UW Biostatistics Working Paper Series

1-21-2011

Non-Homogeneous Markov Process Models with Incomplete Observations: Application to a Dementia Disease Study

Xiao-Hua Zhou
University of Washington, azhou@u.washington.edu

Baojiang Chen
University of Washington, bjchen@u.washington.edu

Non-homogeneous Markov Process Models with Incomplete Observations: Application to a Dementia Disease Study

Baojiang Chen * and Xiao-Hua Zhou **

Department of Biostatistics, University of Washington

Seattle, Washington, US 98195

HSR&D Center of Excellence

VA Puget Sound Health Care System

Seattle, Wwashington, US 98101

*Email: bjchen@u.washington.edu

**Email: azhou@u.washington.edu

Non-homogeneous Markov Process Models with Incomplete Observations: Application to an Dementia Disease Study

Abstract

Multi-state models provide a convenient framework for characterizing disease processes. In many settings, however, individuals are only observed at periodic clinic visits and so the precise times of transitions are not observed. Further, patients may choose when they want to visit the clinics, which creates the incomplete data problem. While Markov process models provide a useful tool for describing disease progression, the literature mainly focuses on time homogeneous processes, and limited tools are available for dealing with non-homogeneity. In this paper we develop methods to deal with non-homogeneous Markov processes with incomplete observations through time scale transformation. Maximum likelihood estimation via an EM algorithm is advocated for parameter estimation. Simulation studies demonstrate that the proposed method works well under a variety of situations. The proposed method is applied to investigate risk factors for transition rates among normal cognition, mildly cognitive impairment (MCI), and death.

KEYWORDS: Likelihood; missing at random; missing not at random; Markov model; non-homogeneous; transition intensity.

1. INTRODUCTION

In studies of Alzheimer's disease (AD), mild cognitive impairment (MCI) is recognized as an important transitional state between normal cognition and dementia. One of the objectives of the study is to model rates and risk factors for the transitions among these states. In clinical studies, cognitive status is often assessed only at periodic follow-up visits, with exact times of transition and trajectory of states visited between observations unknown, which gives rise to interval data. Multi-state models provide a convenient way to characterize the movement of individuals among distinct states. With continuous time multi-state models, transition intensities are often of primary interest, and these are perhaps most widely modeled using Markov models (e.g., Bartholomew 1983; Singer and Spilerman 1976a, 1976b; Wasserman 1980). Various methods based on Markov models have been proposed in the literature, including discrete time (e.g., Albert and Waclawiw 1998) and continuous time models (Andersen et al. 1993). For continuous times models, when observation times are not evenly spaced, Kalbfleish and Lawless (1985, 1989) proposed an effective method for Markov time homogeneous or power function transition intensities. Cook et al. (2004) described a conditional Markov model with multivariate random effects to handle clustered, conditionally Markov, multi-state processes. Motivated by studies of smoking behaviour, Cook, Kalbfleisch and Yi (2002) developed other extensions of Markov models which accommodate heterogeneity in the patterns of movement between states by allowing subject-specific absorbing states.

Most applications assume a homogeneous process; that is, the transition probabilities only depend on the elapsed time between observations. This assumption is not satisfied when transition probabilities depend on time from the process origin. The general method to deal with a non-homogeneous model is straightforward, and limited work has been devoted to deal with non-homogeneous Markov process models. Kalbfleisch and Lawless (1985) proposed a method for modeling non-homogeneous multi-state data under panel

observation, in which the non-homogenous intensity matrix is a product of a baseline homogeneous intensity matrix and a function of time. Gentleman et al. (1994) considered piecewise constant transition intensities to deal with non-homogeneous models. A number of authors have used piecewise homogeneous processes to model temporal homogeneity with applications (e.g., Saint-Pierre et al. 2003; Ocana-Riola 2005; Perez-Ocon, Ruiz-Castro and Gamiz-Perez 2001). Hubbard, Inoue and Fann (2008) considered a time transformation method to deal with the non-homogeneity. This method allows for variation in transition intensities by assuming that the time-varying transition intensity matrix arises from the product of a baseline transition intensity matrix and a scalar function of time.

In cohort studies, clinical assessments may be scheduled before the study, but patients may choose when they want to visit clinics for clinical examination according to their degree of disease activity. Also, there may be other reasons that patients cannot make a visit. This creates a problem somewhat akin to incomplete data arising in longitudinal studies. In this case, data may be missing at random (MAR) (Laird 1988; Little and Rubin 2002) if missing status depends on observed (typically past) responses, or missing not at random (MNAR), where the missing status may depend on the latent disease status. For example, in AD studies, patients may choose to visit the clinics according to their past observed disease status or the present disease status, which creates the MAR or MNAR mechanism.

