5-1-2015

Joint Models of Longitudinal Outcomes and Informative Time

JangDong Seo
University of Northern Colorado

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

JOINT MODELS OF LONGITUDINAL OUTCOMES AND INFORMATIVE TIME

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

JangDong Seo

College of Education and Behavioral Sciences
School of Educational Research Leadership and Technology
Department of Applied Statistics and Research Methods

May 2015

This Dissertation by: JangDong Seo

Entitled: *Joint Models of Longitudinal Outcomes and Informative Time*

has been approved as meeting the requirement for the Degree of Doctor of Philosophy in the College of Education and Behavioral Sciences in the School of Educational Research, Leadership and Technology, Program of Applied Statistics and Research Methods

Accepted by the Doctoral Committee

_____

Khalil Shafie Holighi, Ph.D., Research Advisor

_____

Jay R. Schaffer, Ph.D., Committee Member

_____

Trent L. Lalonde, Ph.D., Committee Member

_____

Heng-Yu Ku, Ph.D., Faculty Representative

Date of Dissertation Defense _____

Accepted by the Graduate School

_____

Linda L. Black, Ed.D., LPC
Dean of the Graduate School and International Admissions

ABSTRACT

Seo, JangDong. *Joint models of longitudinal outcomes and informative time.* Published Doctor of Philosophy Dissertation, University of Northern Colorado, 2015

In longitudinal data analyses, it is commonly assumed that time intervals for collecting outcomes are predetermined – the same across all subjects – and have no information regarding the measured variables. However, in practice researchers might occasionally have irregular time intervals and informative time, which violate the above assumptions. Hence, if traditional statistical methods are used for this situation, the results would be biased.

In this study, as a solution, joint models of longitudinal outcomes and informative time are presented by using joint probability distributions, incorporating the relationships between outcomes and time. The joint models are designed to handle outcome distributions from a member of the exponential family of distributions with informative time following an exponential distribution. For instance, the Poisson probability density function is combined with the exponential distribution for count data, as well as the relations between outcomes and time; the Bernoulli probability density function is combined for binary data; and the Gamma probability density function is combined when the outcome is waiting time or survival time. The maximum likelihood parameter estimates of the joint model are found by using a nonlinear optimization method, and the asymptotic behaviors of the estimators are studied. Moreover, the likelihood ratio test statistic is computed for comparing nested models, and the model selection criteria, such as AIC, AICc, BIC, are found as well.

Through simulation studies, the maximum likelihood parameter estimates of the joint models appeared to be multivariate normal as the number of observations increased. As a result, the likelihood ratio test statistic could be utilized for model comparisons since the asymptotic normality of the maximum likelihood estimators has been verified. Also, AIC, AICc, and BIC scores were calculated as model selection criteria. Furthermore, the computing package using R was developed to handle the joint models and used to analyze the bladder cancer data for demonstration purposes.

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my advisor, Dr. Khalil Shafie, for his excellent guidance, caring, patience, and support, while writing this dissertation. Without his advice on mathematical procedures and developing programming, I would not have succeeded. I would also like to thank my dissertation committee members, Dr. Jay Schaffer, Dr. Trent Lalonde, and Dr. Heng-Yu Ku, for their helpful criticism and provision of the development of this dissertation.

I would like to thank my family, my wife, EunGyeong, and my lovely daughters, GyeongMin and Jean, for all the support they have provided throughout my studies, including this dissertation. My wife was always there cheering me up and stood by me through the good and bad times. Without their patience and support, I would have found it very difficult to overcome the ups and downs during the study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Dedicated To

My Wife,
Lovely Daughters,
and Parents

CHAPTER I

INTRODUCTION

The term "repeated measurements" is used to define outcomes measured at multiple time points on the same subjects or experimental units. When outcomes are collected for a relatively long period of time on the same subjects to evaluate the changes over time, the term "longitudinal data" is a common name instead of repeated measurements. In general, repeated measurement data are treated as a part of longitudinal data (Davis, 2002). Longitudinal study design is believed to be more powerful than traditional cross-sectional design in terms of the ability to capture the within-subject effect, by excluding the between-subject variability and investigating any trend or pattern on the changes for subjects. In addition, due to the exclusion of between-subject variability, the estimate of within-subject effects can be calculated more accurately (Fitzmaurice, Laird, & Ware, 2004; Hedeker & Gibbons, 2006).

Despite the advantages of the longitudinal design over the cross-sectional design, there are some barriers to analyzing the longitudinal data due to its complexity of data structure. Firstly, the measurements are assumed to be correlated, in general, with each other since the outcomes are collected repeatedly at multiple time points from the same subjects. The traditional approaches, such as simple method, the Analysis of Variance (ANOVA), or Multivariate Analysis of

Variance (MANOVA), have an assumption that outcomes are independent of each other, ignoring the correlation among outcomes, which cannot be true in longitudinal designs. Secondly, due to the long duration of a study, the researcher has difficulties controlling the subjects or the whole experiment. Therefore, very often the final data may turn out to be incomplete or unbalanced, or has missing data due to attrition, which is also adding to difficulties in analyzing the data, even though the researcher did not intend it at the beginning of the study.

Many methods have been developed and proposed attempting to handle the problems, but most of the methods are limited to cases with complete data (Hedeker & Gibbons, 2006), or normally distributed outcomes with balanced complete data (Davis, 2002). That is why nowadays the mixed-effects model or the generalized estimating equation (GEE) model is getting popular, but many researchers still also use the traditional methods due to the simplicity of computation, in spite of their drawbacks. Therefore, selection of a method for one's longitudinal study is totally up to the researcher's subjective judgement based on the purpose of the research.

General Notations for Longitudinal Data

Notations to represent longitudinal data used in this study may slightly vary based on the methods introduced in each chapter, but the general notations are similar throughout this study. The number of subjects, or experimental units, is denoted by m and the subject is denoted by i, which are expressed as

$$i = 1, \ldots, m.$$

Time point for the ith subject is denoted by j and the number of outcomes of the ith subject, which are written as

$$j = i, \ldots, n_i.$$

If a dataset is balanced and complete, all subjects will have the same number of outcomes at the same time points, then j can be simplified by $j = 1, \ldots, n$. The advantage of using n_i is that it allows each subject to have a different number of outcomes. The outcomes at the jth time point for the ith subject with n_i outcomes can be expressed by

$$\boldsymbol{y}_{ij} = (y_{i1}, \ldots, y_{in_i})'.$$

A general layout of longitudinal data is described in Table 1 below.

Table 1

General Layout of Longitudinal Data

Subject	Occasions					
1	y_{11}	y_{12}	$\ldots$	y_{1j}	$\ldots$	y_{1n_1}
2	y_{21}	y_{22}	$\ldots$	y_{2j}	$\ldots$	y_{2n_2}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
i	y_{i1}	y_{i2}	$\ldots$	y_{ij}	$\ldots$	y_{in_i}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
m	y_{m1}	y_{m2}	$\ldots$	y_{mj}	$\ldots$	y_{mn_m}

A balanced or complete data structure is a special case of longitudinal data since each subject can have a different number of measurements.

Longitudinal Data Analysis

Many approaches have been developed in order to handle longitudinal data. However, none of those can take care of all potential problems that may arise in the analysis for longitudinal data, so researchers have to make a decision about which approach to use based on their research questions, purpose of study, and data collection method.

For example, a simple method was introduced by Student (1908). In his study there were two outcome variables, such as pretest and posttest, and then a new variable called the summary variable was calculated by taking the difference between the pretest and posttest. Then, the summary variable was tested to determine if the average of the summary variable was equal to zero. This method will be enough when there are only two outcome variables at two different time points. The simple method has been applied by many researchers due to its simplicity of calculation.

Currently, one of the most commonly used methods might be the Analysis of Variance (ANOVA) or Multivariate Analysis of Variance (MANOVA) method. These methods are more advanced techniques than the simple method and easy to apply to longitudinal data. The two methods, however, have unrealistic assumptions, such as equal variance structure and equal predetermined time intervals to all subjects; moreover, data should be complete and balanced.

Other methods called mixed-effects model and generalized estimating equation (GEE) model are also used with highly math-oriented or computer-based

techniques, attempting to handle the special characteristics of longitudinal data: correlation among outcomes, unequal time points for each subject, and unbalanced data structure. Although these methods were developed to obtain more accurate estimations of changes over time, they are known to be computationally very complicated and need more computer programming. Despite computational difficulties, these approaches are being widely used, especially in the fields of biology, pharmaceuticals or physics due to the accuracy and efficiency of the estimation. Moreover, the GEE approach was developed to handle binary and count outcomes as well as continuous outcomes; therefore, the GEE model is very useful when the outcome is binary or count. The common drawback of the GEE and mixed-effects models is that time is still fixed or predetermined by the design. So, when time points are irregular and data are unbalanced, estimators may be biased (Lin, Scharfstein, & Rosenheck, 2004). Researchers are still putting in a lot of effort to find better methods. For instance, when time is informative, which means that the next time points for collecting measurements are adaptively determined based on current outcomes for each subject, the methods mentioned above are not appropriate.

Purpose of the Study

Most of the currently existing methods have now been introduced, and some disadvantages have been described as well as advantages. One common aspect to all the methods above is that time is fixed or predetermined before the study. In general, longitudinal study is conducted for a relatively long period of time to find

if there is any specific trend over time. If time is not fixed or cannot be determined by a researcher – since each subject's next time point is decided based on current outcome, causing irregular time intervals for individuals – estimations by using the methods above may be biased. For instance, individuals may follow different measurement schedules based on their prior health outcomes since patients who have poorer health outcomes will be asked to visit for checkups more often; therefore, patients will not share the common time occasions; instead each patient will follow their own schedule for visiting the clinic depending on their own prior health outcomes (Lipsitz, Fitzmarice, Ibrahim, Gelber, & Lipshultz, 2002). Since the main purpose of this study was to develop a model to handle the situation described above, namely informative schedule data, all the methods mentioned so far do not work for this study.

Fortunately, some researchers have introduced a new approach, named joint model, which is combining longitudinal data and the time-related factor. There are multiple articles under the name of the joint model (Henderson, Diggle, & Dobson, 2000; Kim, Zeng, Chambless, & Li, 2012; Liang, Lu, & Ying, 2009; Lin et al., 2004; Lipsitz et al., 2002; Qiu, Stein, & Elston, 2013). Some of the articles are about the joint model of longitudinal data and survival outcome, and some are about the joint model of longitudinal data and informative time. Fundamentally, the joint model is based on the joint distribution of outcomes and the time related factor with maximum likelihood estimation. As introduced, the joint model can be applied to any kind of situation. Estimation by using the joint model produces more precise results (Qiu et al., 2013). Nonetheless, one fact in common is that

even though they are dealing with the joint models, the situations they are studying are different from each other. For that reason, if any researcher who wants to adapt his/her joint models should be careful of assumptions since all the joint models were built for different situations.

For example, Bronsert (2009) presented a joint model, named Gaussian-Exponential model, in which normally distributed longitudinal responses and intermittent times following an exponential distribution are combined, and he showed the joint model has a very good ability for longitudinal data analysis compared to the mixed-effects model in his simulation study. Bronsert calculated the parameter estimates of the joint model, omitting a procedure checking the properties of the estimators, such as multivariate normality. Later, Lin (2011) extended Bronsert's (2009) study and showed that the parameter estimates maintain the property of multivariate normality, and the joint model can be an alternative method for analyzing longitudinal data. Also, Lin studied on how to calculate the scores for model selection criteria, such as the Akaike information criterion (AIC), the Akaike information criterion with correction (AICc), the Bayesian information criterion (BIC), and likelihood ratio test. A limitation of these studies is that the model incorporates informative time and normally distributed longitudinal responses only.

Research interest of the current study was developing joint models of longitudinal outcomes following the exponential family of distributions and informative time. Thus, this study mainly focused on verifying (1) if the joint model developed by Bronsert and tested by Lin can be extended and (2) if the

parameter estimates of the extended joint models maintain the asymptotic multivariate normality, when the outcome distribution is from a member of the exponential family of distributions.

Definition of Terminology

Terminology used throughout this study are described below.

Longitudinal Data is a set of outcomes measured at multiple time points on the same subjects over a given time duration. In general, time points are predetermined by researchers before outcomes are collected.

Informative Time is used to describe the fact that the next time point for collecting a response is determined by the current outcome. Thus, all subjects may not share the common set of time intervals.

Informative Schedule Data is a set of measurements collected repeatedly from the same subjects for a given time duration. The outcomes are measured based on the informative time for each subject determined by the previous outcomes.

Outcome Process is a set of measurements collected repeatedly at multiple time points for each subject.

Counting Process is a stochastic process with values that are integer, increasing, and positive in which the values represent arbitrary time points for collecting responses for each individual.

Research Questions

Q1 Can the joint model (Gaussian-Exponential model) developed by Bronsert (2009) and tested by Lin (2011) be extended to longitudinal outcomes following the exponential family of distributions?

Q2 Do the maximum likelihood estimators of the joint models obtain asymptotic normality?

Q3 Can the likelihood ratio test be conducted to compare the fit of two models?

Q4 Can the model selection criteria, such as AIC, AICc, or BIC, be developed to compare models?

If the answer to Research Question One says that the joint models are not appropriate for longitudinal outcomes with the exponential family of distributions and informative schedule data other than normally distributed outcomes, then the rest of the research questions can not be answered.

Limitations

The following limitations must be considered before any researcher is willing to take advantage of this study.

1. This study is limited to outcomes with the exponential family of distributions with a single response variable; therefore, the models should not be applied to any study with multivariate responses
2. In this study, time is assumed to be exponentially distributed. If any other distribution other than the exponential distribution is believed to be appropriate, applying the joint models should not be considered with caution.

Conclusion

One of the greatest advantages of longitudinal data analyses is that it can detect changes over time. Most of the analyzing methods introduced above are treating time as a fixed factor that is predetermined by researchers at the beginning of the study. In some cases, time intervals, however, are decided based on the previous outcomes. In that situation, approaches mentioned before cannot be used and will generate biased estimators, if used. Therefore, as a new model, the joint model was presented by Bronsert (2009), combining normally distributed longitudinal responses and informative time. Bronsert used maximum likelihood estimation to compute the parameter estimates. Then, the joint model was extended by Lin (2011). He verified that the maximum likelihood estimators of the joint model obtain multivariate normality. Also, Lin proposed the likelihood ratio test statistic, AIC, AICc, and BIC as model selection criteria. The assumptions in their studies are that longitudinal data are normally distributed and time is exponentially distributed.

Hence, questions arose: what if the outcomes are not assumed to follow a normal distribution; what if the outcomes have any other types, for instance, binary or count outcomes? Also, if the joint model is extended to any other types of outcomes, do the parameter estimates of the extended joint models still keep asymptotic normaltiy? Thus, this study focused on extending their joint model to the exponential family of distributions and checking the multivariate normality assumption of the estimators of the extended joint models.

CHAPTER II

LITERATURE REVIEW

Multiple methods have been developed due to its complexity of analysis and used to analyze longitudinal data. In this chapter, commonly used methods and previously studied joint models combining informative time and outcomes are discussed.

Simple Method Approach

The purpose of longitudinal studies is to detect changes between outcomes measured at different time points on the same subjects. One of the elementary methods was introduced by Student (1908), including outcomes measured at two time points for each subject. In the analysis, the differences between outcomes measured at two different time points on the same subjects were used to determine the change over time. The idea of the simple method is to reduce the multivariate measurements into a single measurement ignoring the correlation between repeated measurements.

A well-known example of this type of method is the dependent two sample t-test. This method can be simply applied when there are two outcomes, but generally in longitudinal studies, multiple observations at multiple time points are collected. However, due to the simplicity of the application, this approach has been commonly used with different names, such as the summary-statistic approach

(Frison & Pocock, 1992; Dawson, 1994), the response feature analysis (Crowder & Hand, 1990), or the derived variable (Diggle, Heagerty, Liang, & Zeger, 2002).

For example, the summary statistic could be the coefficient of the regression line for each subject. The summary statistic is used to test whether or not the average of the summary measurements differs from zero. However, the results of the analysis will be misleading if a wrong summary measurement is selected that does not represent an individual trend (Davis, 2002). That means the success of the analysis depends on the selection of summary measurements. Moreover, when time intervals are not equal across all subjects, the summary variable does not satisfy the homoscedasticity assumption (Diggle et al., 2002). Therefore, this summary-statistic approach is not appropriate to be used in this study with longitudinal informative schedule data.

Repeated Measures Analysis with ANOVA

When outcomes are from a normal distribution satisfying the assumptions of independence and homogeneity of variance, traditional univariate ANOVA can be simply applied. Additionally, for repeated measures ANOVA, the assumption of sphericity must be met also. Sphericity is sometimes called compound symmetry, also known as homogeneity of covariance. Sphericity is that the variance of difference between any two levels of within-subjects factor is the same to any pairwise combination. In general, it is very difficult that longitudinal data meet the assumptions of independence and homogeneity of variance. In order to test sphericity assumption, Mauchly's test is generally used, but this test is not powerful

for small sample sizes and sensitive to non-normality; therefore, Mauchly's test is not practical (Davis, 2002). Another approach using adjusted degrees of freedom has been introduced by Greenhouse and Geisser, and Huynh and Feldt when the sphericity assumption is violated, but these are too conservative, which increases Type II error (Davis, 2002).

In repeated measures ANOVA model, occasions of measurements are treated as within-subject effects, so the model becomes

$$\boldsymbol{Y}_{ij} = \boldsymbol{X}_{ij}^{'} + \boldsymbol{b}_i + \boldsymbol{e}_{ij}, \tag{1}$$

where $\boldsymbol{X}_{ij}^{'}$ is the design matrix, $\boldsymbol{b}_i$ is the random subject-specific effect, and $\boldsymbol{e}_{ij}$ is the individual-specific measurement error (Fitzmaurice et al., 2004). In this model, there are two sources of variation; one is within-subject variation, σ_b^2, and the other is between-subject variation, σ_e^2. Based on the information for variations above, the covariance structure of the model, known as compound symmetry, becomes

$$Cov(\boldsymbol{Y}_i) = \begin{pmatrix} \sigma_b^2 + \sigma_e^2 & \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma_e^2 & \sigma_b^2 & \cdots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 + \sigma_e^2 & \cdots & \sigma_b^2 \\ \vdots & \vdots & \vdots & \cdots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 + \sigma_e^2 \end{pmatrix}. \tag{2}$$

The covariance structure shows that the variance for each occasion is equal to $\sigma_b^2 + \sigma_e^2$. Therefore, the correlation between all pairs of measures is

$$Corr(\boldsymbol{Y}_{ij}, \boldsymbol{Y}_{ik}) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}, \qquad j \neq k. \tag{3}$$

Commonly, correlation between measurements tends to be stronger when they are closer in time, and the correlation decreases when they are further away from each other in time; therefore, constant correlation between two measurements is not realistic. In addition, it is assumed that all subjects have an equal number of measurements at fixed time points since the repeated measures ANOVA approach was designed to handle experimental studies (Fitzmaurice et al., 2004). Due to these unrealistic assumptions, the repeated measures ANOVA model is inappropriate for this study on informative time and non-normal outcomes.

Repeated Measures Analysis with MANOVA

When the assumptions are violated for repeated measures ANOVA, MANOVA based on Hotellings T^2 is an alternative way to handle repeated measures since the sphericity assumption is not necessary. Also, it can handle multiple response variables simultaneously, but MANOVA reduces the power of analysis (Vincent, 2005). A general idea of the MANOVA approach is followed below.

For instance, we have a vector of responses from the ith subject at time j

$$\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})',$$

where $i = i, \ldots, m$ and $j = 1, \ldots, n_i$, and $\boldsymbol{y}_i$ is from $N_t(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. To test the hypothesis of $H_0 : \mu_1 = \mu_2 = \cdots = \mu_t$, $y^*_{ij} = y_{ij} - y_{i,j+1}$ for $j = 1, \ldots, t-1$ is obtained. The $\boldsymbol{y}^*_i = (y^*_{i1}, \ldots, y_{i,t-1})'$ vectors are random samples from $N_{t-1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, where

$$\boldsymbol{\mu}^* = (\mu_1 - \mu_2, \mu_2 - \mu_3, \ldots, \mu_{t-1} - \mu_t)'. \tag{4}$$

By using Hottelling's T^2,

$$T^2 = n\bar{\boldsymbol{y}}^{*'}\boldsymbol{S}^{*-1}\bar{\boldsymbol{y}}^* \sim T^2_{t-1,n-1,n\boldsymbol{\mu}^{*'}\boldsymbol{\Sigma}^{*-1}\boldsymbol{\mu}^*}. \tag{5}$$

The test statistic F,

$$F = \frac{n-t+1}{(n-1)(t-1)}T^2, \tag{6}$$

has a $F_{t-1,n-t+1}$ distribution if $H^*_0 : \boldsymbol{\mu}^* = (0, \ldots, 0)'$ holds (Davis, 2002).

One of the advantages of using MANOVA approach, instead of repeated measures ANOVA, is that it can handle repeated measurements at multiple time points since MANOVA was originally designed to deal with multiple response variables (Fitzmaurice et al., 2004). The other advantage is that the assumption of sphericity is not needed while it still assumes the multivariate normality (Davis, 2002; Hedeker & Gibbons, 2006). However, MANOVA method itself has multiple shortcomings. One drawback is that the covariance matrix with $t \times t$ must be estimated. When t is large, numerous degrees of freedom to estimate covariance parameters, $t(t-1)/2$, will be lost. Therefore, it causes power to decrease, especially when F has small degrees of freedom for the denominator (Davis, 2002). Secondly, like ANOVA, MANOVA approach also can be used only if time points are

fixed across all subjects (Hedeker & Gibbons, 2006). However, in longitudinal studies with informative time, each subject may have different sets of time points. Thirdly, MANOVA cannot handle missing data like ANOVA. Due to these restrictions, subjects with missing data will be excluded in the analysis, which causes loss of a large amount of information in the data (Fitzmaurice et al., 2004; Hedeker & Gibbons, 2006). In other words, MANOVA cannot handle unbalanced or incomplete data. Hedeker et. al.(2006) named it "the Achilles heel of the MANOVA model for repeated measurements" (p. 34). A complete dataset is unrealistic in longitudinal studies; therefore, it can be used only in limited situations. Despite benefits of using MANOVA model over repeated measures ANOVA, it is not considered for this study on informative time and non-normal outcomes because of these drawbacks.

Mixed-Effects Model

Due to the unrealistic assumptions, such as non-missing, sphericity, and complete data structure in ANOVA and MANOVA models the use of these traditional methods are restricted to certain cases and must be interpreted with caution. The mixed-effects model is an univariate regression analysis on correlated responses (Davis, 2002). One main advantage of using the mixed-effects model is that it can handle missing data and incomplete data that are problematic in ANOVA and MANOVA analyses for repeated measures. Due to its flexibility, the mixed-effects model has been studied by and became popular among many statisticians for longitudinal data analyses these days (Lindstrom & Bates, 1990;

Pinheiro & Bates, 1995, 2000). Also, there are a couple more benefits that make the mixed-effects model popular for repeated measures analyses. Firstly, traditional methods, such as ANOVA and MANOVA, assume that all subjects would have the outcomes measured at the same number of fixed time points, but the mixed-effects model does not need this assumption. For this reason, any subjects with missing data or incomplete data can be included in the analysis, which means it will increase statistical power. It is not necessary for each subject to be measured at the same time points since time is treated as a continuous variable in the mixed-effects model (Hedeker & Gibbons, 2006). Secondly, since the mixed-effects model is an extension of repeated measures ANOVA, one can measure the average changes across time. Moreover, by including subject-specific effects in the model as a random effect, the model can detect individual trajectories. For these reasons, the mixed-effects model has become popular for longitudinal data analyses. The average responses across time for each subject is named as a fixed effect, while a subject-specific effect, which is unique to each subject, is called a random effect. Literally, the mixed-effects model combines both fixed effects and random effects (Fitzmaurice et al., 2004).

The mixed-effects model has been used with a variety of different names: random effects model (Diggle et al., 2002; Fitzmaurice et al., 2004; Laird & Ware, 1982), multilevel model (Goldstein, 2011; Nash & Varadhan, 2011), hierarchical model (Lee & Nelder, 1996; Raudenbush & Bryk, 2002), and random coefficient model (Leeuw & Kreft, 1986). Using matrix notations, the mixed-effects model for a vector of $n_i \times 1$ responses for the ith subject can be written as

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i, \tag{7}$$

where

$\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})'$ is the vector of responses for the ith subject at time point j for $j = 1, \ldots, n_i$,

$\boldsymbol{X}_i$ is the $n_i \times p$ covariate matrix for the ith subject,

$\boldsymbol{\beta}$ is the $p \times 1$ vector of fixed effects,

$\boldsymbol{Z}$ is the $n_i \times r$ design matrix for random effects,

$\boldsymbol{\gamma}$ is the $r \times 1$ vector of random effects, and

$\boldsymbol{\epsilon}$ is the $n_i \times 1$ vector for error with the assumptions of

$$\boldsymbol{\epsilon}_i = N(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{n_i}) \quad \text{and} \quad \boldsymbol{\gamma}_i = N(\boldsymbol{0}, \boldsymbol{\Sigma}_\gamma). \tag{8}$$

The variance-covariance matrix of the model becomes

$$\begin{aligned} Cov(\boldsymbol{y}_i) &= Cov(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i) \\ Cov(\boldsymbol{y}_i) &= Cov(\boldsymbol{Z}_i\boldsymbol{\gamma}_i) + Cov(\boldsymbol{\epsilon}_i) \\ Cov(\boldsymbol{y}_i) &= \boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}_i^{'} + Cov(\boldsymbol{\epsilon}_i). \end{aligned} \tag{9}$$

The model can be rewritten as

$$Cov(\boldsymbol{y}_i) = \boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}_i^{'} + \sigma^2\boldsymbol{I}_{n_i}, \tag{10}$$

since $Cov(\boldsymbol{\epsilon}_i) = \sigma^2\boldsymbol{I}_{n_i}$, which has the diagonal elements of the variance-covariance matrix with all zeros for the entries outside the main diagonal (Fitzmaurice et al., 2004). The mixed-effects model can be presented in a multilevel form for a better

explanation of each subject's effect on their observations (Hedeker & Gibbons, 2006).

