

6-7-2020

Joint Models of Longitudinal Outcomes and Informative Time

JangDong Seo

Indiana University, Bloomington, jangseo@iu.edu

Follow this and additional works at: <https://digitalcommons.wayne.edu/jmasm>



Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Seo, JangDong (2020) "Joint Models of Longitudinal Outcomes and Informative Time," *Journal of Modern Applied Statistical Methods*: Vol. 18 : Iss. 1 , Article 32.

DOI: 10.22237/jmasm/1556670000

Available at: <https://digitalcommons.wayne.edu/jmasm/vol18/iss1/32>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

Joint Models of Longitudinal Outcomes and Informative Time

JangDong Seo

Indiana University Bloomington
Bloomington, IN

Longitudinal data analyses commonly assume that time intervals are predetermined and have no information regarding the outcomes. However, there might be irregular time intervals and informative time. Presented are joint models and asymptotic behaviors of the parameter estimates. Also, the models are applied for real data sets.

Keywords: longitudinal data, informative time, joint model, exponential family, outcome and time processes, asymptotic multivariate normality, model selection

Introduction

Longitudinal studies are commonly designed in many research fields in order to see changes over time in which time intervals are prearranged and shared by all participants. The time points are assumed to be independent of responses, providing no information with respect to measurements. In informative schedule data, however, time points are highly related to prior outcomes; therefore, time intervals become irregular, and all subjects have their own unique time points. Due to these reasons, the assumptions of traditional methods are not satisfied.

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 with any other factors that the researcher is interested in (Henderson, Diggle, & Dobson, 2000; Kim, Zeng, Chambless, & Li, 2012; Liang, Lu, & Ying, 2009; Lin & Scharfstein, 2004; Lipsitz, Fitzmarice, Ibrahim, Gelber, & Lipshultz, 2002; Qiu, Stein, & Elston, 2013; Wu, Lin, Yi, & Huang, 2012). 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. Outcomes are measured only at the dropout time.

Lipsitz et al. (2002) presented a joint model for longitudinal data. The joint model assumes time points are not fixed and dependent on previous outcomes, and the repeated measurements are supposed to follow a multivariate normal distribution. It can be applied to normally distributed responses only. Bronsert (2009) presented a joint model of a longitudinal process and informative time schedule data. Lin (2011) studied the validity of the joint model by checking the asymptotic multivariate normality of the maximum likelihood estimators (MLEs) of the joint model, when outcomes are normally distributed and time is exponentially distributed.

The model presented by Bronsert (2009) and Lin (2011), however, is limited to normal distribution; therefore, the current study has two goals. The first is generalizing the joint model to fit members of the exponential family of distributions for outcomes, while assuming that time follows an exponential distribution. For instance, the Poisson, Bernoulli, and Gamma probability density functions are combined with the exponential distribution of time, including the relationship between the outcome and time as well for count data, binary data, and waiting time or survival time, respectively.

The second goal is computing the likelihood ratio test statistic and model selection criteria, such as the Akaike information criterion (AIC), the Akaike information criterion with correction (AICc), and the Bayesian information (BIC), for model comparisons. The underlying assumption of the joint models is time is informative, which means the current time point for collecting an outcome is dependent on the one-step prior outcome. Therefore, time intervals become irregular from subject to subject, and the number of measurements are different across all subjects.

These circumstances are relatively common in public health. For instance, if a patient's health outcome is poor, the doctor will ask the patient to visit the clinic more often than a patient with a good health outcome, because the former patient needs more intensive care. In this situation, all subjects do not share the common time points for checkups as assumed in the traditional methods, such as repeated measures analysis of variance, multivariate analysis of variance, or mixed-effects model. Hence, as an alternative, the current joint models are developed because the results of analyses using the traditional methods will be biased.

Methodology

The Likelihood Function and Parameter Estimation

Three outcome distributions are selected as examples from the exponential family of distributions: the Bernoulli for binary outcome, the Poisson for count data, and the Gamma for waiting time or survival time. So, the presented joint models are named as Bernoulli-Exponential, Poisson-Exponential, and Gamma-Exponential models.

Let y_{ij} be the outcome for the i th subject measured at the j th time point when $j = 1, 2, \dots, n_i$; therefore, the i th subject has $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})'$ collected at $\mathbf{t}_i = (t_{i1}, t_{i2}, \dots, t_{in_i})'$. Then the joint distribution of outcomes and time points given some explanatory variables \mathbf{x}_i becomes

$$f_{\Theta}(\mathbf{y}_i, \mathbf{t}_i | \mathbf{x}_i) = f_{\Theta}(\mathbf{y}_i | \mathbf{t}_i, \mathbf{x}_i) \cdot f_{\Theta}(\mathbf{t}_i | \mathbf{x}_i), \quad (1)$$

where Θ is a vector of unknown parameters. A general model can be assumed for the joint distribution of \mathbf{y}_i and \mathbf{t}_i . Consequently, the general model under the assumption that the current outcome is dependent on the one-step prior outcome (y_{ij-1}) and current time point (t_{ij}) becomes

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

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. To derive joint models, the concept of the generalized linear model is adopted, because the generalized linear model provides a unified class of models of regression analysis, regardless of discrete or continuous outcomes (Dobson, 2001; Fitzmaurice, Laird, & Ware, 2004; McCullagh & Nelder, 1989; Nelder & Wedderburn, 1972). Generalized linear model can be written with p explanatory variables as

$$g(\mu_i) = \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}. \quad (3)$$