Little work in the literature has dealt with incomplete data under the framework of non-homogenous Markov process. Under a MAR or MNAR mechanism, the naïve analysis method such as the complete case analysis can give biased inferences. In this paper, we provide a general method to handle incomplete data for the non-homogeneous Markov processes using the time transformation method (Hubbard, Inoue and Fann 2008). Our proposed method is very appealing in that it can deal with all types of missing data mechanisms and allow variation in transition intensities under the framework of non-homogeneity.

Maximum likelihood methods are used with parameter estimation carried out via an EM algorithm (Dempster et al. 1977), and variance estimation is performed using Louis' method (Louis 1982).

In the literature of AD studies, time-dependent increases in the rate of progression of disease are a prominent feature. Yesavage et al. (2002), and Harezlak, Gao, and Hui (2003) studied the Markov process models using the piecewise constant transition intensities, and Salazar et al. (2007) addressed a discrete time Markov model for AD through the piecewise constant transition intensities. Disadvantages of this approach are that it requires a large study population to obtain precise estimates because transition rates must be estimated separately in each group. Additionally, transition rates must be assumed constant with each groups. Our proposed method differs from the piecewise constant model in its use of a continuous time model, allowing the transition rates change by time. Further, little work devoted to the missing data problem in the AD studies.

The remainder of this paper is organized as follows. Section 2 reviews models and estimation for continuous time models. Section 3 and Section 4 describe methods for parameter estimation when data are MAR and MNAR. Empirical studies including the simulation studies and sensitivity analyses are implemented in Section 5 to study the performance of the proposed method. Data arising from a dementia disease study are analyzed with the proposed method in Section 6. We conclude the paper with a general discussion in Section 7.

2. NOTATION AND MODEL FORMULATION

2.1 Non-homogeneous Markov Process Model via Time Transformation

Suppose there are K states, $1, 2, \ldots, K$, and let $Y(u)$ represent the state occupied at time $u \geq 0$, and $\mathcal{H}(u) = \{Y(v), 0 \leq v < u\}$ denote the history of the response process which records the states occupied over the interval $[0, u)$. The transition probability function is

written generally as $P(Y(u+v)=k|Y(v)=\tilde{k},\mathcal{H}(v))$ for $u,v>0$, but under a Markov model this simplifies to $P(Y(u+v)=k|Y(v)=\tilde{k})$, which we denote compactly as $P_{\tilde{k}k}(v,v+u)$. The Markov transition intensity function at time u for transitions from state $\tilde{k}$ to state k, is

$$\begin{aligned} q^*_{\tilde{k}k}(u) &= \lim_{\Delta u\to 0}\frac{P(Y(u+\Delta u)=k|Y(u)=\tilde{k})}{\Delta u}, \quad \tilde{k}\neq k, \\ q^*_{\tilde{k}\tilde{k}}(u) &= -\sum_{k\neq\tilde{k}} q^*_{\tilde{k}k}(u)\,, \end{aligned}$$

(Cox and Miller 1977). A multi-state model with state space $\{1,2,\ldots,K\}$ can then be described via the transition intensity matrix $Q^*(u)$ with elements $q^*_{\tilde{k}k}(u)$, $\tilde{k},k=1,\ldots,K$.

To model the dependence of the transition intensities on risk factors, we may introduce covariates by expressing the transition intensities as functions of time (in the non-homogeneous case) and the covariates. For a given individual i, we often adopt models of the form

$$q^*_{i\tilde{k}k}(u)=q^*_{0\tilde{k}k}(u)g(X_{i\tilde{k}k};\beta_{\tilde{k}k}),$$

where $q^*_{i\tilde{k}k}(u)$ is the $\tilde{k}k$ element of the transition intensity matrix $Q^*_i(u)$, $q^*_{0\tilde{k}k}(u)$ is the baseline transition intensity, $X_{i\tilde{k}k}$ is the time-invariant covariate vector, and $g(\cdot;\cdot)$ is any positive function. Let $X_i=(X'_{i\tilde{k}k},\tilde{k},k=1,\ldots,K)'$.

