymbol{x}_{i1}) = \frac{e^{\boldsymbol{x}_{i1}\boldsymbol{\beta}_1}}{1+e^{\boldsymbol{x}_{i1}\boldsymbol{\beta}_1}}$ and $p_{i2} = P(Y_{i2} = 1|Y_{i1} = y_{i1}, \boldsymbol{X}_{i2} = \boldsymbol{x}_{i2}) = \frac{e^{\boldsymbol{x}_{i2}\boldsymbol{\beta}_{2.1}}}{1+e^{\boldsymbol{x}_{i2}\boldsymbol{\beta}_{2.1}}}$ where $\boldsymbol{x}_{ij} = (1, x_{ij1})$, $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11})'$ and $\boldsymbol{\beta}_{2.1} = (\beta_{20}, \beta_{21})'$. 200 pairs of (Y_{i1}, Y_{i2}), were generated using the R-package 'bindata' for cases where Y_{i1} and Y_{i2} are (i) independent and identically distributed and (ii) non-identical with varying correlation (ρ = 0, 0.3, 0.5, 0.7) between Y_{i1} and Y_{i2}. For 1000 iterations, the bias, standard error and coverage probability of the 95% confidence interval of the estimates of parameters of the proposed model, $\boldsymbol{\beta} = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21})$, and the same for GEE or ALR, $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1^*)'$, were constructed over a range of scenarios. Note that, we used the notation $\boldsymbol{\beta}^*$ for the vector of parameters of GEE and ALR to distinguish the vector of parameters from the vector of parameters of the proposed marginal-conditional model.

7.5 Results

The estimate of parameters of the proposed model and the same for GEE or ALR along with their bias, standard error and coverage probability under the assumption of identical and independently distributed Y_{i1} and Y_{i2} are shown in Table 7.1. Results in Table 7.1 show that while the distribution of Y_{i1} and Y_{i2} are identical and independent, the estimates of the parameters of the GEE, ALR and the estimates of the parameters of the proposed model at all the time points are exactly same and GEE can be used.

Table 7.1: Parameters(Par), estimates(Est), bias, standard error(SE) and coverage probability(CP) of the estimates of parameters of GEE, ALR and the proposed model for independent and identically distributed Y_{i1} and Y_{i2} (β_{10}=β_{20}=0.5, β_{11}=β_{21}=0.2, ρ=0)
.

Methods	Par	Est	Bias	SE	CP
Proposed Model	β_{10}	0.504	-0.004	0.209	0.960
	β_{11}	0.206	-0.006	0.300	0.968
	β_{20}	0.504		0.209	0.950
	β_{21}	0.201		0.300	0.965
GEE(In)	β_0^*	0.503	-0.003	0.191	0.957
	β_1^*	0.203	-0.003	0.275	0.969
GEE(Ex)	β_0^*	0.503	-0.003	0.191	0.957
	β_1^*	0.203	-0.003	0.275	0.969
ALR(Ex)	β_0^*	0.503	-0.003	0.191	0.957
	β_1^*	0.203	-0.003	0.275	0.969

For non identical and dependent Y_{i1} and Y_{i2}, the estimates of the parameters of the proposed model and GEE and ALR at varying levels of association (ρ = 0.0, 0.3, 0.5 and 0.7 respectively) are shown in Table 7.2 to 7.5 respectively.
The results in Table 7.2 to 7.5 show that when Y_{i1} and Y_{i2} are correlated or when distribution of Y_{i1} and Y_{i2} are not identical, GEE or ALR, being marginal approaches, can not capture the covariate effect on response variable at different time points effectively. The average effect of the covariates at two time points are reflected in GEE and ALR which are not appropriate to present the actual scenario.

Table 7.2: Parameters (Par), estimates (Est), bias, standard error (SE) and coverage probability (CP) of estimates of parameters of GEE, ALR and the proposed model for non-identical Y_{i1} and Y_{i2}, (β_{10}=0.5, β_{11}=0.5, β_{20}=0.2, β_{21}=0.2 and ρ=0)

Methods	Par	Est	Bias from β_1	Bias from β_2	SE	CP for β_1	CP for β_2
Proposed Model	β_{10}	0.501	-0.001		0.209	0.964	
	β_{11}	0.514	-0.014		0.310	0.951	
	β_{20}	0.196		0.004	0.203		0.953
	β_{21}	0.211		-0.011	0.290		0.958
GEE(In)	β_0^*	0.345	0.155	-0.145	0.144	0.813	0.832
	β_1^*	0.347	0.153	-0.147	0.207	0.895	0.905
GEE(Ex)	β_0^*	0.345	0.155	-0.145	0.144	0.813	0.832
	β_1^*	0.347	0.153	-0.147	0.207	0.895	0.905
ALR(Ex)	β_0^*	0.345	0.155	-0.145	0.144	0.813	0.832
	β_1^*	0.347	0.153	-0.147	0.207	0.895	0.905

Table 7.3: Parameters (Par), estimates (Est), bias, standard error (SE) and coverage probability (CP) of estimates of parameters of GEE, ALR and the proposed model for dependent outcomes with non-identical distributions, (β_{10}=0.5, β_{11}=0.5, β_{20}=0.2, β_{21}=0.2 and ρ=0.3)

Methods	Par	Est	Bias from β_1	Bias from β_2	SE	CP for β_1	CP for β_2
Proposed Model	β_{10}	0.496	0.004		0.209	0.957	
	β_{11}	0.522	-0.022		0.311	0.953	
	β_{20}	0.190		0.010	0.204		0.954
	β_{21}	0.215		-0.015	0.290		0.954
GEE(In)	β_0^*	0.340	0.160	-0.140	0.164	0.821	0.881
	β_1^*	0.353	0.147	-0.153	0.237	0.908	0.911
GEE(Ex)	β_0^*	0.340	0.160	-0.140	0.164	0.821	0.881
	β_1^*	0.353	0.147	-0.153	0.237	0.908	0.911
ALR(Ex)	β_0^*	0.340	0.160	-0.140	0.164	0.821	0.881
	β_1^*	0.353	0.147	-0.153	0.237	0.908	0.911

Table 7.4: Parameters(Par), estimates(Est), bias, standard error(SE) and coverage probability(CP) of estimates of parameters of GEE, ALR and the proposed model for dependent outcomes with non-identical distributions, (β_{10}=0.5, β_{11}=0.5, β_{20}=0.2, β_{21}=0.2 and ρ=0.5)

Methods	Par	Est	Bias from β_1	Bias from β_2	SE	CP for β_1	CP for β_2
Proposed Model	β_{10}	0.496	0.004		0.209	0.960	
	β_{11}	0.517	-0.017		0.310	0.952	
	β_{20}	0.195		0.005	0.204		0.950
	β_{21}	0.213		-0.013	0.290		0.950
GEE(In)	β_0^*	0.342	0.158	-0.142	0.176	0.840	0.880
	β_1^*	0.350	0.150	-0.150	0.254	0.902	0.919
GEE(Ex)	β_0^*	0.342	0.158	-0.142	0.176	0.840	0.880
	β_1^*	0.350	0.150	-0.150	0.254	0.902	0.919
ALR(Ex)	β_0^*	0.342	0.158	-0.142	0.176	0.840	0.880
	β_1^*	0.350	0.150	-0.150	0.254	0.902	0.919

Table 7.5: Parameters(Par), estimates(Est), bias, standard error(SE) and coverage probability of estimates of parameters of of GEE, ALR and the proposed model for dependent outcomes with non-identical distributions, (β_{10}=0.5, β_{11}=0.5, β_{20}=0.2, β_{21}=0.2 and ρ=0.7)

Methods	Par	Est	Bias from β_1	Bias from β_2	SE	CP for β_1	CP for β_2
Proposed Model	β_{10}	0.497	0.003		0.209	0.958	
	β_{11}	0.517	-0.017		0.310	0.962	
	β_{20}	0.192		0.008	0.203		0.950
	β_{21}	0.213		-0.013	0.290		0.951
GEE(In)	β_0^*	0.342	0.158	-0.142	0.188	0.860	0.886
	β_1^*	0.351	0.149	-0.151	0.271	0.906	0.926
GEE(Ex)	β_0^*	0.342	0.158	-0.142	0.188	0.860	0.886
	β_1^*	0.351	0.149	-0.151	0.271	0.906	0.926
ALR(Ex)	β_0^*	0.342	0.158	-0.142	0.188	0.860	0.886
	β_1^*	0.351	0.149	-0.151	0.271	0.906	0.926

The bias of the estimates of GEE and ALR from β_1 and β_2 show that the estimates represent neither the parameters of the distribution of Y_{i1} nor the distributions of Y_{i2}. On the other hand, the proposed joint model gives better estimates of the covariate effects on response variables at different time points in terms of bias of the estimates as well as the coverage probability. Clearly it can be said that when the distributions of the outcome variables at different time points are not identical, the parameters of GEE or ALR are inadequate to portray the true covariate effect on dependent variable because GEE or ALR estimate the parameters of a population average model.

In a nutshell, GEE or ALR are appropriate to estimate the covariate effects when Y_{i1} and Y_{i2} are identically and independently distributed. The bias and the standard error of estimates of the proposed model and that of marginal model based GEE and ALR are competitive for longitudinal data when the repeated measures are independent and identical.

Nevertheless, the bias of the estimates of GEE and ALR from β_1 and $\boldsymbol{\beta_2}$ shows that GEE and ALR do not portray the actual relationship between the covariates and the response variables for cases where Y_{i1} and Y_{i2} are correlated and/or have non identical distribution. If the repeated measures are associated or outcome variables have different distributions at different time points, neither the GEE nor ALR are adequate to portray the relationship between the covariates and the response variable.

The simulation study also indicates that there is no noticeable difference among the estimates of GEE under different correlation structures or ALR with exchangeable correlation. So inducing any nuisance correlation structure in the estimation procedure does not contribute to capture the correlation among the responses and the estimates of parameters under different correlation structures are virtually the same. The proposed joint model, on the other hand, captures the dependence among the repeated responses in a built-in nature. As a result,

the proposed model provides better estimates of the covariate and outcome relationship which is portrayed in the bias and coverage probability as shown in the results of the simulation studies.