In the current joint models, this expression is modified to take into account of the relationship of current time and one-step prior outcome, such as

$$g(\mu_i) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \gamma t_{ij} + \varphi y_{ij-1} = \mathbf{X}'_i \beta + \gamma t_{ij} + \varphi y_{ij-1}, \quad (4)$$

where γ is the effect of current time on the mean response, and φ is the effect of the previous outcome on the mean response of the current outcome. Using the generalized linear model form, two different types of linear predictors are used:

$$g(\mu_i) = \mathbf{X}'_i \beta \text{ or } g(\mu_i) = \mathbf{X}'_i \beta + \gamma t_{ij} + \varphi y_{ij-1} \quad (5)$$

where the former link function is used for the mean for the first observation of the i th subject, and the latter is for the j th mean of the i th subject to incorporate the effects of the current time and the one-step prior outcome. Hence, the mean functions can be expressed as, for instance,

$$\mu_{i1} = \exp(\mathbf{X}'_i \beta) \text{ or } \mu_{ij} = \exp(\mathbf{X}'_i \beta + \gamma t_{ij} + \varphi y_{ij-1}), \quad (6)$$

if the log link function is used. To find the MLEs, a nonlinear optimization function, named the `maxLik` in R, is used (Henningsen & Ott, 2010; R Core Team, 2014). In simulation studies, it is tested whether parameter estimators of the joint models follow asymptotic normality as the number of observations increases,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \rightarrow N(\mathbf{0}, \mathbf{I}_{\boldsymbol{\theta}_0}^{-1}), \quad (7)$$

where $\boldsymbol{\theta}$ is the unknown parameter, $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$, $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$, and $\mathbf{I}_{\boldsymbol{\theta}_0}^{-1}$ is the Fisher information matrix. Then, the models are applied to a real-life situation, and three information criteria are used to compare models.

Bernoulli-Exponential Model

This model is for binary data with informative time. The likelihood function for m individuals is expressed as

JOINT MODELS WITH INFORMATIVE TIME

$$L(\Theta, y_1, \dots, y_m) = \prod_{i=1}^m \left\{ \prod_{j=2}^{n_i} \exp \left(y_{i1} \log \left(\frac{\mu_{i1}}{1 - \mu_{i1}} \right) + \log(1 - \mu_{i1}) \right) \cdot \exp \left(\alpha + \delta_{y_{ij-1}} \right) \cdot \exp \left(-e^{\alpha + \delta y_{ij-1}} t_{ij} \right) \right\}. \quad (8)$$

The mean functions, using the logit link function, are

$$\mu_{i1} = \frac{\exp(X'_i \beta)}{1 + \exp(X'_i \beta)} \quad \text{and} \quad \mu_{ij} = \frac{\exp(X'_i \beta + \gamma t_{ij} + \varphi y_{ij-1})}{1 + \exp(X'_i \beta + \gamma t_{ij} + \varphi y_{ij-1})}, \quad (9)$$

where μ_{i1} and μ_{ij} are the means for the first and the j th observation with $j = 2, 3, \dots, n_i$, for the i th subject and y_{i1} and y_{ij} are the first and the j th observation with $j = 2, 3, \dots, n_i$, for the i th subject. The parameters of α and δ in the likelihood function represent the relationship between the prior outcome and current time. By using the notation of n_i , each subject can have a different number of measurements. The coefficients of γ and φ represent the effects of the current time and previous outcome on the mean value, and the X'_i is the i th column of the design matrix used when estimating the parameters of the models. The design matrix can hold categorical variables and covariates, if any. Within the design matrix, a column can be added representing each categorical variable and covariate. Thus, the log-likelihood function for m individuals can be written as the sum of m individuals' log-likelihood functions,

$$\begin{aligned} l &= \sum_{i=1}^m l_i = \sum_{i=1}^m \left(y_{i1} \log \left(\frac{\mu_{i1}}{1 - \mu_{i1}} \right) + \log(1 - \mu_{i1}) \right) \\ &+ \sum_{i=1}^m \sum_{j=2}^{n_i} \left(y_{i1} \log \left(\frac{\mu_{ij}}{1 - \mu_{ij}} \right) + \log(1 - \mu_{ij}) \right) \\ &+ \sum_{i=1}^m \sum_{j=2}^{n_i} \left(\alpha + \delta y_{ij-1} - e^{\alpha + \delta y_{ij-1}} t_{ij} \right). \end{aligned} \quad (10)$$

The function can be separated into two parts: one with the first two rows and the other with the last row. The former is the likelihood for outcome process, which represents the relationships among the current time, prior outcome, and covariates. The latter is the likelihood for time process, which is the relationships between the current time and prior outcome. Because the time process does not have any β terms, the time process can be ignored, while maximizing the log-likelihood function with respect to β , or vice versa. Therefore, in the simulation studies, the multivariate normality test is conducted separately for the outcome and time processes. This concept is applied to the other joint models.