Let $P(v,u+v)$ denote the $K\times K$ transition probability matrix from time v to time $v+u$. For a homogeneous process we assume $q^*_{\tilde{k}k}(u)=q^*_{\tilde{k}k}$, and $P_{\tilde{k}k}(v,v+u)=P_{\tilde{k}k}(u)$ for $\tilde{k},k=1,\ldots,K$. In the matrix form, we have

$$P(u)=\exp(Q^*u).$$

For the non-homogeneous process, we can do some proper time scale transformation such that the process is homogeneous afterwards (Hubbard, Inoue, and Fann 2008). Specifically, let $t=h(u)$ be a time transformation on which the process is homogeneous with

intensity matrix Q, then

$$P(u_1, u_2) = P(t_2 - t_1) = \exp\{Q(t_2 - t_1)\}.$$

It is easy to show that $Q^*(u) = Qdh(u)/du$, which implies that time scale transformations leading to a time homogeneous Markov process are possible if the non-homogeneity in the process is due to a time-varying multiplicative change in the matrix of transition intensities.

Here we assume, after the time scale transformation $t = h(u)$, the transition intensity matrix for subject i does not depend on time. Then, the model becomes

$$q_{i\tilde{k}k}(t) = q_{0\tilde{k}k} g(X_{i\tilde{k}k}; \beta_{\tilde{k}k}),$$

where $q_{i\tilde{k}k}$ is the $\tilde{k}k$th element of the homogeneous intensity matrix for subject i, Q_i. Let $P_{i\tilde{k}k}(t)$ denote the transition probability for subject i from states $\tilde{k}$ to k given the covariate $X_{i\tilde{k}k}$.

The choice of $h(\cdot)$ is very flexible, and we require $h(u) \geq 0$ and $dh(u)/du \geq 0$, since $h(u)$ defines a time scale. Often, $h(\cdot)$ is selected by involving some parameter ϕ. The two examples used widely in practice, provided by Hubbard, Inoue and Fann (2008), are the exponential time transformation $h(u) = u\phi^u$ and the nonparametric time transformation $h(u) = u\phi(u)$, where

$$\phi(u) = \sum_{m=1}^{d} c(u)\phi_m \left\{ \frac{1}{\gamma} K(\frac{u - u_i}{\gamma}) \right\},$$

$$c(u) = \left\{ \sum_{k=1}^{d} \frac{1}{\gamma} K(\frac{u - u_k}{\gamma}) \right\}^{-1}.$$

This kernel smoother has knots at u_k, $k = 1, \ldots, d$; smoothing parameter ϕ satisfies constraints $\phi_k > 0$. To make identifiability, we often assume $\phi_1 = 1$ or $\phi(0) = 0$.

2.2 Maximum Likelihood Estimation under Independent Inspection Process for Complete Data

With continuous time models and observation schemes, the response process $\{Y(u), u > 0\}$ may be observed at any time point u over the period observation. If the time of as-

sessment u does not depend on the state of the underlying response process Y, we can base inference on the response process conditional on the assessment times (Grüger, Kay and Schumacher 1991), and this is typically an implicit assumption in standard analyses. In this paper, we consider the problem in which subjects are scheduled to be examined at pre-specified assessment times denoted $u_1 < u_2 < \cdots < u_J$. This reflects many common clinical settings where patients are expected to return for regular follow-up assessment on annual, say, basis. This enable us to adopt a convenient frame work employed to describe incomplete longitudinal data since it is then only necessary to indicate whether each assessment as made.

Let $Y_i = (Y_i(u_1), \ldots, Y_i(u_J))'$ be a health state vector for subject i at each observation time points, which takes values $1, \ldots, K$, $i = 1, \ldots, n$. Let β denote the parameter vector in the intensity matrix Q_{i0}, as well as the parameter ϕ in the time transformation function $h(u; \phi)$. Using the time transformation and the homogeneity of the process on the new time scale we can express the likelihood as, conditional on the initial states,

$$L(\beta; Y) \;=\; \prod_{i=1}^{n} \prod_{j=2}^{J} P\{Y_i(h(u_j; \phi)) | Y_i(h(u_{j-1}; \phi)), X_i\}$$

given the covariate vector X_i. Maximum likelihood estimation can be obtained via the Fisher-scoring method.

3. ESTIMATION PROCEDURE UNDER THE MISSING AT RANDOM MECHANISM

Here we focus on the likelihood based method in which we need to specify the joint distribution of the observed response variable $Y_i^{(o)}$ and the missing data indicators R_i, given the covariates X_i. Here Y_i is written as $(Y_i^{(o)}, Y_i^{(m)})$ to implicitly indicate observed and missing components, and $R_i = (R_i(u_1), \ldots, R_i(u_J))'$ where $R_i(u_j) = 1$ if the response $Y_i(u_j)$ is observed, and 0 otherwise. The joint density of the observed data $(y_i^{(o)}, r_i)$ can be

written as

$$f(r_i, y_i^{(o)}|X_i;\alpha,\beta) = \int f(r_i|y_i^{(o)}, y_i^{(m)}, X_i;\alpha) f(y_i^{(o)}, y_i^{(m)}|X_i;\beta) dy_i^{(m)},$$

where α and β are the associated parameters for the missing data process and response process respectively. Then the joint likelihood for (α, β) is

$$L(\alpha,\beta;y^{(o)},r) = \prod_{i=1}^{n} \int f(r_i|y_i^{(o)}, y_i^{(m)}, X_i;\alpha) f(y_i^{(o)}, y_i^{(m)}|X_i;\beta) dy_i^{(m)}. \quad (1)$$