For example, when there is time and one factor in the model, the level-1 model, which represents within-subjects, becomes

$$y_{ij} = \pi_{0j} + \pi_{1j} Time_{ij} + \epsilon_{ij}. \tag{11}$$

The level-2 model, which represents between-subjects, is

$$\begin{aligned} \pi_{0j} &= \beta_{00} + \beta_{01} Group_i + \mu_{0j} \\ \pi_{1j} &= \beta_{10} + \beta_{11} Group_i + \mu_{1j}, \end{aligned} \tag{12}$$

where

$$\epsilon_{ij} \sim N(0, \sigma^2) \quad \text{and} \quad \begin{pmatrix} \mu_{0j} \\ \mu_{1j} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11}. \end{pmatrix} \right]. \tag{13}$$

The linear format of the combined level-1 and level-2 models becomes

$$\begin{aligned} y_{ij} =& \beta_{00} + \beta_{01} Group_j + \beta_{10} Time_{ij} + \beta_{11} Group_j Time_{ij} \\ &+ \mu_{0j} + \mu_{1j} Time_{ij} + \epsilon_{ij}. \end{aligned} \tag{14}$$

In the linear model, β_{00}, β_{01}, β_{10}, and β_{11} are the coefficients for the fixed effects, and μ_{0j}, μ_{1j}, and ϵ_{ij} are the random effects. A description of each term in the model is as follows.

y_{ij} is the outcome for subject i measured at time point j,

β_{00} is the average intercept for all subjects,

β_{01} is the average difference in π_{0j} for a unit change in level-2 predictor,

β_{10} is the average of the level-1 slope,

β_{11} is the average difference in π_{1j} for a unit change in level-2 predictor,

π_{0j} is the intercept of the trajectory of subject i,

π_{1j} is the slope of the trajectory of subject i,

μ_{0j} is the unique contribution of each j to the mean response on time point j,

μ_{1j} is the unique contribution of each j to the slope on time point j

(Singer & Willett, 2003). As it can be seen in the model, because time points are from $j = 1$ to n_i, the model can include all outcomes measured at different time points in the analysis. Additionally, the model allows for each subject to be measured at different schedules of time points, which means the model can handle missing data, incomplete data, or unbalanced data (Hedeker & Gibbons, 2006).

Parameters of variance components can be estimated by using maximum likelihood estimation, which needs an iterative numerical solution of a nonlinear optimization procedure (Davis, 2002). There are many statistical computing software that can calculate paramteres of covariance components. Also, these softwares provide a variety of types of covariance structures for the G matrix, such as compound symmetry, unstructured, first-order autoregressive, or Toeplitz, etc., as an initial value of the iteration. An alternative approach instead of the maximum likelihood estimation is the restricted maximum likelihood (REML) approach. This approach was introduced since the maximum likelihood estimation

method generates a somewhat baised solution when the design is unbalanced (Patterson & Thompson, 1971).

Despite many advantages of the mixed-effects model for longitudinal data, this model also has some drawbacks. One is that the covariance matrix structure of the y_i vector is nonstationary. For example, when outcomes are collected from the same subjects at equally spaced time points, $j = i, \ldots, n$, the variance and covariance become respectively

$$\begin{aligned} Var(y_{ij}) &= \sigma_\alpha^2 + 2j\sigma_{\alpha\beta} + j^2\sigma_\beta^2 + \sigma^2, \\ Cov(y_{ij}, y_{ij'}) &= \sigma_\alpha^2 + (j + j')\sigma_{\alpha\beta} + jj\sigma_\beta^2. \end{aligned} \tag{15}$$

Consequently, general trends at time point j are (1) the $Var(y_{ij})$ increases after time j when $j > -\sigma_{\alpha\beta}/\sigma_\beta^2$, and (2) the $Var(y_{ij})$ decreases up to time j when $j < -\sigma_{\alpha\beta}/\sigma_\beta^2$ (Davis, 2002). However, this result is not realistic in longitudinal data. The other drawback shown in a simulation study is that the quality of the mixed-effects model is greatly affected by the choice of variance-covariance matrix structure (Davis, 2002). Moreover, in the mixed-effects model, time is still treated as fixed; therefore, the mixed-effects model is not an appropriate method for this study on informative schedule data.

Marginal Model

Liang and Zeger (1986) introduced the marginal model approach using the generalized estimating equations (GEE) method to analyze repeated measurements, which is an extension of the generalized linear model to longitudinal data analyses using quasilikelihood estimation. The terminology of the marginal model is referred

to as the population-average model, meaning that the model for the mean response is not affected by any random effects or previous responses, and instead solely depends on the covariate. The marginal model does not need any assumptions for the distribution of outcomes; it depends only on assumptions for the mean of responses, so the marginal model can be used for binary or count as well as continuous outcomes. However, in general, the GEE model is very useful for categorical and count outcomes (Hedeker & Gibbons, 2006). The other advantage is that subjects do not need to have the same number of outcomes measured at the same time periods (Fitzmaurice et al., 2004).

A brief introduction to GEE method is as follows. The expected mean of each response given covariate, $\mu_{ij} = E(y_{ij}|X_{ij})$, can be rewritten with a link function as

$$g(\mu_{ij}) = \boldsymbol{X}'_{ij}\boldsymbol{\beta}. \tag{16}$$

The variance of y_{ij} given the covariates becomes

$$Var(y_{ij}) = v(\mu_{ij})\phi, \tag{17}$$

where $v(\mu_{ij})$ is a known variance function, which is the relationship between the mean and the variance, expressing the variance as a function of the mean, and ϕ is a known or to be estimated scale parameter. The link and variance functions for different distributions of outcomes are shown below. The normally distributed outcomes with the identity link function has

$$g(\mu_{ij}) = \mu_{ij} = \boldsymbol{X}'_{ij}\boldsymbol{\beta}, \quad v(\mu_{ij}) = 1, \quad \text{and} \quad Var(y_{ij}) = v(\mu_{ij})\phi = \phi. \tag{18}$$

For binary outcomes,

$$g(\mu_{ij}) = \log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \boldsymbol{X}'_{ij}\boldsymbol{\beta}, \quad Var(y_{ij}) = \mu_{ij}(1-\mu_{ij}), \quad \text{and} \quad \phi = 1. \tag{19}$$

For Poisson outcomes,

$$g(\mu_{ij}) = \log(\mu_{ij}) = \boldsymbol{X}'_{ij}\boldsymbol{\beta}, \quad Var(y_{ij}) = \mu_{ij}, \quad \text{and} \quad \phi = 1. \tag{20}$$

The last component of the GEE model is the working correlation matrix $\boldsymbol{R}_i(\boldsymbol{a})$ for each $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})'$ with dimension of $n_i \times n_i$, which is called the working correlation between y_{ij} and $y_{ij'}$. This correlation structure will vary based on a pattern of relationships of the repeated measurements. For example, when no correlation is assumed, $\boldsymbol{R}(\boldsymbol{a}) = \boldsymbol{I}$, which is the identity matrix with the equal correlations, $\boldsymbol{R}_{ij}(\boldsymbol{a}) = \rho$ for any $i \neq j$, the correlation structure is called exchangeable, also known as compound symmetry. The working variance-covariance matrix for $\boldsymbol{y}_i$ becomes

$$\boldsymbol{V}(\boldsymbol{a}) = \phi \boldsymbol{A}_i^{1/2} \boldsymbol{R}_i(\boldsymbol{a}) \boldsymbol{A}_i^{1/2}, \tag{21}$$

where $\boldsymbol{A}_i$ is a diagonal matrix with $v(\mu_{ij})$ as the jth diagonal element. Finally, the GEE estimates of the parameter vector $\boldsymbol{\beta}$ can be obtained by solving

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right) [\boldsymbol{V}_i]^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \boldsymbol{0}_p \tag{22}$$

(Singer & Willett, 2003). To get the vector of parameters $\boldsymbol{\beta}$, a numerical iterative method is needed by using the quasilikelihood method (Davis, 2002).

Since there is no distributional assumption, the GEE model provides a very flexible approach to estimate the mean and pairwise correlations among repeated measures; also, it can handle missing data and unbalanced data (Fitzmaurice et al., 2004). The GEE model also has some drawbacks. Firstly, the estimation of $\boldsymbol{\beta}$ used in the GEE model is less efficient compared to the maximum likelihood-based estimation due to no assumption on the distribution (Fitzmaurice et al., 2004). Secondly, the GEE model is not sensitive to misselection of variance-covariance structure; therefore, when research questions are not about variance-covariance structure, the GEE model can be a good option to the researcher. However, if the researcher's interest is in the estimation of the variance-covariance structure, the GEE model will not be a good selection. Thirdly, even though complete data across time for subjects is not required in the GEE model, the model assumes all time points are fixed, and if there are any missing responses, it must be missing completely at random (MCAR) (Hedeker & Gibbons, 2006). The assumption of fixed time points is unrealistic and does not fit this study on informative schedule data. Fourthly, parameter estimates of $\hat{\boldsymbol{\beta}}$ are not consistent to estimate $\boldsymbol{\beta}$ when time-varying covariates are involved in the regression model (Pepe & Anderson, 1994). Despite all the advantages of the GEE model, in this study, time points are considered to be not fixed. That is why the GEE model is not considered to be a potential candidate for this study.

Joint Model

Currently, a new approach named the joint model is getting popular in an attempt to handle irregular measurement occasions in the analysis, which is combining longitudinal data and time or any other factors that the researcher is interested in (Henderson et al., 2000; Kim et al., 2012; Liang et al., 2009; Lin et al., 2004; Lipsitz et al., 2002; Qiu et al., 2013; Wu, Liu, Yi, & Huang, 2012). The joint model uses maximum likelihood estimators of its joint distribution.

For example, Liang et al., (2009) presented a joint model of longitudinal data with informative observation times via latent variables to handle highly irregular time points and longitudinal outcomes. Their model is for longitudinal outcomes and censoring or dropout time under the assumption that censoring time is non-informative; in addition, outcomes are measured only at the dropout time. That is the difference from this current study that involves multiple time points.

Lipsitz et al. (2002) presented a joint model for longitudinal data. The joint model assumes that time points are not fixed and dependent on previous outcomes, and the repeated measurements are supposed to follow a multivariate normal distribution. The likelihood function of the joint model consists of two components: one for the counting process, the other for the outcome process. The outcome process is determined by $y_i(t) = \mu_i(t) + \epsilon_i(t)$, where $\mu_i(t)$ is the marginal mean at time t and $\epsilon_i(t)$ is a Gaussian process. The counting process is the number of measurements on the ith subject by a continuous time point. The joint model was developed based on the idea that time points are dependent on previously

observed data only, not the times where outcomes are measured. The advantage of the separation of likelihood function, such as the counting process and the outcome process, is the simplicity of the calculation of the maximum likelihood estimators, since the counting process can be ignored when using the likelihood based estimation for the outcome process, or vice versa.

Bronsert (2009) presented his joint model to handle informative time and normally distributed longitudinal outcomes. Bronsert's Gaussian-Exponential model comes with assumptions of a normal distribution for the outcome process and an exponential distribution for informative time. The main difference between Bronsert's joint model and the joint models above is that repeated outcomes are dependent on previous outcomes and current time points. That is given by

$$\begin{aligned} f_{\theta_i}(\boldsymbol{y}_i, \boldsymbol{t}_i) =& \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(y_{i1} - \boldsymbol{X}'_{i1}\boldsymbol{\beta})^2}{\sigma^2}\right) \times f(t_{i1}) \\ & \times \prod_{j=2}^{n_i} \left\{ \frac{1}{\sqrt{2\pi\sigma^2\sqrt{1-\rho_i^2}}} \exp\left(-\frac{1}{2}\frac{(y_{ij} - \gamma t_{ij} - \varphi_i y_{ij-1} - \boldsymbol{X}'_{ij}\boldsymbol{\beta})^2}{\sigma^2(1-\rho_i^2)}\right) \right. \\ & \left. \cdot \exp(\alpha + \delta_i y_{ij-1}) \cdot \exp(-e^{\alpha+\delta_i y_{ij-1}} t_{ij}) \right\}, \end{aligned} \tag{23}$$

where

$\boldsymbol{\beta}$ is the effect of the independent variables on outcomes,

$f(t_{i1})$ is the initial time point for the ith subject,

φ is the effect of the previous outcome on the mean response of the current outcome,

γ is the effect of current time on the mean response,

α is the constant parameter for the time process,

δ is the effect of the previous outcome on the mean time, and

$\boldsymbol{X}$ is the design matrix.

$f(t_{i1})$ is a distribution for an initial time point, and it is assumed that the distribution of $f(t_{i1})$ does not affect any parameters. The maximum likelihood estimators $\boldsymbol{\Theta}$ were calculated from the log-likelihood function of the Gaussian-Exponential model by using the nonlinear optimization algorithms called NLPDD Call in SAS/IML that uses the double dogleg optimization method. Bronsert's study showed only parameter estimates without testing if the estimators can be usable for hypothesis testing, which assumes the normality assumption.

Later, Lin (2011) adapted and modified Bronsert's Gaussian-Exponential model. Lin's modified model is

$$\begin{aligned} f_{\theta_i}(\boldsymbol{y}_i, \boldsymbol{t}_i) = & \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(y_{i1} - \boldsymbol{X}'_{i1}\boldsymbol{\beta})^2}{\sigma^2}\right) \times f(t_{i1}) \\ & \times \prod_{j=2}^{n_i} \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(y_{ij} - \gamma t_{ij} - \varphi_i y_{ij-1} - \boldsymbol{X}'_{ij}\boldsymbol{\beta})^2}{\sigma^2}\right) \right. \\ & \left. \cdot \exp(\alpha + \delta_i y_{ij-1}) \cdot \exp(-e^{\alpha+\delta_i y_{ij-1}} t_{ij}) \right\}. \end{aligned} \tag{24}$$

The only difference is that ρ_i^2 is not included in Lin's model. The term of ρ_i^2 was used to account for the relationships between two outcome variables at two time points in Bronsert's model. Lin believed that it is redundant since there is another term to take care of the correlation, so ρ_i^2 is not included in the modified model. The modified model was used in this study to test if the model can be extended

when the responses have the exponential family of distributions with informative time.

Conclusion

In general, the main purpose of longitudinal data analyses is to capture the changes over time. Currently, multiple methods are being used to fulfill this purpose, and most of them are working well in finely-designed experiments, especially when all subjects share the common fixed time points. One common assumption to all the models, except the joint models discussed in this chapter, is that time intervals are predetermined by the researcher or previous studies. In practice, sometimes time points must be determined based on prior outcomes, which means individuals may have different sets of time points. As a result of irregular measurement occasions for each subject, traditional methods may not be the best for this longitudinal design with informative schedule data, since traditional methods assume time to be fixed. The joint model does not require time to be fixed. Thus, in the current study, the joint model by Bronsert and Lin was investigated to find if the model could be extended when repeated outcomes have the exponential family of distributions.

CHAPTER III

METHODOLOGY

The traditional models can be used for informative schedule data, but the results of analyses may not be usable because the traditional approaches assume time is fixed. The joint model by Bronsert (2009) and Lin (2011) was developed under the assumption that outcomes follow a normal distribution and time follows an exponential distribution. In this study, Bronsert and Lin's joint model was adapted and modified to find if the parameter estimates of the extended joint models satisfy the normality assumption when the distribution of outcomes is a member of the exponential family of distributions.

Notation and Joint Model

The outcome for the ith subject measured at the jth time point refers to y_{ij}; therefore, the ith subject has $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})'$ collected at $\boldsymbol{t}_i = (t_{i1}, t_{i2}, \ldots, t_{in_i})'$. The joint distribution of outcomes and time points becomes

$$f_{\boldsymbol{\Theta}}(y_i, t_i) = f_{\boldsymbol{\Theta}}(\boldsymbol{y}_i | \boldsymbol{t}_i) \cdot f_{\boldsymbol{\Theta}}(\boldsymbol{t}_i), \tag{25}$$

where $\boldsymbol{\Theta}$ is a vector of unknown parameters. A general model can be derived by using this joint distribution of $\boldsymbol{y}_i$ and $\boldsymbol{t}_i$. Therefore, the general model under the assumptions that the current outcome is dependent on the one-step prior outcome (y_{ij-1}), current outcome (y_{ij}), and current time point (t_{ij}) becomes

$$f_{\boldsymbol{\Theta}}(\boldsymbol{y}_i, \boldsymbol{t}_i) = f_{\boldsymbol{\Theta}}(y_{i1}|t_{i1}) \cdot f_{\boldsymbol{\Theta}}(t_{i1}) \cdot \prod_{j=2}^{n_i} f_{\boldsymbol{\Theta}}(y_{ij}|t_{ij}, y_{ij-1}) \cdot f_{\boldsymbol{\Theta}}(t_{ij}|y_{ij-1}). \quad (26)$$

Based on this general model, a joint model was developed for each member of the exponential family of distributions, while assuming time to follow an exponential distribution.

Generalized Linear Model

The purpose of this study was to develop joint models that can handle outcomes from the exponential family of distributions. The generalized linear model provides a unified class of models of regression analysis, regardless of discrete or continuous outcomes (Dobson, 2001; Fitzmaurice et al., 2004; McCullagh & Nelder, 1983, 1989; Nelder & Wedderburn, 1972). The generalized linear model has three components: a random component, a systematic component, and a link function. The random component identifies the distribution of outcome variable.

The generalized linear model assumes that the outcome variable has a probability distribution from the exponential family of distributions. For example, when the outcome is binary, such as yes or no, a binomial or Bernoulli distribution is assumed. When the outcome is count, a Poisson distribution is assumed. The variance of the outcome can be written as a product of a single scale or dispersion parameter, ϕ, and it is called the variance function:

$$Var(\boldsymbol{Y}) = \phi v(\mu). \quad (27)$$

The systematic component of the generalized linear model identifies explanatory variables. These explanatory variables are combined into a linear

format, and it is called the linear predictor:

$$\eta_i = \boldsymbol{X}_i'\boldsymbol{\beta} = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}, \tag{28}$$

where $X_{i1} = 1$ for all i, and β_1 is the intercept. $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is the vector of unknown parameters (Fitzmaurice et al., 2004). The link function is a function that connects the linear predictor with the mean of the probability distribution. So, the function $g(\cdot)$ connects a random component to a systematic component, which can be written as

$$g(\eta_i) = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} \tag{29}$$

(Fitzmaurice et al., 2004). As an example, the variance and link function for the Gaussian, Bernoulli, Poisson, and Gamma distributions are as follows (Fitzmaurice et al., 2004).

Table 2

Variances and link functions

Distribution	Variance Function	Link Function	Mean Function
Gaussian	$v(\mu) = 1$	Identity: $\mu = \eta$	$\mu = \eta$
Bernoulli	$v(\mu) = \mu(1-\mu)$	Logit: $\log(\mu/1-\mu) = \eta$	$\mu = \frac{\exp(\eta)}{1-\exp(\eta)}$
Poisson	$v(\mu) = \mu$	Log: $\log(\mu) = \eta$	$\mu = \exp(\eta)$
Gamma	$v(\mu) = \mu^2$	Log: $\log(\mu) = \eta$	$\mu = \exp(\eta)$

All the distributions from the exponential family can be expressed as

$$f(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right), \tag{30}$$

where θ is a canonical parameter and ϕ is a scale or dispersion parameter (Fitzmaurice et al., 2004). The commonly used distributions of the outcomes are the Gaussian for normally distributed outcomes, the Bernoulli for binary outcomes, the Poisson for count outcomes, and the Gamma distribution for survival time or waiting time. The probability density function of the Gaussian distribution can be rewritten in an exponential family form of

$$\begin{aligned} f(y;\mu,\sigma^2) &= (2\pi\sigma^2)^{-1/2}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \\ &= \exp\left(-\frac{1}{2}\log(2\pi\sigma^2)\right)\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \\ &= \exp\left(-\frac{(y^2-2y\mu+\mu^2)}{2\sigma^2}-\frac{1}{2}\log(2\pi\sigma^2)\right) \\ &= \exp\left(\frac{y\mu-\mu^2/2}{\sigma^2}-\frac{1}{2}\left(\frac{y^2}{\sigma^2}+\log(2\pi\sigma^2)\right)\right), \end{aligned} \tag{31}$$

with a canonical parameter and a dispersion parameter of

$$\theta = \mu \quad \text{and} \quad a(\phi) = \sigma^2. \tag{32}$$

The probability density function of the Bernoulli distribution can be rewritten as

$$\begin{aligned} f(y;\mu) &= \mu^y(1-\mu)^{1-y} \\ &= \exp(y\log(\mu)+(1-y)\log(1-\mu)) \\ &= \exp\left(y\log\left(\frac{\mu}{1-\mu}\right)+\log(1-\mu)\right), \end{aligned} \tag{33}$$

with a canonical parameter and a dispersion parameter of

$$\theta = \log\left(\frac{\mu}{1-\mu}\right) = \text{logit}(\mu) \quad \text{and} \quad a(\pi) = 1. \tag{34}$$

The probability density function of the Poisson distribution can be rewritten as

$$\begin{aligned} f(y;\mu) &= \frac{e^{-\mu}\mu^{y}}{y!} \\ &= \exp(y\log(\mu) - \mu - \log(y!)), \end{aligned} \tag{35}$$

with a canonical parameter and a dispersion parameter of

$$\theta = \log(\mu) \quad \text{and} \quad a(\phi) = 1 \tag{36}$$

(Fitzmaurice et al., 2004). Lastly, the probability density function of the Gamma distribution can be rewritten as

$$\begin{aligned} f(y;\lambda,v) &= \frac{y^{v-1}\lambda^{v}e^{-y\lambda}}{\Gamma(v)} \\ &= \exp(-y\lambda + v\log(\lambda) + (v-1)\log(y) - \log(\Gamma(v)) \\ &= \exp\left(\frac{y(-\lambda/v) - \log(\lambda)}{1/v} + (v-1)\log(y) - \log(\Gamma(v)\right) \end{aligned} \tag{37}$$

with a canonical parameter and a dispersion parameter of

$$\theta = -\frac{\lambda}{v} \quad \text{and} \quad a(\phi) = \frac{1}{v}. \tag{38}$$

Joint Model

In this section, three joint models are presented with different outcome types as examples, but time is still assumed to follow an exponential distribution. Examples of the outcome distributions include the Bernoulli for binary, the Poisson for count, and the Gamma distribution for survival time or waiting time. Also, the mean function and the link function for each distribution are modified to take

account of the dependency among the current outcome, one-step prior outcome, and current time into the model.

Bernoulli-Exponential Model

This joint model is for binary outcomes and informative time. The Bernoulli distribution, in an exponential family form, can be expressed to handle repeated outcomes with given dependency

$$\begin{aligned} f(y_{ij}|y_{ij-1}, t_{ij}, X) &= \mu_{ij}^{y_{ij}}(1-\mu_{ij})^{1-y_{ij}} \\ &= \exp\left(y_{ij}\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) + \log(1-\mu_{ij})\right), \end{aligned} \tag{39}$$

where $\mu_i = E(Y_i) = P(Y_i = 1)$. The parameter is

$$\theta_{ij} = \log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \text{logit}(\mu_{ij}). \tag{40}$$

The link function becomes

$$\begin{aligned} \log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) &= \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \gamma t_{ij} + \varphi y_{ij-1} \\ &= \boldsymbol{X}_i^{'}\boldsymbol{\beta} + \gamma t_{ij} + \varphi y_{ij-1}. \end{aligned} \tag{41}$$

Then, the mean function can be expressed as

$$\mu_{ij} = \frac{\exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta} + \gamma t_{ij} + \varphi y_{ij-1})}{1 + \exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta} + \gamma t_{ij} + \varphi y_{ij-1})}. \tag{42}$$

Hence, the mean function for the initial value for the ith subject and the mean function after the initial value can be expressed as

$$\mu_{i1} = \frac{\exp(\boldsymbol{X}_i'\boldsymbol{\beta})}{1 + \exp(\boldsymbol{X}_i'\boldsymbol{\beta})} \quad \text{and} \quad \mu_{ij} = \frac{\exp(\boldsymbol{X}_i'\boldsymbol{\beta} + \gamma t_{ij} + \varphi y_{ij-1})}{1 + \exp(\boldsymbol{X}_i'\boldsymbol{\beta} + \gamma t_{ij} + \varphi y_{ij-1})}. \tag{43}$$

Finally, the joint model for the Bernoulli-Exponential model can be written as

$$f_{\boldsymbol{\Theta}}(\boldsymbol{y}_i, \boldsymbol{t}_i) = \exp\left(y_{i1}\log\left(\frac{\mu_{i1}}{1-A_{i1}}\right) + \log(1-\mu_{i1})\right) \times \prod_{j=2}^{n_i}\left\{\exp\left(y_{ij}\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) + \log(1-\mu_{ij})\right) \cdot \exp(\alpha + \delta y_{ij-1}) \cdot \exp(-e^{\alpha+\delta y_{ij-1}} t_{ij})\right\}. \tag{44}$$

The likelihood function, which is the product of the density functions for m individuals, is

$$L(\boldsymbol{\Theta}, y_i, \ldots, y_m) = \prod_{i=1}^{m}\left\{\exp\left(y_{i1}\log\left(\frac{\mu_{i1}}{1-\mu_{i1}}\right) + \log(1-\mu_{i1})\right) \times \prod_{j=2}^{n_i}\exp\left(y_{ij}\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) + \log(1-\mu_{ij})\right) \cdot \exp(\alpha + \delta y_{ij-1}) \cdot \exp(-e^{\alpha+\delta y_{ij-1}} t_{ij})\right\}. \tag{45}$$

Generally, the log-likelihood function is used to find the parameter estimates for convenience. The log-likelihood function for the ith individual in the model becomes

$$\begin{aligned} l_i =& \log\left\{\exp\left(y_{i1}\log\left(\frac{\mu_{i1}}{1-\mu_{i1}}\right) + \log(1-\mu_{i1})\right) \right. \\ & \left. \times \prod_{j=2}^{n_i}\exp\left(y_{ij}\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) + \log(1-\mu_{ij})\right) \cdot \exp(\alpha + \delta y_{ij-1}) \cdot \exp(-e^{\alpha+\delta y_{ij-1}} t_{ij})\right\} \\ =& \left(y_{i1}\log\left(\frac{\mu_{i1}}{1-\mu_{i1}}\right) + \log(1-\mu_{i1})\right) \\ & + \sum_{j=2}^{n_i}\left\{(y_{ij}\log\left(\frac{\mu_{ij}}{1-\mu_{ij}} + \log(1-\mu_{ij}\right) + \alpha + \delta y_{ij-1} - e^{\alpha+\delta y_{ij-1}} t_{ij})\right\}. \end{aligned} \tag{46}$$

The log-likelihood function for all individuals is the sum of m individuals' log-likelihood functions, as shown below.