Poisson-Exponential Model

This model is developed to handle count outcomes and informative time. The likelihood function of the model for m individuals can be written as

$$L(\Theta, y_1, \dots, y_m) = \prod_{i=1}^m \left\{ \begin{array}{l} \exp(y_{i1} \log(\mu_{i1}) - \mu_{i1} - \log(y_{i1}!)) \\ \times \prod_{j=2}^{n_i} \exp(y_{ij} \log(\mu_{ij}) - \mu_{ij} - \log(y_{ij}!)) \\ \cdot \exp(\alpha + \delta_{y_{ij-1}}) \cdot \exp(-e^{\alpha + \delta_{y_{ij-1}}} t_{ij}) \end{array} \right\}, \quad (11)$$

where the mean functions, using the log link function, for the initial and the j th outcome of the i th subject are expressed as

$$\mu_{i1} = \exp(X_i' \beta) \text{ and } \mu_{ij} = \exp(X_i' \beta + \gamma t_{ij} + \phi y_{ij-1}). \quad (12)$$

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

$$\begin{aligned} l &= \sum_{i=1}^m l_i = \sum_{i=1}^m (y_{i1} \log(\mu_{i1}) - \mu_{i1} - \log(y_{i1}!)) \\ &\quad + \sum_{i=1}^m \sum_{j=2}^{n_i} (y_{ij} \log(\mu_{ij}) - \mu_{ij} - \log(y_{ij}!)) \\ &\quad + \sum_{i=1}^m \sum_{j=2}^{n_i} (\alpha + \delta_{y_{ij-1}} - e^{\alpha + \delta_{y_{ij-1}}} t_{ij}). \end{aligned} \quad (13)$$

Gamma-Exponential Model

This model is used for waiting or survival time and informative time. The likelihood function becomes the product of the density functions for all individuals:

$$L(\Theta, y_i, \dots, y_m) = \prod_{i=1}^m \left\{ \exp \left(\frac{y_{i1}(1/\mu_{i1}) - \log(\mu_{i1})}{-1/v_{i1}} + v_{i1} \log(v_{i1}) \right) \right. \\ \left. \times \prod_{j=2}^m \exp \left(\frac{y_{ij}(1/\mu_{ij}) - \log(\mu_{ij})}{-1/v_{ij}} + v_{ij} \log(v_{ij}) \right) \right\} \cdot \exp(\alpha + \delta y_{ij-1}) \cdot \exp(-e^{\alpha + \delta y_{ij-1}} t_{ij}) \quad (14)$$

with the mean functions, using the log link function, are

$$\mu_{i1} = \exp(X_i' \beta) \quad \text{and} \quad \mu_{ij} = \exp(X_i' \beta + \gamma t_{ij} + \phi y_{ij-1}). \quad (15)$$

The log link function is used to avoid any negative mean values. However, if the design allows negative means, the identity or inverse link function can be used. Then, the log-likelihood function for all individuals can be written as

$$\begin{aligned}
 l = \sum_{i=1}^m l_i = \sum_{i=1}^m & \left(\frac{y_{i1}(1/\mu_{i1}) - \log(\mu_{i1})}{-1/\nu_{i1}} + \nu_{i1} \log(\nu_{i1}) \right) \\
 & + \sum_{i=1}^m \sum_{j=2}^{n_i} \left(\frac{y_{ij}(1/\mu_{ij}) - \log(\mu_{ij})}{-1/\nu_{ij}} + \nu_{ij} \log(\nu_{ij}) \right) \\
 & + \sum_{i=1}^m \sum_{j=2}^{n_i} \left(\alpha + \delta y_{ij-1} - e^{\alpha + \delta y_{ij-1}} t_{ij} \right).
 \end{aligned} \tag{16}$$

Parameter Estimation

The log-likelihood function is commonly used to find the MLEs for convenience. Generally, computing the MLEs needs an iterative method that can be performed in the statistical computing software (Fitzmaurice et al., 2004). Many nonlinear optimization methods are provided in commonly used statistical computing software, such as SAS or R (Nash & Varadhan, 2011; SAS Institute, 2008). However, none of them seems to be superior to any other methods in any situations. In this study, a non-linear optimization routine, named the `maxLik` function in R, is chosen because it provides a single, unified interface to a variety of optimization routines, providing methods handling the likelihood-specific properties of the estimates, including standard errors (Henningsen & Ott, 2010; R Core Team, 2014). The Newton-Raphson maximization algorithm is used by default in the `maxLik` function.

Parameter Testing and Model Selection

There are three ways of using the likelihood function: (1) the Wald test, (2) the score test, and (3) the likelihood ratio test. They are used for hypothesis testing, to determine the significance of the parameter estimators or to determine the confidence intervals (Agresti, 2007). The likelihood ratio test uses the ratio of two maximized log-likelihood functions for comparing two nested models: (1) the maximized log-likelihood value for the null hypothesis denoted by \hat{l}_{red} for the reduced model and (2) the maximized log-likelihood value for the alternative hypothesis denoted by \hat{l}_{full} . Hence, if the MLEs are asymptotically normal, the likelihood ratio test statistic used is

JOINT MODELS WITH INFORMATIVE TIME

$$2\left(\hat{l}_{full} - \hat{l}_{red}\right) \sim \chi_{df_{full}-df_{red}}^2, \quad (17)$$

where df_{full} is the degrees of freedom for the full model and df_{red} is the degrees of freedom for the reduced model. Then, the value is compared to a 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 the reduced model is inappropriate (Fitzmaurice et al., 2004).