7.6 Application to HRS Data

To illustrate the proposed method, we used, as an example data, the longitudinal data from the Health and Retirement Study (HRS) conducted by the University of Michigan [73] which is a nationally representative sample data of older Americans. The study started in 1992 on American individuals over the age of 50 years and their spouses and the subjects are observed every two years. The initial HRS cohort, recruited in 1992, consisted of persons born in 1931 to 1941 (then aged 51 to 61) and their spouses of any age. The data on activities in daily living from the initial cohort was selected for this study

Activities in Daily Living Data from HRS

Difficulties in activities of daily living is a common phenomena among elderly individuals often resulting in specific physical and mental conditions [26, 70]. The term Activities of daily living (ADLs or ADL) is used in health care to refer to individuals daily self care activities. ADL has been added to and refined by a variety of researchers since it was introduced by Sidney Katz and his team [46] in 1950s [63]. Basic ADLs consist of self-care tasks that include bathing and showering, personal hygiene and grooming (including brushing/combing/styling hair), dressing, toilet hygiene, transferring, as measured by the ability to walk, get in and out of bed, and get into and out of a chair and self-feeding [46, 78]. While basic definitions of ADLs have been suggested, what specifically constitutes a particular ADL for each individual may vary. Identification of group of individuals with difficulty in performing ADL is very

important to ensure proper assistance and care for elderly people. For illustration of the proposed joint model using quasi-likelihood method and comparison of the proposed method with other selected methods, in this study, we considered ADL data from Health and Retirement Study [72]. The subset of HRS data with respondents from round 10, round 11 and round 12 from the initial HRS cohort (recruited in 1992, consisted of persons born in $1931 - 41$) was considered. In round 10, the sample size of the HRS cohort was 13593. The sample size was reduced to 7889 due to dropping of cases with missing values of outcome variable at round 10. The sample size was further reduced to 7124 at round 11 due to dropping cases with missing values of outcome variables at round 11 and and to 6246 at round 12 after dropping cases with missing values of outcome variables at round 12. Seven more cases were dropped due to missing values in the covariates. Finally the complete panel data of size 6239 from the rounds for $2010, 2012$ and 2014 have been used to illustrate the proposed quasi likelihood method and to compare with selected methods in this study. In HRS data [73], the variables on activities of daily living (ADL) included dressing, walking across room, bathing, eating, getting in/out of bed and using toilet. We constructed a binary outcome variable named Difficulty in Activities of Daily Living or DADL with values 0 and 1 (No difficulty = 0 and at least one difficulty $= 1$). The covariates are age, gender (Male$= 1$), marital status (Married or partnered$= 1$), ethnicity: White (White$= 1$) and education in years. The estimates of covariate effects on difficulties in activities of daily living using GEE, ALR and the proposed model are obtained and the findings are discussed as follow.

The estimates of covariate effects using GEE and ALR under different correlation structures are shown in Table 7.6 and estimates of covariate effects using the proposed model are shown in Table 7.7.

Table 7.6: GEE and ALR for estimating covariate effects on difficulty in activities of daily living using HRS data

Covariates	GEE(Independent)			GEE(Exchangeable)		
	$\hat{\beta}$	$SE(\hat{\beta})$	p-value	$\hat{\beta}$	$SE(\hat{\beta})$	p-value
Intercept	-2.400	0.419	0.000	-3.549	0.410	0.000
Age	0.039	0.005	0.000	0.055	0.005	0.000
Gender	-0.083	0.061	0.175	-0.143	0.061	0.019
Marital Status	-0.457	0.056	0.000	-0.368	0.053	0.000
Ethnicity: White	-0.277	0.068	0.000	-0.297	0.069	0.000
Education	-0.123	0.009	0.000	-0.122	0.009	0.000
Covariates	GEE(Autoregressive)			ALR(Exchangeable)		
	$\hat{\beta}$	SE($\hat{\beta}$)	p-value	$\hat{\beta}$	SE($\hat{\beta}$)	p-value
Intercept	-3.256	0.407	0.000	-3.438	0.420	0.000
Age	0.051	0.005	0.000	0.053	0.005	0.000
Gender	-0.153	0.061	0.013	-0.130	0.061	0.043
Marital Status	-0.360	0.053	0.000	-0.375	0.053	0.000
Ethnicity: White	-0.294	0.069	0.000	-0.299	0.069	0.000
Education	-0.125	0.009	0.000	-0.121	0.009	0.000

The GEE model under the assumption of independence detects that age, marital status, white ethnicity and education were significantly associated with DADL. GEE model under the assumption of exchangeable and autoregressive correlation and the ALR model under the exchangeable correlation showed that all the selected variables, including gender, were significantly associated with DADL. Clearly, the GEE models model the average relationship among covariates and outcome over different follow ups.

The table 7.7 described the effects of covariates on the difficulty in daily activities at different follow ups. We observed that the effects of the covariates were not the same on DADL. In the 10th and 11th round of HRS, age, marital status, White ethnicity and education were significantly associated with difficulty in activities of daily living (DADL). Age increased the risk of DADL. Married people or who had a partner had less risk of DADL as compared to their single counterparts; White people had less risk of DADL and more were the respondents educated, less they experienced DADL.

Table 7.7: Marginal-conditional model for estimating covariate effects on difficulty in activities of daily living using HRS data

	$\hat{\beta}_1$	$SE(\hat{\beta}_1)$	p-value	$\hat{\beta}_{2.1}$	$SE(\hat{\beta}_{2.1})$	p-value	$\hat{\beta}_{3.12}$	$SE(\hat{\beta}_{3.12})$	p-value
Intercept	-1.313	0.480	0.006	-2.035	0.480	0.000	-2.643	0.455	0.000
Age	0.026	0.006	0.000	0.032	0.006	0.000	0.044	0.006	0.000
Gender	-0.091	0.079	0.245	0.001	0.076	0.988	-0.124	0.070	0.075
Marital Status	-0.407	0.076	0.000	-0.465	0.074	0.000	-0.472	0.067	0.000
Ethnicity: White	-0.351	0.085	0.000	-0.293	0.084	0.000	-0.203	0.078	0.010
Education	-0.136	0.011	0.000	-0.114	0.010	0.000	-0.122	0.010	0.000

In the 10th and 11th rounds, gender had no significant association with DADL. In round 12, along with age, marital status, white ethnicity and education, gender also had significant association with DADL. Nevertheless, effect of all covariates differ at different follow ups as shown by the varying parameter estimates at different round. For example, the estimates of regression parameter for covariate age in 10th 11th and 12th round are 0.026, 0.032 and 0.044 respectively.

These findings confirm our assertion that the GEE based models which are extensively used fails to specify the covariate effects adequately when the covariate effects are possibly different at different time points in case of a longitudinal data and might give us misleading conclusions. A joint model based on marginal-conditional approach explains the covariate effects more explicitly and more meaningfully and is suggested in such cases.

7.7 Conclusion

In this chapter we proposed a joint model using a marginal-conditional approach in the quasi-likelihood set up for modelling correlated binary data along with necessary related tests. In a marginal-conditional model, the relationship among covariates and outcome variables are studied at different time points and hence the models are expected to portray the outcome-covariate relationship in a more meaningful and explicit way as compared to GEE or ALR based on marginal approaches. Unlike GEE or GEE based approaches, the proposed method do not need to estimate any correlation parameter but takes care of the probable correlation among repeated outcomes in a built in nature and estimates the covariate effects on the response variable more effectively. This model can be extended for any number of repeated measures without complicating the theory. For outcome variables with known distributions, the proposed method can simply be used with likelihood based estimation procedures as shown in Chap-

ter 5 and Chapter 6.

The simulation studies showed that estimates of parameters of GEE and ALR performed well in terms of bias and coverage probability when the outcome variables are independent and identically distributed and the estimates of the proposed model are competitive to GEE and ALR. When the outcome variables are correlated or the distribution of outcome variables are not identical, estimates of the proposed method has less bias and a better coverage probability than GEE or ALR and hence would be a better choice than those methods in analysing correlated binary data. From the simulation study and the application of GEE, ALR and the proposed method on HRS data, it is evident that the proposed method is expected to provide more specified model in a simpler set up and the results are expected to be more useful to the theoreticians and the practitioners.

Chapter 8

Conclusion

8.1 Introduction

The repeated responses on each individual are expected to be correlated which is a major feature of the longitudinal data and this makes the analysis of such data challenging. Two common approaches for analysing correlated binary outcomes are marginal and conditional modelling. In the theoretical and applied literature of statistical data analysis, there is an apparent agreement that the selection of a model must depend on the question under study, but the disagreement over when to choose which model is yet to be settled [60]. This study is an attempt to contribute in this area by proposing a joint model based on marginal conditional approach for analysing longitudinal data, using likelihood methods when applicable (Chapter 5). We studied the properties of the proposed joint

models in terms of bias and 95% coverage probability of the estimates (Chapter 6). We also developed a new model for repeated measures data with unknown distribution under the set up of a quasi-likelihood method (Chapter 7).

8.2 Major Findings

The first objective of this study was to examine selected popular methods for analysing repeated binary data in order to figure out the advantages and limitations of the approaches and to examine the selected methods to study how these methods addressed the dependence relationship among the repeated responses. We discussed on the marginal models, GEE and ALR in details (Chapter 3 and Chapter 4) in terms of assumptions, correlation structures, estimation techniques and the advantages and limitations of the approaches. We diagnosed that a major limitation of the marginal models lie in the way they address the correlation among repeated responses. We showed that

- GEE, ALR and the model suggested by Zeger et al. [80] are based on induced correlations which may be far from the true scenario (section 4.3, 4.4 and 4.5, Chapter 4);
- inducing a correlation structure in GEE contradicts with the basic assumptions of GEE unless the correlation structure considered is independent correlation (section 4.3, Chapter 4);
- the induced correlation used in Zeger et al. [80] does not reflect the true correlation (section 4.5, Chapter 4);
- Darlington and Farewell [16]'s model was close to address the correlation among the repeated responses, however, their method could not address the correlation completely as they did not consider all possible transition or conditional probabilities in second or higher follow ups (section 4.6, Chapter 4).

The second objective was to propose a joint model as an alternative to GEE or ALR for the analysis of longitudinal binary data incorporating the true dependence of repeated outcomes using likelihood based methodologies.

In Chapter 5 we proposed, as alternatives to GEE or ALR, three joint models based on marginal conditional approach. Among the three models, the first one, $(MCM1)$, is an extension of Darlington and Farewell [16]'s model, second one, $MCM2$, is the joint model proposed by Islam et al. [43]. The third one, $MCM3$, is a joint model based on an extended regressive model Islam et al. [42].

Note that the joint models based on a marginal conditional approach is not a new approach from the technical point of view. But the joint models of the earlier works mainly focused on the estimates of marginal and transition probabilities, testing the association parameter and/or order of dependence among repeated outcomes [3, 35, 36, 38, 40, 43].

Although the estimation of parameters of the models to study the dependence relationship among covariates and outcome variables are available in literature, but the properties of the estimates were not. The following theoretical exploration was done in the study.

- The idea of using a marginal conditional model to estimate the covariate effects on repeated measures data is proposed in Chapter 5 and the properties of the joint models in terms of bias, standard error and 95% coverage probability of the estimates are demonstrated in Chapter 6.
- The proposed model 2 $(MCM2)$ for longitudinal binary data can easily be extended for exponential family members. The likelihood function, the score equations and information matrix as well as the relevant test procedure to test the overall model and the individual parameters were described for exponential family.

- Since the number of parameters of proposed model 2 increases rapidly with increase in the number of repeated outcomes on one individual, we proposed the use of a joint model using extended regressive model approach [42] which we denoted as proposed model 3 or $MCM3$.