$$
\begin{aligned}
l = \sum_{i=1}^{m} l_i = & \sum_{i=1}^{m} \left(y_{i1} \log \left(\frac{\mu_{i1}}{1-\mu_{i1}} \right) + \log(1-\mu_{i1}) \right) \\
& + \sum_{i=1}^{m} \sum_{j=2}^{n_i} \left(y_{ij} \log \left(\frac{\mu_{ij}}{1-\mu_{ij}} \right) + \log(1-\mu_{ij}) \right) \\
& + \sum_{i=1}^{m} \sum_{j=2}^{n_i} (\alpha + \delta y_{ij-1} - e^{\alpha + \delta y_{ij-1}} t_{ij}).
\end{aligned} \tag{47}
$$

Poisson-Exponential Model

This joint model is for count outcomes and informative time with an exponential distribution. Since the Poisson distribution is a member of the exponential family, it can be rewritten as

$$
\begin{aligned}
f(y_{ij} | y_{ij-1}, t_{ij}, X) &= \frac{e^{-\mu_{ij}} \mu_{ij}^{y_{ij}}}{y_{ij}!} \\
&= \exp(y_{ij} \log(\mu_{ij}) - \mu_{ij} - \log(y_{ij}!)).
\end{aligned} \tag{48}
$$

The parameter is

$$
\theta_{ij} = \log(\mu_{ij}). \tag{49}
$$

The canonical link function becomes

$$
\begin{aligned}
\log(\mu_{ij}) &= \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \gamma t_{ij} + \varphi y_{ij-1} \\
&= \boldsymbol{X}_i^{'} \boldsymbol{\beta} + \gamma t_{ij} + \varphi y_{ij-1}.
\end{aligned} \tag{50}
$$

Then, the mean function is

$$
\mu_{ij} = \exp(\boldsymbol{X}_i^{'} \boldsymbol{\beta} + \gamma t_{ij} + \varphi y_{ij-1}). \tag{51}
$$

Hence, the mean function for the initial value for the ith subject and the mean function after the initial value can be expressed as

$$\mu_{i1} = \exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta}) \quad \text{and} \quad \mu_{ij} = \exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta} + \gamma t_{ij} + \varphi y_{ij-1}). \tag{52}$$

Then, the joint model of the Poisson-Expoential model becomes

$$\begin{aligned} f_{\boldsymbol{\Theta}}(\boldsymbol{y}_i, \boldsymbol{t}_i) &= \exp(y_{i1}\log(\mu_{i1}) - \mu_{i1} - \log(y_{i1}!)) \\ &\times \prod_{j=2}^{n_i} \exp(y_{ij}\log(\mu_{ij}) - \mu_{ij} - \log(y_{ij}!)) \cdot \exp(\alpha + \delta y_{ij-1}) \cdot \exp(-e^{\alpha+\delta y_{ij-1}} t_{ij}). \end{aligned} \tag{53}$$

The likelihood function for m individuals is

$$\begin{aligned} L(\boldsymbol{\Theta}, y_i, \ldots, y_m) = \prod_{i=1}^{m} \Bigg\{ &\exp(y_{i1}\log(\mu_{i1}) - \mu_{i1} - \log(y_{i1}!)) \\ &\times \prod_{j=2}^{n_i} \exp(y_{ij}\log(\mu_{ij}) - \mu_{ij} - \log(y_{ij}!)) \cdot \exp(\alpha + \delta y_{ij-1}) \cdot \exp(-e^{\alpha+\delta y_{ij-1}} t_{ij}) \Bigg\}. \end{aligned} \tag{54}$$

The log-likelihood function for the ith individual in the model becomes

$$\begin{aligned} l_i =& \log\Bigg\{ \exp(y_{i1}\log(\mu_{i1}) - \mu_{i1} - \log(y_{i1}!)) \\ &\times \prod_{j=2}^{n_i} \exp(y_{ij}\log(\mu_{ij}) - \mu_{ij} - \log(y_{ij}!)) \cdot \exp(\alpha + \delta y_{ij-1}) \cdot \exp(-e^{\alpha+\delta y_{ij-1}} t_{ij}) \Bigg\} \\ =& (y_{i1}\log(A_{i1}) - \mu_{i1} - \log(y_{i1}!)) \\ &+ \sum_{j=2}^{n_i} \left\{ y_{ij}\log(\mu_{ij}) - \mu_{ij} - \log(y_{ij}!)) + \alpha + \delta y_{ij-1} - e^{\alpha+\delta y_{ij-1}} t_{ij} \right\}. \end{aligned} \tag{55}$$

The log-likelihood function for all individuals, which is the sum of m individuals' log-likelihood functions, is

$$
\begin{aligned}
l = \sum_{i=1}^{m} l_i = & \sum_{i=1}^{m} (y_{i1}\log(\mu_{i1}) - \mu_{i1} - \log(y_{i1}!)) \\
& + \sum_{i=1}^{m}\sum_{j=2}^{n_i} (y_{ij}\log(\mu_{ij}) - \mu_{ij} - \log(y_{ij}!)) \\
& + \sum_{i=1}^{m}\sum_{j=2}^{n_i} (\alpha + \delta y_{ij-1} - e^{\alpha+\delta y_{ij-1}} t_{ij}).
\end{aligned} \tag{56}
$$

Gamma-Exponential Model

This joint model is used for waiting or survival time and informative time with an exponential distribution. The density function of a random variable Y with a Gamma distribution can be rewritten in an exponential family form:

$$
\begin{aligned}
& f(y_{ij}|y_{ij-1}, t_{ij}, X) = \frac{y_{ij}^{v_{ij-1}} \lambda_{ij}^{v_{ij}} e^{-y_{ij}\lambda_{ij}}}{\Gamma(v_{ij})} \\
& = \exp\left(\frac{y_{ij}(-1/\mu_{ij}) - \log(\mu_{ij})}{1/v_{ij}} + v_{ij}\log(v_{ij}) + (v_{ij}-1)\log(y_{ij}) - \log(\Gamma(v_{ij})) \right).
\end{aligned} \tag{57}
$$

The θ becomes $1/\mu_{ij}$, so the canonical link function and dispersion parameter are

$$
g(\mu_{ij}) = -\frac{1}{\mu_{ij}} \quad \text{and} \quad a(\phi) = \frac{1}{v_{ij}}. \tag{58}
$$

For the Gamma distribution, there are three link functions: (1) inverse link ($g(\mu) = 1/\mu$), (2) log link ($g(\mu) = \log(\mu)$), and (3) identity link function ($g(\mu) = \mu$). The log link function was used in this study. So, the mean function with the log link function becomes

$$
\mu_{ij} = \exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta} + \gamma t_{ij} + \varphi y_{ij-1}). \tag{59}
$$

Hence, the mean function for the initial value for the ith subject and the mean function after the initial value can be expressed as

$$\mu_{i1} = \exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta}) \quad \text{and} \quad \mu_{ij} = \exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta} + \gamma t_{ij} + \varphi y_{ij-1}). \tag{60}$$

As a consequence, the joint model for the Gamma distribution becomes

$$\begin{aligned} f_{\boldsymbol{\Theta}}(\boldsymbol{y}_i, \boldsymbol{t}_i) &= \exp\left(\frac{y_{i1}(-1/\mu_{i1}) - \log(\mu_{i1})}{1/v_{ij}} + v_{ij}\log(v_{ij}) + (v_{ij}-1)\log(y_{i1}) - \log(\Gamma(v_{ij}))\right) \\ &\times \prod_{j=2}^{n_i}\left(\exp\left(\frac{y_{ij}(-1/\mu_{ij}) - \log(\mu_{ij})}{1/v_{ij}} + v_{ij}\log(v_{ij}) + (v_{ij}-1)\log(y_{i1}) - \log(\Gamma(v_{ij}))\right)\right. \\ &\left.\cdot \exp(\alpha + \delta y_{ij-1}) \cdot \exp(-e^{\alpha+\delta y_{ij-1}} t_{ij})\right). \end{aligned} \tag{61}$$

The likelihood function is the product of the density functions for m individuals.

$$\begin{aligned} &L(\boldsymbol{\Theta}, y_i, \ldots, y_m) = \\ &\prod_{i=1}^{m}\left\{\exp\left(\frac{y_{i1}(-1/\mu_{i1}) - \log(\mu_{i1})}{1/v_{ij}} + v_{ij}\log(v_{ij}) + (v_{ij}-1)\log(y_{i1}) - \log(\Gamma(v_{ij}))\right)\right. \\ &\times \prod_{j=2}^{n_i}\exp\left(\frac{y_{ij}(-1/\mu_{ij}) - \log(\mu_{ij})}{1/v_{ij}} + v_{ij}\log(v_{ij}) + (v_{ij}-1)\log(y_{ij}) - \log(\Gamma(v_{ij}))\right) \\ &\left.\cdot \exp(\alpha + \delta y_{ij-1}) \cdot \exp(-e^{\alpha+\delta y_{ij-1}} t_{ij})\right\}. \end{aligned} \tag{62}$$

The log-likelihood function for the ith individual in the model becomes

$$\begin{aligned} l_i =& \log\Bigg\{ \exp\left(\frac{y_{i1}(-1/\mu_{i1}) - \log(\mu_{i1})}{1/v_{ij}} + v_{ij}\log(v_{ij}) + (v_{ij}-1)\log(y_{i1}) - \log(\Gamma(v_{ij})) \right) \\ & \times \prod_{j=2}^{n_i} \Bigg(\exp\left(\frac{y_{ij}(-1/\mu_{ij}) - \log(\mu_{ij})}{1/v_{ij}} + v_{ij}\log(v_{ij}) + (v_{ij}-1)\log(y_{ij}) - \log(\Gamma(v_{ij})) \right) \\ & \cdot \exp(\alpha + \delta y_{ij-1}) \cdot \exp(-e^{\alpha+\delta y_{ij-1}} t_{ij}) \Bigg) \Bigg\} \\ =& \left(\frac{y_{i1}(-1/\mu_{i1}) - \log(\mu_{i1})}{1/v_{ij}} + v_{ij}\log(v_{ij}) + (v_{ij}-1)\log(y_{i1}) - \log(\Gamma(v_{ij})) \right) \\ & + \sum_{j=2}^{n_i} \Bigg(\frac{y_{i1}(-1/\mu_{ij}) - \log(\mu_{ij})}{1/v_{ij}} + v_{ij}\log(v_{ij}) + (v_{ij}-1)\log(y_{ij}) - \log(\Gamma(v_{ij})) \\ & + \alpha + \delta y_{ij-1} - e^{\alpha+\delta y_{ij-1}} t_{ij} \Bigg). \end{aligned} \tag{63}$$

Finally, the log-likelihood function for all individuals becomes

$$\begin{aligned} l =& \sum_{i=1}^{m} l_i = \sum_{i=1}^{m} \left(\frac{y_{i1}(-1/\mu_{i1}) - \log(\mu_{i1})}{1/v_{ij}} + v_{ij}\log(v_{ij}) + (v_{ij}-1)\log(y_{i1}) - \log(\Gamma(v_{ij})) \right) \\ & + \sum_{i=1}^{m} \sum_{j=2}^{n_i} \left(\frac{y_{ij}(-1/\mu_{ij}) - \log(\mu_{ij})}{1/v_{ij}} + v_{ij}\log(v_{ij}) + (v_{ij}-1)\log(y_{ij}) - \log(\Gamma(v_{ij})) \right) \\ & + \sum_{i=1}^{m} \sum_{j=2}^{n_i} \left(\alpha + \delta y_{ij-1} - e^{\alpha+\delta y_{ij-1}} t_{ij} \right). \end{aligned} \tag{64}$$

Parameter Estimation

The maximum likelihood estimators are defined as the values that maximize the joint probability evaluated at their observed data (Fitzmaurice et al., 2004). Generally, the log-likelihood function is used rather than the likelihood function for joint probability because it is simpler to find the parameter estimates. However, as can be seen, maximization of the log-likelihood function is not an easy job;

therefore, mathematical iterative methods are needed. In the two previous studies, the nonlinear optimization function called NLPDD in SAS/IML was used to find the maximum likelihood estimators from the log-likelihood function. The NLPDD function is a nonlinear optimization function using a double dogleg method, which combines the quasi-Newton and trust-region methods. Many nonlinear optimization methods are provided in SAS or R (Nash & Varadhan, 2011; SAS Institute, 2008). None of them seems to be superior to any other methods in any situation. The nonlinear optimization function, called maxLik in R, is chosen in the current study since it provides a single, unified interface to various optimization routines, offering easy access to likelihood-specific features. The Newton-Raphson maximization algorithm is used by default in the maxLik function.

Parameter Testing

There are three ways of using the likelihood function: the Wald test, the score test, and the likelihood ratio test. These are used for hypothesis testing, to determine the significance of the parameter estimates, or to determine confidence intervals (Agresti, 2007). The likelihood ratio test uses the ratio of two maximized log-likelihood functions for two nested models: (1) the maximized log-likelihood value for the null hypothesis denoted by $\hat{l}_{red}$; and (2) the maximized log-likelihood value for the alternative hypothesis denoted by $\hat{l}_{full}$. Hence, the likelihood ratio test statistic is

$$2(\hat{l}_{full} - \hat{l}_{red}) \sim \chi^2_{df_{full} - df_{red}}, \tag{65}$$

where df_{full} is the degrees of freedom for the full model and df_{red} is the degrees of freedom for the reduced model. The ratio is compared to the chi-squared distribution with the degrees of freedom equal to the difference between the two models' number of parameters. When the difference gets larger, it shows that the reduced model is inappropriate (Fitzmaurice et al., 2004). This likelihood ratio test is used to maintain the consistency from the previous study and ease of calculation.

Model Selection

Model selection criteria used in this study are the Akaike information criterion (AIC), the Akaike information criterion with correction (AICc), and the Bayesian information criterion (BIC) to compare non-nested models.

The AIC measures relative quality of a model to provide a method on model selection with given data. The AIC is defined as

$$AIC = 2k - 2ln(L), \tag{66}$$

where k is the number of parameters in the model and $ln(L)$ is the maximized value of the log-likelihood function of the model. The AIC takes into account both the statistical goodness of fit and the number of parameters to be estimated. The model with a low AIC value is preferred, which has the fewest number of parameters with adequate fit to the data (Everitt, 2006).

For example, since $ln(L)$ is the maximized log-likelihood at $\hat{\boldsymbol{\Theta}}$, the AIC for the Bernoulli-Exponential model becomes

$$\begin{aligned} AIC = 2k - 2\Bigg\{ & \sum_{i=1}^{m} y_{i1} \boldsymbol{X}_i^{'} \hat{\boldsymbol{\beta}} + \log\left(\frac{1}{1+\exp(\boldsymbol{X}_i^{'}\hat{\boldsymbol{\beta}})}\right) \\ & + \sum_{i=1}^{m}\sum_{j=2}^{n_i}\left(y_{ij}(\boldsymbol{X}_i^{'}\hat{\boldsymbol{\beta}} + \hat{\gamma} t_{ij} + \hat{\varphi} y_{ij-1}) + \log\left(\frac{1}{1+\exp(\boldsymbol{X}_i^{'}\hat{\boldsymbol{\beta}} + \hat{\gamma} t_{ij} + \hat{\varphi} y_{ij-1})}\right)\right) \\ & + \sum_{i=1}^{m}\sum_{j=2}^{n_i}\left(\hat{\alpha} + \hat{\delta} y_{ij-1} - e^{\hat{\alpha}+\hat{\delta} y_{ij-1}} t_{ij}\right)\Bigg\}. \end{aligned} \tag{67}$$

The Poisson-Exponential model has the AIC of

$$\begin{aligned} AIC = & 2k - 2\Bigg\{ \sum_{i=1}^{m}\left(y_{i1}\boldsymbol{X}_i^{'}\hat{\boldsymbol{\beta}} - \exp(\boldsymbol{X}_i^{'}\hat{\boldsymbol{\beta}}) - \log(y_{ij}!)\right) \\ & + \sum_{i=1}^{m}\sum_{j=2}^{n_i}\left(y_{ij}(\boldsymbol{X}_i^{'}\hat{\boldsymbol{\beta}} + \hat{\gamma} t_{ij} + \hat{\varphi} y_{ij-1}) - \exp(\boldsymbol{X}_i^{'}\hat{\boldsymbol{\beta}} + \hat{\gamma} t_{ij} + \hat{\varphi} y_{ij-1}) - \log(y_{ij}!)\right) \\ & + \sum_{i=1}^{m}\sum_{j=2}^{n_i}\left(\hat{\alpha} + \hat{\delta} y_{ij-1} - e^{\hat{\alpha}+\hat{\delta} y_{ij-1}} t_{ij}\right)\Bigg\}. \end{aligned} \tag{68}$$

The AIC of the Gamma-Exponential model becomes

$$\begin{aligned} AIC = & 2k - 2\Bigg\{ \sum_{i=1}^{m}\left(\frac{y_{i1}(-1/\exp(\boldsymbol{X}_i^{'}\hat{\boldsymbol{\beta}})) - \boldsymbol{X}_i^{'}\hat{\boldsymbol{\beta}}}{1/\hat{v}_{ij}} + \hat{v}_{ij}\log(\hat{v}_{ij}) + (\hat{v}_{ij}-1)\log(y_{i1}) - \log(\Gamma(\hat{v}_{ij}))\right) \\ & + \sum_{i=1}^{m}\sum_{j=2}^{n_i}\Bigg(\frac{y_{ij}(-1/\exp(\boldsymbol{X}_i^{'}\hat{\boldsymbol{\beta}} + \hat{\gamma} t_{ij} + \hat{\varphi} y_{ij-1})) - (\boldsymbol{X}_i^{'}\hat{\boldsymbol{\beta}} + \hat{\gamma} t_{ij} + \hat{\varphi} y_{ij-1})}{1/\hat{v}_{ij}} \\ & + \hat{v}_{ij}\log(\hat{v}_{ij}) + (\hat{v}_{ij}-1)\log(y_{i1}) - \log(\Gamma(\hat{v}_{ij}))\Bigg) \\ & + \sum_{i=1}^{m}\sum_{j=2}^{n_i}\left(\hat{\alpha} + \hat{\delta} y_{ij-1} - e^{\hat{\alpha}+\hat{\delta} y_{ij-1}} t_{ij}\right)\Bigg\}. \end{aligned} \tag{69}$$

The AICc is the AIC with a correction for finite sample sizes and is defined as

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}, \tag{70}$$

where k is the number of parameters and n is the sample size. The AIC performs poorly with small sample sizes and AICc converges to AIC as n increases, so AICc is recommended (Burnham & Anderson, 2002).

Bayesian information criterion (BIC) is given by

$$BIC = -2ln(L) + k\log(n), \tag{71}$$

where k is the number of parameters and n is the sample size. The penalty associated with BIC is more severe than that of AIC because the sample size is included in the function. Like AIC, a model with the lowest BIC value is preferred. The use of BIC is not recommended due to a high risk of selecting a model that is too simple or parsimonious (Fitzmaurice et al., 2004). However, in a simulation study, the BIC outperformed the AIC as sample size increased (McQuarrie, Shumway, & Tsai, 1997). Therefore, BIC is included since this study uses the log-likelihood function in the process of finding parameter estimates.

Data Simulation

This study is an extension of the studies of Bronsert (2009) and Lin (2011). To maintain consistency, all of the parameter values in Table 3 and simulation conditions in Table 4 were adopted from Lin's study. The joint model developed by the two researchers can be used only when outcomes follow a normal distribution and time follows an exponential distribution. Since the purpose of this study was to

develop joint models for members of the exponential family of distributions for outcomes with an exponential distribution for time, three outcome distributions were used in this study as examples. These are the Bernoulli for binary outcomes, the Poisson distribution for counts, and the Gamma distribution for waiting time or survival time.

In the two previous studies, Monte Carlo simulations were utilized using SAS/IML. Lin used the Henze-Zirkler multivariate normality test statistic computed in the PROC MODEL procedure. During this study, a program error was found in that procedure in the software. Consequently, a SAS problem note was posted on their website (http://support.sas.com/kb/51/281.html) on October 11, 2013. The SAS company said that the Henze-Zirkler test should not be used to test for multivariate normality. Therefore, this study retested the simulation results of the previous study done by Lin, by using R instead of SAS.

The basic structure of simulated data has two categorical variables with three levels each and two continuous variables. The values of the dependent variables are generated from the Bernoulli, the Poisson, and the Gamma distributions respectively. The first outcome is generated from each distribution, then the next outcome is calculated based on the relationship between the previous outcome and the previous time to predict the average outcome with fixed parameter values in Table 3. All parameter values are assumed to be equal across subjects for simplicity of model form in simulation studies. The scale parameter values (v) in Table 3 are used for the Gamma distribution only.

Table 3

Parameter Values for Simulations

v	β_0	β_1	β_2	β_3	β_4	β_5	β_6	φ	γ	α	δ
1	0.4	0.2	0.3	0.1	0.3	0.4	0.9	0.8	0.1	2	0.01
1	0.4	0.2	0.3	0.1	0.3	0.4	0.9	0.8	0.1	1	0.02
1	0.4	0.2	0.3	0.1	0.3	0.4	0.9	0.8	0.1	2	0.01
1.2	0.4	0.2	0.3	0.1	0.3	0.4	0.9	0	0.1	1	0.02
1.2	0.4	0.2	0.3	0.1	0.3	0.4	0.9	0	0.1	2	0.01
1.2	0.4	0.2	0.3	0.1	0.3	0.4	0.9	0.8	0.1	1	0.02

To test multivariate normality of the estimators of the joint models, a diversity of sample sizes and design structures (balanced or unbalanced) were included in the simulations. As described in Table 4, five sample sizes are combined with four types of design structures with a different number of observations. Because sample sizes range from 18 to 180, simulation studies are believed to be enough to see if the multivariate normality test shows a trend as sample size increases. In addition, a different number of observations were included in each sample size to check if there is a certain pattern as the number of observations increases.

Some researchers used 1,000 replications (Lipsitz et al., 2002; Qiu et al., 2013), and some used 500 replications (Liang et al., 2009). The number of replications was 5,000 times in Lin's study. In this study, each simulation design was run 1,000 times. Each outcome distribution had 120 simulation designs.

Therefore, the total number of simulation designs was 480 with the six parameter schemes, five sample sizes, and four different numbers of observations with four distributions.

Table 4

Simulation Designs

Scheme Number	Sample Size	Number of Observations	Design Structure	Total number of Observations
1	18	10	Balanced	180
2		5 & 3	Unbalanced	72
3		10 & 5	Unbalanced	135
4		20 & 6	Unbalanced	234
5	36	10	Balanced	360
6		5 & 3	Unbalanced	144
7		10 & 5	Unbalanced	180
8		20 & 6	Unbalanced	288
9	54	10	Balanced	540
10		5 & 3	Unbalanced	216
11		10 & 5	Unbalanced	405
12		20 & 6	Unbalanced	702
13	90	10	Balanced	900
14		5 & 3	Unbalanced	360
15		10 & 5	Unbalanced	675
16		20 & 6	Unbalanced	1170
17	180	10	Balanced	1800
18		5 & 3	Unbalanced	720
19		10 & 5	Unbalanced	1350
20		20 & 6	Unbalanced	2340

As shown in Table 4, each sample size with a different number of observations was simulated, and estimators were calculated. Those two procedures were replicated 1,000 times, and multivariate normality was tested with 1,000 sets of the estimators by using the Henze-Zirkler test. For example, when the sample size is 54 and the number of observations is 5 & 3, 27 subjects have 5 outcomes each and 27 subjects have 3 outcomes each, which makes the total number of observations 216.

This study used the Henze-Zirkler multivariate normality test for checking the property of the estimators of the joint models by using R instead of SAS due to the error. The built-in function name is HZ.test in the MVN package. After showing the test results, AIC, AICc, and BIC scores were calculated for model selection purposes. Finally, the computing program package using R was applied to a real dataset, bladder cancer data, and the output is presented.

CHAPTER IV

RESULTS

Two main purposes of this study were (1) to build joint models with longitudinal outcomes and informative time and verify if the asymptotic normality of the maximum likelihood estimators holds for the joint models and (2) to establish a computing program package that can handle the joint models by using R. A partial result of the analysis of the bladder cancer data using the package is presented at the end of this chapter to demonstrate the performance of the package. To verify the asymptotic normality of the maximum likelihood estimators of the joint models, four outcome distributions were selected as examples, and the simulation results are presented in this chapter. The chosen outcome distributions are the Gaussian, Bernoulli, Poisson, and Gamma distributions. The simulation results of the Gaussian distribution are presented since an error was found in the Henze-Zirkler multivariate normality test in SAS used by the previous researcher.

A general description of the simulation procedures is as follows. Firstly, a design matrix is generated with two continuous and two categorical variables with three levels each. Secondly, a dataset is created based on the fixed parameter values shown in Table 3, in addition to the relations among previous and current outcomes and previous time. Thirdly, the maximum likelihood estimators are computed using a nonlinear parameter optimization function called maxLik in R. Fourthly, the

parameter estimates are standardized by the standard error from the Hessian matrix. The above four steps are repeated 1,000 times. The Henze-Zirkler test is then used to test the multivariate normality of the 10,000 sets of parameter estimates. The test was conducted separately for the parameter estimates of the outcome and time processes. Simulation designs are based on five sample sizes, four different observations, and six parameter schemes. Each outcome distribution has 120 (5 * 4 * 6) simulation conditions, which makes 480 simulation designs in total for all four outcome distributions.

While simulating data, the logit link function for the Bernoulli, log link function for the Poisson and Gamma distributions, and identity link function for the Gaussian are used. The mean function of the Bernoulli becomes $\mu = \frac{\exp(\eta)}{1-\exp(\eta)}$, and the mean function of the Poisson and Gamma is $\mu = \exp(\eta)$. In addition, to generate time points, another exponential function is used. Due to the two exponential functions, oftentimes outcomes and time points quickly became impractical. To avoid generating unrealistic outcomes and time points, the parameter values in Table 3 on page 46 were multiplied by 1/10 to reduce the magnitude.

For example, when $\boldsymbol{X}_i'\boldsymbol{B} = 3$ for the ith subject, $\mu_i = \exp(\boldsymbol{X}_i'\boldsymbol{B})$ becomes 20.08. This value is passed to the second exponential function to generate a time point, which is $\exp(20.08) = 528,491,311$ without including any other terms in the computation. Moreover, this value is plugged into a function to generate a random number from the Poisson distribution. As a result, an outcome of 528,538,500 is

generated, which is very unlikely to have in practice. This problem was resolved by reducing the magnitude of the parameter values.