The likelihood ratio test is used because the log-likelihood functions for the joint models are already found, so it is easy to extract the information. Model selection criteria computed are the AIC, the AICc, and the BIC to compare nested and non-nested models. The AIC measures a relative quality of a model to provide a method on model selection with given data. The AIC is defined as

$$AIC = 2k - 2\ln(L), \quad (18)$$

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. A 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, because $\ln(L)$ is the maximized log-likelihood at $\hat{\Theta}$, the AIC of the Bernoulli- Exponential model becomes

$$AIC = 2k - 2 \left\{ \begin{array}{l} \sum_{i=1}^m y_{i1} \mathbf{X}'_i \hat{\beta} + \log \left(\frac{1}{1 + \exp(\mathbf{X}'_i \hat{\beta})} \right) \\ + \sum_{i=1}^m \sum_{j=2}^{n_i} \left(y_{ij} \left(\mathbf{X}'_i \hat{\beta} + \hat{\gamma} t_{ij} + \hat{\phi} y_{ij-1} \right) \right. \\ \left. + \log \left(1 + \frac{\exp(\mathbf{X}'_i \hat{\beta} + \hat{\gamma} t_{ij} + \hat{\phi} y_{ij-1})}{1 + \exp(\mathbf{X}'_i \hat{\beta} + \hat{\gamma} t_{ij} + \hat{\phi} y_{ij-1})} \right) \right) \\ \left. + \sum_{i=1}^m \sum_{j=2}^{n_i} \left(\hat{\alpha} + \hat{\delta} y_{ij-1} - e^{\hat{\alpha} + \hat{\delta} y_{ij-1}} t_{ij} \right) \right\}. \quad (19)$$

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

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1}, \quad (20)$$

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

BIC is given by

$$BIC = -2\ln(L) + k\log(n), \quad (21)$$

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. However, the use of BIC is not recommended due to a high risk of selecting a too simple model (Fitzmaurice et al., 2004). Nevertheless, in a simulation study, BIC outperformed AIC as sample size increased (McQuarrie, Shumway, & Tsai, 1997). Therefore, BIC is included as a model selection criterion because this study uses the log-likelihood function to find parameter estimators.

Simulation

The purpose of the simulation is to verify the asymptotic multivariate normality of the MLEs of the joint models as the number of observations increases. To examine the property of the MLEs of the joint models, the Bernoulli, Poisson, and Gamma distributions are selected as explained before.

To simulate data, six parameter schemes, five sample sizes (18, 36, 54, 90, and 180), and four different numbers of measurements are combined together in three selected distributions. Therefore, simulations were implemented with 360 conditions ($6 \times 5 \times 4 \times 3$) with 1,000 replications each. This simulation design was developed by Bronsert (2009) and Lin (2011) and adopted in this study to maintain consistency. With 1,000 sets of estimators in each condition, the Henze-Zirkler multivariate normality test statistic is computed for the outcome and time processes, respectively, to test the asymptotic normality of the estimators by using the `maxLik` function. To increase practical settings, two categorical variables with three levels each and two continuous variables were included when creating the design matrix. As described in Table 1, in each distribution, 20 schemes were designed with different sample sizes and design structures, resulting in different numbers of total measurements. For example, in scheme 1, 10 responses were simulated from 18

JOINT MODELS WITH INFORMATIVE TIME

subjects, resulting in 180 observations in total. In scheme 2, 5 responses were generated from 9 subjects and 3 responses from the remaining 9 subjects, resulting in 72 outcomes in total.

Table 1. Simulation Designs

Scheme Num	Sample Size	Num. of Observations	Design Structure	Total Num. of Observations
1	18	10	Balanced	180
2		5 & 3	Unbalanced	72
3		10 & 5	Unbalanced	135
4		20 & 6	Unbalanced	234
⋮	⋮	⋮	⋮	⋮
17	180	10	Balanced	1800
18		5 & 3	Unbalanced	720
19		10 & 5	Unbalanced	1350
20		20 & 6	Unbalanced	2340

During the simulation, the first outcome was 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. For instance, in the Bernoulli-Exponential model, a first binary outcome was generated, then a second outcome was generated based on the fixed parameter values, such as $\beta_0 = 0.4$, $\beta_1 = 0.2$, $\beta_2 = 0.3$, $\beta_3 = 0.1$, $\beta_4 = 0.3$, $\beta_5 = 0.4$, $\beta_6 = 0.9$, $\varphi = 0.8$, $\gamma = 0.1$, $\alpha = 2$, and $\delta = 0.01$. Then, it is tested if $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \rightarrow N(\mathbf{0}, \mathbf{I}_{\boldsymbol{\theta}_0}^{-1})$.

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 of the simulation results are displayed in Figure 1 through Figure 3. As can be seen, the outcome and the time processes obtain the asymptotic multivariate normality in three selected outcome distributions, when sample size, in general, goes beyond about 18. Also, it is found that the data structures, unbalanced or balanced, do not affect the property of the asymptotic normality. In addition, the time process tends to obtain the normality faster than the outcome process.

Analysis of Bladder Cancer Data

The proposed joint models are applied to the bladder cancer data provided in R, and the outputs of the analyses are presented in this section. R provides two different datasets, `bladder` and `bladder1` in the package called `survival` (Andrews & Herzberg, 1985; Wei, Lin, & Weissfeld, 1989). The `bladder1` is the full dataset with 118 subjects, and the `bladder` is a subset of the `bladder1` with 85 subjects and the reduced number of variables. The bladder cancer dataset has been studied by many methodologists, such as Cai, Wenbin, & Zhang (2012), Sun & Wei (2000), Sun, Park, Sun, & Zhao, (2005), and Zhang (2002). The `bladder` dataset is most commonly used by many researchers for recurrent event modeling (Andrews & Herzberg, 1985; Wei et al., 1989). The variable “stop” in both datasets measures the time interval (in months) because 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. However, two treatment types are included in the `bladder`, and three treatments in the `bladder1`. The `bladder` dataset is used for the Bernoulli-Exponential model with the variable “rx” as an outcome variable, and the `bladder1` dataset with the variable “treatment” as an outcome variable is applied to the Poisson-Exponential model.