We conducted a number of simulation studies to examine the properties of the proposed joint models based on marginal-conditional approach and to compare them with GEE and ALR in terms of bias, standard error and 95% coverage probability. The simulation studies are conducted under different conditions: repeated outcomes are (i) independent and identically distributed, (ii) independent but not identically distributed and (iii) associated (section 6.3, Chapter 6). We observed that

- the biases of estimates of the proposed joint models are very small and the coverage probability of the estimates are more than 90%.
- the estimates of the parameters of GEE and ALR performed well in terms of bias and coverage probability when the outcome variables are identically and independently distributed and all three proposed joint models show competitive bias and coverage probability.
- When the outcome variables are not identical and/or are correlated, the estimates of the parameters of the proposed models perform better than GEE or ALR with smaller bias and better coverage probability.

Another important objective of this study was to propose a new marginal conditional model under a quasi-likelihood setup for the analysis of longitudinal binary data when the distribution of the repeated outcomes are assumed to be unknown. This study showed the development of a new model based on quasi-likelihood approach. The corresponding quasi-likelihood functions, quasi-estimating equations for parameters of the proposed models along with score equations and information matrix, and related tests of hypothesis are also developed (Chapter 7).

To study the properties of the estimates of the parameters (bias, 95% coverage probability) of the proposed new model under quasi likelihood set up, and to compare the proposed model with GEE and ALR, we conducted another set of simulation studies under varying conditions. We found that

- the estimates of the parameters of GEE and ALR performed well in terms of bias and coverage probability only when the outcome variables are identically and independently distributed. Note that the proposed new model using quasi-likelihood approach shows competitive results (similar bias and coverage probability).
- When the outcome variables are not identical and/or are correlated, the estimates of the proposed method has less bias and a better coverage probability than GEE or ALR.

Finally, we used HRS data [73] to illustrate the proposed models and showed that the proposed method is expected to provide more specified model in a simpler set up as compared to popular GEE or ALR.

8.3 Recommendations

It is evident from the study that, marginal models can be useful, given that the scientific question explicitly requires such a model formulation. But a joint model based on a marginal-conditional approach is a more logical choice to explain how covariates are associated with a nonnormal response at different time points. We recommend the use of proposed marginal conditional models (Proposed models 1, 2 and 3 or $MCM1$, $MCM2$ and $MCM3$), based on likelihood methods and the newly developed marginal conditional model based on quasi-likelihood approach as better alternatives of GEE or related approaches for correlated binary outcomes because

- the proposed models are simple and easy to fit, understand and interpret;

- they consider the true correlation among the repeated measures in a built-in nature and there is no need to estimate any correlation parameter;
- they can be extended to any number of follow ups without complicating the theory.
- performance of the proposed models is similar to GEE or ALR in terms of bias of the estimators and coverage probability when the repeated outcomes are independent and identically distributed;
- the proposed models perform better than GEE or ALR when the repeated outcomes are not independent and/or identically distributed at different time points.

8.4 Further Scope of Study

An extension of this study could be developing the procedure of missing data analysis in marginal conditional models (MCM). Further scopes of this study may also include a comparison of the proposed marginal conditional models with generalized linear mixed models.