The computing package using R presented in Appendix D was designed to handle outcome distributions, such as the Gaussian, Bernoulli, Poisson, and Gamma, with informative time, which follows an exponential distribution. The package uses the identity link function for the Gaussian, logit link function for the Bernoulli, and log link function for the Poisson and Gamma distributions, by default. But for the Gamma distribution, the inverse and identity link functions can be used as well, in addition to the log link function. Once all the information needed to utilize the package is provided, the package computes the estimators, AIC, AICc, BIC, and likelihood ratio test statistic with a corresponding p-value for a model. For demonstration purposes, the package was applied to the bladder cancer data; then, the outputs generated by the package are shown in Appendix A and B.

Simulation Results

The results of the multivariate normality tests for all simulation conditions from the four selected outcome distributions are presented.

Gaussian-Exponential Model

As can be seen in Figure 1 and 4 and Table 5, multivariate normality is not stable when the sample size is 18 for the outcome process. However, as sample size increases, most of the cases show multivariate normality. In these figures, filled circles represent the unbalanced simulation design, and unfilled circles are for the

balanced simulation design. Figure 2 and 4 show the multivariate normality test results at the significance level of $\alpha = 0.05$ for each parameter scheme. When a p-value of the test is less than 0.05, it is categorized into "Non-normal," but otherwise it is "Normal." By looking at the plots, the effects of the design structures are not clear. The Gaussian-Exponential model has ten parameters for the outcome process and two for the time process.

For the time process, the multivariate normality test results can be seen in Figure 5 and 8 and Table 6. Most of the test results show multivariate normality even with the sample size of 18. Also, the design structures do not seem to affect the multivariate normality of the estimators. The estimators of the time process show a pattern in achieving normality slightly faster than the outcome process.

The previous researcher stated that "there was a tendency of achieving multivariate normality as the number of subjects exceeds 54" for the outcome and time processes (Lin, 2011, p. 60 & 65). However, in this study, most of the simulation results for the Gaussian model obtained normality as sample size is greater than 18 for both the outcome and time processes. It is assumed that the Henze-Zirkler multivariate normality test in the PROC MODEL procedure in SAS used by the previous researcher did not compute the test statistic correctly due to the error in the software.

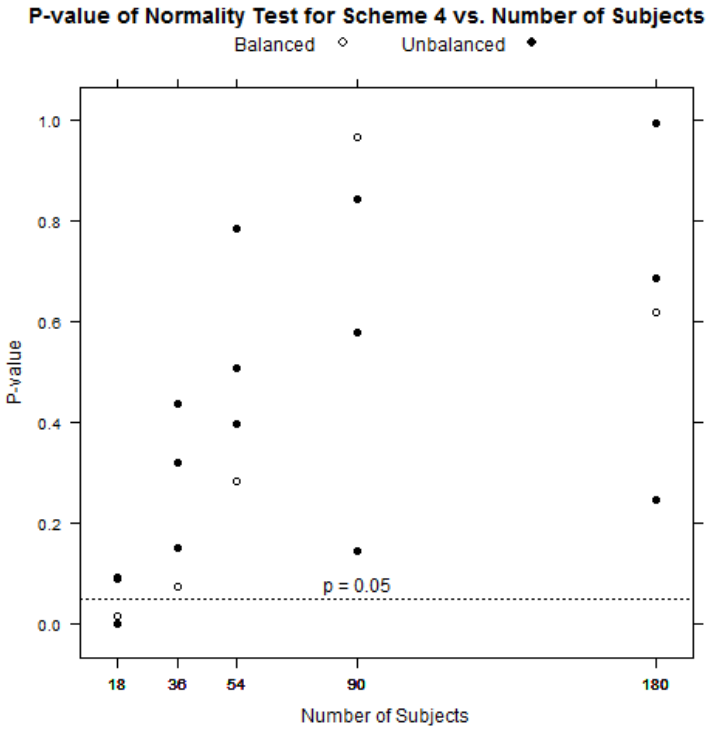

Figure 1. Outcome Process of the Gaussian-Exponential Model (Subjects)

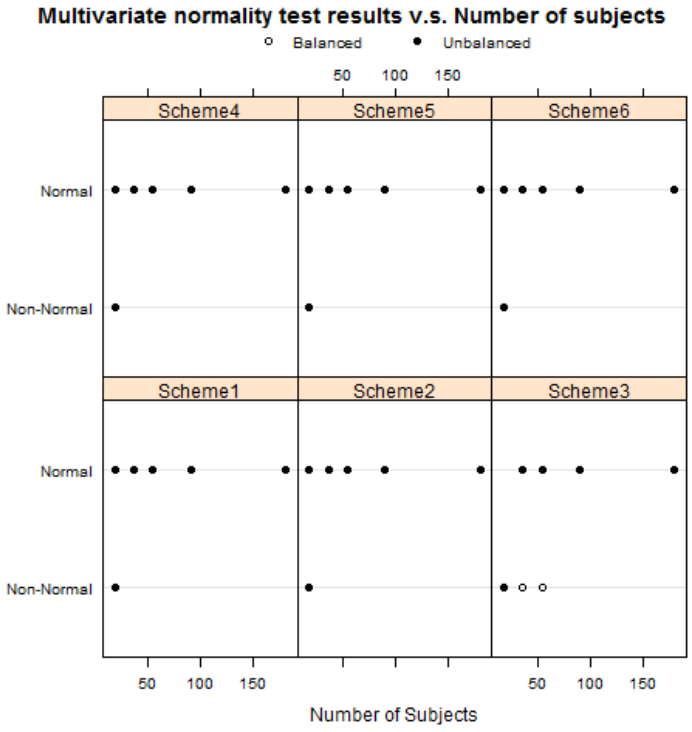

Figure 2. Outcome Process of the Gaussian-Exponential Model (Subjects)

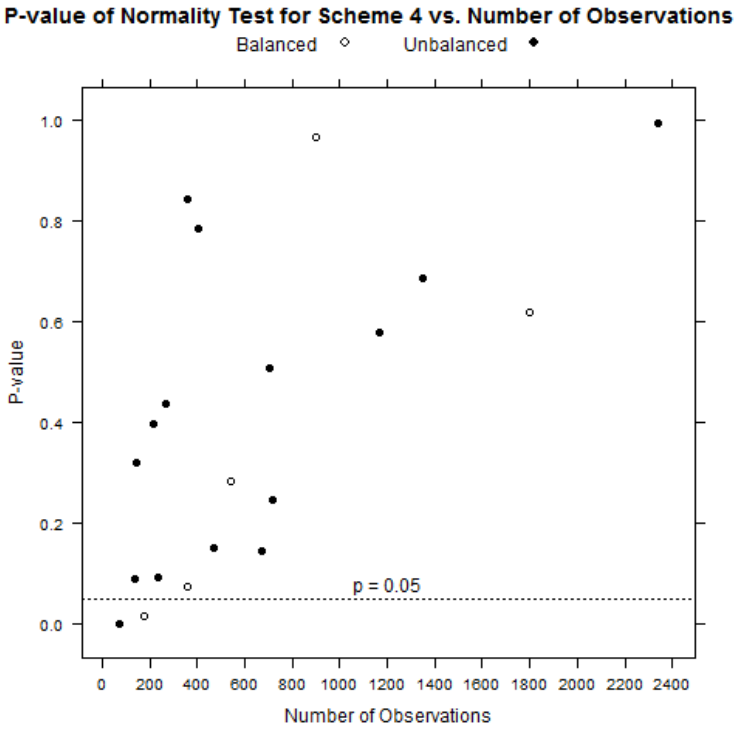

Figure 3. Outcome Process of the Gaussian-Exponential Model (Observations)

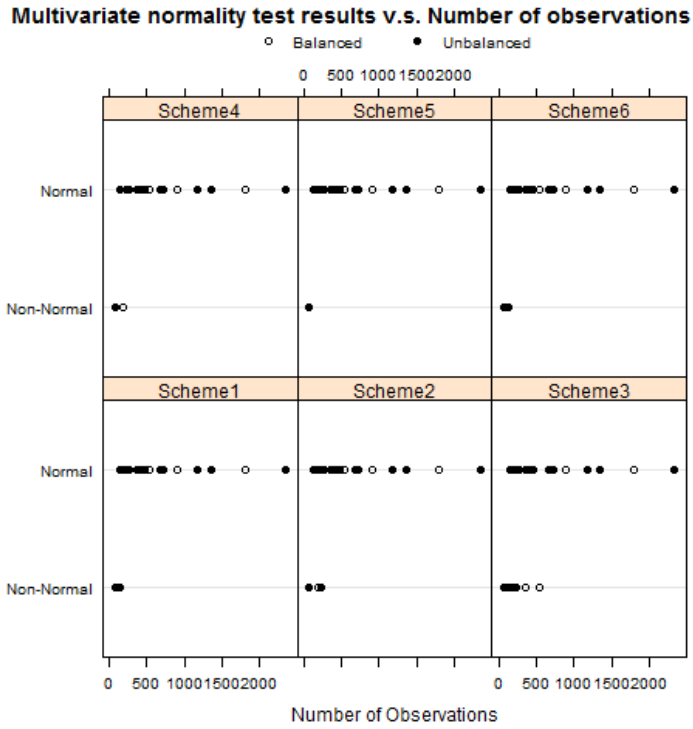

Figure 4. Outcome Process of the Gaussian-Exponential Model (Observations)

Table 5

Outcome Process of the Gaussian-Exponential Model: p-values of the Multivariate Normality Test for different parameter schemes and sample sizes

			Parameter Scheme					
Sample Scheme	Sample Size	# of Obs	1 p-value	2 p-value	3 p-value	4 p-value	5 p-value	6 p-value
1	18	180	0.29971	0.03930	0.00001	0.01529	0.53645	0.17374
2		72	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001
3		135	0.01769	0.14771	0.00001	0.09001	0.22121	0.02722
4		234	0.15295	0.01213	0.00001	0.09044	0.61176	0.19633
5	36	360	0.32246	0.10253	0.00001	0.07367	0.12644	0.27061
6		144	0.34162	0.36361	0.24073	0.31887	0.20862	0.51200
7		180	0.37339	0.09288	0.89520	0.43668	0.18142	0.72612
8		288	0.11841	0.20394	0.78541	0.15052	0.94910	0.18182
9	54	540	0.51538	0.93902	0.01182	0.28265	0.60469	0.74613
10		216	0.36149	0.51311	0.45283	0.39490	0.15466	0.91270
11		405	0.65701	0.84358	0.53248	0.78497	0.61603	0.89729
12		702	0.33136	0.89321	0.79814	0.50627	0.66203	0.93771
13	90	900	0.20652	0.45802	0.40458	0.96591	0.17377	0.71388
14		360	0.29366	0.75052	0.35323	0.84260	0.54679	0.89042
15		675	0.41352	0.74227	0.69382	0.14320	0.36336	0.86635
16		1170	0.63985	0.06333	0.67474	0.57609	0.23243	0.68908
17	180	1800	0.06868	0.37988	0.53598	0.61698	0.46444	0.82726
18		720	0.24276	0.80939	0.76237	0.24585	0.83869	0.55219
19		1350	0.41318	0.32841	0.91741	0.68461	0.51855	0.47480
20		2340	0.58655	0.23763	0.78124	0.99267	0.69329	0.98776

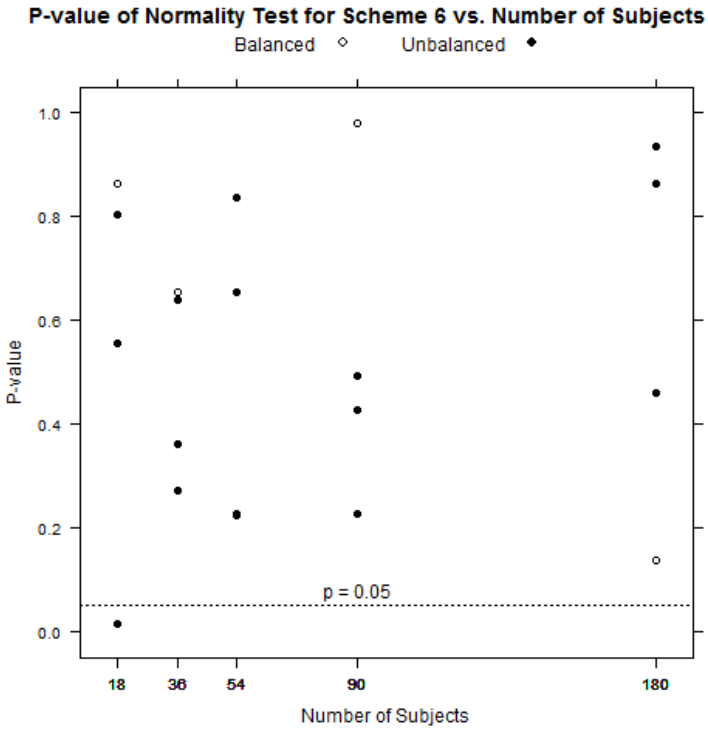

Figure 5. Time Process of the Gaussian-Exponential Model (Subjects)

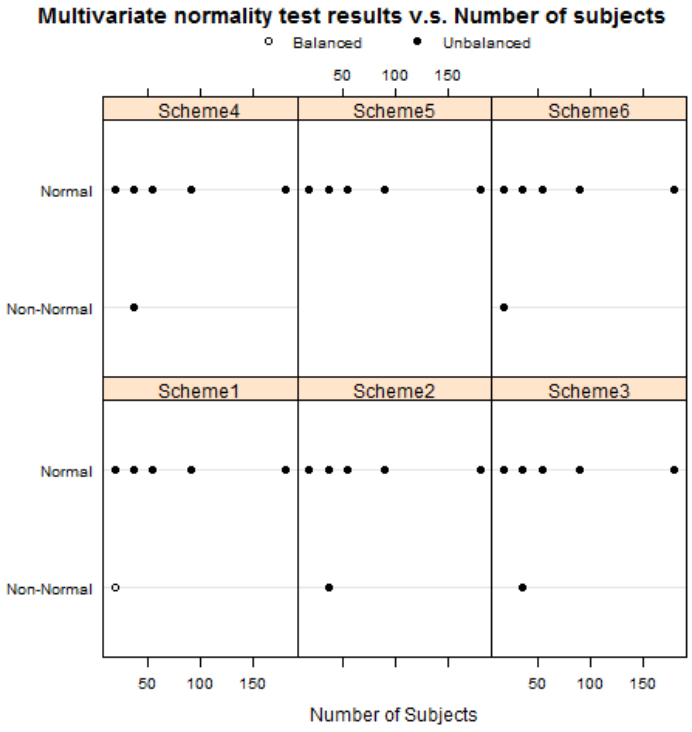

Figure 6. Time Process of the Gaussian-Exponential Model (Subjects)

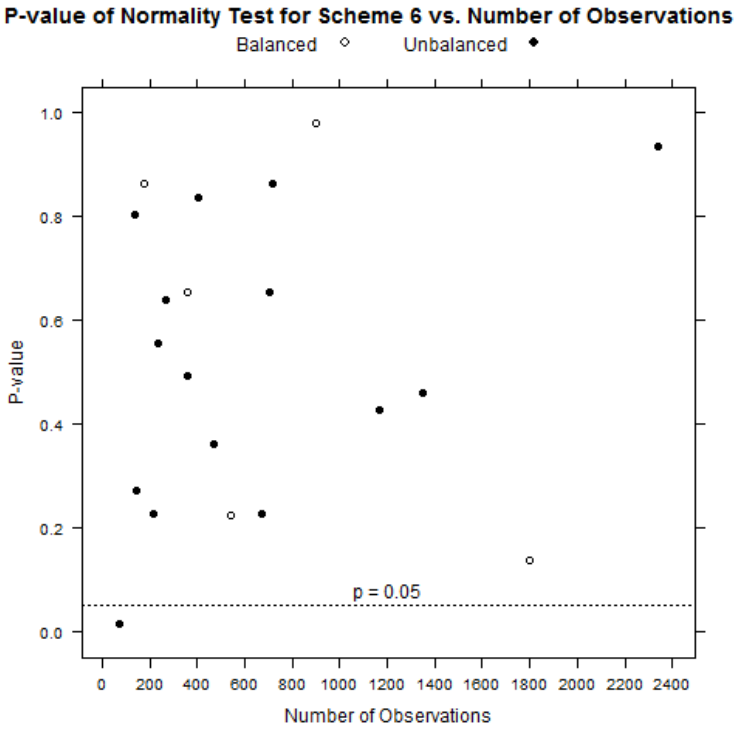

Figure 7. Time Process of the Gaussian-Exponential Model (Observations)

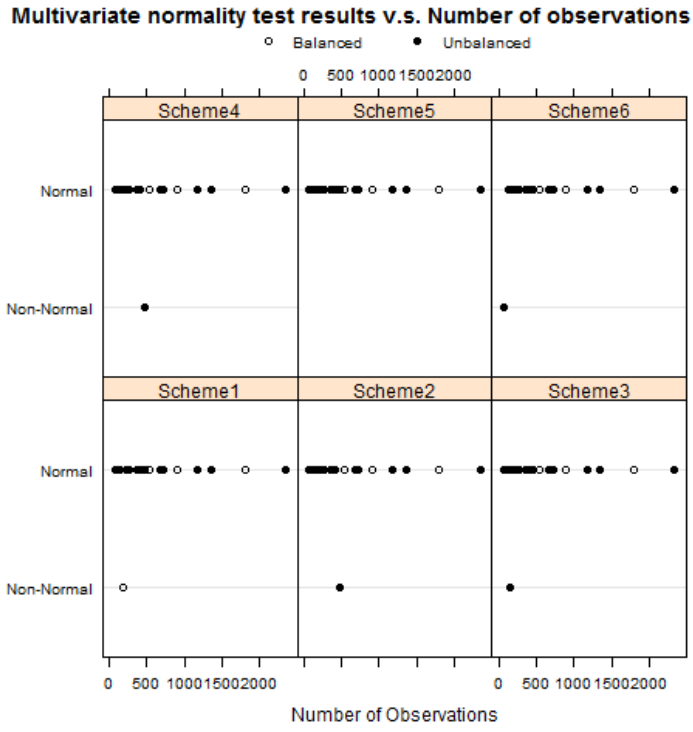

Figure 8. Time Process of the Gaussian-Exponential Model (Observations)

Table 6

Time Process of the Gaussian-Exponential Model: p-values of the Multivariate Normality Test for different parameter schemes and sample sizes

			Parameter Scheme					
Sample Scheme	Sample Size	# of Obs	1 p-value	2 p-value	3 p-value	4 p-value	5 p-value	6 p-value
1	18	180	0.00536	0.62492	0.44094	0.27630	0.06168	0.86184
2		72	0.09875	0.39235	0.26561	0.29329	0.23040	0.01546
3		135	0.79393	0.43889	0.61088	0.22100	0.40391	0.80296
4		234	0.96929	0.06662	0.29046	0.25676	0.11720	0.55487
5	36	360	0.37019	0.59933	0.38449	0.19781	0.05106	0.65543
6		144	0.09881	0.08263	0.04383	0.64623	0.08940	0.27226
7		180	0.61501	0.65657	0.51750	0.96478	0.07434	0.63833
8		288	0.42738	0.01507	0.32936	0.01262	0.85221	0.36209
9	54	540	0.77590	0.94485	0.70976	0.10574	0.53873	0.22542
10		216	0.25769	0.62343	0.52139	0.36188	0.80927	0.22598
11		405	0.52672	0.41992	0.23673	0.41124	0.62807	0.83770
12		702	0.39817	0.19285	0.45714	0.20934	0.56195	0.65525
13	90	900	0.82082	0.51221	0.10389	0.95954	0.71905	0.97961
14		360	0.37800	0.33752	0.23034	0.88490	0.70013	0.49251
15		675	0.59168	0.11903	0.63875	0.34989	0.77778	0.22766
16		1170	0.90701	0.47944	0.65285	0.09478	0.70765	0.42658
17	180	1800	0.35810	0.95003	0.20777	0.29659	0.81707	0.13618
18		720	0.25920	0.75638	0.38214	0.86188	0.42393	0.86280
19		1350	0.24551	0.70440	0.27935	0.77896	0.26871	0.46085
20		2340	0.78797	0.57098	0.44826	0.43375	0.66599	0.93396

Bernoulli-Exponential Model

The simulation results of the multivariate normality tests for the outcome process are shown in Figure 9 and 12 and Table 7. As can be seen, some simulation designs do not obtain multivariate normality when the sample size is 18. However, when sample size goes beyond 18, it has a tendency of obtaining multivariate normality in most of the cases. The test results for the time process are shown in Figure 13 and 16 and Table 8. Like the Gaussian and Poisson models, design structures do not seem to affect the test results. Compared to the Gaussian model, the Bernoulli model is slightly slower to obtain multivariate normality in both outcome and time processes. The Bernoulli model shows, however, the same pattern for obtaining multivariate normality as sample size increases.

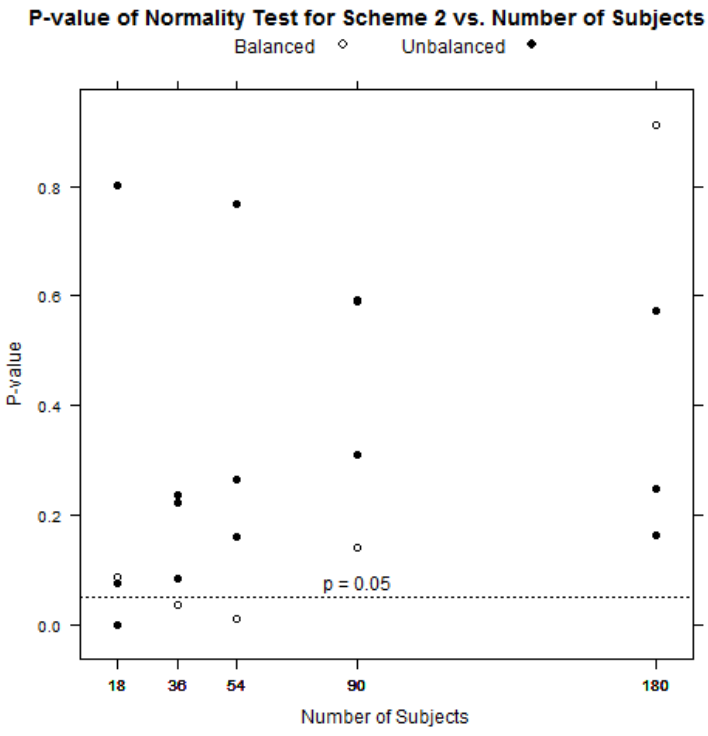

Figure 9. Outcome Process of the Bernoulli-Exponential Model (Subjects)

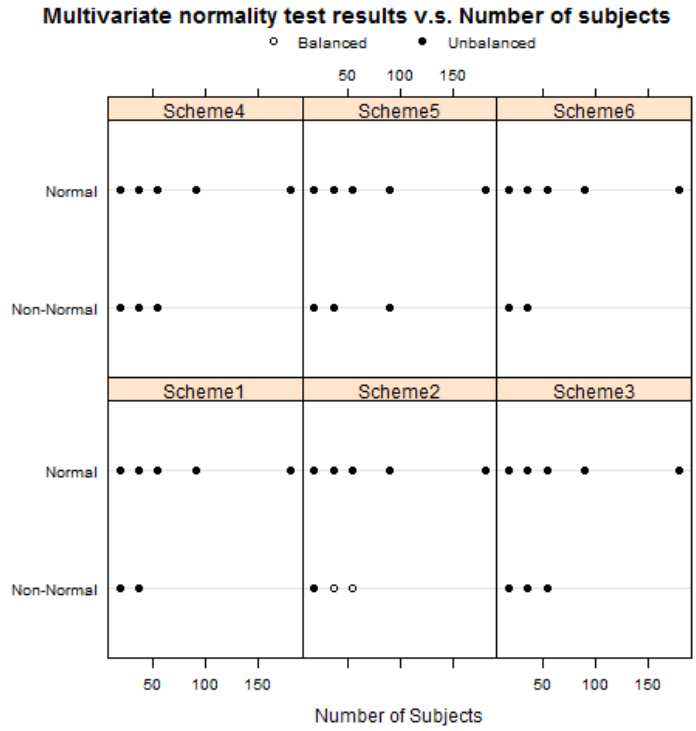

Figure 10. Outcome Process of the Bernoulli-Exponential Model (Subjects)

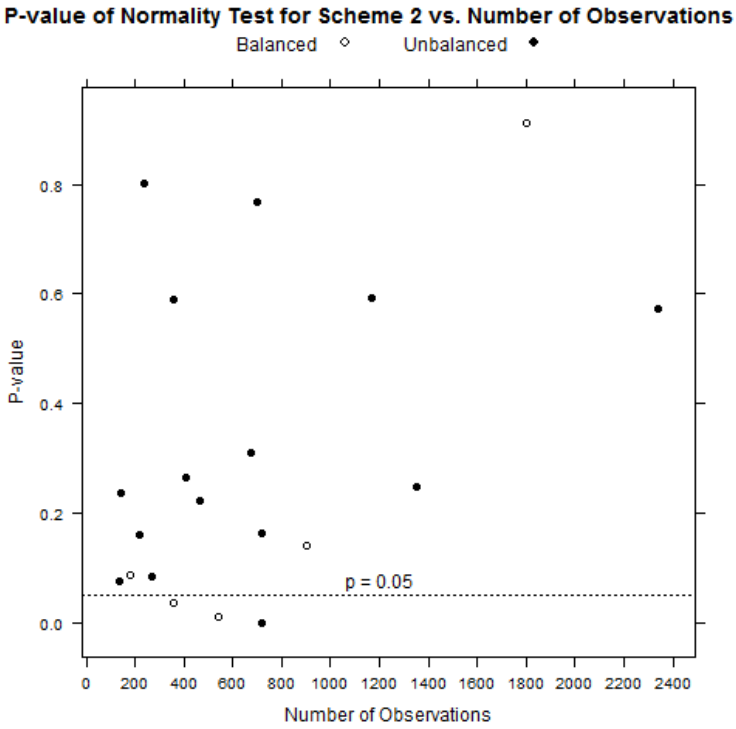

Figure 11. Outcome Process of the Bernoulli-Exponential Model (Observations)