When the missing data mechanism is MCAR or MAR, this likelihood becomes

$$\begin{aligned} L(\alpha,\beta;y^{(o)},r) &= \prod_{i=1}^{n} \left\{ f(r_i|y_i^{(o)}, X_i;\alpha) \int f(y_i^{(o)}, y_i^{(m)}|X_i;\beta) dy_i^{(m)} \right\} \\ &= \prod_{i=1}^{n} \left\{ f(r_i|y_i^{(o)}, X_i;\alpha) f(y_i^{(o)}|X_i;\beta) \right\}. \end{aligned}$$

Assuming the parameters α and β are functionally independent, then likelihood inference for β from the likelihood $L(\alpha, \beta; y^{(o)}, r)$ is equivalent to a likelihood inference for β from the observed likelihood

$$L(\beta; y^{(o)}) = \prod_{i=1}^{n} f(y_i^{(o)}|X_i;\beta). \quad (2)$$

Here for incomplete data with missing at random (MAR), we develop two inference methods in the following two subsections.

3.1 A Fisher-scoring Method

Let $j_1 < j_2 < \cdots < j_{m_i}$ be the ordered observed assessment time points index for subject i. We assume data are always available in the baseline assessment, so $j_1 = 1$. The observed likelihood (2) can be written as

$$\begin{aligned} L(\beta; y^{(o)}) &= \prod_{i=1}^{n} \prod_{m=2}^{m_i} P\{Y_i(h(u_{j_m};\phi))|Y_i(h(u_{j_{m-1}};\phi)), X_i\} \\ &= \prod_{i=1}^{n} \prod_{m=2}^{m_i} \prod_{\tilde{k},k=1}^{K} \{P_{i\tilde{k}k}(h(u_{j_m}) - h(u_{j_{m-1}}))\}^{I(Y_i(h(u_{j_{m-1}}))=\tilde{k}, Y_i(h(u_{j_m}))=k)}, \end{aligned}$$

leading to the score function

$$S(\beta) = \sum_{i=1}^{n}\sum_{m=2}^{m_i}\sum_{\tilde{k},k=1}^{K} \frac{I(Y_i(h(u_{j_{m-1}}))=\tilde{k}, Y_i(h(u_{j_m}))=k)}{P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}}))} \frac{\partial P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}}))}{\partial\beta},$$

and the second derivative

$$\begin{aligned}
&\frac{\partial^2 \log\, L(\beta; y^{(o)})}{\partial\beta\partial\beta'} \\
&= \sum_{i=1}^{n}\sum_{m=2}^{m_i}\sum_{\tilde{k},k=1}^{K} \left\{ \frac{I(Y_i(h(u_{j_{m-1}}))=\tilde{k}, Y_i(h(u_{j_m}))=k)}{P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}}))} \frac{\partial^2 P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}}))}{\partial\beta\partial\beta'} \right. \\
&\quad - \frac{I(Y_i(h(u_{j_{m-1}}))=\tilde{k}, Y_i(h(u_{j_m}))=k)}{[P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}}))]^2} \frac{\partial P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}}))}{\partial\beta} \\
&\quad \left. \frac{\partial P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}}))}{\partial\beta'} \right\}.
\end{aligned}$$

To solve $S(\beta) = 0$ in order to obtain the estimate $\hat{\beta}$, we may, in principal, apply the Newton-Raphson algorithm by using the observed information matrix. This requires the availability of the second derivatives of the log-likelihood. However, the second derivatives of the log-likelihood are tedious to derivative and program. Here we develop a Fisher-scoring method which obviates need for the second derivatives (Kalbfleisch and Lawless, 1985) when the covariates are discrete.