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 are recorded. And, the status variable “event” for recurrence of tumors has 1 for recurrence and 0 for everything else (including death for any reason). Therefore, the Bernoulli-Exponential model is applied with the “event” as an outcome variable. A chosen research interest is to study the effects of the treatment, initial number of tumors, and size of the largest initial tumor on the tumor recurrence. In the `bladder` dataset all patients are measured four times with nonzero follow-up. 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 2 can be used to test if all β s but β_0 are 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. In the best model, the prior

JOINT MODELS WITH INFORMATIVE TIME

outcome and current time have significant effects on the number of tumors ($\varphi = 4.336$ and $\gamma = -0.142$ with p-value of < 0.001 for both terms). More detailed information is provided in Table 3.

Table 2. Model Selection Criteria for the Bernoulli-Exponential Model

Model	AIC	AICc	BIC	LR Ratio	P-value
1. event ~ rx + number + size	2652.85	2654.74	2692.39	199.10	<0.001
2. event ~ rx + number	2651.44	2652.90	2668.54	198.51	<0.001
3. event ~ rx + size	2651.64	2653.09	2668.73	198.31	<0.001
4. event ~ rx	2650.60	2651.67	2665.25	197.35	<0.001

Table 3. Summary of Analysis for Cancer Recurrence ($N = 85$)

Variable	Model 1			Model 2		
	β	SE	P	β	SE	P
Intercept	-0.550	0.244	< 0.001	0.077	0.174	0.657
rx2	0.276	0.215	0.199	0.273	0.215	0.204
number	-0.051	0.057	0.375	-0.060	0.056	0.283
size	0.056	0.073	0.441			
Prior size	4.349	0.582	< 0.001	4.336	0.571	< 0.001
Current time	-0.142	0.021	< 0.001	-0.142	0.020	< 0.001

Variable	Model 3			Model 4		
	β	SE	P	β	SE	P
Intercept	-0.179	0.201	0.374	-0.037	0.138	0.790
rx2	0.257	0.215	0.231	0.248	0.215	0.248
number						
size	0.070	0.071	0.329			
Prior size	4.349	0.569	< 0.001	4.331	0.606	< 0.001
Current time	-0.142	0.020	< 0.001	-0.142	0.021	< 0.001

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”, which is the number of tumors found at the time of recurrence. The Poisson-Exponential model is 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 are used with the variable “rtumor” as an outcome variable. Based on the information criteria presented in Table 4, 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 ($\beta_2 = -0.275$ and $\varphi = 0.079$ with p-value of 0.024 and < 0.001 , respectively). Also, the likelihood ratio test statistic and the corresponding p-value for each model are presented in Table 4. More information about the analysis for each model is provided in Table 5.

Table 4. Model Selection Criteria for the Poisson-Exponential Model

Model	AIC	AICc	BIC	LR Ratio	P-value
1. rtumor ~ trt + number + size	2220.32	2223.78	2239.47	45.50	<0.001
2. rtumor ~ trt + number	2219.59	2222.31	2236.61	44.23	<0.001
3. rtumor ~ trt + size	2218.47	2221.19	2235.49	45.35	<0.001
4. rtumor ~ trt	2218.13	2220.21	2233.02	43.69	<0.001

Note: trt = treatment

Table 5. Summary of Analysis for Number of tumors ($N = 118$)

Variable	Model 1			Model 2		
	β	SE	P	β	SE	P
Intercept	1.003	0.098	< 0.001	0.933	0.076	< 0.001
Trt2	-0.006	0.098	0.954	-0.010	0.102	0.921
Trt3	-0.301	0.102	0.002	-0.200	0.127	0.018
Number	0.009	0.023	0.700	0.016	0.022	0.461
Size	-0.027	0.024	0.264			
Prior Obs	0.077	0.018	< 0.001	0.078	0.018	< 0.001
Current time	0.001	0.003	0.802	0.001	0.003	0.839

Variable	Model 3			Model 4		
	β	SE	P	β	SE	P
Intercept	1.027	0.075	< 0.001	0.968	0.059	< 0.001
Trt2	-0.006	0.102	0.956	-0.010	0.101	0.924
Trt3	-0.289	0.122	0.018	-0.275	0.121	0.024
Number						
Size	-0.030	0.023	0.202			
Prior Obs	0.077	0.018	< 0.001	0.079	0.018	< 0.001
Current time	0.001	0.003	0.797	0.001	0.003	0.837

JOINT MODELS WITH INFORMATIVE TIME

The second model, “rtumor \sim treatment + number,” in Table 4, was studied by Cai et al. (2012). Their research interest was to study the effects of the treatment and the 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 in Table 6.

Table 6. Estimates 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 7. Estimates of the Poisson-Exponential Model

	Estimate	Std. error	t value	Pr (> t)
Intercept	0.933	0.076	12.301	< 0.001
Treatment.pyridoxine	-0.010	0.102	-0.099	0.921
Treatment.thiotepa	-0.300	0.127	-2.371	0.018
Number	0.016	0.022	0.736	0.461
Prior Outcome	0.078	0.018	4.285	< 0.001
Current Time	0.001	0.003	0.203	0.839

Even though two models have different concepts, both models found that the treatment has a negative effect on the number of tumors. The output in Table 7 shows that the second treatment level, thiotepa, has a significant negative effect on the number of tumors, and that the effect of the number of initial tumors is not significant. The results based on the model by Cai et al. (2012) show similar findings for the estimation of the effects of the treatment and the number of initial tumors. As can be seen in Table 6 and 7, 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 by Cai et al. (2012) computed the overall treatment effect, ignoring the effect of each treatment level, while 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 because “the more often the

patients visited the clinic and received the treatment, the less chance they will have tumor recurrence” (Cai et al., 2012, see p. 10). This interpretation makes sense, but does not specify which treatment is the most effective to reduce the tumor occurrence. Yet the Poisson-Exponential model specifically pointed out that the thiotep treatment only can reduce the tumor occurrence.