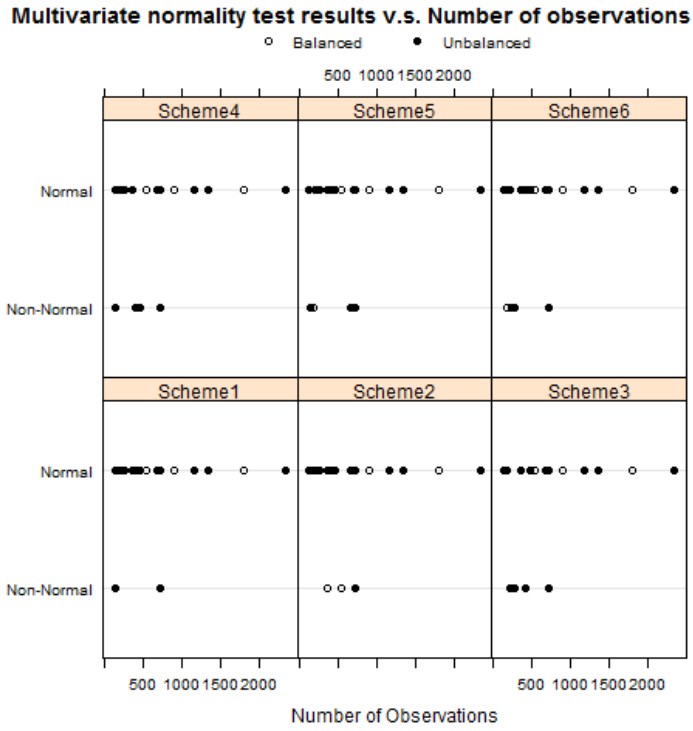

Figure 12. Outcome Process of the Bernoulli-Exponential Model (Observations)

Table 7

Outcome Process of the Bernoulli-Exponential Model: p-values of the Multivariate Normality Test for different parameter schemes and sample sizes

			Parameter Scheme					
Sample Scheme	Sample Size	# of Obs	1 p-value	2 p-value	3 p-value	4 p-value	5 p-value	6 p-value
1	18	180	0.05675	0.08736	0.55261	0.40705	0.02049	0.02893
2		72	0.00170	0.00001	0.00443	0.00005	0.00001	0.00001
3		135	0.08779	0.07709	0.36283	0.00722	0.75259	0.44204
4		234	0.23295	0.80118	0.01941	0.44327	0.11254	0.02502
5	36	360	0.75219	0.03635	0.10186	0.51965	0.69102	0.57244
6		144	0.03161	0.23727	0.05807	0.30133	0.03666	0.55714
7		180	0.17731	0.08543	0.00142	0.55424	0.10476	0.00169
8		288	0.79914	0.22383	0.81402	0.02987	0.74556	0.12857
9	54	540	0.92032	0.01005	0.67330	0.08488	0.28485	0.38808
10		216	0.30015	0.16119	0.04276	0.63300	0.93464	0.21757
11		405	0.44539	0.26563	0.01529	0.01394	0.12901	0.23853
12		702	0.33028	0.76880	0.05700	0.53018	0.42188	0.33184
13	90	900	0.15870	0.14108	0.12986	0.88998	0.10659	0.65001
14		360	0.48833	0.58988	0.55276	0.07854	0.12428	0.73702
15		675	0.12926	0.31151	0.05007	0.72414	0.03409	0.49010
16		1170	0.89959	0.59373	0.72635	0.80065	0.56804	0.53755
17	180	1800	0.18810	0.91157	0.57165	0.72655	0.94775	0.35183
18		720	0.07356	0.16238	0.78948	0.49637	0.87328	0.65930
19		1350	0.17788	0.24911	0.25517	0.64486	0.16041	0.13796
20		2340	0.23908	0.57153	0.28496	0.28281	0.13885	0.15042

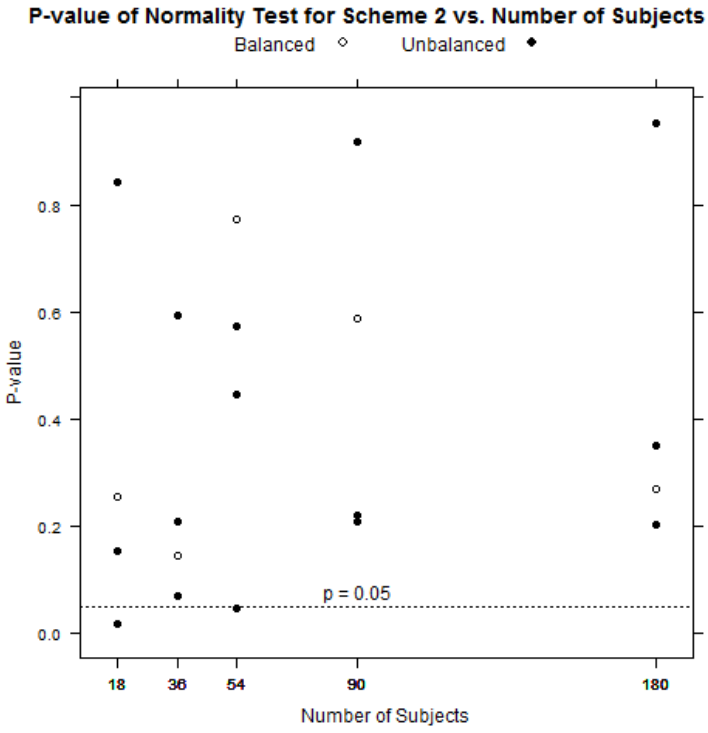

Figure 13. Time Process of the Bernoulli-Exponential Model (Subjects)

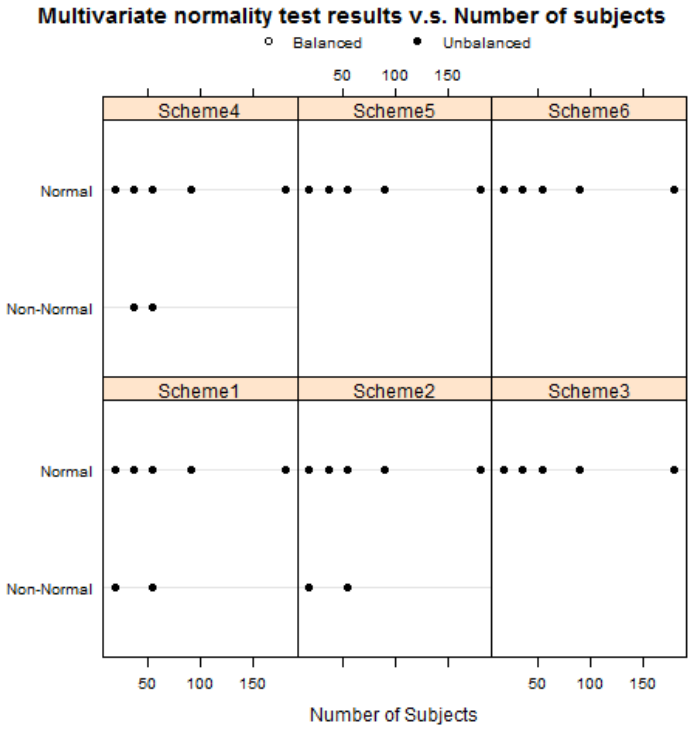

Figure 14. Time Process of the Bernoulli-Exponential Model (Subjects)

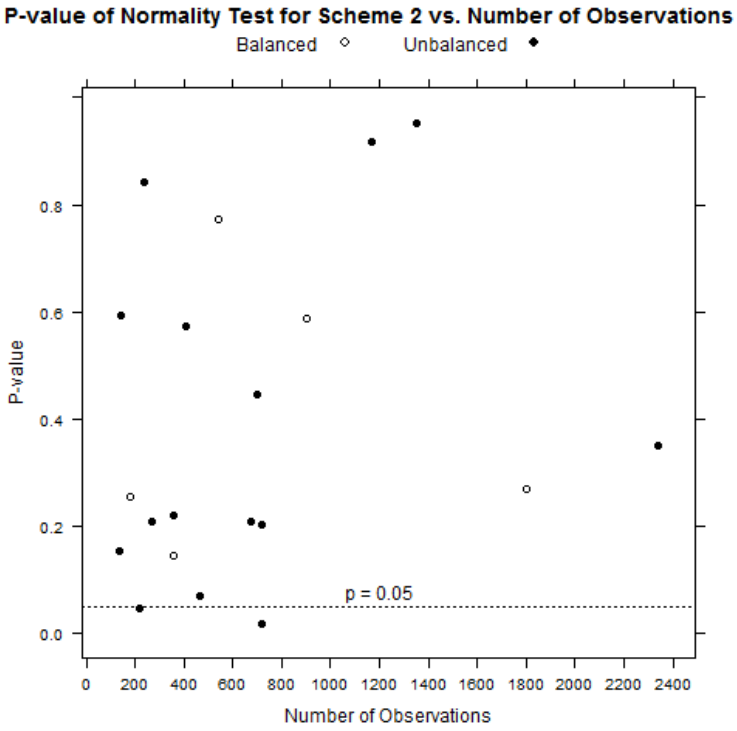

Figure 15. Time Process of the Bernoulli-Exponential Model (Observations)

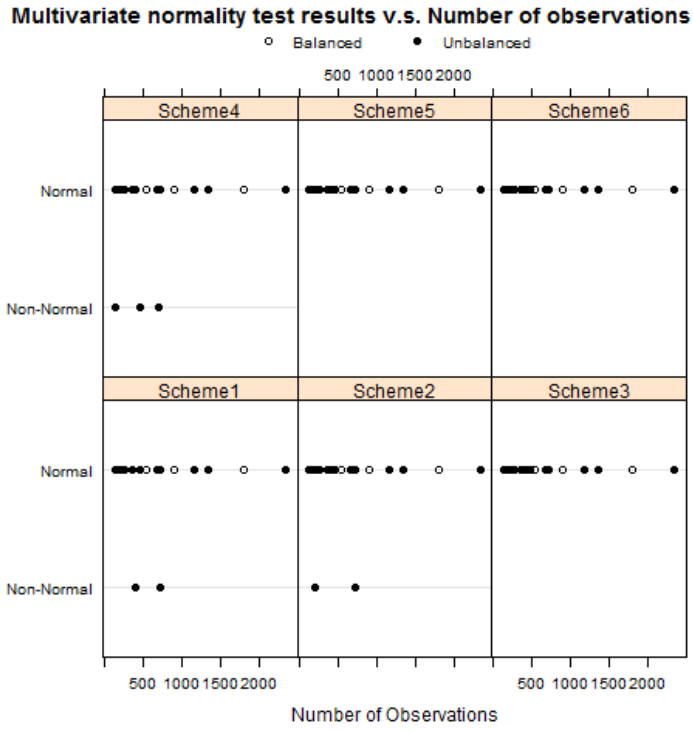

Figure 16. Time Process of the Bernoulli-Exponential Model (Observations)

Table 8

Time Process of the Bernoulli-Exponential Model: p-values of the Multivariate Normality Test for different parameter schemes and sample sizes

			Parameter Scheme					
Sample Scheme	Sample Size	# of Obs	1 p-value	2 p-value	3 p-value	4 p-value	5 p-value	6 p-value
1	18	180	0.63656	0.25638	0.83782	0.88686	0.80580	0.39971
2		72	0.02153	0.01852	0.09288	0.74398	0.14944	0.20743
3		135	0.11022	0.15489	0.08093	0.13597	0.56070	0.46692
4		234	0.21111	0.84092	0.54910	0.79137	0.54731	0.94560
5	36	360	0.23178	0.14502	0.60016	0.28552	0.65038	0.27444
6		144	0.82361	0.59250	0.49520	0.00784	0.35823	0.05031
7		180	0.51769	0.20966	0.16625	0.24893	0.27654	0.60556
8		288	0.48287	0.07160	0.79540	0.00947	0.77695	0.07273
9	54	540	0.65554	0.77282	0.66345	0.81803	0.09565	0.49507
10		216	0.23119	0.04746	0.16934	0.10530	0.33313	0.17070
11		405	0.04573	0.57223	0.53258	0.63360	0.52933	0.72817
12		702	0.85915	0.44562	0.57071	0.01737	0.84832	0.25854
13	90	900	0.35063	0.58724	0.38891	0.75017	0.99741	0.79655
14		360	0.82732	0.22178	0.37066	0.77456	0.18029	0.30316
15		675	0.87315	0.20865	0.85218	0.72115	0.38335	0.62810
16		1170	0.13283	0.91835	0.68403	0.05189	0.61365	0.77964
17	180	1800	0.60475	0.27057	0.91102	0.11225	0.40229	0.48817
18		720	0.52796	0.20331	0.86455	0.42649	0.34244	0.05321
19		1350	0.38633	0.95209	0.05218	0.74395	0.61988	0.34718
20		2340	0.55747	0.35095	0.18420	0.92866	0.66420	0.17732

Poisson-Exponential Model

The results of the multivariate normality tests for the outome process are shown in Figure 17 and 20 and Table 9. Figure 21 and 24 and Table 10 are presented for the time process. When the sample size is 18, some of the p-values are smaller than 0.05. But when sample size goes over 18, most cases obtain normality in the outcome and time processes. The effects of design structures are not clear. The outcome and time process of the Poisson model show a similar pattern to the Bernoulli model in obtaining multivariate normality.

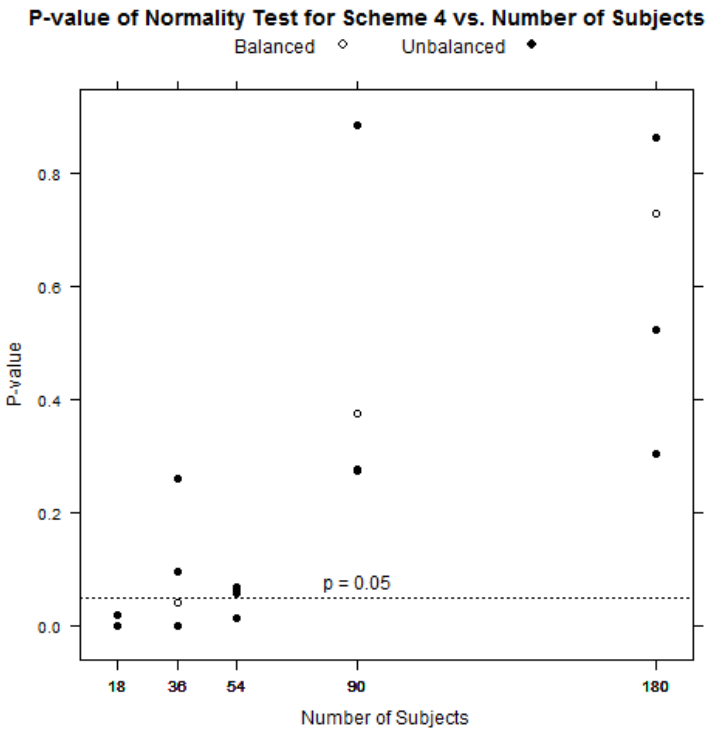

Figure 17. Outcome Process of the Poisson-Exponential Model (Subjects)

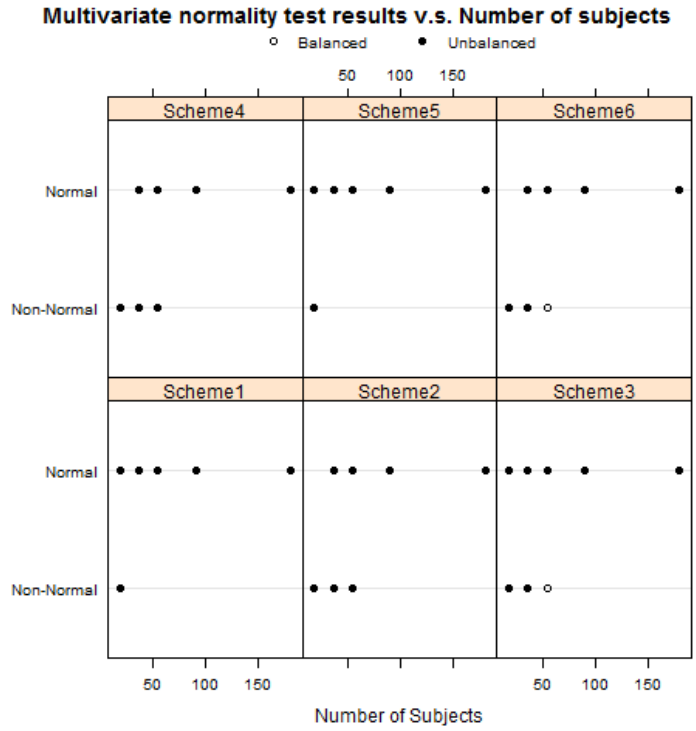

Figure 18. Outcome Process of the Poisson-Exponential Model (Subjects)

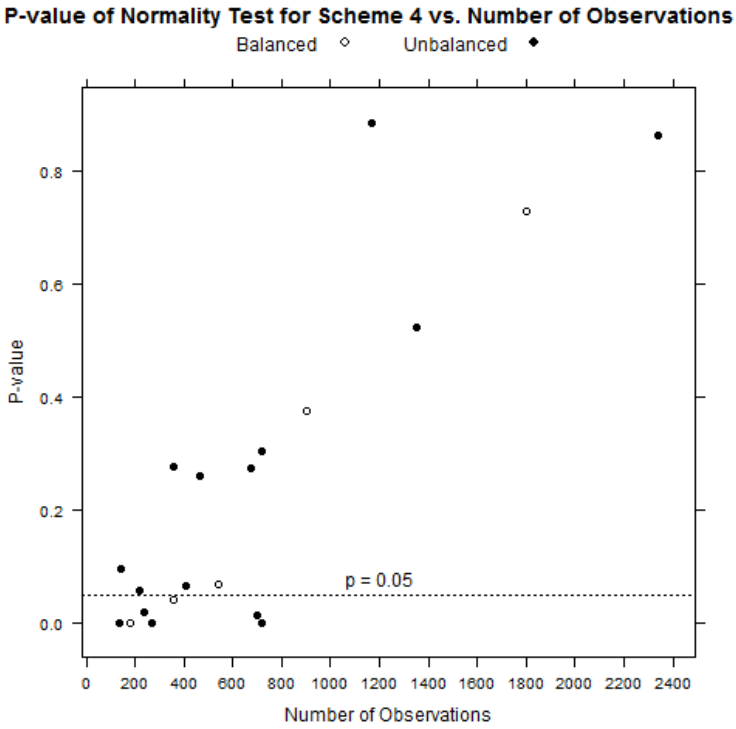

Figure 19. Outcome Process of the Poisson-Exponential Model (Observations)

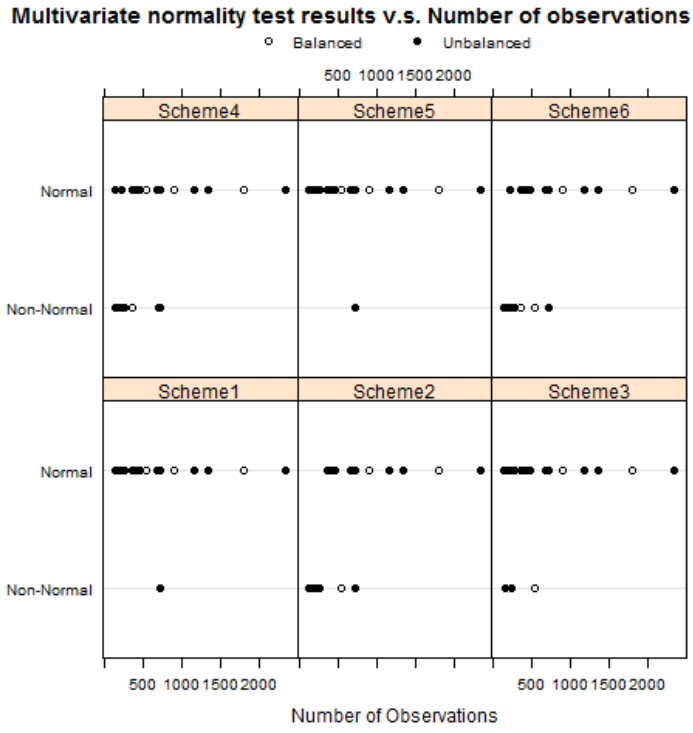

Figure 20. Outcome Process of the Poisson-Exponential Model (Observations)

Table 9

Outcome Process of the Poisson-Exponential Model: p-values of the Multivariate Normality Test for different parameter schemes and sample sizes

			Parameter Scheme					
Sample Scheme	Sample Size	# of Obs	1 p-value	2 p-value	3 p-value	4 p-value	5 p-value	6 p-value
1	18	180	0.10057	0.02895	0.18157	0.00001	0.16507	0.00514
2		72	0.00021	0.00001	0.22112	0.00001	0.02607	0.00001
3		135	0.36130	0.00001	0.16868	0.00001	0.07339	0.00001
4		234	0.67790	0.00163	0.00080	0.01920	0.44829	0.00001
5	36	360	0.08865	0.38247	0.64722	0.04076	0.27216	0.00293
6		144	0.11726	0.00986	0.00362	0.09490	0.05695	0.00001
7		180	0.15288	0.02433	0.23448	0.00053	0.23166	0.00086
8		288	0.28679	0.26373	0.09063	0.26016	0.23803	0.05345
9	54	540	0.31881	0.00039	0.01431	0.06875	0.51118	0.02247
10		216	0.41920	0.01423	0.21104	0.05748	0.39067	0.06158
11		405	0.44330	0.15155	0.82205	0.06632	0.22130	0.09995
12		702	0.28225	0.41316	0.47929	0.01236	0.16226	0.16811
13	90	900	0.27823	0.11821	0.11337	0.37641	0.45065	0.24115
14		360	0.28681	0.18526	0.12320	0.27628	0.53792	0.27771
15		675	0.23369	0.52214	0.07221	0.27438	0.18044	0.77676
16		1170	0.45542	0.61720	0.74317	0.88513	0.61144	0.42771
17	180	1800	0.56882	0.41468	0.49275	0.72791	0.62766	0.56920
18		720	0.20004	0.68145	0.42954	0.30436	0.31401	0.42366
19		1350	0.60901	0.10406	0.21374	0.52314	0.43834	0.25464
20		2340	0.90283	0.18755	0.19736	0.86215	0.52802	0.42036

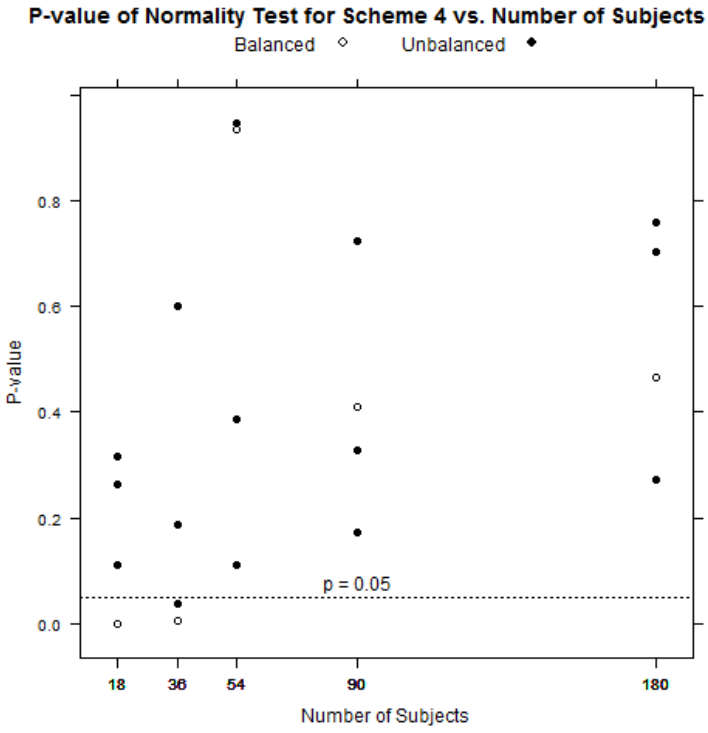

Figure 21. Time Process of the Poisson-Exponential Model (Subjects)

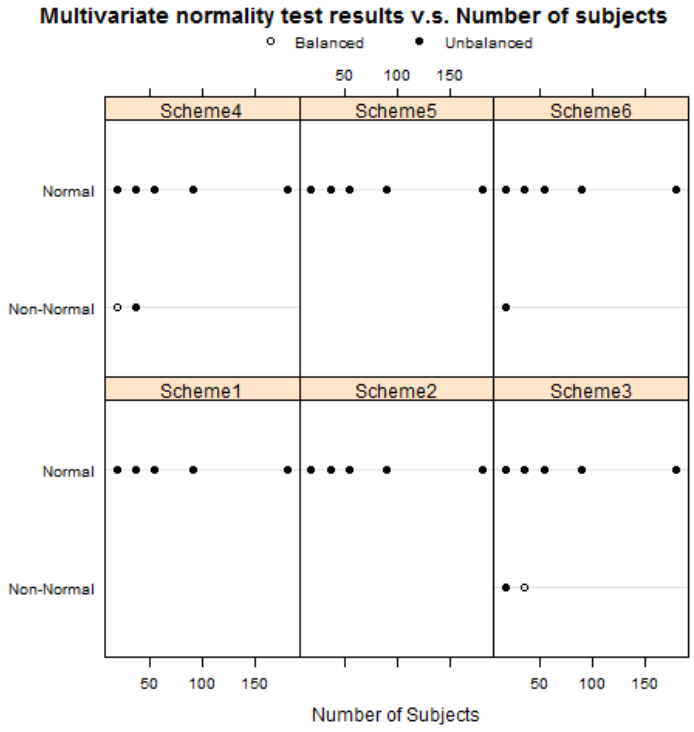

Figure 22. Time Process of the Poisson-Exponential Model (Subjects)

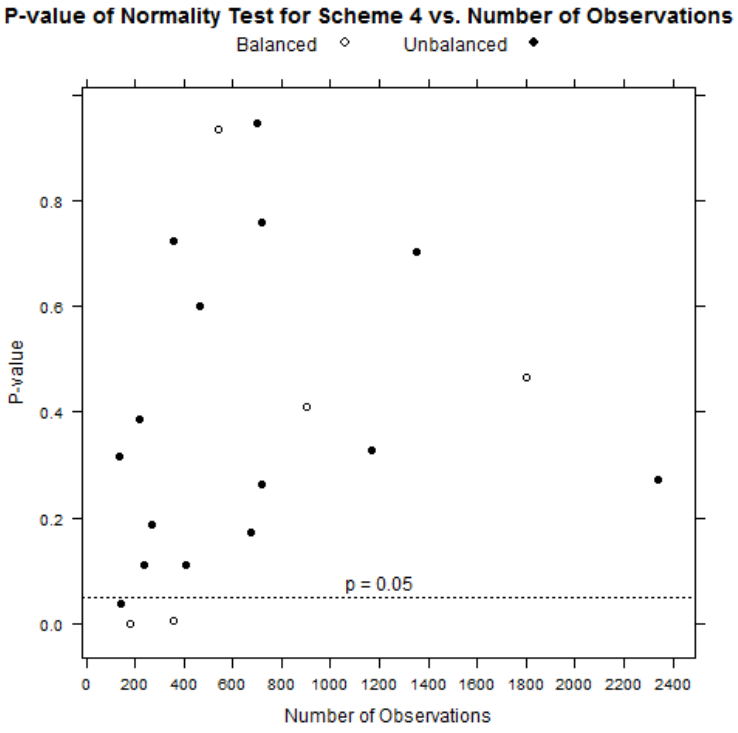

Figure 23. Time Process of the Poisson-Exponential Model (Observations)