Appendix

A1. R Codes for MCM2, GEE and ALR in Chapter 6

```
############################### FUNCTIONS ##########################
#################### Generating data Y1 and Y2  ####################
modsim<-function(N,intercept1,intercept2,beta1,beta2, gamma1,x1){
## simulate y1
id <- 1:N
xbeta <- intercept1 +  beta1 *x1
proba1 <- exp(xbeta)/(1 + exp(xbeta))
Y1 <- ifelse(runif(N,0,1) < proba1,1,0)
# print(table(Y1))  # print(proba1)

# simulate y2
xbeta2 <- intercept2 +  beta2 *x1 + gamma1 * Y1
proba2 <- exp(xbeta2)/(1 + exp(xbeta2))
Y2 <- ifelse(runif(N,0,1) < proba2,1,0)
# print(table(Y2))

dat <- data.frame(id,Y1,Y2,x1,x1)
alp<-cor(dat[2:3])
pij<-data.frame(proba1, proba2)

sdatacor<-list(alp, dat, pij)
# print(sdatacor)
return(sdatacor)
}
################## simulation ######################
# Data is generated for the models:
# Y1 = intercept1 + beta1 * x1;
# Y2 = intercept2 + beta2 * x2 + gamma1*Y1;
# initial values required are :
# totsim: total number of simulation,
# t.seed: seed,
```

```
# N: sample size,
# intercept1 :
# intercept2 :
# beta1:
# beta2:
# x1:
newsimbr<-function(totsim,t.seed,N,intercept1,
intercept2,beta1,beta2,gamma1,x1)
{for (i in 1:totsim){
cat("i=",i,"\n")
dd<- modsim(N,intercept1,intercept2,beta1,beta2,
gamma1,x1)
sdata<-data.frame(dd[2])
colnames(sdata)<-c("id", "Y1","Y2","x1","x2")
pij<-data.frame(dd[3])
#correlation matrix of Y1, Y2
cmatr<-data.frame(dd[1])
#correlation between Y1 and Y2
calpha<-cmatr[1,2]
sdata01<-subset(sdata,Y1==0)
sdata11<-subset(sdata,Y1==1)

###### Fitting Models with parameters ##############
##### beta1, beta2, beta01, and beta11 #############
print("++++++++models+++++++++")
# Fitting model with parameters beta1
mod1 <-glm(Y1~x1,family=binomial,data=sdata)
#### Fitting model with parameters beta2
mod2 <-glm(Y2~x1,family=binomial,data=sdata)
#### Fitting model with parameters beta01
mod01<-glm(Y2~x1,family=binomial,data=sdata01)
#### Fitting model with parameters beta11
mod11<-glm(Y2~x1,family=binomial,data=sdata11)

###### Add models for GEE
###### Rearrange the data
dat1<-sdata[,c(1,2,4)]
dat2<-sdata[,c(1,3,4)]
colnames(dat1)<-c("id","Y","x1")
colnames(dat2)<-c("id","Y","x1")

dat<-rbind(dat1,dat2)
dat<-arrange(dat,id)
#print(head(dat))
colnames(dat)<-c("id", "Y", "x1")

## Fitting GLM for beta
mod<-glm(Y~x1, family=binomial, data=dat)

geeI<-geeglm(Y~x1,family=binomial("logit"),id=id,
corstr="independence",std.err="san.se",data=dat)
print(summary(geeI))
```

```
geeE<-geeglm(Y~x1,family=binomial(link="logit"),id=id,
corstr="exchangeable",std.err="san.se",data=dat)
print(summary(geeE))

geeA<-geeglm(Y~x1,family=binomial("logit"),
id=id, corstr="ar1",std.err="san.se",data=dat)
print(summary(geeA))

# Add models for ALR
alrmod <- alr(dat$Y ~ dat$x1, id=dat$id,
depm="exchangeable", ainit=0.2)
print(summary(alrmod))
alralpha<-alrmod$alpha

#### Test for the equality of parameters of
#### marginal and conditional models.
dpt<- dtest(trt01=mod01,trt11=mod11,trt2=mod2)
# print(dpt)

#  Add results of GEE and ALR in allres
# "allres" is the collection of all results

allres<-data.frame(cbind(
matrix(mod$coefficients,nrow=1),
matrix(summary(mod1)$coeff[,4],nrow=1),
matrix(mod1$coefficients,nrow=1),
matrix(summary(mod1)$coeff[,4],nrow=1),
matrix(mod2$coefficients,nrow=1),
matrix(summary(mod2)$coeff[,4],nrow=1),
matrix(mod01$coefficients,nrow=1),
matrix(summary(mod01)$coeff[,4],nrow=1),
matrix(mod11$coefficients,nrow=1),
matrix(summary(mod11)$coeff[,4],nrow=1),
matrix(geeI$coefficients,nrow=1),
matrix(summary(geeI)$coeff[,4],nrow=1),
matrix(geeE$coefficients,nrow=1),
matrix(summary(geeE)$coeff[,4],nrow=1),
matrix(geeA$coefficients,nrow=1),
matrix(summary(geeA)$coeff[,4],nrow=1),
matrix(alrmod$coefficients, nrow=1), alralpha,
matrix(summary(mod)$coefficients[,2], nrow=1),
matrix(summary(mod1)$coefficients[,2], nrow=1),
matrix(summary(mod2)$coefficients[,2], nrow=1),
matrix(summary(mod01)$coefficients[,2], nrow=1),
matrix(summary(mod11)$coefficients[,2], nrow=1),
matrix(summary(geeI)$coefficients[,2], nrow=1),
matrix(summary(geeE)$coefficients[,2], nrow=1),
matrix(summary(geeA)$coefficients[,2], nrow=1),
matrix(summary(alrmod)$coefficients[,2], nrow=1),
intercept1,intercept2,beta1,beta2,gamma1,
cbind(dpt[1,c(2,4)],dpt[2,c(2,4)],dpt[3,c(2,4)],
```

```
dpt[1,3]), t.seed
))
if(i==1){
myres<-allres
}
if(i>1){
myres<-rbind(myres,allres)
}}  ## end of simulation

# Add column names to the results

colnames(myres)<-c("mb0","mb1","mp0","mp1",
"m1b0","m1b1","m1p0","m1p1", "m2b0","m2b1", "m2p0","m2p1",
"m01b0","m01b1","m01p0","m01p1", "m11b0","m11b1","m11p0","m11p1",
"geeIb0", "geeIb1","geeIb0p0","geeIb1p1",
"geeEb0", "geeEb1","geeEb0p0","geeEb1p1",
"geeAb0", "geeAb1","geeAb0p0","geeAb1p1",
"alrb0", "alrb1", "alralpha",
"mb0se", "mb1se","m1b0se", "m1b1se",
"m2b0se", "m2b1se","m01b0se", "m01b1se",
"m11b0se", "m11b1se","geeIb0se", "geeIb1se",
"geeEb0se", "geeEb1se","geeAb0se", "geeAb1se",
"alrb0se", "alrb1se",
"intercept1","intercept2","beta1","beta2","gamma1",
"b01b11ch","pv","b01b2ch","pv","b11b2ch","pv","df",
"t.seed")
res<-list(myres)
return(res)
}   #  end function

########Function for testing ######################
########the equality of beta 1 and beta 2###########

dtest<-function(trt01,trt11,trt2){
co1 <-matrix(trt01$coefficients,nrow=1) #Model 01
co2 <-matrix(trt11$coefficients,nrow=1) #Model 11
co4 <-matrix(trt2$coefficients,nrow=1)  #Model 2(y2)
nvr <-dim(co4)[2]
#  cat("nvr")  # Number of variables
#   print(nvr)
sc1  <- vcov(trt01)
sc2  <- vcov(trt11)
sc4  <- vcov(trt2)
#cat("sc1") #   print(sc1) # print(sc2)
#   print(dim(sc1)) # print(dim(sc2)) # print(sc4)

## testing beta01 vs. beta11

pab1 <- sc1+sc2
tt2  <- matrix((co1-co2),nrow=1)
ch3  <- tt2%*%solve(pab1)%*%t(tt2)
chp3 <- 1-pchisq(ch3, nvr,ncp=0,
```

```
lower.tail = TRUE, log.p = FALSE)

## testing beta01 vs. beta2

pab4 <- sc1+sc4
tt3  <- matrix((co1-co4),nrow=1)
ch4  <- tt3 %*%solve(pab4)%*%t(tt3)
chp4 <- 1-pchisq(ch4, nvr,ncp=0,
lower.tail = TRUE, log.p = FALSE)

## testing beta11 vs. beta2

pab5 <- sc2+sc4
tt4  <- matrix((co2-co4),nrow=1)
ch5  <- tt4 %*%solve(pab5)%*%t(tt4)
chp5 <- 1-pchisq(ch5, nvr,ncp=0,
lower.tail = TRUE, log.p = FALSE)

chtv<-rbind(ch3,ch4,ch5)
chpv<-rbind(chp3,chp4,chp5)
rrn<-c("Beta_01 VS Beta_11:",
"Beta_01 VS Beta_2:","Beta_11 VS Beta_2:")
deres<-data.frame(Test=rrn,chi_square=chtv,
d.f.=nvr,p_value=chpv)
return(deres)
}   # end of function

#########################################################################
## Codes for the Simulation Studies for MCM2, GEE and ALR ########
## in Chapter 6 ######################
#########################################################################
library(MASS)
library(magic)
library(geepack)
library(dplyr)
library(alr)
t.seed     <- 3456
set.seed(t.seed)

totsim  <- 1000 ## Number of Simulations
N       <- 200 ## Sample Size
t       <- 2 ## Number of observations per subject

#TRIAL1# Identical population, non-correlated data

intercept1 <- 0.5; intercept2 <- 0.5; beta1 <- 0.2
beta2      <- 0.2; gamma1     <- 0.0

#TRIAL2# Identical population, correlated data

intercept1 <- 0.5; intercept2 <- 0.5; beta1 <- 0.2
beta2      <- 0.2; gamma1     <- 1.0
```

```
#TRIAL3# Non-Identical, non-correlated data

intercept1 <- 0.5; intercept2 <- 0.2; beta1 <- 0.2
beta2      <- 0.7; gamma1     <- 0.0

#TRIAL 4# Non-Identical, correlated data

intercept1 <- 0.5; intercept2 <- 0.2;  beta1      <- 0.2
beta2      <- 0.7; gamma1     <- 1.0
################################################
x0<- rep(1, N)
x1 <- rbinom(N, 1, .5)
init <-c(0.4, 25, 0.25)
bta<-c(init[1], init[2], init[3])
source("G:/phd/R_codes/model.R")
mysim2<- newsimbr(totsim,t.seed,N,intercept1,
intercept2,beta1,beta2,gamma1,x1)

mysim<-data.frame(mysim2)
i<-1
for(i in 1: length(mysim$m01b0))
{  if(mysim$m01b1se[i]> 2.0)
{mysim<-mysim[-i,]}
i<-i+1
}
totsim<-length(mysim$m01b0)
i<-1
for(i in 1: length(mysim$m01b0))
{  if(mysim$m11b1se[i]> 2.0)
{mysim<-mysim[-i,]}
i<-i+1
}
totsim<-length(mysim$m01b0)

est.mb0<-mean(mysim$mb0)
est.mb1<-mean(mysim$mb1)
bias1.mb0<-mean(intercept1-mysim$mb0)
#Bias from intercept1
bias1.mb1<-mean(beta1-mysim$m1b1)
#Bias from beta1
bias2.mb0<-mean(intercept2-mysim$mb0)
#Bias from intercept2
bias2.mb1<-mean(beta2-mysim$m1b1)
#Bias from beta2
lcimb0<-mysim$mb0 - 1.96*mysim$mb0se
ucimb0<-mysim$mb0 + 1.96*mysim$mb0se
lcimb1<-mysim$mb1 - 1.96*mysim$mb1se
ucimb1<-mysim$mb1 + 1.96*mysim$mb1se

### CP for intercept1 and beta1 ### Model P(Y_ij=1|X)
CP1mb0<-rep(0, totsim)
```

```
CP1mb1<-rep(0, totsim)
for(i in 1: length(lcimb0)){
if(intercept1<=ucimb0[i] & intercept1>=lcimb0[i])
{CP1mb0[i]<-1}
if(beta1<ucimb1[i] & beta1>lcimb1[i])
{CP1mb1[i]<-1}
}
CP1m.b0<-mean(CP1mb0)
CP1m.b1<-mean(CP1mb1)
se.mb0<-mean(mysim$mb0se)
se.mb1<-mean(mysim$mb1se)

### CP for intercept2 and beta2 ## Model P(Y_ij=1|X)
CP2mb0<-rep(0, totsim)
CP2mb1<-rep(0, totsim)
for(i in 1: length(lcimb0)){
if(intercept2<=ucimb0[i] & intercept2>=lcimb0[i])
{CP2mb0[i]<-1}
if(beta2<ucimb1[i] & beta2>lcimb1[i])
{CP2mb1[i]<-1}
}
CP2m.b0<-mean(CP2mb0)
CP2m.b1<-mean(CP2mb1)

est.m1b0<-mean(mysim$m1b0)
est.m1b1<-mean(mysim$m1b1)
bias.m1b0<-mean(intercept1-mysim$m1b0)
bias.m1b1<-mean(beta1-mysim$m1b1)
lcim1b0<-mysim$m1b0 - 1.96*mysim$m1b0se
ucim1b0<-mysim$m1b0 + 1.96*mysim$m1b0se
lcim1b1<-mysim$m1b1 - 1.96*mysim$m1b1se
ucim1b1<-mysim$m1b1 + 1.96*mysim$m1b1se

### CP of intercept1 and beta1 ## Model P(Y1=1|X)
CPm1b0<-rep(0, totsim)
CPm1b1<-rep(0, totsim)
for(i in 1: length(lcim1b0)){
if(intercept1<=ucim1b0[i] & intercept1>=lcim1b0[i])
{CPm1b0[i]<-1}
if(beta1<ucim1b1[i] & beta1>lcim1b1[i])
{CPm1b1[i]<-1}
}
CPm1.b0<-mean(CPm1b0)
CPm1.b1<-mean(CPm1b1)
se.m1b0<-mean(mysim$m1b0se)
se.m1b1<-mean(mysim$m1b1se)

est.m2b0<-mean(mysim$m2b0)
est.m2b1<-mean(mysim$m2b1)
bias.m2b0<-mean(intercept2-mysim$m2b0)
bias.m2b1<-mean(beta2-mysim$m2b1)
lcim2b0<-mysim$m2b0 - 1.96*mysim$m2b0se
```

```
ucim2b0<-mysim$m2b0 + 1.96*mysim$m2b0se
lcim2b1<-mysim$m2b1 - 1.96*mysim$m2b1se
ucim2b1<-mysim$m2b1 + 1.96*mysim$m2b1se

### CP for intercept2 and beta2 ## Model P(Y2=1|X)
CPm2b0<-rep(0, totsim)
CPm2b1<-rep(0, totsim)
for(i in 1: length(lcim2b0)){
if(intercept2<ucim2b0[i] & intercept2>lcim2b0[i])
CPm2b0[i]<-1
if(beta2<ucim2b1[i] & beta2>lcim2b1[i])
CPm2b1[i]<-1
}
CPm2.b0<-mean(CPm2b0)
CPm2.b1<-mean(CPm2b1)
se.m2b0<-mean(mysim$m2b0se)
se.m2b1<-mean(mysim$m2b1se)

est.m01b0<-mean(mysim$m01b0)
est.m01b1<-mean(mysim$m01b1)
bias.m01b0<-mean(intercept2-mysim$m01b0)
bias.m01b1<-mean(beta2-mysim$m01b1)
lcim01b0<-mysim$m01b0 - 1.96*mysim$m01b0se
ucim01b0<-mysim$m01b0 + 1.96*mysim$m01b0se
lcim01b1<-mysim$m01b1 - 1.96*mysim$m01b1se
ucim01b1<-mysim$m01b1 + 1.96*mysim$m01b1se

### CP of intercept2 and beta2 # Model 0-1 P(Y2=1|Y1=0,X)
CPm01b0<-rep(0, totsim)
CPm01b1<-rep(0, totsim)
for(i in 1: length(lcim01b0)){
if(intercept2<ucim01b0[i] & intercept2>lcim01b0[i])
CPm01b0[i]<-1
if(beta2<ucim01b1[i] & beta2>lcim01b1[i])
CPm01b1[i]<-1
}
CPm01.b0<-mean(CPm01b0)
CPm01.b1<-mean(CPm01b1)
se.m01b0<-mean(mysim$m01b0se)
se.m01b1<-mean(mysim$m01b1se)

est.m11b0<-mean(mysim$m11b0)
est.m11b1<-mean(mysim$m11b1)
bias.m11b0<-mean((intercept2+gamma1)-mysim$m11b0)
bias.m11b1<-mean((beta2)-mysim$m11b1)
lcim11b0<-mysim$m11b0 - 1.96*mysim$m11b0se
ucim11b0<-mysim$m11b0 + 1.96*mysim$m11b0se
lcim11b1<-mysim$m11b1 - 1.96*mysim$m11b1se
ucim11b1<-mysim$m11b1 + 1.96*mysim$m11b1se

### CP of intercept2 and beta2 # Model 1-1 P(Y2=1|Y1=1,X)
CPm11b0<-rep(0, totsim)
```

```
CPm11b1<-rep(0, totsim)
for(i in 1: totsim){
if(intercept2+gamma1<ucim11b0[i]
& intercept2+gamma1>lcim11b0[i])
CPm11b0[i]<-1
if(beta2<ucim11b1[i] & beta2>lcim11b1[i])
CPm11b1[i]<-1
}
CPm11.b0<-mean(CPm11b0)
CPm11.b1<-mean(CPm11b1)
se.m11b0<-mean(mysim$m11b0se)
se.m11b1<-mean(mysim$m11b1se)

est.geeIb0<-mean(mysim$geeIb0)
est.geeIb1<-mean(mysim$geeIb1)
bias1.geeIb0<-mean(intercept1-mysim$geeIb0)
bias1.geeIb1<-mean(beta1-mysim$geeIb1)
bias2.geeIb0<-mean(intercept2-mysim$geeIb0)
bias2.geeIb1<-mean(beta2-mysim$geeIb1)
lcigeeIb0<-mysim$geeIb0 - 1.96*mysim$geeIb0se
ucigeeIb0<-mysim$geeIb0 + 1.96*mysim$geeIb0se
lcigeeIb1<-mysim$geeIb1 - 1.96*mysim$geeIb1se
ucigeeIb1<-mysim$geeIb1 + 1.96*mysim$geeIb1se

### CP for intercept1 and beta1 ## Model geeI

CP1geeIb0<-rep(0, totsim)
CP1geeIb1<-rep(0, totsim)
for(i in 1: totsim){
if(intercept1<ucigeeIb0[i] & intercept1>lcigeeIb0[i])
CP1geeIb0[i]<-1
if(beta1<ucigeeIb1[i] & beta1>lcigeeIb1[i])
CP1geeIb1[i]<-1
}
CP1geeI.b0<-mean(CP1geeIb0)
CP1geeI.b1<-mean(CP1geeIb1)

### CP for intercept2 and beta2 ## Model geeI
CP2geeIb0<-rep(0, totsim)
CP2geeIb1<-rep(0, totsim)
for(i in 1: totsim){
if(intercept2<ucigeeIb0[i] & intercept2>lcigeeIb0[i])
CP2geeIb0[i]<-1
if(beta2<ucigeeIb1[i] & beta2>lcigeeIb1[i])
CP2geeIb1[i]<-1
}
CP2geeI.b0<-mean(CP2geeIb0)
CP2geeI.b1<-mean(CP2geeIb1)
se.geeIb0<-mean(mysim$geeIb0se)
se.geeIb1<-mean(mysim$geeIb1se)

est.geeEb0<-mean(mysim$geeEb0)
```

```
est.geeEb1<-mean(mysim$geeEb1)
bias1.geeEb0<-mean(intercept1-mysim$geeEb0)
bias1.geeEb1<-mean(beta1-mysim$geeEb1)
bias2.geeEb0<-mean(intercept2-mysim$geeEb0)
bias2.geeEb1<-mean(beta2-mysim$geeEb1)
lcigeeEb0<-mysim$geeEb0 - 1.96*mysim$geeEb0se
ucigeeEb0<-mysim$geeEb0 + 1.96*mysim$geeEb0se
lcigeeEb1<-mysim$geeEb1 - 1.96*mysim$geeEb1se
ucigeeEb1<-mysim$geeEb1 + 1.96*mysim$geeEb1se

### CP for intercept1 and beta1 ## Model geeE
CP1geeEb0<-rep(0, totsim)
CP1geeEb1<-rep(0, totsim)
for(i in 1: totsim){
if(intercept1<ucigeeEb0[i] & intercept1>lcigeeEb0[i])
CP1geeEb0[i]<-1
if(beta1<ucigeeEb1[i] & beta1>lcigeeEb1[i])
CP1geeEb1[i]<-1
}
CP1geeE.b0<-mean(CP1geeEb0)
CP1geeE.b1<-mean(CP1geeEb1)

### CP for intercept2 and beta2 ### Model geeE
CP2geeEb0<-rep(0, totsim)
CP2geeEb1<-rep(0, totsim)
for(i in 1: totsim){
if(intercept2<ucigeeEb0[i] & intercept2>lcigeeEb0[i])
CP2geeEb0[i]<-1
if(beta2<ucigeeEb1[i] & beta2>lcigeeEb1[i])
CP2geeEb1[i]<-1
}
CP2geeE.b0<-mean(CP2geeEb0)
CP2geeE.b1<-mean(CP2geeEb1)
se.geeEb0<-mean(mysim$geeEb0se)
se.geeEb1<-mean(mysim$geeEb1se)

est.geeAb0<-mean(mysim$geeAb0)
est.geeAb1<-mean(mysim$geeAb1)
bias1.geeAb0<-mean(intercept1-mysim$geeAb0)
bias1.geeAb1<-mean(beta1-mysim$geeAb1)
bias2.geeAb0<-mean(intercept2-mysim$geeAb0)
bias2.geeAb1<-mean(beta2-mysim$geeAb1)
lcigeeAb0<-mysim$geeAb0 - 1.96*mysim$geeAb0se
ucigeeAb0<-mysim$geeAb0 + 1.96*mysim$geeAb0se
lcigeeAb1<-mysim$geeAb1 - 1.96*mysim$geeAb1se
ucigeeAb1<-mysim$geeAb1 + 1.96*mysim$geeAb1se

### CP for intercept1 and beta1 #### Model geeA
CP1geeAb0<-rep(0, totsim)
CP1geeAb1<-rep(0, totsim)
for(i in 1: totsim){
if(intercept1<ucigeeAb0[i] & intercept1>lcigeeAb0[i])
```

```
CP1geeAb0[i]<-1
if(beta1<ucigeeAb1[i] & beta1>lcigeeAb1[i])
CP1geeAb1[i]<-1
}
CP1geeA.b0<-mean(CP1geeAb0)
CP1geeA.b1<-mean(CP1geeAb1)

### CP for intercept2 and beta_2 ## Model geeA
CP2geeAb0<-rep(0, totsim)
CP2geeAb1<-rep(0, totsim)

for(i in 1: totsim){
if(intercept2<ucigeeAb0[i] & intercept2>lcigeeAb0[i])
CP2geeAb0[i]<-1
if(beta2<ucigeeAb1[i] & beta2>lcigeeAb1[i])
CP2geeAb1[i]<-1
}
CP2geeA.b0<-mean(CP2geeAb0)
CP2geeA.b1<-mean(CP2geeAb1)
se.geeAb0<-mean(mysim$geeAb0se)
se.geeAb1<-mean(mysim$geeAb1se)

est.alrb0<-mean(mysim$alrb0)
est.alrb1<-mean(mysim$alrb1)
bias1.alrb0<-mean(intercept1-mysim$alrb0)
bias1.alrb1<-mean(beta1-mysim$alrb1)
bias2.alrb0<-mean(intercept2-mysim$alrb0)
bias2.alrb1<-mean(beta2-mysim$alrb1)
lcialrb0<-mysim$alrb0 - 1.96*mysim$alrb0se
ucialrb0<-mysim$alrb0 + 1.96*mysim$alrb0se
lcialrb1<-mysim$alrb1 - 1.96*mysim$alrb1se
ucialrb1<-mysim$alrb1 + 1.96*mysim$alrb1se

### CP for intercept1 and beta1 ## Model  alr
CP1alrb0<-rep(0, totsim)
CP1alrb1<-rep(0, totsim)
for(i in 1: totsim){
if(intercept1<ucialrb0[i] & intercept1>lcialrb0[i])
CP1alrb0[i]<-1
if(beta1<ucialrb1[i] & beta1>lcialrb1[i])
CP1alrb1[i]<-1
}
CP1alr.b0<-mean(CP1alrb0)
CP1alr.b1<-mean(CP1alrb1)

## CP for intercept2 and beta_2 ### Model alr
CP2alrb0<-rep(0, totsim)
CP2alrb1<-rep(0, totsim)
for(i in 1: totsim){
if(intercept2<ucialrb0[i] & intercept2>lcialrb0[i])
CP2alrb0[i]<-1
if(beta2<ucialrb1[i] & beta2>lcialrb1[i])
```

```
CP2alrb1[i]<-1
}
CP2alr.b0<-mean(CP2alrb0)
CP2alr.b1<-mean(CP2alrb1)

se.alrb0<-mean(mysim$alrb0se)
se.alrb1<-mean(mysim$alrb1se)

#######   Simulation Outputs
# 1. Mean of estimates # 2. Bias # 3. Cov Prob
tab1<-rbind(
cbind(est.mb0, bias1.mb0, bias2.mb0, se.mb0,
CP1m.b0, CP2m.b0),
cbind(est.mb1, bias1.mb1, bias2.mb1, se.mb1,
CP1m.b1, CP2m.b1),
cbind(est.m1b0, bias.m1b0, , se.m1b0, CPm1.b0,),
cbind(est.m1b1, bias.m1b1, ,se.m1b1, CPm1.b1,),
cbind(est.m2b0,, bias.m2b0, se.m2b0, , CPm2.b0),
cbind(est.m2b1,,bias.m2b1,se.m2b1, , CPm2.b1),
cbind(est.m01b0,,bias.m01b0,se.m01b0, ,CPm01.b0),
cbind(est.m01b1,,bias.m01b1,se.m01b1, ,CPm01.b1),
cbind(est.m11b0,,bias.m11b0,se.m11b0,,CPm11.b0),
cbind(est.m11b1,,bias.m11b1,se.m11b1,,CPm11.b1),
cbind(est.geeIb0, bias1.geeIb0, bias2.geeIb0,
se.geeIb0, CP1geeI.b0, CP2geeI.b0),
cbind(est.geeIb1,bias1.geeIb1, bias2.geeIb1,
se.geeIb1, CP1geeI.b1, CP2geeI.b1),
cbind(est.geeEb0, bias1.geeEb0, bias2.geeEb0,
se.geeEb0, CP1geeE.b0, CP2geeE.b0),
cbind(est.geeEb1,bias1.geeEb1, bias2.geeEb1 ,
se.geeEb1, CP1geeE.b1, CP2geeE.b1),
cbind(est.geeAb0, bias1.geeAb0, bias2.geeAb0,
se.geeAb0, CP1geeA.b0, CP2geeA.b0),
cbind(est.geeAb1,bias1.geeAb1, bias2.geeAb1,
se.geeAb1, CP1geeA.b1, CP2geeA.b1),
cbind(est.alrb0,bias1.alrb0, bias2.alrb0,
se.alrb0, CP1alr.b0,CP2alr.b0),
cbind(est.alrb1, bias1.alrb1, bias2.alrb1,
se.alrb1, CP1alr.b1, CP2alr.b1))
write.csv(data.frame(tab1), "tab1.csv")
```