Discussion

The purpose of this study was to build models that can fit outcome distributions belonging to the exponential family of distributions. The simulation verified the MLEs of the joint models obtained the asymptotic multivariate normality and that accordingly model selection methods are valid in the joint models, which are based on the normality assumption of MLEs. In addition, the computing package using R was developed to handle the joint models and used to analyze the bladder cancer data to demonstrate the performance of the models. The Poisson-Exponential model was compared with the latent effect model proposed by Cai et al. (2012) and showed its computational capability for testing treatment effect. The differences of the two models and the advantages of the Poisson-Exponential model over the latent model were detailed in the previous section.

This study was based on two assumptions: (1) time is informative and (2) current outcomes are dependent on one-step prior outcomes. So this study is strictly limited to the situation satisfying those two conditions. However, 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 log-normal distribution. If that is the situation, the appropriate distribution can be applied to the time process; then, the MLEs 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 dataset. 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 give simultaneous tests for separate response, in addition to a single response’s analysis. All of those

JOINT MODELS WITH INFORMATIVE TIME

mentioned above are technically possible and can be further explored by researchers in order to improve the joint models.

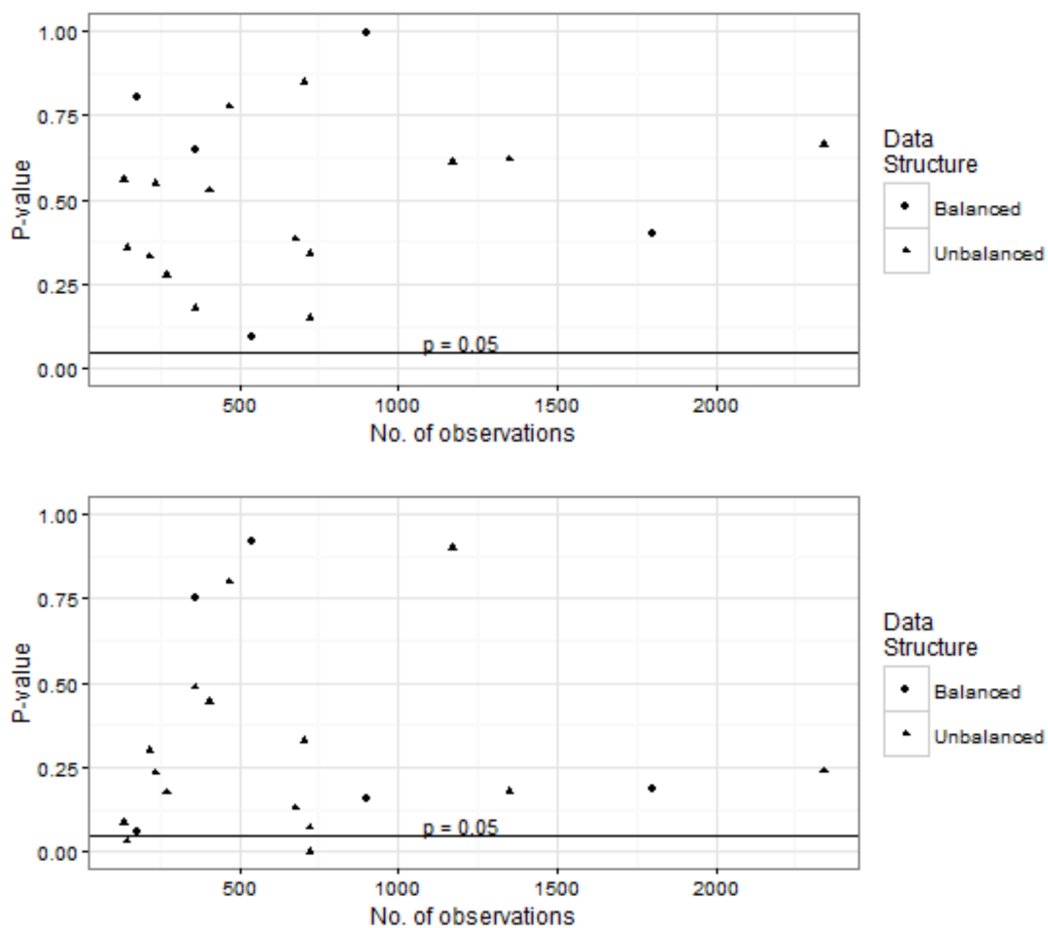


Figure 1. Bernoulli-Exponential Model: Multivariate normality test results. The upper graph is for time process and the bottom graph for outcome process. And the filled circle is for balanced data and the triangle for unbalanced data.