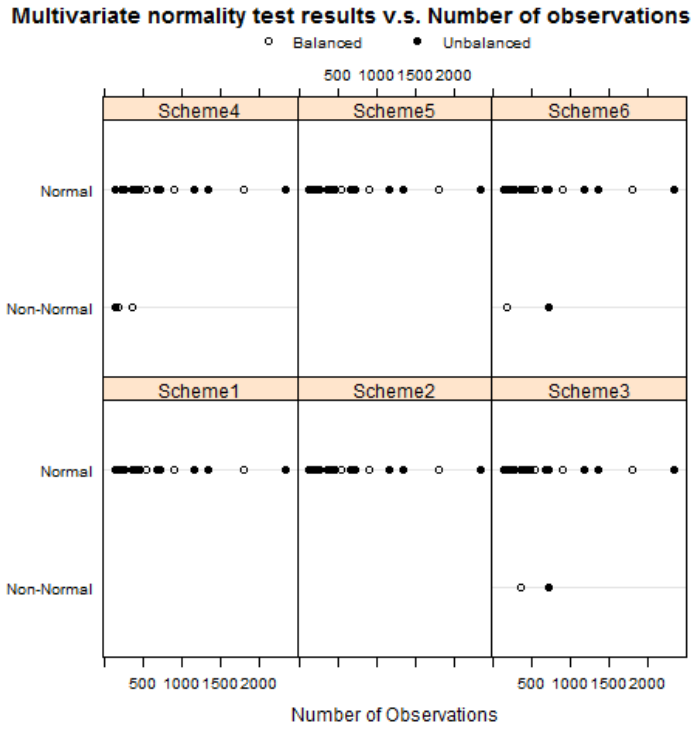

Figure 24. Time Process of the Poisson-Exponential Model (Observations)

Table 10

Time Process of the Poisson-Exponential Model: p-values of the Multivariate Normality Test for different parameter schemes and sample sizes

			Parameter Scheme					
Sample Scheme	Sample Size	# of Obs	1 p-value	2 p-value	3 p-value	4 p-value	5 p-value	6 p-value
1	18	180	0.36968	0.05920	0.56300	0.00001	0.17918	0.04218
2		72	0.11893	0.44316	0.00997	0.26382	0.36458	0.00143
3		135	0.69557	0.47282	0.80654	0.31771	0.77710	0.70368
4		234	0.11195	0.33696	0.52556	0.11119	0.06275	0.55649
5	36	360	0.63561	0.51381	0.03773	0.00701	0.94172	0.59264
6		144	0.76581	0.12487	0.86573	0.03754	0.43839	0.36994
7		180	0.59204	0.21128	0.78980	0.18652	0.08187	0.86942
8		288	0.99272	0.39981	0.84805	0.60154	0.66689	0.85443
9	54	540	0.59380	0.84378	0.98316	0.93627	0.87752	0.92402
10		216	0.11246	0.16843	0.69691	0.38750	0.27582	0.17288
11		405	0.74750	0.55241	0.49931	0.10998	0.39248	0.85567
12		702	0.57741	0.07716	0.17194	0.94672	0.86231	0.62319
13	90	900	0.43914	0.09881	0.60773	0.41156	0.21464	0.31899
14		360	0.21623	0.88285	0.12009	0.72366	0.71011	0.25202
15		675	0.28960	0.52017	0.06290	0.17220	0.84827	0.12679
16		1170	0.72832	0.62466	0.52392	0.32749	0.66726	0.83152
17	180	1800	0.21276	0.32669	0.77498	0.46644	0.44065	0.74302
18		720	0.57254	0.12714	0.84636	0.75986	0.15928	0.41047
19		1350	0.10377	0.82882	0.68000	0.70381	0.58446	0.32260
20		2340	0.47510	0.11559	0.30508	0.27282	0.40995	0.86716

Gamma-Exponential Model

The test results of the multivariate normality tests for the outcome process are shown in Figure 25 and 28 and Table 11. The test results for the time process are presented in Figure 29 and 32 and Table 12. Compared to other models, the parameter estimates of the Gamma-Exponential model obtain normality a little slowler, but the estimators of the model still show the same pattern.

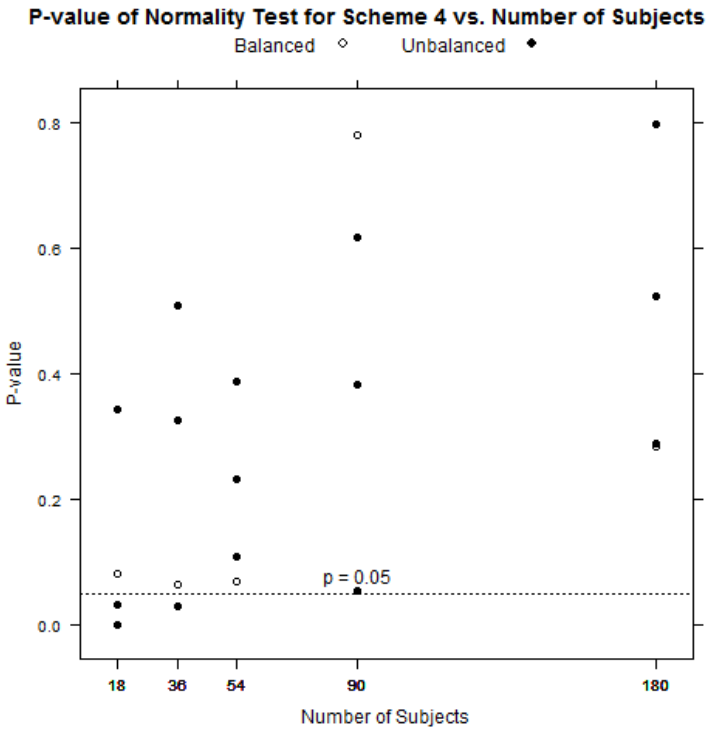

Figure 25. Outcome Process of the Gamma-Exponential Model (Subjects)

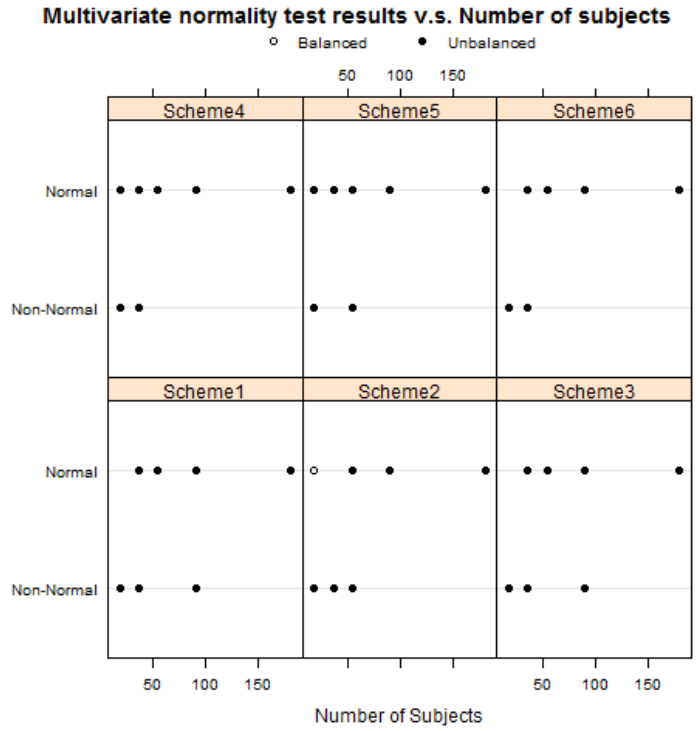

Figure 26. Outcome Process of the Gamma-Exponential Model (Subjects)

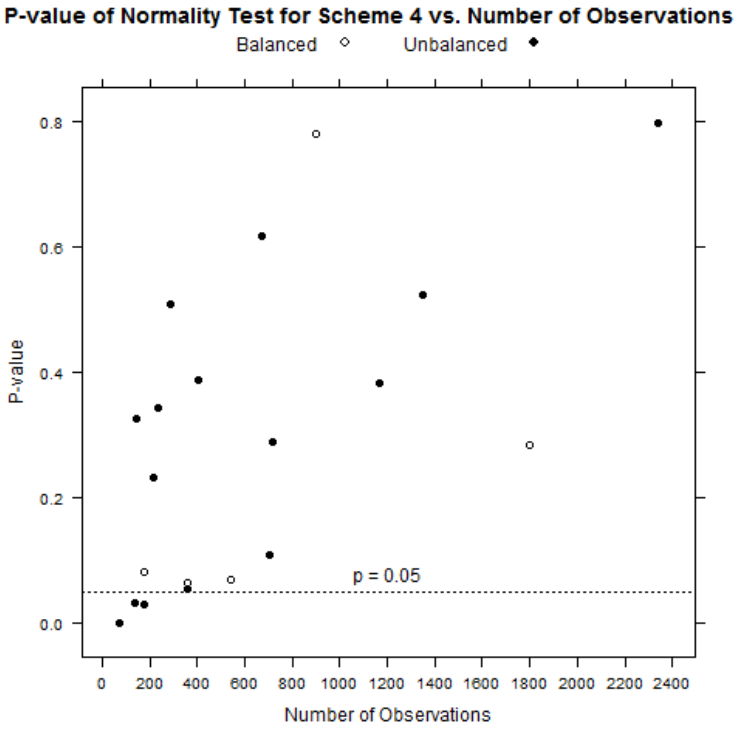

Figure 27. Outcome Process of the Gamma-Exponential Model (Observations)

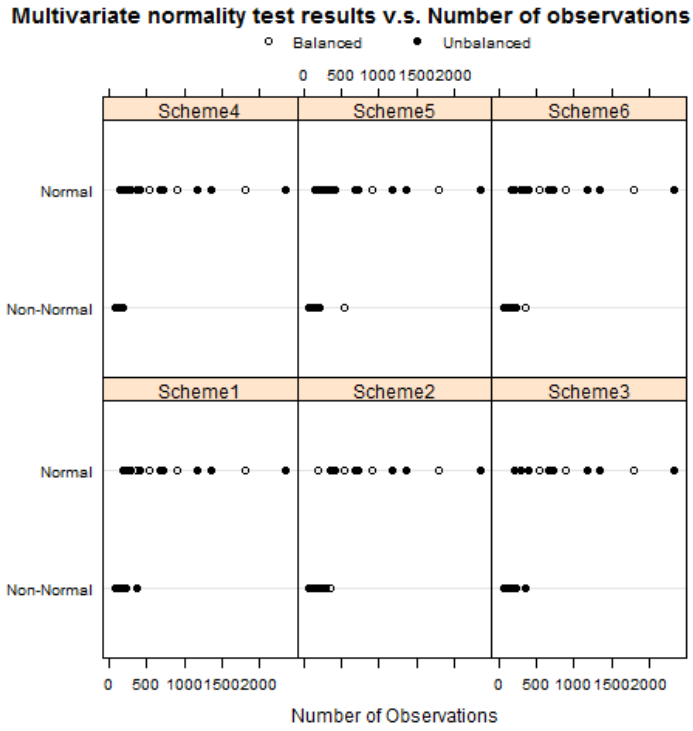

Figure 28. Outcome Process of the Gamma-Exponential Model (Observations)

Table 11

Outcome Process of the Gamma-Exponential Model: p-values of the Multivariate Normality Test for different parameter schemes and sample sizes

			Parameter Scheme					
Sample Scheme	Sample Size	# of Obs	1 p-value	2 p-value	3 p-value	4 p-value	5 p-value	6 p-value
1	18	180	0.01510	0.17792	0.00003	0.08083	0.00001	0.00262
2		72	0.00001	0.00001	0.01978	0.00031	0.00001	0.00001
3		135	0.00210	0.00001	0.01978	0.03275	0.00001	0.00001
4		234	0.00103	0.00128	0.02076	0.34407	0.11764	0.00001
5	36	360	0.35421	0.00001	0.01028	0.06356	0.40382	0.00825
6		144	0.00722	0.00019	0.00001	0.32497	0.22074	0.00046
7		180	0.08193	0.02348	0.01526	0.02878	0.22582	0.05744
8		288	0.23467	0.02400	0.05475	0.50811	0.09404	0.29712
9	54	540	0.49101	0.75876	0.05001	0.06823	0.00321	0.40774
10		216	0.21903	0.04334	0.34122	0.23272	0.01000	0.31520
11		405	0.05726	0.38631	0.13727	0.38797	0.66817	0.28811
12		702	0.33324	0.12490	0.81706	0.10886	0.14504	0.10684
13	90	900	0.96265	0.32650	0.23116	0.78075	0.31541	0.30844
14		360	0.00042	0.84341	0.02709	0.05386	0.52406	0.06959
15		675	0.28685	0.22844	0.30391	0.61856	0.74784	0.48915
16		1170	0.49667	0.99965	0.22644	0.38226	0.51399	0.12637
17	180	1800	0.11085	0.17630	0.98062	0.28320	0.10254	0.32523
18		720	0.12046	0.23472	0.13587	0.28952	0.27900	0.71762
19		1350	0.77981	0.47663	0.07182	0.52340	0.63300	0.11060
20		2340	0.97624	0.05044	0.89781	0.79816	0.97604	0.43067

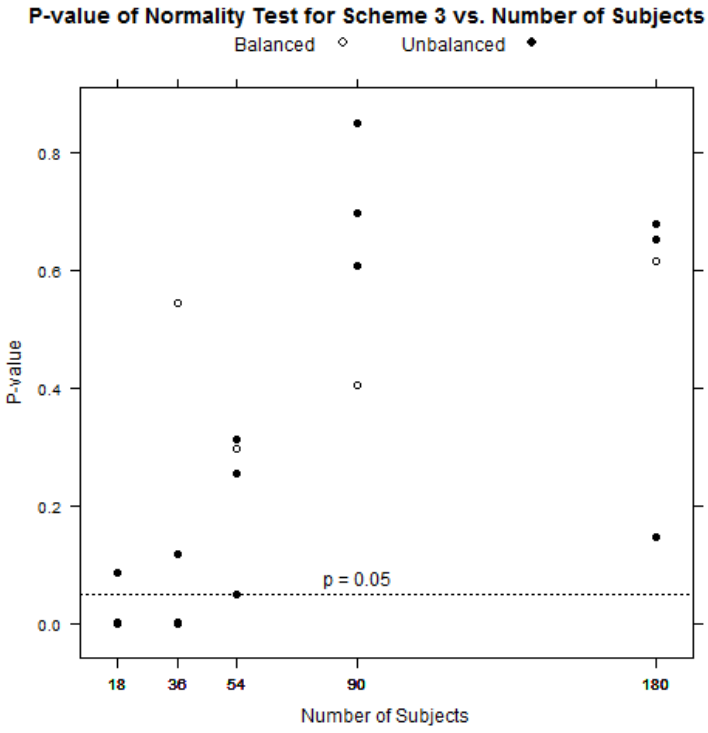

Figure 29. Time Process of the Gamma-Exponential Model (Subjects)

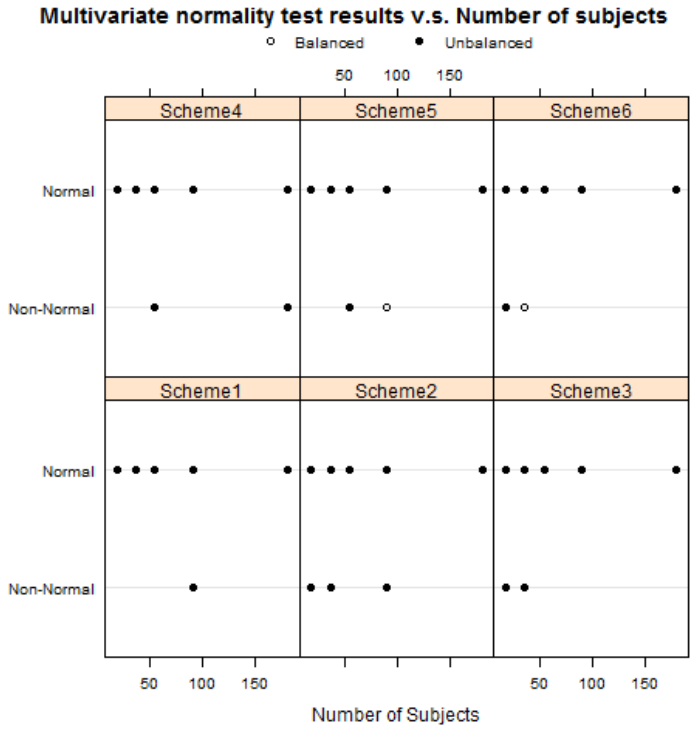

Figure 30. Time Process of the Gamma-Exponential Model (Subjects)

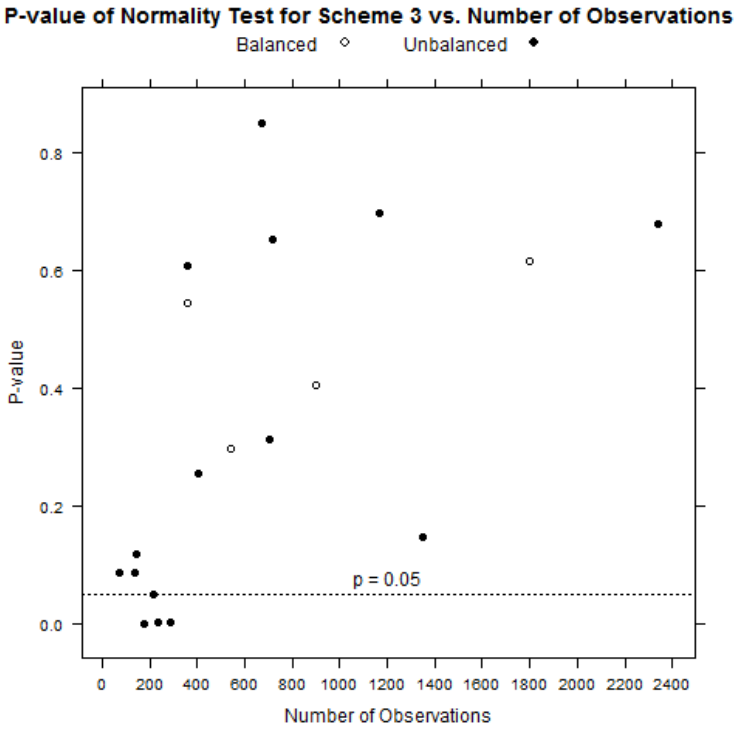

Figure 31. Time Process of the Gamma-Exponential Model (Observations)

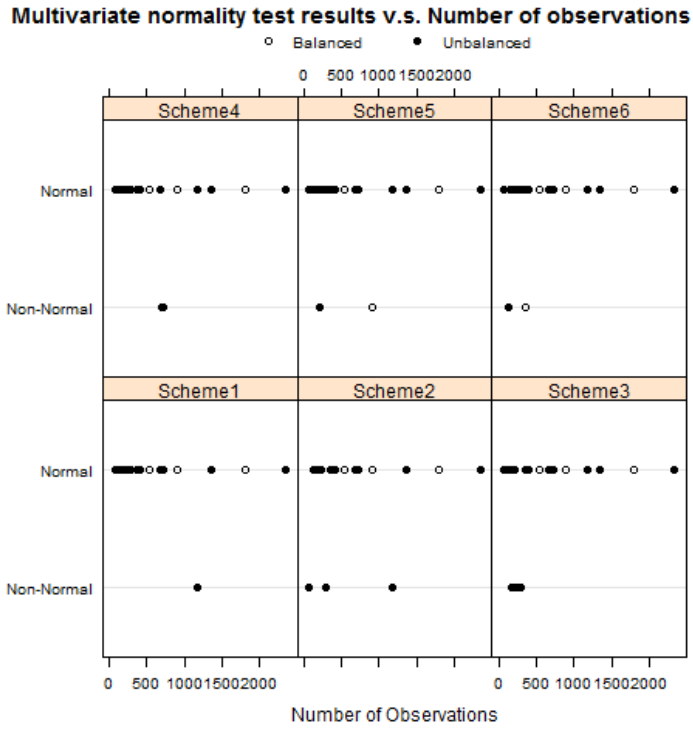

Figure 32. Time Process of the Gamma-Exponential Model (Observations)

Table 12

Time Process of the Gamma-Exponential Model: p-values of the Multivariate Normality Test for different parameter schemes and sample sizes

			Parameter Scheme					
Sample Scheme	Sample Size	# of Obs	1 p-value	2 p-value	3 p-value	4 p-value	5 p-value	6 p-value
1	18	180	0.86572	0.37131	0.00100	0.67151	0.60584	0.27504
2		72	0.29841	0.00741	0.08825	0.27548	0.09334	0.09979
3		135	0.77992	0.16504	0.08825	0.72134	0.32923	0.01447
4		234	0.35246	0.64602	0.00409	0.48813	0.11194	0.09114
5	36	360	0.07756	0.49069	0.54544	0.74836	0.19391	0.03679
6		144	0.22799	0.40778	0.11894	0.12180	0.98087	0.14305
7		180	0.63773	0.67727	0.00129	0.85998	0.06022	0.17012
8		288	0.15240	0.04646	0.00432	0.18650	0.54985	0.77054
9	54	540	0.71251	0.67581	0.29709	0.33731	0.41851	0.94068
10		216	0.74534	0.85472	0.05161	0.23272	0.01336	0.89112
11		405	0.63544	0.88261	0.25709	0.28075	0.12749	0.13077
12		702	0.73240	0.81992	0.31355	0.02845	0.44592	0.79969
13	90	900	0.74591	0.26546	0.40587	0.31457	0.03915	0.80940
14		360	0.19773	0.68945	0.60848	0.29627	0.41734	0.49496
15		675	0.22578	0.20036	0.85129	0.79418	0.34258	0.50440
16		1170	0.01130	0.02301	0.69937	0.96062	0.06400	0.31042
17	180	1800	0.53021	0.47935	0.61597	0.11075	0.16306	0.57966
18		720	0.55779	0.74150	0.65422	0.04132	0.87813	0.47186
19		1350	0.81460	0.61568	0.14774	0.23289	0.06044	0.36606
20		2340	0.76442	0.15216	0.68042	0.49321	0.58991	0.73169

Overall, the design structures, balanced or unbalanced, do not seem to affect multivariate normality. And the maximum likelihood estimators of the outcome and time processes of the joint models gain asymptotic multivariate normality as sample size goes beyond 18. Also, p-values of the tests have an increasing trend as sample size increases. In addition, the estimators of the time process obtain asymptotic multivariate normality faster than the outcome process.

Likelihood Ratio Test

For Research Question Three, the simulation study shows that the parameter estimates of the joint models have the asymptotic normal distribution; therefore, the likelihood ratio test statistic can be computed and used for comparing nested models. The likelihood ratio test assumes that the asymptotic normality of the parameter estimates is satisfied.

The likelihood ratio test statistic is twice the difference in the two nested models' maximized log-likelihoods; thus, it can be written as

$$2(\hat{l}_{full} - \hat{l}_{red}) \sim \chi^2_{df_{full} - df_{red}}.$$

The test statistic is compared to a chi-squared distribution. The common likelihood ratio test is for testing if all parameter estimates but the intercept (β_0) are zero,

$$H_0 : \begin{bmatrix} \boldsymbol{\beta} \\ \gamma \\ \varphi \end{bmatrix} = 0.$$

For instance, the maximized log-likelihood of the full model of the Bernoulli-Exponential model is

$$\hat{l}_{full} = \sum_{i=1}^{m} \left(y_{i1} \log \left(\frac{\mu_{i1}}{1-\mu_{i1}} \right) + \log(1-\mu_{i1}) \right) + \sum_{i=1}^{m} \sum_{j=2}^{n_i} \left(y_{ij} \log \left(\frac{\mu_{ij}}{1-\mu_{ij}} \right) + \log(1-\mu_{ij}) \right), \tag{72}$$

where

$$\mu_{i1} = \frac{\exp(\boldsymbol{X}_i' \hat{\boldsymbol{\beta}})}{1+\exp(\boldsymbol{X}_i' \hat{\boldsymbol{\beta}})} \quad \text{and} \quad \mu_{ij} = \frac{\exp(\boldsymbol{X}_i' \hat{\boldsymbol{\beta}} + \hat{\gamma} t_{ij} + \hat{\varphi} y_{ij-1})}{1+\exp(\boldsymbol{X}_i' \hat{\boldsymbol{\beta}} + \hat{\gamma} t_{ij} + \hat{\varphi} y_{ij-1})}. \tag{73}$$

The maximized log-likelihood of the reduced model of the Bernoulli-Exponential model is

$$\hat{l}_{red} = \sum_{i=1}^{m} \left(y_{i1} \log \left(\frac{\mu_{i1}}{1-\mu_{i1}} \right) + \log(1-\mu_{i1}) \right) + \sum_{i=1}^{m} \sum_{j=2}^{n_i} \left(y_{ij} \log \left(\frac{\mu_{ij}}{1-\mu_{ij}} \right) + \log(1-\mu_{ij}) \right), \tag{74}$$

where

$$\mu_{i1} = \frac{\exp(\boldsymbol{X}_i' \cdot \boldsymbol{0})}{1+\exp(\boldsymbol{X}_i' \cdot \boldsymbol{0})} \quad \text{and} \quad \mu_{ij} = \frac{\exp(\boldsymbol{X}_i' \cdot \boldsymbol{0} + 0 \cdot t_{ij} + 0 \cdot y_{ij-1})}{1+\exp(\boldsymbol{X}_i' \cdot \boldsymbol{0} + 0 \cdot t_{ij} + 0 \cdot y_{ij-1})}. \tag{75}$$

Thus, the likelihood ratio test statistic can be calculated by taking the difference between the two maximized log-likelihoods of the two models multiplied by two. This concept is applied to the other joint models.

Information Criteria

For Research Question Four, the AIC, AICc, and BIC are computed as model selection criteria. The AIC is expressed as $2k - 2\hat{l}$, where $\hat{l}$ is the maximized log-likelihood evaluated at $\hat{\boldsymbol{\theta}}$. The log-likelihood functions for different outcome distributions studied were described in Chapter III. The AICc has the form of

$AIC + \frac{2k(k-1)}{n-k-1}$ and the BIC is $-2\hat{l} + k\log(n)$, which both are the extention of the AIC. The computing program package calculates those model selection criteria, along with the likelihood ratio test statistic.