A2. R codes for MCM3, GEE and ALR in Chapter 6

```
######################## FUNCTIONS ######################################
############ Generating data Y1 , Y2, Y3 and Y4   ######################

modsimbonney<-function(N,intercept, beta1, gamma1, gamma2, gamma3, x1){
## simulate y1
id <- 1:N
xbeta <- intercept +  beta1 *x1
proba1 <- exp(xbeta)/(1 + exp(xbeta))
Y1 <- ifelse(runif(N,0,1) < proba1,1,0)
```

```
# simulate y2
xbeta2 <- intercept +  beta1 *x1 + gamma1 * Y1
proba2 <- exp(xbeta2)/(1 + exp(xbeta2))
Y2 <- ifelse(runif(N,0,1) < proba2,1,0)

# simulate y3
xbeta3 <- intercept +  beta1 *x1 + gamma1* Y1+ gamma2*Y2
proba3 <- exp(xbeta3)/(1 + exp(xbeta3))
Y3 <- ifelse(runif(N,0,1) < proba3,1,0)

# simulate y4
xbeta4 <- intercept +  beta1 *x1 + gamma1* Y1+ gamma2*Y2 + gamma3*Y3
proba4 <- exp(xbeta4)/(1 + exp(xbeta4))
Y4 <- ifelse(runif(N,0,1) < proba4,1,0)

dat <- data.frame(id,Y1,Y2,Y3,Y4,x1,x1,x1, x1)
alp<-cor(dat[2:5])
pij<-data.frame(proba1, proba2, proba3, proba4)

sdatacor<-list(alp, dat, pij)
# print(sdatacor)
return(sdatacor)
}
newsimbonney<-function(totsim,t.seed,N,intercept,beta1,
gamma1,gamma2,gamma3,x1){
for (i in 1:totsim){
cat("i=",i,"\n")
dd<- modsimbonney(N,intercept,beta1,gamma1,gamma2,gamma3, x1)
sdata<-data.frame(dd[2])
colnames(sdata)<-c("id", "Y1","Y2","Y3","Y4", "x1","x2", "x3", "x4")
#  sdata<-data.frame(sdata)
pij<-data.frame(dd[3])
cmatr<-data.frame(dd[1])# correlation matrix of Y1 Y2 and Y3
calpha12<-cmatr[1,2]  # correlation between Y1 and Y2
calpha13<-cmatr[1,3] # correlation between Y1 and Y3
calpha23<-cmatr[2,3] # correlation between Y2 and Y3

print("++++++++SSSSS+++++++++")

## Regressive Models
dat1<-sdata[,c(1,2,6)]
dat2<-sdata[,c(1,3,7)]
dat3<-sdata[,c(1,4,8)]
dat4<-sdata[,c(1,5,9)]
colnames(dat1)<-c("id","Y","x1")
colnames(dat2)<-c("id","Y","x1")
colnames(dat3)<-c("id","Y","x1")
colnames(dat4)<-c("id","Y","x1")
dat<-rbind(dat1,dat2,dat3, dat4)
dat<-arrange(dat,id)
#print(head(dat))
```

```
#colnames(dat)<-c("id", "Y", "x1")
geeI<-geeglm(Y~x1,family=binomial("logit"),id=id,
corstr="independence",std.err="san.se",data=dat)
# print(summary(geeI)); #print(head(dat))
geeE<-geeglm(Y~x1,family=binomial(link="logit"),id=id,
corstr="exchangeable",std.err="san.se",data=dat)
# print(summary(geeE))
geeA<-geeglm(Y~x1,family=binomial("logit"),id=id,
corstr="ar1",std.err="san.se",data=dat)
# print(summary(geeA))

# Add models for ALR
alrmod <- alr(dat$Y ~ dat$x1, id=dat$id, depm="exchangeable", ainit=0.2)
# print(summary(alrmod))
alralpha<-alrmod$alpha
## Fitting model with parameters beta and gamma
mbon<-glm(Y4~x1+Y1+Y2+Y3,family=binomial,data=sdata)

allres<-cbind(
matrix(mbon$coefficients,nrow=1),matrix(summary(mbon)$coeff[,4],nrow=1),
matrix(geeI$coefficients,nrow=1),matrix(summary(geeI)$coeff[,4],nrow=1),
matrix(geeE$coefficients,nrow=1),matrix(summary(geeE)$coeff[,4],nrow=1),
matrix(geeA$coefficients,nrow=1),matrix(summary(geeA)$coeff[,4],nrow=1),
matrix(alrmod$coefficients, nrow=1), alralpha,
matrix(summary(mbon)$coefficients[,2], nrow=1 ),
matrix(summary(geeI)$coefficients[,2], nrow=1 ),
matrix(summary(geeE)$coefficients[,2], nrow=1 ),
matrix(summary(geeA)$coefficients[,2], nrow=1 ),
matrix(summary(alrmod)$coefficients[,2], nrow=1),
intercept,beta1,gamma1,gamma2,gamma3, t.seed)
if(i==1){
myres<-allres
}
if(i>1){
myres<-rbind(myres,allres)
}
}  ## end of simulation

# Add column names of results matrix
colnames(myres)<-c( "mbonb0", "mbonb1","mbong1","mbong2","mbong3",
"mbonb0p0","mbonb1p1","mbong1p1","mbong2p2","mbong3p3",
"geeIb0","geeIb1","geeIb0p0","geeIb1p1",
"geeEb0","geeEb1","geeEb0p0","geeEb1p1",
"geeAb0","geeAb1","geeAb0p0","geeAb1p1",
"alrb0", "alrb1", "alralpha",
"mbonb0se", "mbonb1se","mbong1se","mbong2se","mbong3se",
"geeIb0se", "geeIb1se", "geeEb0se", "geeEb1se",
"geeAb0se", "geeAb1se", "alrb0se", "alrb1se",
"intercept1", "beta1", "gamma1","gamma2","gamma3","t.seed")
res<-list(myres)
return(res)
}   #  end function
```

```
#########################################################################
## Codes for Running the Simulation Study for MCM3, GEE and ALR ##
#########################################################################
library(MASS)
library(magic)
library(geepack)
library(dplyr)
library(alr)
t.seed     <- 3456
set.seed(t.seed)
totsim     <- 1000
N          <- 500
t          <- 4

## TRIAL 1    # non-Identical population, non-correlated data

intercept <- 0.2
beta1      <- 0.7
gamma1     <- 0.0
gamma2     <- 0.0
gamma3     <- 0.0
x0<- rep(1, N)
x1 <- rbinom(N, 1, .5)
init <-c(0.4, 25, 0.25)
bta<-c(init[1], init[2], init[3])

source("G:/phd/R_codes/model_bonney4.R")
myresultsbn<- newsimbonney(totsim,t.seed,N,
intercept,beta1,gamma1,gamma2,gamma3, x1)
mysim<-data.frame(myresultsbn)

est.mbonb0<-mean(mysim$mbonb0)
est.mbonb1<-mean(mysim$mbonb1)
est.mbong1<-mean(mysim$mbong1)
est.mbong2<-mean(mysim$mbong2)
est.mbong3<-mean(mysim$mbong3)
bias.mbonb0<-mean(intercept-mysim$mbonb0)
bias.mbonb1<-mean(beta1-mysim$mbonb1)
bias.mbong1<-mean(gamma1-mysim$mbong1)
bias.mbong2<-mean(gamma2-mysim$mbong2)
bias.mbong3<-mean(gamma3-mysim$mbong3)
lcimbonb0<-mysim$mbonb0 - 1.96*mysim$mbonb0se
ucimbonb0<-mysim$mbonb0 + 1.96*mysim$mbonb0se
lcimbonb1<-mysim$mbonb1 - 1.96*mysim$mbonb1se
ucimbonb1<-mysim$mbonb1 + 1.96*mysim$mbonb1se
lcimbong1<-mysim$mbong1 - 1.96*mysim$mbong1se
ucimbong1<-mysim$mbong1 + 1.96*mysim$mbong1se
lcimbong2<-mysim$mbong2 - 1.96*mysim$mbong2se
ucimbong2<-mysim$mbong2 + 1.96*mysim$mbong2se
lcimbong3<-mysim$mbong3 - 1.96*mysim$mbong3se
ucimbong3<-mysim$mbong3 + 1.96*mysim$mbong3se
```

```
# Computing CP for b0 and b1 of model 1, P(Y1=1|X)
CPm1b0<-rep(0, totsim)
CPm1b1<-rep(0, totsim)
CPm1g1<-rep(0, totsim)
CPm1g2<-rep(0, totsim)
CPm1g3<-rep(0, totsim)
for(i in 1: length(lcimbonb0)){
if(intercept<=ucimbonb0[i] & intercept>=lcimbonb0[i])
{CPm1b0[i]<-1}
if(beta1<ucimbonb1[i] & beta1>lcimbonb1[i])
{CPm1b1[i]<-1}
if(gamma1<ucimbong1[i] & gamma1>lcimbong1[i])
{CPm1g1[i]<-1}
if(gamma2<ucimbong2[i] & gamma2>lcimbong2[i])
{CPm1g2[i]<-1}
if(gamma3<ucimbong3[i] & gamma3>lcimbong3[i])
{CPm1g3[i]<-1}
}
CPm1.b0<-mean(CPm1b0)
CPm1.b1<-mean(CPm1b1)
CPm1.g1<-mean(CPm1g1)
CPm1.g2<-mean(CPm1g2)
CPm1.g3<-mean(CPm1g3)
se.mbonb0<-mean(mysim$mbonb0se)
se.mbonb1<-mean(mysim$mbonb1se)
se.mbong1<-mean(mysim$mbong1se)
se.mbong2<-mean(mysim$mbong2se)
se.mbong3<-mean(mysim$mbong3se)

est.geeIb0<-mean(mysim$geeIb0)
est.geeIb1<-mean(mysim$geeIb1)
bias.geeIb0<-mean(intercept-mysim$geeIb0)
bias.geeIb1<-mean(beta1-mysim$geeIb1)
lcigeeIb0<-mysim$geeIb0 - 1.96*mysim$geeIb0se
ucigeeIb0<-mysim$geeIb0 + 1.96*mysim$geeIb0se
lcigeeIb1<-mysim$geeIb1 - 1.96*mysim$geeIb1se
ucigeeIb1<-mysim$geeIb1 + 1.96*mysim$geeIb1se

## Computing CP for b0 and b1 of model geeI P(Y2=1|Y1=1,X)
CPgeeIb0<-rep(0, totsim)
CPgeeIb1<-rep(0, totsim)

for(i in 1: totsim){
if(intercept<ucigeeIb0[i] & intercept>lcigeeIb0[i])
CPgeeIb0[i]<-1
if(beta1<ucigeeIb1[i] & beta1>lcigeeIb1[i])
CPgeeIb1[i]<-1
}
CPgeeI.b0<-mean(CPgeeIb0)
CPgeeI.b1<-mean(CPgeeIb1)
se.geeIb0<-mean(mysim$geeIb0se)
```

```
se.geeIb1<-mean(mysim$geeIb1se)

est.geeEb0<-mean(mysim$geeEb0)
est.geeEb1<-mean(mysim$geeEb1)
bias.geeEb0<-mean(intercept-mysim$geeEb0)
bias.geeEb1<-mean(beta1-mysim$geeEb1)
lcigeeEb0<-mysim$geeEb0 - 1.96*mysim$geeEb0se
ucigeeEb0<-mysim$geeEb0 + 1.96*mysim$geeEb0se
lcigeeEb1<-mysim$geeEb1 - 1.96*mysim$geeEb1se
ucigeeEb1<-mysim$geeEb1 + 1.96*mysim$geeEb1se

## Computing CP for b0 and b1 of model geeE P(Y2=1|Y1=1,X)
CPgeeEb0<-rep(0, totsim)
CPgeeEb1<-rep(0, totsim)
for(i in 1: totsim){
if(intercept<ucigeeEb0[i] & intercept>=lcigeeEb0[i])
{CPgeeEb0[i]<-1}
if(beta1<ucigeeEb1[i] & beta1>lcigeeEb1[i])
{CPgeeEb1[i]<-1}
}
CPgeeE.b0<-mean(CPgeeEb0)
CPgeeE.b1<-mean(CPgeeEb1)
se.geeEb0<-mean(mysim$geeEb0se)
se.geeEb1<-mean(mysim$geeEb1se)

est.geeAb0<-mean(mysim$geeAb0)
est.geeAb1<-mean(mysim$geeAb1)
bias.geeAb0<-mean(intercept-mysim$geeAb0)
bias.geeAb1<-mean(beta1-mysim$geeAb1)
lcigeeAb0<-mysim$geeAb0 - 1.96*mysim$geeAb0se
ucigeeAb0<-mysim$geeAb0 + 1.96*mysim$geeAb0se
lcigeeAb1<-mysim$geeAb1 - 1.96*mysim$geeAb1se
ucigeeAb1<-mysim$geeAb1 + 1.96*mysim$geeAb1se

## Computing CP for b0 and b1 of model geeE P(Y2=1|Y1=1,X)
CPgeeAb0<-rep(0, totsim)
CPgeeAb1<-rep(0, totsim)
for(i in 1: totsim){
if(intercept<ucigeeAb0[i] & intercept>lcigeeAb0[i])
CPgeeAb0[i]<-1
if(beta1<ucigeeAb1[i] & beta1>lcigeeAb1[i])
CPgeeAb1[i]<-1
}
CPgeeA.b0<-mean(CPgeeAb0)
CPgeeA.b1<-mean(CPgeeAb1)
se.geeAb0<-mean(mysim$geeAb0se)
se.geeAb1<-mean(mysim$geeAb1se)

est.alrb0<-mean(mysim$alrb0)
est.alrb1<-mean(mysim$alrb1)
bias.alrb0<-mean(intercept-mysim$alrb0)
bias.alrb1<-mean(beta1-mysim$alrb1)
```

```
lcialrb0<-mysim$alrb0 - 1.96*mysim$alrb0se
ucialrb0<-mysim$alrb0 + 1.96*mysim$alrb0se
lcialrb1<-mysim$alrb1 - 1.96*mysim$alrb1se
ucialrb1<-mysim$alrb1 + 1.96*mysim$alrb1se
CPalrb0<-rep(0, totsim)
CPalrb1<-rep(0, totsim)

for(i in 1: totsim){
if(intercept<ucialrb0[i] & intercept>lcialrb0[i])
CPalrb0[i]<-1
if(beta1<ucialrb1[i] & beta1>lcialrb1[i])
CPalrb1[i]<-1
}
CPalr.b0<-mean(CPalrb0)
CPalr.b1<-mean(CPalrb1)
se.alrb0<-mean(mysim$alrb0se)
se.alrb1<-mean(mysim$alrb1se)
#######   Simulation OUtputs
bonres<- data.frame(rbind(
cbind(est.mbonb0, bias.mbonb0, se.mbonb0, CPm1.b0),
cbind(est.mbonb1, bias.mbonb1,se.mbonb1, CPm1.b1),
cbind(est.mbong1, bias.mbong1, se.mbong1, CPm1.g1),
cbind(est.mbong2, bias.mbong2, se.mbong2, CPm1.g2),
cbind(est.mbong3, bias.mbong3, se.mbong3, CPm1.g3),
cbind(est.geeIb0, bias.geeIb0, se.geeIb0,CPgeeI.b0),
cbind(est.geeIb1, bias.geeIb1, se.geeIb1, CPgeeI.b1),
cbind(est.geeEb0, bias.geeEb0, se.geeEb0,CPgeeE.b0),
cbind(est.geeEb1, bias.geeEb1, se.geeEb1, CPgeeE.b1),
cbind(est.geeAb0, bias.geeAb0, se.geeAb0,CPgeeA.b0),
cbind(est.geeAb1, bias.geeAb1, se.geeAb1, CPgeeA.b1),
cbind(est.alrb0, bias.alrb0, se.alrb0, CPalr.b0),
cbind(est.alrb1, bias.alrb1, se.alrb1, CPalr.b1)))

write.csv(bonres, 'G:\\phd\\bonres.csv')
```