Taking the expectation with respect to the conditional distribution of the response vectors given the covariates, we obtain

$$\begin{aligned}
E\left\{ -\frac{\partial^2 \log\, L(\beta; y^{(o)})}{\partial\beta\partial\beta'} \right\} &= \sum_{i=1}^{n}\sum_{m=2}^{m_i}\sum_{\tilde{k},k=1}^{K} \left\{ \frac{P(Y_i(h(u_{j_{m-1}}))=\tilde{k}|X_i)}{[P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}}))]} \right. \\
&\quad \left. \cdot \frac{\partial P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}}))}{\partial\beta} \frac{\partial P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}}))}{\partial\beta'} \right\}
\end{aligned}$$

by noting that

$$E[I(Y_i(h(u_{j_{m-1}}))=\tilde{k}, Y_i(h(u_{j_m}))=k)] = P(Y_i(h(u_{j_{m-1}}))=\tilde{k}|X_i)\cdot P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}}))$$

and

$$\sum_{k=1}^{K} P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}})) = 1.$$

This expectation can be estimated by

$$M(\beta) = \sum_{i=1}^{n}\sum_{m=2}^{m_i}\sum_{\tilde{k},k=1}^{K}\left\{\frac{p_{ij_{m-1}\tilde{k}}}{[P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}}))]}\cdot\frac{\partial P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}}))}{\partial\beta}\right.$$
$$\left.\cdot\frac{\partial P_{i\tilde{k}k}(h(u_{j_m})-h(u_{j_{m-1}}))}{\partial\beta'}\right\}$$

when covariates X_i are discrete, where $p_{ij_{m-1}\tilde{k}}$ is the proportion of the subjects with covariate x_i in state $\tilde{k}$ at the (j_{m-1})th time point.

Then an updated estimate is obtained as

$$\beta^{(h+1)} = \beta^{(h)} + [M(\beta^{(h)})]^{-1}S(\beta^{(h)}), \quad h = 0, 1, \ldots,$$

where $M(\beta^{(h)})$ is assumed nonsingular. The iteration is cycled through until convergence of $\beta^{(h+1)}$. Let $\hat{\beta}$ denote the corresponding limit. Under the regularity conditions for maximum likelihood estimators, $\sqrt{n}(\hat{\beta}-\beta) \to N(0, \Sigma^{-1})$ as the sample size n approaches infinity. Here $\Sigma = E[-\partial^2 \log L_i(\beta; y_i^{(o)})/\partial\beta\partial\beta']$ with $L_i(\beta; y_i^{(o)})$ the observed likelihood for subject i. We can estimate the asymptotic covariance matrix of $\hat{\beta}$ by $[M(\hat{\beta})]^{-1}$.

3.2 An EM Algorithm

When the number K of states is relatively small, the Fisher-scoring method described above works well. However, if K is large, the Fisher-scoring method may become computationally burdensome. In this subsection we describe an alternative method–the EM algorithm, which is simple to implement. The complete data likelihood for subject i is given by

$$L_i(\beta; y_i) = \prod_{j=2}^{J}\prod_{\tilde{k},k=1}^{K}\{P_{i\tilde{k}k}(h(u_j)-h(u_{j-1}))\}^{I(Y_i(h(u_{j-1}))=\tilde{k}, Y_i(h(u_j))=k)},$$

leading to the complete data log-likelihood

$$\ell_i(\beta; y_i) = \sum_{j=2}^{J}\sum_{\tilde{k},k=1}^{K} I(Y_i(h(u_{j-1}))=\tilde{k}, Y_i(h(u_j))=k)\log\{P_{i\tilde{k}k}(h(u_j)-h(u_{j-1}))\}.$$

In the expectation step (E-step), with the estimate $\beta^{(h)}$ at iteration h, we construct the conditional expectation $Q(\beta;\beta^{(h)}) = \sum_{i=1}^{n} Q_i(\beta;\beta^{(h)})$, where $Q_i(\beta;\beta^{(h)}) = E[\ell_i(\beta;y_i)|y_i^{(o)}, \beta^{(h)}] = \sum_{y_i^{(m)}} w_i(y_i;\beta^{(h)}) \cdot \ell_i(\beta;y_i)$, and

$$w_i(y_i;\beta^{(h)}) = \frac{L_i(\beta^{(h)};y_i^{(o)},y_i^{(m)})}{\sum_{y_i^{(m)}} L_i(\beta^{(h)};y_i^{(o)},y_i^{(m)})}\,.$$