Analysis of Bladder Cancer Data

The proposed joint models were applied to the bladder cancer data provided in R, and the outputs generated by the computing package are presented in this section and Appendix A and B. R provides two different datasets, bladder and bladder1 in the package called survival. The bladder1 is the full dataset with 118 subjects, and the bladder is the subset of the bladder1 with 85 subjects and a reduced number of variables. The bladder cancer dataset has been studied by many methodologists, Cai, Lu, and Zhang (2012), Sun and Wei (2000), Sun, Park, Sun, and Zhao (2005), and Zhang (2002). The bladder dataset is most commonly used by many researchers for recurrent event modeling (R Core Team, 2014). To demonstrate the performance of the computing package, the Bernoulli-Exponential model was applied to the bladder dataset, and the Poisson-Exponential model was applied to the bladder1 dataset. The variable "stop" in both datasets measure the time interval (in months) since the last visit. Moreover, the next visiting time is scheduled depending on the recurrence of bladder tumor at the time of measurement. Therefore, time becomes informative, and time intervals become irregular across all subjects. The variable "rx" in the bladder or "treatment" in the bladder1 represents treatment types, such as placebo, pyridoxine, and thiotepa.

Two treatment types are included in the bladder, and three treatments in the bladder1.

Bernoulli-Exponential Model

The bladder cancer dataset is composed of 85 subjects with bladder tumors who were assigned to either thiotepa or placebo treatment group. For each patient, the recurrence of tumors, treatment, initial number of tumors, size (cm) of the largest initial tumor, and visiting time (in months) since the last visit were recorded. The status variable "event" for the recurrence of tumors has 1 for recurrence and 0 for everything else (including death for any reason). Therefore, the Bernoulli-Exponential model was applied with the "event" as an outcome variable. The chosen research interest is to study the effects of the treatment, initial number of tumors, and size of the largest initial tumor on tumor recurrence. In the bladder dataset, all patients were measured four times.

The placebo treatment group has 47 randomly selected patients, and the thiotepa group has 38 patients. The likelihood ratio test statistic and the corresponding p-value for each model shown in Table 13 can be used to test if all βs but β_0 equal to zero. Based on the information criteria, AIC, AICc, and BIC, the best fitting model is the one with the treatment, prior outcome, and current time as predictors. Based on the output in the Appendex A says that the treatment (= rx) has a non-significant effect on cancer recurrence ($\beta_1 = 0.2480$ with p-value of 0.248). However, the prior outcome has a significant effect ($\varphi = 4.3312$ with p-value of $< .0001$), and the current time has a significant effect ($\gamma = -0.1412$ with p-value

of $<$ .0001) as well. More descriptions of the data and the output for each model are provided in Appendix A.

Table 13

Model Selection Criteria for the Bernoulli-Exponential Model

Model	AIC	AICc	BIC	LR Test	P-value
$event \sim rx + number + size$	2652.85	2654.74	2692.39	199.10	< 0.0001
$event \sim rx + number$	2651.44	2652.90	2668.54	198.51	< 0.0001
$event \sim rx + size$	2651.64	2653.09	2668.73	198.31	< 0.0001
$event \sim rx$	2650.60	2651.67	2665.25	197.35	< 0.0001

Poisson-Exponential Model

The bladder1 dataset is the full data set of the study for 118 patients, and the maximum observed number of recurrences is 9. The dataset contains all three treatments, placebo, pyridoxine, and thiotepa, with a variable "rtumor", the number of tumors found at the time of recurrence. The Poisson-Exponential model was then applied to model the number of tumors with predictors, such as the treatments, initial number of tumors, and size of the largest initial tumor. The same predictors used in the Bernoulli-Exponential model were used with the variable "rtumor" as an outcome variable. Based on the information criteria presented in Table 14, the model with the treatment, prior outcome, and current time as predictors is selected as the best fitting model. In the model, the "thiotepa" in the treatment and the prior outcome have significant effects on the

number of tumors ($\beta_2 = -0.274$ and $\varphi = 0.078$ with p-value of 0.023 and $< .0001$, respectively). Also, the likelihood ratio test statistic and the corresponding p-value for each model are presented in Table 14. More information about the data and the output for each model are presented in Appendix B.

Table 14

Model Selection Criteria for the Poisson-Exponential Model

Model	AIC	AICc	BIC	LR Test	P-value
$rtumor \sim trt + number + size$	2220.32	2223.78	2239.47	45.50	< 0.0001
$rtumor \sim trt + number$	2219.59	2222.31	2236.61	44.23	< 0.0001
$rtumor \sim trt + size$	2218.47	2221.19	2235.49	45.35	< 0.0001
$rtumor \sim trt$	2218.13	2220.21	2233.02	43.69	< 0.0001

The term "*trt*" in the above table is a shortened word for treatment. The second model, "rtumor $\sim$ treatment + number," in the table above, was studied by Cai et al. (2012). Their research interest was to study the effects of the treatment and number of initial tumors on tumor recurrence. The difference is that those researchers studied a time-varying latent effect model with time-independent covariates. The model they found is shown below.

Table 15

Coefficients of the Latent Effect Model

	Est.	SE
Treatment	-0.152	0.042
Number of initial tumors	0.205	0.050

For model comparisons, the coefficients of the Poisson-Exponential model are provided below.

Table 16

Coefficients of the Poisson-Exponential Model

Coefficient	Estimate	Std. error	t value	$\Pr(>\|t\|)$
Intercept	0.9328	0.0758	12.3010	$< .0001$
treatment.pyridoxine	-0.0100	0.1015	-0.0993	0.9208
treatment.thiotepa	-0.3001	0.1266	-2.3709	0.0177
number	0.0160	0.0218	0.7364	0.4614
Prior Outcome	0.0784	0.0183	4.2848	$< .0001$
Current Time	0.0006	0.0030	0.2032	0.8389

Even though these two models have different concepts, both models found that the treatment has a negative effect on the number of tumors. The output above shows that the third treatment level, thiotepa, has a significant negative effect on the number of tumors ($\beta_2 = -0.3001$ with p-value of 0.0177), and that the number of

initial tumors is not significant on the tumor occurrence ($\beta_3 = 0.0160$ with p-value of 0.4614).

CHAPTER V

CONCLUSION AND DISCUSSIONS

Conclusion

The purpose of this study was to test if the Gaussian-Exponential model (Bronsert, 2009; Lin, 2011) can be extended to outcomes belonging to the exponential family of distributions. The simulation studies were conducted with six parameter schemes, two design structures, and five sample sizes, to test the asymptotic multivariate normality of the maximum likelihood estimators of the extended joint models. Outcome distributions considered in this study were the Gaussian, Bernoulli, Poisson, and Gamma distributions. In all of the simulation designs, the maximum likelihood parameter estimates of the joint models appeared to be multivariate normal as the number of observations increased. As a result, the likelihood ratio test statistic could be utilized for model comparisons since the asymptotic normality of the maximum likelihood estimators has been verified. Also, AIC, AICc, and BIC scores were calculated as model selection criteria. Furthermore, the computing package using R was developed to handle the joint models and used to analyze the bladder cancer data for demonstration purposes.

Also, a part of this study was to retest the simulation results for the Gaussian-Exponential model, since an error was found in the Henze-Zirkler test in SAS software, used by the previous researcher, at the beginning of this study. Lin

(2011) suggested that the minimum sample size to be applied to the Gaussian-Exponential model should be greater than 54 subjects, each with at least 20 observations. However, in this study, most of the simulation results for the Gaussian-Exponential model attained asymptotic normality as sample size exceeded 18 for both the outcome and time processes. Consequently, the Gaussian-Exponential model can be valid for data with a sample size of greater than 18, instead of 54.

Overall, based on the multivariate normality test from the simulation study, the maximum likelihood parameter estimates of the joint models obtain the asymptotic normality when sample size is greater than 18. Accordingly, the model selection methods are valid in the joint models, which are based on the normality assumption of parameter estimates. For those reasons, the extended joint models can be recommended to use when sample size is greater than 18. In addition, the maximum likelihood parameter estimates of the joint models obtain asymptotic multivariate normality in both balanced and unbalanced designs.

The joint models presented in this study rely on the relation among the one-step prior outcome, current time, and potential covariates. Also, it is assumed that time and covariates are independent of each other, and that time should be informative and exponentially distributed. If any of these assumptions is not satisfied, the joint models proposed in this study should be considered with caution.

Application of the Joint Models

The proposed joint models were applied to real datasets to demonstrate the performance of the models. The Poisson-Exponential model was compared with the latent effect model proposed by Cai et al. (2012). The results based on the model of Cai et al. (2012) showed similar findings, for the estimation of the effects of the treatment and the number of initial tumors. As can be seen in Table 15 and 16, both methods found that the treatment has a negative effect on the tumor occurrence, and that the number of initial tumors has a positive effect on the tumor occurrence. The difference is that the model of Cai et al. (2012) computed the overall treatment effect, ignoring the effect of each treatment level. However, the Poisson-Exponential model calculated the estimator for each treatment level, like other regression analyses normally do. Cai et al. (2012) stated that the treatment has a negative association since "the more often the patients visited the clinic and received the treatment, the less chance they will have tumor recurrence" (p. 10). This interpretation makes sense, but does not specify which treatment is most effective to reduce the tumor occurrence. However, the Poisson-Exponential model specifically pointed out that the thiotep treatment only can reduce the tumor occurrence.

Discussions

In this study, the asymptotic multivariate normality of the maximum likelihood estimators of the joint models has been verified by simulation studies with the arbitrary chosen simulation designs. Proving the asymptotic normality

with mathematical theories is beyond the scope of the current study. This undeveloped step can be studied by a theorist in the future.

As described in the last paragraph of the previous section, the joint models have multiple assumptions, which limit the use of the joint models. If those assumptions are relaxed, the joint models can be easily expanded to be more flexible.

For example, firstly, the current response is assumed to be dependent upon the one-step prior outcome. In some experiments, it is possible that the current response depends on the two-step prior outcome or three-step prior outcome, etc. In that case, the joint models can be modified to accommodate those terms in the models by simply replacing y_{t-1} by y_{t-2} or y_{t-3}. Secondly, time is assumed to follow an exponential distribution. The distribution of time can be different based on a research design, for example, a normal distribution. If that is the situation, the appropriate distribution can be applied to the time process; then, the maximum likelihood parameter estimates from the time process can be obtained. Thirdly, currently time and covariates are assumed to be independent of each other. If they are related, another term can be added to define the relations between them in the models. Fourthly, the current joint models have a single response variable in a data set. If multiple response variables are included in the analysis, the joint models should be able to take correlations among those into account; furthermore, the joint models should be able to give simultaneous tests for separate responses, in addition to a single responses analysis. All of assumption relaxations mentioned above are

technically possible and can be further explored by a researcher in order to improve the joint models.

REFERENCES

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Bronsert, M. R. (2009). *A joint model of a longitudinal process and informative time schedule data* (Unpublished doctoral dissertation). University of Northern Colorado.

Burnham, K. P., & Anderson, D. (2002). *Model selection and multimodel inference* (2nd ed.). Verlag, New York: Springer.

Cai, N., Lu, W., & Zhang, H. H. (2012). Time-varying latent model for longitudinal data with informative observation times. *Biometrics*, *64*, 1093-1102.

Crowder, M. J., & Hand, D. J. (1990). *Analysis of repeated measures.* Chapman & Hall.

Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements* (2nd ed.). Verlag, NY: Springer.

Dawson, J. D. (1994). Comparing treatment groups on the basis of slopes, areas-under-the-curve, and other summary measures. *Drug Information Journal*, *28*, 723-732.

Diggle, P. J., Heagerty, P., Liang, K.-Y., & Zeger, S. L. (2002). *Analysis of longitudinal data* (2nd ed.). Oxford, UK: Oxford University Press.

Dobson, A. J. (2001). *An introduction to generalized linear models.* London: Chapman and Hall/CRC Press.

Everitt, B. S. (2006). *The cambridge dictionary of statistics* (3rd ed.). Cambridge, UK: Oxford University Press.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons. Inc.

Frison, L., & Pocock, S. J. (1992). Repeated measures in clinical trials: Analysis using mean summary statistics and its implications for design. *Statistics in Medicine*, *11*, 1685-1704.

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). University of Bristol, UK: John Wiley & Sons. Inc.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis.* Hoboken, NJ: John Wiley & Sons, Inc.

Henderson, R., Diggle, P., & Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, *1*, 465-480.

Kim, S., Zeng, D., Chambless, L., & Li, Y. (2012). Joint models of longitudinal data and recurrent events with informative terminal event. *Stat Biosci*, *4*, 262-281.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963-974.

Lee, Y., & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society*, *58*, 619-678.

Leeuw, J. D., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, *11*, 57-85.

Liang, K., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13-22.

Liang, Y., Lu, W., & Ying, Z. (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics*, *65*, 377-384.

Lin, H., Scharfstein, D. O., & Rosenheck, R. A. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society*, *66*, 791-813.

Lin, Y. (2011). *Hypothesis testing for the Gaussian-Exponential longitudinal model* (Unpublished doctoral dissertation). University of Northern Colorado.

Lindstrom, M. J., & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, *46*, 673-687.

Lipsitz, S. R., Fitzmarice, G. M., Ibrahim, J. G., Gelber, R., & Lipshultz, S. (2002). Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics*, *58*, 621-630.

McCullagh, P., & Nelder, J. A. (1983). *Generalized linear models.* London: Chapman and Hall/CRC Press.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall/CRC Press.

McQuarrie, A., Shumway, R., & Tsai, C.-L. (1997). The model selection criterion AICu. *Statistics & Probability Letters*, *34*, 285-292.

Nash, J. C., & Varadhan, R. (2011). Unifying optimization algorithms to aid software system users. *Journal of Statistical Software*, *43*, 1-14.

Nelder, J. A., & Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, *135*, 370-384.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, *58*, 545-554.

Pepe, M. S., & Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics*, *23*, 939-951.

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, *4*, 12-35.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Verlag, New York: Springer.

Qiu, F., Stein, C. M., & Elston, R. C. (2013). Joint modeling of longitudinal data and discrete-time survival outcome. *Statistical Methods in Medical Research*, 1-15.

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

SAS Institute. (2008). *SAS 9.2: User's Guide*. SAS.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurence*. New York, New York: Oxford University Press Inc.

Sun, J., Park, D.-H., Sun, L., & Zhao, X. (2005). Semiparametric regression analysis of longitudinal data with informative observation times. *Journal of the American Statistical Association*, *100*, 882-889.

Sun, J., & Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society*, *62*, 293-302.

Vincent, W. J. (2005). *Statistics in kinesiology* (3rd ed.). Champaign, IL: Human Kinetics.

Wu, L., Liu, W., Yi, G. Y., & Huang, Y. (2012). Analysis of longitudinal and survival data: Joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 1-17.

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika*, *89*, 39-48.

Appendix A

OUTPUT OF THE BERNOULLI-EXPONENTIAL MODEL

Bladder Cancer Data

This analysis uses the status variable, "event", as a outcome, "rx" as a grouping variable, and "number" and "size" as covariates. This dataset consists of 85 patients.

1. id: Patient id
2. event: Recurrence of tumors (1 = recurrence and 0 = everything else)
3. rx: Treatment 1=placebo 2=thiotepa
4. number: Initial number of tumors (8 = 8 or more)
5. size: Size (cm) of the largest initial tumor
6. stop: Recurrence or censoring time
7. enum: Which recurrence (up to 4)

Model 1 : $event \sim rx + number + size$

```
$Model
[1] "Call:  event ~ rx + number + size"

$Coefficients    Estimate Std. error     t value      Pr(> t)
(Intercept)   -0.05502464 0.24446703  -0.2250800 8.219170e-01
rx2            0.27649473 0.21521080   1.2847623 1.988754e-01
number        -0.05055372 0.05694457  -0.8877707 3.746641e-01
size           0.05607314 0.07282951   0.7699233 4.413454e-01
prior.outcome  4.34943175 0.58220294   7.4706455 7.980240e-14
current.time  -0.14184232 0.02058932  -6.8891199 5.613863e-12
alpha         -3.38082136 0.07960693 -42.4689304 0.000000e+00
delta          0.20276751 0.12872836   1.5751581 1.152200e-01

$AIC
[1] 2652.851

$AICc
[1] 2654.745

$BIC
[1] 2672.392

$LogLik
[1] 199.1082

$LogLikPval
[1] 0
```

Model 2 : $event \sim rx + number$

```
$Model
[1] "Call:  event ~ rx + number"

$Coefficients    Estimate Std. error     t value      Pr(> t)
(Intercept)    0.07737720 0.17404807   0.4445737 6.566278e-01
rx2            0.27293430 0.21504153   1.2692167 2.043638e-01
number        -0.05978083 0.05571470  -1.0729812 2.832795e-01
prior.outcome  4.33597152 0.57614103   7.5258856 5.236416e-14
current.time  -0.14179764 0.02035379  -6.9666444 3.245898e-12
alpha         -3.38082133 0.07977955 -42.3770393 0.000000e+00
delta          0.20276748 0.12856084   1.5772103 1.147471e-01

$AIC
[1] 2651.446

$AICc
[1] 2652.901

$BIC
[1] 2668.545

$LogLik
[1] 198.5126

$LogLikPval
[1] 0
```

Model 3 : $event \sim rx + size$

```
$Model
[1] "Call:  event ~ rx + size"

$Coefficients    Estimate Std. error     t value      Pr(> t)
(Intercept)   -0.17922106 0.20141067  -0.8898290 3.735577e-01
rx2            0.25728127 0.21496375   1.1968589 2.313616e-01
size           0.06963772 0.07131094   0.9765363 3.287988e-01
prior.outcome  4.34895831 0.56927249   7.6395020 2.180635e-14
current.time  -0.14216743 0.02026683  -7.0147848 2.303031e-12
alpha         -3.38082134 0.07954964 -42.4995167 0.000000e+00
delta          0.20276751 0.12837268   1.5795223 1.142163e-01

$AIC
[1] 2651.641

$AICc
[1] 2653.095

$BIC
[1] 2668.739

$LogLik
[1] 198.3183

$LogLikPval
[1] 0
```

Model 4 : $event \sim rx$

```
$Model
[1] "Call:  event ~ rx"

$Coefficients    Estimate Std. error     t value      Pr(> t)
(Intercept)   -0.03676028 0.13805297  -0.2662767 7.900261e-01
rx2            0.24806938 0.21487572   1.1544784 2.483041e-01
prior.outcome  4.33127194 0.60620489   7.1448977 9.006294e-13
current.time  -0.14218783 0.02123201  -6.6968610 2.129440e-11
alpha         -3.38082134 0.07983721 -42.3464353 0.000000e+00
delta          0.20276751 0.12875661   1.5748124 1.152997e-01

$AIC
[1] 2650.602

$AICc
[1] 2651.679

$BIC
[1] 2665.258

$LogLik
[1] 197.3565

$LogLikPval
[1] 0
```

Appendix B

OUTPUT OF THE POISSON-EXPONENTIAL MODEL

Bladder1 Cancer Data

This analysis uses "rtumor" as a outcome, "treatment" as a grouping variable, and "number" and "size" as covariates. This dataset consists of 118 patients.

1. id: Patient id
2. treatment: Placebo, pyridoxine (vitamin B6), or thiotepa
3. number: Initial number of tumors (8 = 8 or more)
4. size: Size (cm) of the largest initial tumor
5. recur: Number of recurrences
6. start: The start time of each time interval
7. stop: The end time of each time interval
8. status: End of interval code, 0 = censored, 1 = recurrence, 2 = death from bladder disease, 3 = death other/unknown cause
9. rtumor: Number of tumors found at the time of a recurrence
10. rsize: Size of largest tumor at a recurrence

Model 1 : $rtumor \sim treatment + number + size$

```
$Model
[1] "Call:  rtumor ~ treatment + number + size"

$Coefficients            Estimate  Std. error     t value      Pr(> t)
(Intercept)          1.0028856783 0.098013784  10.2320882 1.424149e-24
treatmentpyridoxine -0.0059002387 0.101527477  -0.0581147 9.536573e-01
treatmentthiotepa   -0.3009050218 0.126373379  -2.3810792 1.726200e-02
number               0.0087449771 0.022726800   0.3847870 7.003952e-01
size                -0.0271937970 0.024349235  -1.1168235 2.640698e-01
prior.outcome        0.0769869072 0.018345161   4.1965785 2.709775e-05
current.time         0.0007712331 0.003072309   0.2510272 8.017931e-01
alpha               -3.1450823925 0.150181259 -20.9419099 2.223643e-97
delta               -0.0110809301 0.040064259  -0.2765789 7.821034e-01

$AIC
[1] 2220.327

$AICc
[1] 2223.788

$BIC
[1] 2239.471

$LogLik
[1] 45.50186

$LogLikPval
[1] 3.719609e-08
```

Model 2 : $rtumor \sim treatment + number$

```
$Model
[1] "Call:  rtumor ~ treatment + number"

$Coefficients            Estimate  Std. error      t value      Pr(> t)
(Intercept)          0.9328033958 0.075831065  12.30107206 8.937990e-35
treatmentpyridoxine -0.0100911560 0.101561128  -0.09936042 9.208521e-01
treatmentthiotepa   -0.3001890441 0.126609851  -2.37097699 1.774114e-02
number               0.0160542137 0.021800670   0.73640918 4.614817e-01
prior.outcome        0.0784415210 0.018306650   4.28486490 1.828503e-05
current.time         0.0006219553 0.003059826   0.20326490 8.389280e-01
alpha               -3.1450823966 0.150016250 -20.96494482 1.370798e-97
delta               -0.0110809300 0.040020304  -0.27688270 7.818702e-01

$AIC
[1] 2219.597

$AICc
[1] 2222.314

$BIC
[1] 2236.614

$LogLik
[1] 44.23187

$LogLikPval
[1] 2.078272e-08
```

Model 3 : $rtumor \sim treatment + size$

```
$Model
[1] "Call:  rtumor ~ treatment + size"

$Coefficients            Estimate  Std. error      t value      Pr(> t)
(Intercept)          1.0273170858 0.074565269  13.77742079 3.485021e-43
treatmentpyridoxine -0.0055632814 0.101657424  -0.05472578 9.563569e-01
treatmentthiotepa   -0.2885945888 0.122187411  -2.36190116 1.818149e-02
size                -0.0298282354 0.023377995  -1.27591075 2.019871e-01
prior.outcome        0.0769599487 0.018326143   4.19946234 2.675495e-05
current.time         0.0007911878 0.003071865   0.25755945 7.967469e-01
alpha               -3.1450824118 0.150107129 -20.95225208 1.789646e-97
delta               -0.0110809278 0.040013072  -0.27693270 7.818318e-01

$AIC
[1] 2218.474

$AICc
[1] 2221.191

$BIC
[1] 2235.492

$LogLik
[1] 45.35418

$LogLikPval
[1] 1.229156e-08
```

Model 4 : $rtumor \sim treatment$

```
$Model
[1] "Call:  rtumor ~ treatment"

$Coefficients            Estimate  Std. error      t value      Pr(> t)
(Intercept)          0.9679454154 0.058686375  16.49352880 4.083858e-61
treatmentpyridoxine -0.0096904843 0.101409612  -0.09555785 9.238717e-01
treatmentthiotepa   -0.2747982336 0.121497045  -2.26176887 2.371169e-02
prior.outcome        0.0786729618 0.018264445   4.30743786 1.651565e-05
current.time         0.0006277268 0.003056412   0.20538030 8.372750e-01
alpha               -3.1450823926 0.150789979 -20.85737002 1.306619e-96
delta               -0.0110809306 0.040152939  -0.27596810 7.825726e-01

$AIC
[1] 2218.136

$AICc
[1] 2220.21

$BIC
[1] 2233.026

$LogLik
[1] 43.69226

$LogLikPval
[1] 7.432897e-09
```

Appendix C

R PROGRAM FOR SIMULATIONS

1. Gaussian-Exponential Model

```
#####################################################################
#                 Packages                                          #
#####################################################################

install.packages('MVN')  # hzTest, roystonTest
install.packages('MASS')
install.packages('maxLik') #maxLik
install.packages('AlgDesign') # gen.factorial
install.packages('mefa') # provide rep(dat,times)

library(MVN)
library(MASS)
library(maxLik)
library(AlgDesign)
library(mefa)

#####################################################################
#         Parameter Setting (Pscheme: 1 to 6)                       #
#####################################################################

parameter = matrix(c(1,1,2,2,0.5,0.5,              #1:sigma
                     0.4,0.4,0.4,0.4,0.4,0.4,      #2:beta0
                     0.2,0.2,0.2,0.2,0.2,0.2,      #3:beta1
                     0.3,0.3,0.3,0.3,0.3,0.3,      #4:beta2
                     0.1,0.1,0.1,0.1,0.1,0.1,      #5:beta3
                     0.3,0.3,0.3,0.3,0.3,0.3,      #6:beta4
                     0.4,0.4,0.4,0.4,0.4,0.4,      #7:beta5
                     0.9,0.9,0.9,0.9,0.9,0.9,      #8:beta6
                     0.8,0.8,0.8,0.0,0.0,0.8,      #9:phi
                     0.1,0.1,0.1,0.1,0.1,0.1,      #10:gamma
                     2,1,2,1,2,1,                  #11:alpha
                     0.01,0.02,0.01,0.02,0.01,0.02),#12:delta
                     nrow=6)

#####################################################################
#     create design matrix (X) with two cat & two cont vars         #
#####################################################################

design=function(level=c(3,3),m=18,p=2){
catg=gen.factorial(levels=level,center=FALSE,factors='all')
ext=rep(catg,m/(prod(level)))
des=model.matrix(~.,data=ext) #'~.' is supported by {AlgDesign}
cont=data.frame(matrix(NA,nrow=m,ncol=p))
```

```
for (i in 1:p){
cont[i]=rnorm(m)
}
xmatrix=as.matrix(cbind(des,cont))
xmatrix
}

#####################################################################
#                Create Data:  c('outcome','time','subject')   #
#####################################################################

outcome<- function(m=m,num=num,parm=parm){
if (num == 1) {n1 = 10; n2=10}
if (num == 2) {n1 = 5; n2=3}
if (num == 3) {n1 = 10; n2=5}
if (num == 4) {n1 = 20; n2=6}

ndesign = matrix(c(rep(n1,m/2),rep(n2,m/2)),byrow=T)
nn=cumsum(c(1,ndesign[-length(ndesign)]))
raw = matrix(NA,sum(ndesign),3)  #Null matrix
mu = xmatrix %*% parm[2:8]   # mu is matrix
raw[nn,1]= mu + rnorm(m)*parm[1]
raw[nn,2] = rexp(m)

for (i in 1:m){
for (j in 2:ndesign[i]){
yjmin1 = raw[nn[i] - 1 + j - 1,1]
raw[nn[i] - 1 + j,2] = rexp(1)*
exp(parm[11] +parm[12] * yjmin1)
raw[nn[i] - 1 + j,1] =mu[i] + yjmin1 * parm[9] +
raw[nn[i]-1+j,2]*parm[10]+rnorm(1)*parm[1]
raw[nn[i],3]=i
raw[nn[i]-1+j,3]=i
} #j
}#i
result=list(raw=raw,nn=nn,ndesign=ndesign)
result
} #outcome

#############################################################
#            Log-Likelihood Function                        #
#############################################################

loglikfn<- function(parms){
```

```
y1=y[nn,1] #initial obs for every subjects
f1=sum(-0.5 * log(parms[1]^2)-0.5*
(y1-xmatrix %*% parms[2:8])^2/parms[1]^2)

f2=0;f3=0
for (i in 1:m){
yi=y[(y[,3]==i), 1] # all obs for ith subject
ti=y[(y[,3]==i), 2] # all time points for ith subject
yi1=yi[-ndesign[i]] #previous obs
tti=ti[-1] #current time
yi2=yi[-1] #current obs

f2=sum(-0.5 * log(parms[1]^2)-0.5 *
(yi2-parms[10]*tti-parms[9]*yi1-xmatrix[i,]%*%parms[2:8])^2/
parms[1]^2)+f2

f3=sum(parms[11]+parms[12]*yi1-
exp(parms[11]+parms[12]*yi1)*tti)+f3
} #i
(m+f1+f2+f3)
}  # loglike

############################################################
#         Simulation                                       #
############################################################

Pschem = 1  # parameter setting, 1 to 6
m=36
num = 2 # design structure 1(10,10), 2(5,3),3(10,5),4(20,6)
rep=1000 #number of replications
out = matrix(NA,rep,ncol(parameter))
parm = parameter[Pschem,]
xmatrix=design(level=c(3,3),m=m,p=2)

for (r in 1:rep){
#compute some info to be used in optimization
result=outcome(m=m,num=num,parm=parm)
y=result$raw
nn=result$nn
ndesign=result$ndesign
mle=maxLik(logLik = loglikfn, start = parm)
diff=coef(mle)-parm
out[r,]=sqrt(sum(ndesign))*diff/summary(mle)$estimate[,2]
}
```

```
hzTest(out[,1:9])   #Outcome process
hzTest(out[,10:11])  #Time process
```