A3. R Codes for MCMQL, GEE and ALR in Chapter 7

```
### Generating data##
id<-c(1:N)
gendat<-function(N, id, intercept1, intercept2, beta1, beta2, ro12){
# STEP 1: generate x
x0<- rep(1, N)
x1 <- rbinom(N, 1, .5)

## STEP 2: Calculate marginal probabilities for population values

xbeta1 <- intercept1 +  beta1 *x1
proba1 <- exp(xbeta1)/(1 + exp(xbeta1))
xbeta2 <- intercept2 +  beta2 *x1
proba2 <- exp(xbeta2)/(1 + exp(xbeta2))
pij<-cbind(proba1, proba2)
simdat<-matrix(0, nrow=N, ncol=2)
```

```
for(i in 1: N){
mm<-cbind(c(1, ro12), c(ro12, 1))# Correlation matrix
comprob<-bincorr2commonprob(c(pij[i,1], pij[i, 2]), mm)
simdat[i,]<-rmvbin(1, c(pij[i,1], pij[i,2]), comprob)
}
dat<-cbind(id, simdat, x1, x1)
# estimating rho from simulated data
dat1<-list(dat, pij)
return(dat1)
}
newsimqsi<-function(totsim,tsd, N,intercept1,intercept2,beta1,beta2,ro12)
{ for (i in 1:totsim){
cat("i=",i,"\n")
dd<- gendat(N,id, intercept1,intercept2,beta1,beta2,ro12)
sdata<-data.frame(dd[[1]])
colnames(sdata)<-c("id", "Y1","Y2","x1","x2")
Y1<-sdata[,2]
Y2<-sdata[,3]
#Estimating correlation of simulated data
freqtab<-table(Y1, Y2)
f00<-freqtab[1,1]
f11<-freqtab[2,2]
f10<-freqtab[2,1]
f01<-freqtab[1,2]
f0.<-f00+f01
f1.<-f10+f11
f.0<-f00+f10
f.1<-f01+f11
r12<-(f00*f11 - f10*f01)/(sqrt(f0.*f1.*f.0*f.1))# corr bet Y1 and Y2

sdata01<-subset(sdata,Y1==0)
sdata11<-subset(sdata,Y1==1)

#   Proposed Model for beta1 and beta2|1
mod1 <-glm(Y1~x1,family=quasibinomial,data=sdata)
mod2 <-glm(Y2~x1,family=quasibinomial,data=sdata)

# Add models for GEE
### Rearrange the data
dat1<-sdata[,c(1,2,4)]
dat2<-sdata[,c(1,3,4)]
colnames(dat1)<-c("id","Y","x1")
colnames(dat2)<-c("id","Y","x1")
dat<-rbind(dat1,dat2)
dat<-arrange(dat,id)
#print(head(dat))
colnames(dat)<-c("id", "Y", "x1")
geeI<-geeglm(Y~x1,family=binomial("logit"),id=id,
corstr="independence",std.err="san.se",data=dat)
print(summary(geeI))
#print(head(dat))
geeE<-geeglm(Y~x1,family=binomial(link="logit"),id=id,
```

```
corstr="exchangeable",std.err="san.se",data=dat)
print(summary(geeE))
geeA<-geeglm(Y~x1,family=binomial("logit"),id=id,
corstr="ar1",std.err="san.se",data=dat)
print(summary(geeA))

# Add models for ALR
alrmod <- alr(dat$Y ~ dat$x1, id=dat$id, depm="exchangeable", ainit=0.2)
print(summary(alrmod))
alralpha<-alrmod$alpha
# dpt<- dtest(trt01=mod01,trt11=mod11,trt2=mod2)
# print(dpt)
#  Add results of GEE and ALR in allres
# "allres" is the collection of all results
allres<-data.frame(cbind(
matrix(mod1$coefficients,nrow=1),matrix(summary(mod1)$coeff[,4],nrow=1),
matrix(mod2$coefficients,nrow=1),matrix(summary(mod2)$coeff[,4],nrow=1),
matrix(geeI$coefficients,nrow=1),matrix(summary(geeI)$coeff[,4],nrow=1),
matrix(geeE$coefficients,nrow=1),matrix(summary(geeE)$coeff[,4],nrow=1),
matrix(geeA$coefficients,nrow=1),matrix(summary(geeA)$coeff[,4],nrow=1),
matrix(alrmod$coefficients, nrow=1), alralpha,
#           self0, self1, selfalpha,
#      gra.est, sigma.gra,
matrix(summary(mod1)$coefficients[,2], nrow=1 )
, matrix(summary(mod2)$coefficients[,2], nrow=1 )
, matrix(summary(geeI)$coefficients[,2], nrow=1 ),
matrix(summary(geeE)$coefficients[,2], nrow=1 ),
matrix(summary(geeA)$coefficients[,2], nrow=1 ),
matrix(summary(alrmod)$coefficients[,2], nrow=1),
#          matrix(summary(bi.est)$par[,2], nrow=1),
intercept1,intercept2,beta1,beta2, r12,
#      cbind(dpt[1,c(2,4)],dpt[2,c(2,4)],dpt[3,c(2,4)],dpt[1,3]),
t.seed
))
if(i==1){
myres<-allres
}
if(i>1){
myres<-rbind(myres,allres)
}
}  ## end of simulation

# Add results of GEE and ALR in colnames
colnames(myres)<-c(
"m1b0","m1b1","m1p0","m1p1","m2b0","m2b1", "m2p0","m2p1",
"geeIb0", "geeIb1","geeIb0p0","geeIb1p1",
"geeEb0", "geeEb1","geeEb0p0","geeEb1p1",
"geeAb0", "geeAb1","geeAb0p0","geeAb1p1",
"alrb0", "alrb1", "alralpha",
"m1b0se", "m1b1se", "m2b0se", "m2b1se",
"geeIb0se", "geeIb1se","geeEb0se", "geeEb1se",
"geeAb0se", "geeAb1se","alrb0se", "alrb1se",
```

```
"intercept1","intercept2","beta1","beta2", "rho",
#  "b01b11ch","pv","b01b2ch","pv","b11b2ch","pv", "df",
"t.seed")
res<-list(myres)
return(res)
}   #  end function
```