The maximization step (M-step) maximizes the function $Q(\beta;\beta^{(h)})$ with respect to the parameter β, and a Newton-Raphson algorithm can be used for this purpose. Alternatively, we may employ the Fisher-scoring method discussed in Section 3.1. Specifically, the score and the expectation of the second derivative are given by

$$S(\beta;\beta^{(h)}) = \sum_{i=1}^{n}\sum_{y_i^{(m)}} w_i(y_i;\beta^{(h)}) \cdot \partial\ell_i(\beta;y_i)/\partial\beta,$$

and

$$M(\beta;\beta^{(h)}) = E\Big\{-\frac{\partial^2 Q(\beta;\beta^{(h)})}{\partial\beta\partial\beta'}\Big\} = \sum_{i=1}^{n}\sum_{y_i^{(m)}} w_i(y_i;\beta^{(h)}) \cdot E[-\partial^2\ell_i(\beta;y_i)/\partial\beta\partial\beta'],$$

where

$$\frac{\partial\ell_i(\beta;y_i)}{\partial\beta} = \sum_{j=2}^{J}\sum_{\tilde{k},k=1}^{K} \frac{I(Y_i(h(u_{j-1}))=\tilde{k}, Y_i(h(u_j))=k)}{P_{i\tilde{k}k}(h(u_j)-h(u_{j-1})))} \frac{\partial P_{i\tilde{k}k}(h(u_j)-h(u_{j-1}))}{\partial\beta},$$

and

$$E\Big\{-\frac{\partial^2\ell_i(\beta;y_i)}{\partial\beta\partial\beta'}\Big\} = \sum_{j=2}^{J}\sum_{\tilde{k},k=1}^{K}\Big\{\frac{P(Y_i(h(u_{j-1}))=\tilde{k}|X_i)}{P_{i\tilde{k}k}(h(u_j)-h(u_{j-1}))}\frac{\partial P_{i\tilde{k}k}(h(u_j)-h(u_{j-1}))}{\partial\beta} \cdot\frac{\partial P_{i\tilde{k}k}(h(u_j)-h(u_{j-1}))}{\partial\beta'}\Big\}.$$

If the covariate vector X_i are discrete, we can replace $P(Y_i(h(u_{j-1})) = \tilde{k}|X_i)$ by the proportion of subjects that are in state $\tilde{k}$ at time point u_{j-1} with covariate X_i.

To obtain the variance estimates for $\hat{\beta}$, we may apply the Louis' method (Louis, 1982):

$$\Sigma(\hat{\beta}) = M(\hat{\beta};\hat{\beta}) - \sum_{i=1}^{n}\sum_{Y_i^{(m)}} w_i(Y_i;\hat{\beta})\left(\frac{\partial\ell_i(\hat{\beta})}{\partial\beta}\right)\left(\frac{\partial\ell_i(\hat{\beta})}{\partial\beta}\right)' + \sum_{i=1}^{n}\left(\frac{\partial Q_i(\hat{\beta};\hat{\beta})}{\partial\beta}\right)\left(\frac{\partial Q_i(\hat{\beta};\hat{\beta})}{\partial\beta}\right)',$$

then $[\Sigma(\hat{\beta})]^{-1}$ is the estimate of the asymptotic covariance matrix of $\hat{\beta}$.

4. ESTIMATION PROCEDURE UNDER THE MISSING NOT AT RANDOM MECHANISM

When data are MNAR, the missing data model must be specified to make valid inference because the likelihood of the observed response $Y^{(o)}$ and the missing process R, (1), can not be simplified. To exploit (1), we must characterize the missing data process and estimate the associated parameters. We address this in the following subsection.

4.1 Modeling for the Missing Data Process

A multinomial missing data model introduced by Ibrahim et al. (2001) is often used to specify the joint distribution of R_i through a sequence of one-dimensional conditional distributions:

$$P(R_i|Y_i, X_i; \alpha) = \prod_{j=2}^{J} P(R_i(u_j)|\bar{R}_i(u_j), Y_i, X_i; \alpha) \cdot P(R_i(u_1)|Y_i, X_i; \alpha), \tag{3}$$

where $\bar{R}_i(u_j) = \{R_i(u_{j-1}), \ldots, R_i(u_1)\}$ and α is a vector of associated parameters for the conditional distribution. This accommodates nonmonotone patterns of missing data, and provides a natural way to specify the joint distribution of the missing data indicators when knowledge about the missingness of one response affects the probability of missingness of another. Let $\lambda_i(u_j) = P(R_i(u_j) = 1|\bar{L}_i(u_j); \alpha)$ denote the conditional probability that the response is observed at time u_j given $\bar{L}_i(u_j) = \{\bar{R}_i(u_j), Y_i, X_i\}$. Typically a logistic link may relate a linear function of $\bar{L}_i(u_j)$ to the probability of being observed at time u_j for subject i, i.e. we can specify $\text{logit}\, \lambda_i(u_j) = \eta(\bar{L}_i(u_j); \alpha)$ for certain function $\eta(\cdot; \cdot)$. For example, we may consider the following model, although other forms may be well adopted for specific applications,

$$\text{logit}\, \lambda_i(u_j) = \alpha_0 + \alpha_1 r_i(u_{j-1}) + \alpha_2 r_i(u_{j-1}) y_i(u_{j-1}) + \alpha_3 y_i(u_j) + \alpha_x' X_i. \tag{4}$$