2. Bernoulli-Exponential Model

```
#####################################################################
#         Parameter Setting (Pscheme: 1 to 6)                       #
#####################################################################

parameter = matrix(c(0.4,0.4,0.4,0.4,0.4,0.4,       #1:beta0
                     0.2,0.2,0.2,0.2,0.2,0.2,       #2:beta1
                     0.3,0.3,0.3,0.3,0.3,0.3,       #3:beta2
                     0.1,0.1,0.1,0.1,0.1,0.1,       #4:beta3
                     0.3,0.3,0.3,0.3,0.3,0.3,       #5:beta4
                     0.4,0.4,0.4,0.4,0.4,0.4,       #6:beta5
                     0.9,0.9,0.9,0.9,0.9,0.9,       #7:beta6
                     0.8,0.8,0.8,0.0,0.0,0.8,       #8:phi
                     0.1,0.1,0.1,0.1,0.1,0.1,       #9:gamma
                     2,1,2,1,2,1,                   #10:alpha
                     0.01,0.02,0.01,0.02,0.01,0.02),#11:delta
                     nrow=6)

#####################################################################
#                Create Data:  c('outcome','time','subject')   #
#####################################################################

outcome<- function(m=m,num=num,parm=parm){
if (num == 1) {n1 = 10; n2=10}
if (num == 2) {n1 = 5; n2=3}
if (num == 3) {n1 = 10; n2=5}
if (num == 4) {n1 = 20; n2=6}

ndesign = matrix(c(rep(n1,m/2),rep(n2,m/2)),byrow=T)
nn=cumsum(c(1,ndesign[-length(ndesign)]))
raw = matrix(NA,sum(ndesign),3)  #Null matrix
mu=xmatrix %*% parm[1:7]
raw[nn,1]=rbinom(m,1,ui1)
raw[nn,2] = rexp(m,rate=1)

for (i in 1:m){
for (j in 2:ndesign[i]){
yjmin1 = raw[nn[i]-1+j-1,1]
raw[nn[i]-1+j,2]= rexp(1)*exp(parm[10]+parm[11]*yjmin1)
uij=1/(1+exp(-(mu[i]+raw[nn[i]-1+j,2]*parm[9]+yjmin1 * parm[8])))
raw[nn[i]-1+j,1]=rbinom(1,1,uij)
raw[nn[i],3]=i
raw[nn[i]-1+j,3]=i
} #j
```

```
}#i
result=list(raw=raw,nn=nn,ndesign=ndesign)
result
} #outcome

############################################################
#            Log-Likelihood Function                       #
############################################################

loglikfn<-function(parms){
y1=y[nn,1] #initial obs for every subjects
mu=xmatrix %*% parms[1:7]
ui1=1/(1+exp(-mu))
f1=sum(y1*log(ui1/(1-ui1))+log(1-ui1))

f2=0;f3=0
for(i in 1:m){
yi=y[(y[,3]==i), 1] # all obs for ith subject
ti=y[(y[,3]==i), 2] # all time points for ith subject
yi1=yi[-ndesign[i]] #previous obs
tti=ti[-1] #current time
yi2=yi[-1] #current obs
uij=1/(1+exp(-(mu[i]+parms[9]*tti+parms[8]*yi1)))
f2=sum((1-yi2)*log(1-uij)+ yi2*log(uij))+f2

f3=sum(parms[10]+parms[11]*yi1-
exp(parms[10]+parms[11]*yi1)*tti)+f3
} #i
(m+f1+f2+f3)
} #loglikfn

############################################################
#         Simulation                                       #
############################################################

Pschem = 1  # parameter setting, 1 to 6
m=36
num = 2 # design structure 1(10,10), 2(5,3),3(10,5),4(20,6)
rep=1000 #number of replications
out = matrix(NA,rep,ncol(parameter))
parm = parameter[Pschem,]*0.1
xmatrix=design(level=c(3,3),m=m,p=2)

for (r in 1:rep){
result=outcome(m=m,num=num,parm=parm)
```

```
y=result$raw
nn=result$nn
ndesign=result$ndesign
mle=maxLik(logLik = loglikfn, start = parm)
diff=coef(mle)-parm
out[r,]=sqrt(sum(ndesign))*diff/summary(mle)$estimate[,2]
}

hzTest(out[,1:9])   #Outcome process
hzTest(out[,10:11])  #Time process
```

3. Poisson-Exponential Model

```
#####################################################################
#         Parameter Setting (Pscheme: 1 to 6)                       #
#####################################################################

parameter = matrix(c(0.4,0.4,0.4,0.4,0.4,0.4,       #1:beta0
                     0.2,0.2,0.2,0.2,0.2,0.2,       #2:beta1
                     0.3,0.3,0.3,0.3,0.3,0.3,       #3:beta2
                     0.1,0.1,0.1,0.1,0.1,0.1,       #4:beta3
                     0.3,0.3,0.3,0.3,0.3,0.3,       #5:beta4
                     0.4,0.4,0.4,0.4,0.4,0.4,       #6:beta5
                     0.9,0.9,0.9,0.9,0.9,0.9,       #7:beta6
                     0.8,0.8,0.8,0.0,0.0,0.8,       #8:phi
                     0.1,0.1,0.1,0.1,0.1,0.1,       #9:gamma
                     2,1,2,1,2,1,                   #10:alpha
                     0.01,0.02,0.01,0.02,0.01,0.02),#11:delta
                     nrow=6)

#####################################################################
#                 Create Data:  c('outcome','time','subject')   #
#####################################################################

outcome<- function(m=m,num=num,parm=parm){
if (num == 1) {n1 = 10; n2=10}
if (num == 2) {n1 = 5; n2=3}
if (num == 3) {n1 = 10; n2=5}
if (num == 4) {n1 = 20; n2=6}

ndesign = matrix(c(rep(n1,m/2),rep(n2,m/2)),byrow=T)
nn=cumsum(c(1,ndesign[-length(ndesign)]))
raw = matrix(NA,sum(ndesign),3)  #Null matrix
mu=xmatrix %*% parm[1:7]

ui1=exp(mu)
raw[nn,1]=rpois(m,ui1)
raw[nn,2] = rexp(m)

for (i in 1:m){
for (j in 2:ndesign[i]){
yjmin1 = raw[nn[i]-1+j-1,1]
raw[nn[i]-1+j,2]= rexp(1)*(exp(parm[10]+parm[11]*yjmin1))
uij=exp(mu[i]+raw[nn[i]-1+j,2]*parm[9]+yjmin1*parm[8])
raw[nn[i]-1+j,1]=rpois(1,uij)
raw[nn[i],3]=i
```

```
raw[nn[i]-1+j,3]=i
} #j
}#i
result=list(raw=raw,nn=nn,ndesign=ndesign)
result
} #outcome

############################################################
#            Log-Likelihood Function                       #
############################################################

loglikfn<- function(parms){
y1=y[nn,1] #initial obs for every subjects
mu=xmatrix %*% parms[1:7]
ui1=exp(mu)
f1=sum(y1*log(ui1)-ui1-log(factorial(y1)))

f2=0;f3=0
for( i in 1:m){
yi=y[(y[,3]==i), 1] # all obs for ith subject
ti=y[(y[,3]==i), 2] # all time points for ith subject

yi1=yi[-ndesign[i]] #previous obs
tti=ti[-1] #current time
yi2=yi[-1] #current obs
uij=exp(mu[i]+parms[9]*tti+parms[8]*yi1)
f2=sum(yi2*log(uij)-uij-log(factorial(yi2)))+f2

f3=sum(parms[10]+parms[11]*yi1-exp(parms[10]+parms[11]*yi1)*tti)+f3
} #i
(m+f1+f2+f3)
} #loglikfn

############################################################
#         Simulation                                       #
############################################################

Pschem = 1  # parameter setting, 1 to 6
m=36
num = 2 # design structure 1(10,10), 2(5,3),3(10,5),4(20,6)
rep=1000 #number of replications
out = matrix(NA,rep,ncol(parameter))
parm = parameter[Pschem,]*0.1
xmatrix=design(level=c(3,3),m=m,p=2)
```

```
for (r in 1:rep){
result=outcome(m=m,num=num,parm=parm)
y=result$raw
nn=result$nn
ndesign=result$ndesign
mle=maxLik(logLik = loglikfn, start = parm)
diff=coef(mle)-parm
out[r,]=(sqrt(sum(ndesign))*diff)/summary(mle)$estimate[,2]
}

hzTest(out[,1:9])   #Outcome process
hzTest(out[,10:11])  #Time process
```

4. Gamma-Exponential Model

```
#####################################################################
#         Parameter Setting (Pscheme: 1 to 6)                       #
#####################################################################

parameter = matrix(c(10,10,10,12,12,12,          #1:v
                   0.4,0.4,0.4,0.4,0.4,0.4,      #2:beta0
                   0.2,0.2,0.2,0.2,0.2,0.2,      #3:beta1
                   0.3,0.3,0.3,0.3,0.3,0.3,      #4:beta2
                   0.1,0.1,0.1,0.1,0.1,0.1,      #5:beta3
                   0.3,0.3,0.3,0.3,0.3,0.3,      #6:beta4
                   0.4,0.4,0.4,0.4,0.4,0.4,      #7:beta5
                   0.9,0.9,0.9,0.9,0.9,0.9,      #8:beta6
                   0.8,0.8,0.8,0.0,0.0,0.8,      #9:phi
                   0.1,0.1,0.1,0.1,0.1,0.1,      #10:gamma
                   2,1,2,1,2,1,                  #11:alpha
                   0.01,0.02,0.01,0.02,0.01,0.02),#12:delta
                   nrow=6)

#####################################################################
#                Create Data:  c('outcome','time','subject')   #
#####################################################################

outcome<- function(m=m,num=num,parm=parm){
if (num == 1) {n1 = 10; n2=10}
if (num == 2) {n1 = 5; n2=3}
if (num == 3) {n1 = 10; n2=5}
if (num == 4) {n1 = 20; n2=6}

ndesign = matrix(c(rep(n1,m/2),rep(n2,m/2)),byrow=T)
nn=cumsum(c(1,ndesign[-length(ndesign)]))
raw = matrix(NA,sum(ndesign),3)  #Null matrix
mu=xmatrix %*% parm[2:8]

ui1=exp(mu)
raw[nn,1]=rgamma(n=m,shape=ui1*parm[1],scale=parm[1])
raw[nn,2] = rexp(m)

for (i in 1:m){
for (j in 2:ndesign[i]){
yjmin1 = raw[nn[i]-1+j-1,1]
raw[nn[i]-1+j,2]= rexp(1)*(exp(parm[11]+parm[12]*yjmin1))

uij=exp(mu[i]+raw[nn[i]-1+j,2]*parm[10]+yjmin1*parm[9])
```

```
raw[nn[i]-1+j,1]=rgamma(1,shape=uij*parm[1],scale=parm[1])
raw[nn[i],3]=i
raw[nn[i]-1+j,3]=i
} #j
}#i
result=list(raw=raw,nn=nn,ndesign=ndesign)
result
} #outcome

############################################################
#            Log-Likelihood Function                       #
############################################################

loglikfn<- function(parms){
y1=y[nn,1] #initial obs for every subjects
mu=xmatrix %*% parms[2:8]
ui1=exp(mu)
f1=sum((-y1/ui1-log(ui1))*parms[1]+parms[1]*log(parms[1])+
(parms[1]-1)*log(y1)-log(gamma(parms[1])))

f2=0;f3=0
for(i in 1:m){
yi=y[(y[,3]==i), 1] # all obs for ith subject
ti=y[(y[,3]==i), 2] # all time points for ith subject
yi1=yi[-ndesign[i]] #previous obs
tti=ti[-1] #current time
yi2=yi[-1] #current obs
uij=exp(mu[i]+parms[10]*tti+parms[9]*yi1)
f2=sum((-yi2/uij-log(uij))*parms[1]+parms[1]*log(parms[1])+
(parms[1]-1)*log(yi2)-log(gamma(parms[1])))+f2

f3=sum(parms[11]+parms[12]*yi1-
exp(parms[11]+parms[12]*yi1)*tti)+f3
} #i
(m+f1+f2+f3)
} #loglikfn

############################################################
#         Simulation                                       #
############################################################

Pschem = 1  # parameter setting, 1 to 6
m=18
num = 2 # design structure 1(10,10), 2(5,3),3(10,5),4(20,6)
rep=1000 #number of replications
```

```
out = matrix(NA,rep,ncol(parameter))
parm = parameter[Pschem,]*0.1
xmatrix=design(level=c(3,3),m=m,p=2)

for (r in 1:rep){
result=outcome(m=m,num=num,parm=parm)
y=result$raw
nn=result$nn
ndesign=result$ndesign
mle=maxLik(logLik = loglikfn, start = parm)
diff=coef(mle)-parm
out[r,]=sqrt(sum(ndesign))*diff/summary(mle)$estimate[,2]
}

hzTest(out[,1:9])   #Outcome process
hzTest(out[,10:11])  #Time process
```

Appendix D

R PROGRAM OF THE JOINT MODELS

```
#####################################################################
#                        Joint Model
#####################################################################
# DESCRIPTION:
#  JointModel is used to fit longitudinal outcomes and informative
#  time under the assumptions that the current outcome is dependent
#  on the one-step prior outcome, and that time follows an
#  exponential  distribution. The outcome distributions that can be
#  analyzed by this function are the Gaussian, Bernoulli, Poisson,
#  and Gamma distributions. The function computes the effects of
#  the prior outcome, current time, alpha, and delta; therefore,
#  these terms do not need to be specified in the model.
#
#  USAGE:
#  JointModel(formula, data, id, time, family, link = log)
#   - formula: a symbolic description of the model to be fitted.
#   - data:    dataset name to be analyzed
#   - id:      id variable name in the dataset to be analyzed
#   - time:    time variable name in a dataset to be analyzed
#   - family:  distribution of outcome variable,
#              such as gaussian, bernoulli, poisson, and gamma.
#   - link:    name of the link function. The default link function
#              for gaussian is identity, logit for bernoulli, log for
#              poisson and gamma. The inverse or identify can be used
#              for gamma.
#####################################################################

#########################################################
#        A Example of Data Structure to be analyzed
#########################################################
# data with two continous vars and one factor
# subj: id, y:outcome, t:time, f1:factor, c1 & c2: continuous vars
#
#   subj y         t f1          c1          c2
#1     1 0 0.4352370  1  0.13333636  1.08576936
#2     1 0 0.2634203  1  0.13333636  1.08576936
#3     1 1 1.4733225  1  0.13333636  1.08576936
#4     1 1 0.7620299  1  0.13333636  1.08576936
#5     2 0 1.2376036  1  0.80418951 -0.69095384
#6     2 1 4.4239342  1  0.80418951 -0.69095384
#7     2 1 1.0545432  1  0.80418951 -0.69095384
#8     2 1 1.0352439  1  0.80418951 -0.69095384
#9     3 1 1.8760352  2 -0.05710677 -1.28459935
#10    3 0 0.6547466  2 -0.05710677 -1.28459935
#########################################################
```

```
JointModel <- function(formula, data, id, time, family, link = log) {
wants <- c('maxLik', 'formula.tools', 'MASS')
has <- wants %in% rownames(installed.packages())
if (any(!has)) install.packages(wants)
library(maxLik)
library(formula.tools)
library(MASS)

arguments <- as.list(match.call())
id <- eval(arguments$id, data)
time <- eval(arguments$time, data)
get.names <- get.vars(formula)

id.time <- data.frame(cbind(id=id,time=time))
df <- data[,get.names]
df <- cbind(id.time, df)
df <- df[complete.cases(df),]

df[,1] <- as.numeric(as.factor(df[,1])) #1=id
xdesign <- model.matrix(formula, data = df)
coef.names <- colnames(xdesign)
ndesign <- as.numeric(table(df[,1])) #1=id
m <- tail(df[,1],1) #1=id
nn <- cumsum(c(1,ndesign[-length(ndesign)]))

df$lag.out <- c(NA, df[-nrow(df), 3])
for (i in 1:(nrow(df)-1)) {
if(df[,1][i]!= df[,1][i+1]) df$lag.out[i+1] <- NA
}

time.initial <- lm(df[,3]~lag.out,data=df) #3=outcome
time.coef <- as.numeric(coef(time.initial))

x.name <- c(get.names[-1],'lag.out','df[,2]')
ff <- as.formula(paste('df[,3]~',paste(x.name,collapse='+')))
out.initial <- lm(ff,data=df)
out.coef <- as.numeric(coef(out.initial))

if (arguments$family == 'poisson'| arguments$family== 'bernoulli') {
initial <- c(out.coef, time.coef)
} else if (arguments$family== 'gaussian') {
initial <- c(out.coef, time.coef, sd(df[,3]))
} else {
v <- fitdistr(df[,3], "Gamma")
```

```
v <- as.numeric(a$estimate[1])
initial <- c(out.coef,time.coef, v)
}

if(arguments$family== 'gaussian') {
LikeFn <- function(parms) {
y1= df[nn,as.character(lhs(formula))]

mu=xdesign %*% parms[1:ncol(xdesign)]
f1=sum(-0.5 * log(parms[ncol(xdesign)+5]^2)-0.5*
(y1-mu)^2/parms[ncol(xdesign)+5]^2)

f2=0;f3=0
for (i in 1:m) {
yi=df[(df[,'id']==i), as.character(lhs(formula))] #1=id
ti=df[(df[,'id']==i), as.character(lhs(formula))] #1=id

yi1=yi[-ndesign[i]] #previous obs
tti=ti[-1] #current time
yi2=yi[-1] #current obs

f2=sum(-0.5 * log(parms[ncol(xdesign)+5]^2)-0.5 *
(yi2-parms[ncol(xdesign)+2]*tti-parms[ncol(xdesign)+1]*
yi1-mu[i])^2/parms[ncol(xdesign)+5]^2)+f2
f3=sum(parms[ncol(xdesign)+3]+parms[ncol(xdesign)+4]*yi1-
exp(parms[ncol(xdesign)+3]+ parms[ncol(xdesign)+4]*yi1)*tti)+f3
} #i
(m+f1+f2+f3)
}  # LikeFn
} #gaussian

if(arguments$family== 'bernoulli') {
LikeFn <-function(parms) {
y1 <- df[nn, as.character(lhs(formula))]

mu <- xdesign %*% parms[1:ncol(xdesign)]
ui1 <- 1/(1+exp(-mu))
f1 <- sum(y1*(log(ui1)-log(1-ui1))+log(1-ui1))

f2 <- 0; f3 <- 0
for(i in 1:m) {
yi <- df[(df[,'id']==i), as.character(lhs(formula))] #1=id
ti <- df[(df[,'id']==i), 2]
yi1 <- yi[-ndesign[i]] #previous obs
tti <- ti[-1] #current time
```

```
yi2 <- yi[-1] #current obs

uij <- 1/(1+exp(-(mu[i]+parms[ncol(xdesign)+2]*tti+
parms[ncol(xdesign)+1]*yi1)))
f2 <- sum((1-yi2)*log(1-uij)+ yi2*log(uij))+f2
f3 <- sum(parms[ncol(xdesign)+3]+parms[ncol(xdesign)+4]*yi1-
exp(parms[ncol(xdesign)+3]+parms[ncol(xdesign)+4]*yi1)*tti)+f3
} #i
(m+f1+f2+f3)
} #LikeFn
} #binomial

if(arguments$family== 'poisson') {
LikeFn <- function(parms) {
y1 <- df[nn, as.character(lhs(formula))]
mu <- xdesign %*% parms[1:ncol(xdesign)]
ui1 <- exp(mu)
f1 <- sum(y1*log(ui1)-ui1-log(factorial(y1)))

f2 <- 0; f3 <- 0
for (i in 1:m) {
yi <- df[(df[,'id']==i), as.character(lhs(formula))]
ti <- df[(df[,'id']==i), 2] #2=time

yi1 <- yi[-ndesign[i]] #previous obs
tti <- ti[-1] #current time
yi2 <- yi[-1] #current obs

uij <- exp(mu[i]+parms[ncol(xdesign)+2]*tti+
parms[ncol(xdesign)+1]*yi1)
f2 <- sum(yi2*log(uij)-uij-log(factorial(yi2)))+f2
f3 <- sum(parms[ncol(xdesign)+3]+parms[ncol(xdesign)+4]*yi1-
exp(parms[ncol(xdesign)+3]+
parms[ncol(xdesign)+4]*yi1)*tti)+f3
} #i
(m+f1+f2+f3)
} #LikeFn
} #poisson

if(arguments$family== 'gamma') {
LikeFn <-function(parms) {
y1 <- df[nn, as.character(lhs(formula))]
mu <- xdesign %*% parms[1:ncol(xdesign)]
ui1 <- exp(mu)
if (arguments$link =='inverse') ui1 <- 1/mu
```

```
if (arguments$link ='indentity') ui1 <- mu
f1 <- sum((-y1/ui1-log(ui1))*parms[ncol(xdesign)+5]+
parms[ncol(xdesign)+5]*log(parms[ncol(xdesign)+5])+
(parms[ncol(xdesign)+5]-1)*log(y1)-
log(gamma(parms[ncol(xdesign)+5])))

f2 <- 0; f3 <- 0
for (i in 1:m) {
yi <- df[(df[,'id']==i), as.character(lhs(formula))]
ti <- df[(df[,'id']==i), time]

yi1 <- yi[-ndesign[i]] #previous obs
tti <- ti[-1] #current time
yi2 <- yi[-1] #current obs
uij <- exp(mu[i]+parms[ncol(xdesign)+2]*tti+
parms[ncol(xdesign)+1]*yi1)
if (arguments$link=='inverse') uij <- 1/uij
if (arguments$link=='indentity') uij <- uij
f2 <- sum((-yi2/uij-log(uij))*parms[ncol(xdesign)+5]+
parms[ncol(xdesign)+5]*log(parms[ncol(xdesign)+5])+
(parms[ncol(xdesign)+5]-1)*log(yi2)-
log(gamma(parms[ncol(xdesign)+5])))+f2

f3 <- sum(parms[ncol(xdesign)+3]+parms[ncol(xdesign)+4]*yi1-
exp(parms[ncol(xdesign)+3]+
parms[ncol(xdesign)+4]*yi1)*tti)+f3
} #i
(m+f1+f2+f3)
} #LikeFn
} #gamma

if(arguments$family== 'gamma') {
mlefull <- maxLik(logLik=LikeFn,start=initial)
betas <- c(coef.names, 'prior.outcome','current.time',
'alpha','delta','shape')
names(mlefull$estimate) <- betas

mlered <- maxLik(logLik=LikeFn,
start=c(initial[1],rep(0,ncol(xdesign)+3),v),
activePar=c(T,rep(F,ncol(xdesign)+1),rep(T,2),T)
)
} else if (arguments$family=='gaussian') {
mlefull <- maxLik(logLik=LikeFn,start=initial)
betas <- c(coef.names,'prior.outcome','current.time',
'alpha','delta','sigma')
```

```
names(mlefull$estimate) <- betas

mlered <- maxLik(logLik=LikeFn,
start=c(initial[1],rep(0,ncol(xdesign)+3),
initial[length(initial)]),
activePar=c(T,rep(F,ncol(xdesign)+1), rep(T,2),T)
)
} else {
mlefull <- maxLik(logLik=LikeFn,start=initial)
betas <- c(coef.names,'prior.outcome','current.time','alpha','delta')
names(mlefull$estimate) <- betas

mlered <- maxLik(logLik=LikeFn,
start=c(initial[1],rep(0,ncol(xdesign)+3)),
activePar=c(T,rep(F,ncol(xdesign)+1), rep(T,2))
)
}

parm <- summary(mlefull)
est <- summary(mlefull)$estimate
AIC <- AIC(mlefull)
AICc <- AIC+2*parm$NActivePar*(parm$NActivePar+1)/
(m-parm$NActivePar-1)
BIC <- -2*parm$loglik+parm$NActivePar*log(m)

ratio <- 2*(logLik(mlefull)-logLik(mlered))
dfred <- summary(mlered)$NActivePar
dffull <- summary(mlefull)$NActivePar
dfchi <- dffull-dfred
Pr <- 1-pchisq(ratio,dfchi)

list(Model = paste('Call: ', formula )),
Coefficients = est,
AIC = AIC, AICc = AICc, BIC = BIC,
LogLik = ratio, LogLikPval = Pr)
} #JointModel
```