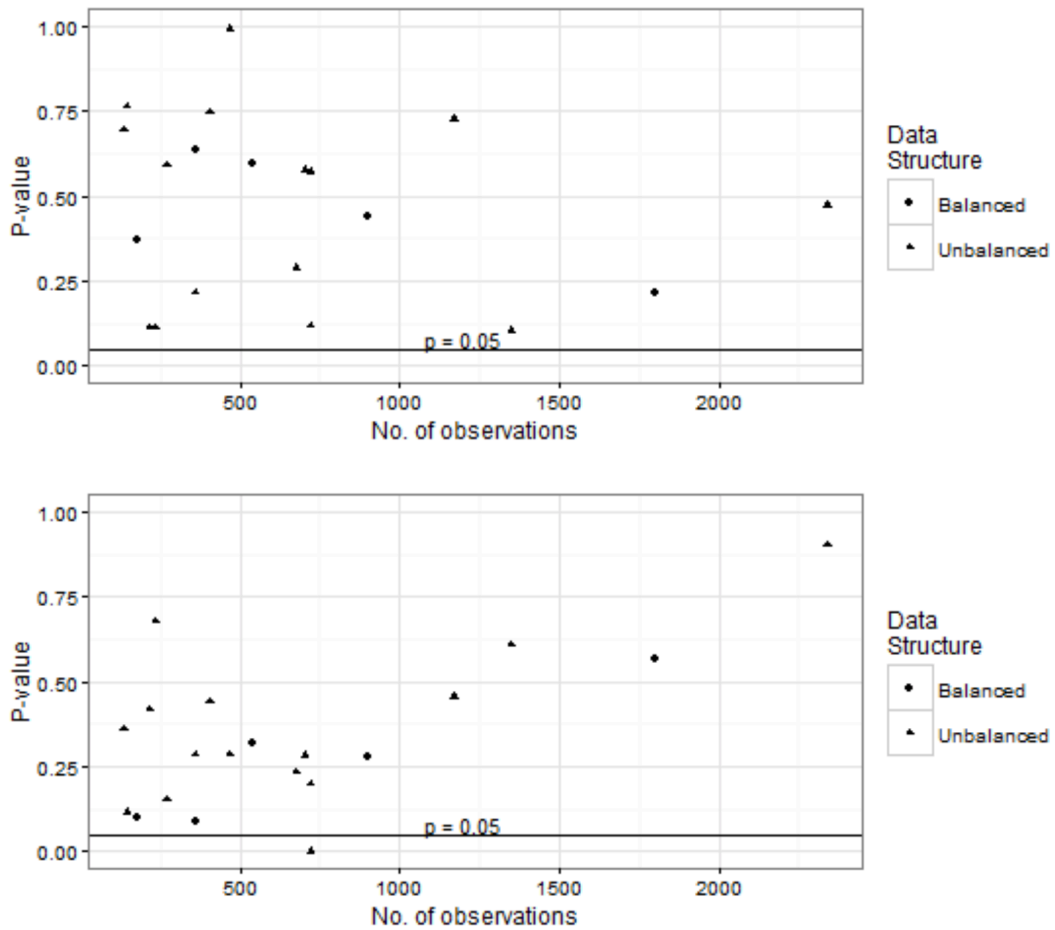


Figure 2. Poisson-Exponential Model: Multivariate normality test results. The upper graph is for time process and the bottom graph for outcome process. And the filled circle is for balanced data and the triangle for unbalanced data.