If $\alpha_3 \neq 0$, the missing mechanism specified by (4) leads to a MNAR mechanism; $\alpha_3 = 0$ but $\alpha_2 \neq 0$ lead to a MAR mechanism; both $\alpha_3 = 0$ and $\alpha_2 = 0$ leads to a missing complete

at random (MCAR). Let $\theta = (\beta', \alpha')'$.

4.2 An EM Algorithm

Here we advocate the EM algorithm for parameter estimation under a MNAR mechanism. Conditional on the initial state, the complete data likelihood for subject i is written as

$$\begin{aligned} L_i(\theta; y_i, r_i) &= \prod_{j=2}^{J} \Big\{ (\lambda_i(u_j))^{r_i(u_j)} (1-\lambda_i(u_j))^{1-r_i(u_j)} \\ &\quad \cdot \prod_{\tilde{k},k=1}^{K} \{P_{i\tilde{k}k}(h(u_j) - h(u_{j-1}))\}^{I(Y_i(h(u_{j-1}))=\tilde{k},\, Y_i(h(u_j))=k)} \Big\}. \end{aligned} \tag{5}$$

leading to the log-likelihood

$$\begin{aligned} \ell_i(\theta; y_i, r_i) &= \sum_{j=2}^{J} \Big\{ r_i(u_j) \log(\lambda_i(u_j)) + (1-r_i(u_j)) \log(1-\lambda_i(u_j)) \\ &\quad + \sum_{\tilde{k},k=1}^{K} I(Y_i(h(u_{j-1})) = \tilde{k},\ Y_i(h(u_j)) = k) \log\left(P_{i\tilde{k}k}(h(u_j) - h(u_{j-1}))\right) \Big\} \end{aligned} \tag{6}$$

In the E-step, we calculate the expectation of the complete data likelihood given the observed data and the hth iteration parameter $\theta^{(h)}$. That is, $Q(\theta; \theta^{(h)}) = \sum_{i=1}^{n} Q_i(\theta; \theta^{(h)})$, where $Q_i(\theta; \theta^{(h)}) = E[\ell_i(\theta; y_i, r_i) | y_i^{(o)}; \theta^{(h)}] = \sum_{y_i^{(m)}} w_i(y_i; \theta^{(h)}) \cdot \ell_i(\theta; y_i, r_i)$, and

$$w_i(y_i; \theta^{(h)}) = \frac{L_i(\theta^{(h)}; y_i^{(m)}, y_i^{(o)}, r_i)}{\sum_{y_i^{(m)}} L_i(\theta^{(h)}; y_i^{(m)}, y_i^{(o)}, r_i)}.$$

In the M-step, we maximize $Q(\theta; \theta^{(h)})$ to get the estimate $\theta^{(h+1)}$. Iterate E and M step until convergence. Denote the limit $\hat{\theta}$. We comment that we can still use the Fisher-scoring method introduced in Section 3 to obtain the estimate when the covariate X_i is discrete, which we do not need to calculate the second derivatives of the transition probabilities.

Standard errors for these parameter estimates can be calculated using the Louis' method (Louis 1982), which partitions the complete data information into two parts: the information associated with the observed data and the one associated with the missing data. The

estimated observed information matrix of θ based on Louis' method is given by

$$\begin{aligned}\Sigma(\hat{\theta}) &= -\frac{\partial^2 Q(\hat{\theta};\hat{\theta})}{\partial\theta\partial\theta'} - \sum_{i=1}^{n}\sum_{y_i^{(m)}} w_i(y_i;\hat{\theta})\left(\frac{\partial \ell_i(\hat{\theta};y_i,r_i)}{\partial\theta}\right)\left(\frac{\partial \ell_i(\hat{\theta};y_i,r_i)}{\partial\theta}\right)' \\ &\quad + \sum_{i=1}^{n}\left(\frac{\partial Q_i(\hat{\theta};\hat{\theta})}{\partial\theta}\right)\left(\frac{\partial Q_i(\hat{\theta};\hat{\theta})}{\partial\theta}\right)'.\end{aligned}$$

The estimate of the asymptotic covariance matrix of $\hat{\beta}$ is the left upper $p \times p$ block of $[\Sigma(\hat{\theta})]^{-1}$, where p is the dimension of β.