Bibliography

[1] Altham, P. M. E. (1978). Two generalizations of the binomial distribution. *Applied Statistics*, **27**(2):162–167.

[2] Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about markov chains. *The Annals of Mathematical Statistics*, **28**(1):89–110.

[3] Azzalini, A. (1994). Logistic regression for autocorrelated data with application to repeated measures. *Biometrika*, **81**(4):767–775.

[4] Bahadur, R. R. (1961). *A representation of the joint distribution of responses to n dichotomous items.* In Studies in item analysis and prediction, H. Solomon (ed), 158-168. Stanford Mathematical Studies in the Social Sciences VI. Stanford University Press, California.

[5] Billingsley, P. (1961). Statistical methods in markov chains. *The Annals of Mathematical Statistics*, **32**(1):12–40.

[6] Blazer, D. G. (2003). Depression in late life: review and commentary. *The Journals of Gerontology: Series A*, **58**(3):249–265.

[7] Bonney, G. E. (1986). Regressive logistic models for familial disease and other binary traits. *Biometrics*, **42**(3):611–625.

[8] Bonney, G. E. (1987). Logistic regression for dependent binary observations. *Biometrics*, **43**(4):951–973.

[9] Carey, V., Zeger, S. L., and Diggle, P. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, **80**(3):517–526.

[10] Chen, J. and Lazar, N. A. (2012). Selection of working correlation structure in generalized estimating equations via empirical likelihood. *Journal of Computational and Graphical Statistics*, **21**(1):18–41.

[11] Cho, J. S. and White, H. (2007). Testing for regime switching. *Econometrica*, **75**(6):1671–1720.

[12] Costa, A. G., Colosimo, E. A., Vaz, A. B., Silva, J. L. P., and Amorim, L. D. (2017). Marginal models for the association structure of hierarchical binary responses. *Journal of Applied Statistics*, **44**(10):1827–1838.