JOINT MODELS WITH INFORMATIVE TIME

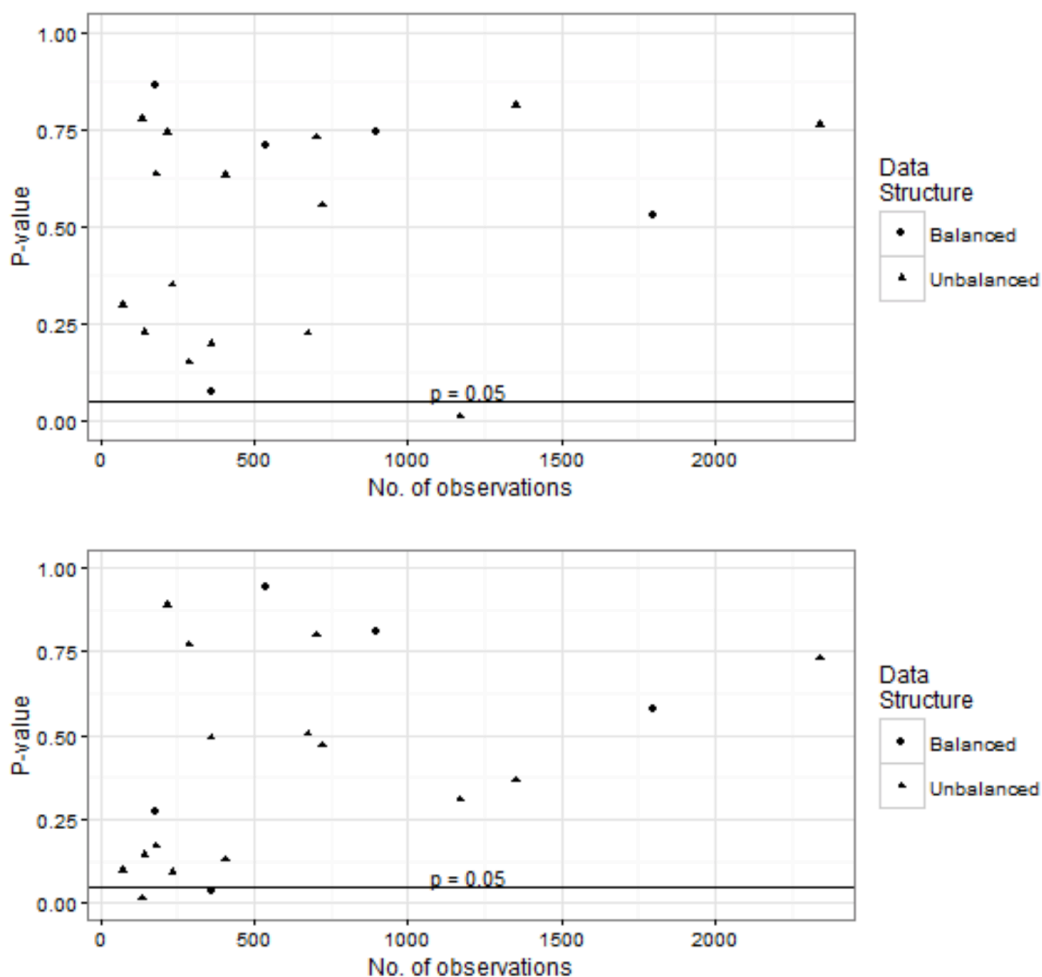


Figure 3. Gamma-Exponential Model: Multivariate normality test results. The upper graph is for time process and the bottom graph for outcome process. And the filled circle is for balanced data and the triangle for unbalanced data.

Acknowledgments

This paper is developed in part from the author's unpublished dissertation under the same title (Seo, 2015), and elements throughout are derived directly from that work.

References

- Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.
- Andrews, D. F., & Herzberg, A. M. (1985). *Data: A collection of problems from many fields for the student and research worker*. Springer-Verlag.
- Bronsert, M. F. (2009). *A joint model of a longitudinal process and informative time schedule data*. (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., Wenbin, L., & Zhang, H. H. (2012). Time-varying latent model for longitudinal data with informative observation times. *Biometrics*, 68(4), 1093-1102. <https://doi.org/10.1111/j.1541-0420.2012.01794.x>
- Dobson, A. J. (2001). *An introduction to generalized linear models*. London: Chapman and Hall/CRC Press. <https://doi.org/10.1201/9781420057683>
- Everitt, B. S. (2006). *The cambridge dictionary of statistics*. 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.
- Henderson, R., Diggle, P., & Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4), 465-480. <https://doi.org/10.1093/biostatistics/1.4.465>
- Henningsen, A., & Ott, T. (2010). maxLik: A package for maximum likelihood estimation in R. *Computational Statistics*, 26(3), 443-458. <https://doi.org/10.1007/s00180-010-0217-1>
- Kim, S., Zeng, D., Chambless, L., & Li, Y. (2012). Joint models of longitudinal data and recurrent events with informative terminal event. *Statistics in Biosciences*, 4(2), 262-281. <https://doi.org/10.1007/s12561-012-9061-x>
- Liang, Y., Lu, W., & Ying, Z. (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics*, 65(2), 377-384. <https://doi.org/10.1111/j.1541-0420.2008.01104.x>
- Lin, H., & Scharfstein, D. O. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3), 791-813. <https://doi.org/10.1111/j.1467-9868.2004.b5543.x>

JOINT MODELS WITH INFORMATIVE TIME

- Lin, Y. (2011). *Hypothesis testing for the Gaussian-Exponential longitudinal model*. (Doctoral Dissertation), University of Northern Colorado.
- Lipsitz, S., Fitzmarice, G., Ibrahim, J., Gelber, R., & Lipshultz, S. (2002). Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics*, 58(3), 621-630. <https://doi.org/10.1111/j.0006-341x.2002.00621.x>
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall/CRC Press. <https://doi.org/10.1007/978-1-4899-3242-6>
- McQuarrie, A., Shumway, R., & Tsai, C.-L. (1997). The model selection criterion AICu. *Statistics & Probability Letters*, 34(3), 285-292. [https://doi.org/10.1016/s0167-7152\(96\)00192-7](https://doi.org/10.1016/s0167-7152(96)00192-7)
- Nash, J. C., & Varadhan, R. (2011). Unifying optimization algorithms to aid software system users. *Journal of Statistical Software*, 43(9), 1-14. <https://doi.org/10.18637/jss.v043.i09>
- Nelder, J. A., & Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370-384. <https://doi.org/10.2307/2344614>
- Qiu, F., Stein, C. M., & Elston, R. C. (2013). Joint modeling of longitudinal data and discrete-time survival outcome. *Statistical Methods in Medical Research*, 25(4), 1512-1526. <https://doi.org/10.1177/0962280213490342>
- R Core Team. (2014). R: A Language and Environment for Statistical Computing. Retrieved from R Foundation for Statistical Computing: <http://www.R-project.org/>
- SAS Institute. (2008). *SAS 9.2: User's guide*. SAS.
- Seo, J. (2015). *Joint Models of Longitudinal Outcomes and Informative Time*. (Unpublished doctoral dissertation). University of Northern Colorado, Greeley, CO.
- 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: Series B (Statistical Methodology)*, 62(2), 293-302. <https://doi.org/10.1111/1467-9868.00232>
- 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(471), 882-889. <https://doi.org/10.1198/016214505000000060>
- Wei, L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, 84(408), 1065-1073. <https://doi.org/10.1080/01621459.1989.10478873>

JANGDONG SEO

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*, 2012, 1-17. <https://doi.org/10.1155/2012/640153>

Zhang, Y. (2002). A semiparametric pseudo-likelihood estimation method for panel count data. *Biometrika*, 89(1), 39-48. <https://doi.org/10.1093/biomet/89.1.39>