[13] Cox, D. R. (1958). Two further applications of a model for binary regression. *Biometrika*, **45**(3-4):562–565.

[14] Cox, D. R. (1970). *Analysis of binary data*. Methuen.

[15] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, **34**(2):187–202.

[16] Darlington, G. and Farewell, V. (1992). Binary longitudinal data analysis with correlation a function of explanatory variables. *Biometrical Journal*, **34**(8):899–910.

[17] Diggle, P. (1992). Discussion of paper by K.Y. Liang, S.L. Zeger and B. Qaqish. *Journal of Royal Statistical Society B*, **54**(1):28–29.

[18] Diggle, P., Heagerty, P., Liang, K. Y., and Zeger, S. (2002). *Analysis of longitudinal data*. Oxford University Press, New York.

[19] Diggle, P., Liang, K. Y., and Zeger, S. L. (1994). *Longitudinal data analysis*. Oxford University Press, New York.

[20] Evans, M. and Mottram, P. (2000). Diagnosis of depression in elderly patients. *Advances in Psychiatric Treatment*, **6**(1):49–56.

[21] Firth, D. (1992). Discussion on multivariate regression analysis for categorical data(by k.-y. liang, sl zeger and b. qaqish). *Journal of Royal Statistical Society B*, **54**(1):24–26.

[22] Fitzmaurice, G., Molenberghs, G., Davidian, M., and Verbeke, G. (2008). Advances in longitudinal data analysis. In *Longitudinal data analysis*, pages 13–38. Chapman and Hall/CRC, New York.

[23] Fitzmaurice, G. M. and Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, **80**(1):141–151.

[24] Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012). *Applied longitudinal analysis*. John Wiley & Sons, New Jersey.

[25] Fu, L., Hao, Y., and Wang, Y. G. (2018). Working correlation structure selection in generalized estimating equations. *Computational Statistics*, **33**(2):983–996.

[26] Gill, T. M., Allore, H. G., and Han, L. (2012). Bathing disability and the risk of long-term admission to a nursing home. *Journal of Gerontology, Series A*, **61**(8):821–825.

[27] Glonek, G. F. V. and McCullagh, P. (1995). Multivariate logistic models. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**(3):533–546.

[28] Grizzle, J. E., Starmer, C. F., and Koch, G. G. (1969). Analysis of categorical data by linear models. *Biometrics*, **25**(3):489–504.

[29] Guerra, M. W., Shults, J., Amsterdam, J., and Ten Have, T. (2012). The analysis of binary longitudinal data with time-dependent covariates. *Statistics in medicine*, **31**(10):931–948.

[30] Gumbel, E. J. (1961). Bivariate logistic distributions. *Journal of the American Statistical Association*, **56**(294):335–349.

[31] Hardin, J. W. and Hilbe, J. M. (2002). *Generalized estimating equations*. Chapman and Hall/CRC, New York.

[32] Heyde, C. C. (2008). *Quasi-likelihood and its application: a general approach to optimal parameter estimation*. Springer-Verlag, New York.

[33] Imori, S. (2015). Consistent selection of working correlation structure in GEE analysis based on steins loss function. *Hiroshima Mathematical Journal*, **45**(1):91–107.

[34] Islam, M. A., Alzaid, A. A., Chowdhury, R. I., and Sultan, K. S. (2013a). A generalized bivariate bernoulli model with covariate dependence. *Journal of Applied Statistics*, **40**(5):1064–1075.

[35] Islam, M. A. and Chowdhury, R. I. (2006). A higher order markov model for analyzing covariate dependence. *Applied Mathematical Modelling*, **30**(6):477–488.

[36] Islam, M. A. and Chowdhury, R. I. (2007). *First and higher order transition models with covariate dependence.* In Progress in Applied Mathematical Modeling, 153-198. Nova Science Publisher Inc., New York.

[37] Islam, M. A. and Chowdhury, R. I. (2010). Prediction of disease status: A regressive model approach for repeated measures. *Statistical Methodology*, **7**(5):520–540.

[38] Islam, M. A. and Chowdhury, R. I. (2017). *Analysis of Repeated Measures Data*. Springer, Singapore.

[39] Islam, M. A., Chowdhury, R. I., and Alzaid, A. A. (2012a). Tests for dependence in binary repeated measures data. *Journal of Statistical Research*, **46**(2):203–217.

[40] Islam, M. A., Chowdhury, R. I., and Briollais, L. (2012b). A bivariate binary model for testing dependence in outcomes. *Bulletin of Malaysian Mathematical Sciences Society*, **35**(4):845–858.

[41] Islam, M. A., Chowdhury, R. I., and Huda, S. (2013b). A multistate transition model for analyzing longitudinal depression data. *Bulletin of the Malaysian Mathematical Sciences Society*, **36**(3):637–655.

[42] Islam, M. A., Chowdhuryb, R. I., Bae, S., and Singh, K. P. (2014). Assessing the association in the repeated measures of depression. *Advances in Applications in Statistics*, **42**(2):83–93.

[43] Islam, M. A., Sultan, K. S., and Chowdhury, R. I. (2009). Estimation and tests for a longitudinal regression model based on the markov chain. *Statistical Methodology*, **6**(5):478–489.

[44] Kalbfleisch, J. D. and Lawless, J. F. (1985). The analysis of panel data under a markov assumption. *Journal of the American Statistical Association*, **80**(392):863–871.

[45] Karakus, M. C. and Patton, L. C. (2011). Depression and the onset of chronic illness in older adults: A 12-year prospective study. *The Journal of Behavioral Health Services & Research*, **38**(3):373–382.

[46] Katz, S., Ford, A., Moskowitz, R., Jackson, B., and Jaffe, M. (1963). Studies of illness in the aged. the index of adl: a standardized measure of biological and psychological function. *Journal of American Medical Association*, **185**(12):914–9.

[47] Koch, G. G., Landis, J. R., Freeman, J. L., Freeman Jr, D. H., and Lehnen, R. G. (1977). A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics*, **33**(1):133–158.

[48] Korn, E. L. and Whittemore, A. S. (1979). Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics*, **35**(4):795–802.

[49] Kupper, L. L. and Haseman, J. K. (1978). The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics*, **34**(1):69–76.

[50] Le Cessie, S. and Van Houwelingen, J. (1994). Logistic regression for correlated binary data. *Applied Statistics*, **43**(1):95–108.

[51] Lee, Y. and Nelder, J. A. (2004). Conditional and marginal models: Another view. *Statistical Science*, **19**(2):219–228.

[52] Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**(1):13–22.

[53] Liang, K. Y., Zeger, S. L., and Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, **54**(1):3–40.

[54] Lindsey, J. K. and Lambert, P. (1998). On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in medicine*, 17(4):447–469.

[55] Lipsitz, S. R., Laird, N. M., and Harrington, D. P. (1990). Maximum likelihood regression methods for paired binary data. *Statistics in Medicine*, **9**(12):1517–1525.

[56] Lipsitz, S. R., Laird, N. M., and Harrington, D. P. (1991). Generalized estimating equations for correlated binary data using the odds ratio as a measure of association. *Biometrika*, **78**(1):153–160.

[57] McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*. CRC press, New York.

[58] McDonald, B. W. (1993). Estimating logistic regression parameters for bivariate binary data. *Journal of the Royal Statistical Society. Series B*, **55**(2):391–397.

[59] Muenz, L. R. and Rubinstein, L. V. (1985). Markov models for covariate dependence of binary sequences. *Biometrics*, **41**(1):91–101.

[60] Muff, S., Held, L., and Keller, L. F. (2016). Marginal or conditional regression models for correlated non-normal data? *Methods in Ecology and Evolution*, **7**(12):1514–1524.

[61] Nelder, J. and Lee, Y. (1992). Likelihood, quasi-likelihood and pseudolikelihood: some comparisons. *Journal of the Royal Statistical Society. Series B (Methodological)*, **54**(1):273–284.

[62] Nikoloulopoulos, A. K. (2016). Correlation structure and variable selection in generalized estimating equations via composite likelihood information criteria. *Statistics in medicine*, **35**(14):2377–2390.

[63] Noelker, L. and Browdie, R. (2013). Sidney katz, md: A new paradigm for chronic illness and long-term care. *The Gerontologist*, **54**(1):13–20.

[64] Pardo, M. C. and Alonso, R. (2017). Working correlation structure selection in GEE analysis. *Statistical Papers*, **60**(5):1447–1467.

[65] Pitt, M. K., Chatfield, C., and Walker, S. G. (2002). Constructing first order stationary autoregressive models via latent processes. *Scandinavian Journal of Statistics*, **29**(4):657–663.

[66] Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **44**(4):1033–1048.

[67] Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, **44**(1):50-57.

[68] Shults, J., Sun, W., Tu, X., Kim, H., Amsterdam, J., Hilbe, J. M., and Ten Have, T. (2009). A comparison of several approaches for choosing between working correlation structures in generalized estimating equation analysis of longitudinal binary data. *Statistics in medicine*, **28**(18):2338–2355.

[69] Steffick, D. (2000). Documentation of affective functioning measures in the Health and Retirement Study. An online report accessed on January 18, 2018 from http://hrsonline.isr.umich.edu/sitedocs/userg/dr-005.pdf.

[70] Stineman, M., Xie, D., Pan, Q., Kurichi, J., Saliba, D., and Streim, J. (2011). Activity of daily living staging, chronic health conditions, and perceived lack of home accessibility features for elderly people living in the community. *Journal of American Geriatric Society*, **59**(3):454–62.

[71] Tuma, N. B., Hannan, M. T., and Groeneveld, L. P. (1979). Dynamic analysis of event histories. *American Journal of Sociology*, **84**(4):820–854.

[72] University of Michigan (2014a). Health and Retirement Study. Accessed on January 20, 2018, from https://www.rand.org/well-being/social-and-behavioral-policy/centers/aging/dataprod/hrs-data.html.

[73] University of Michigan (2014b). Health and Retirement Study Data. Accessed on January 20, 2018, from http://hrsonline.isr.umich.edu/data /index.html.

[74] Wang, M. (2014). Generalized estimating equations in longitudinal data analysis: a review and recent developments. In *Advances in Statistics* Accessed on January 20, 2018, from http://dx.doi.org/10.1155/2014/303728.

[75] Wang, Y. G. and Carey, V. (2003). Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. *Biometrika*, **90**(1):29–41.

[76] Wang, Y. G. and Fu, L. (2017). Selection of working correlation structure in generalized estimating equations. *Statistics in medicine*, **36**(14):2206–2219.

[77] Wedderburn, R. W. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika*, **61**(3):439–447.

[78] Williams, B. (2014). *Current Diagnosis and Treatment: Geriatrics, Second Edition*. McGraw-Hill, New York.

[79] Zeger, S. L. and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**(1):121–130.

[80] Zeger, S. L., Liang, K. Y., and Self, S. G. (1985). The analysis of binary longitudinal data with time-independent covariates. *Biometrika*, **72**(1):31–38.

[81] Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series a quasi-likelihood approach. *Biometrics*, **44**(4):1019–1031.