5. NUMERICAL STUDIES

5.1 Model Assessment for the Proposed Method

Here we consider a three-state transition process with transition intensity matrix

$$Q = \begin{pmatrix} -0.3 & 0.2 & 0.1 \\ 0.2 & -0.3 & 0.1 \\ 0 & 0 & 0 \end{pmatrix}.$$

We assume the time points are uniformly on $(0.5, 2)$ with equal space interval 0.5. At the first observation times points, subject are equally likely to be in state one or two. The number of subjects is 200. we assume the true transformation function $h(u) = u\phi^u$ with true parameter $\phi = 1.0$.

The missing data model is

$$\text{logit}\,\lambda_i(u_j) = \alpha_0 + \alpha_1 r_i(u_{j-1}) + \alpha_2 r_i(u_{j-1}) y_i(u_{j-1}) + \alpha_3 y_i(u_j) \tag{7}$$

for $j = 2, 3, \ldots$. The true values are $\alpha_0 = \log(0.4)$, $\alpha_1 = \log(1.2)$, $\alpha_2 = \log(1.5)$, and we change α_3 to adjust the missing mechanism and missing proportions. One thousand simulations are run for each parameter configuration.

Here we compare two methods. One is the proposed method, and the second is the complete case analysis. Tables 1 and 2 report the result for sample size 200 and 1000

respectively, where BIAS is the relative bias; ASE is the average standard error based on the Louis' formula for the proposed method and model based for the naive method; ESE is the empirical standard derivation; CP is the 95% coverage probability. It is seen that the proposed method gives satisfactory results with ignorable finite sample biases and good coverage probabilities. However, the complete case method yields large biases and poor coverage probabilities.

Insert Tables 1 and 2 here

5.2 Sensitivity Analysis and Model Diagnosis

The validity of this approach relies on correct specification of the structure for the intensity and correct modeling of the missing data process. Here we investigate the impact of model misspecification for the missing data process. The model setup for the transition intensity and transformation of time scale is the same as that in the simulation studies.

We first evaluate the asymptotic biases induced by misspecifying the link function in the missing data model. We assume the true missing data model is (7), while we use model

$$\text{probit}\, \lambda_i(u_j) = \alpha_0 + \alpha_1 r_i(u_{j-1}) + \alpha_2 r_i(u_{j-1}) y_i(u_{j-1}) + \alpha_3 y_i(u_j)$$

in the estimation procedure, where probit$(\cdot)$ is the cumulative distribution function of the standard normal distribution. Figure 1 contains plots of the asymptotic percent relative bias of the regression coefficients, for example, q_{12}, q_{13} and ϕ. It can be seen that as the absolute value of α_3 increases, the relative biases increases; however, the relative bias is very small, which indicates that the estimates are less sensitive to the misspecification of the link function.

Next we consider the impact of misspecifying the missing data process by assuming it is MAR when in fact it is MNAR. We assume the true missing data model is (7), while we

use model

$$\text{logit}\, \lambda_i(u_j) = \alpha_0^* + \alpha_1^* r_i(u_{j-1}) + \alpha_2^* r_i(u_{j-1}) y_i(u_{j-1})$$

in the estimation procedure. Figure 2 contains plots of the asymptotic percent relative bias of the regression coefficient q_{12}, q_{13} and ϕ. It can be seen that as the absolute value of α_3 increases, the relative biases increases; the transition intensity parameters are more sensitive for the misspecification of the missing data models, while the transformation parameter ϕ is less sensitive.

Insert Figures 1 and 2 here

As demonstrated above, estimation of β is sensitive to misspecificatoin of the model for the missing data process. Therefore careful assessment of the missing data model is warranted. In the framework of likelihood, likelihood ratio test provides a useful tool to test two nested models. Specifically, the likelihood ratio test statistic $2\{\log\ L(\theta) - \log\ L_R(\theta)\} \rightarrow \chi^2(k)$ as $n \rightarrow \infty$, where $L_R(\theta)$ is the likelihood of the reduced model, and k is the difference of the number of parameters between the full model and the reduced model. For non-nested models, people may consider Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), etc.

In many applications, however, the data may not be sufficient to distinguish between alternative nonignorable missing data mechanisms. For this reason, sensitivity analyses are often advocated (e.g., Ibrahim et al. 2005). Several authors have proposed the use of global and local influence tools to sensitivity analyses in missing data contexts (e.g., Verbeke et al. 2001; Verbeke and Molenberghs 2000; Molenberghs and Verbeke 2005; Zhu and Lee 2001; Van Steen, Molenberghs, and Thjis 2001; Molenberghs, Kenward, and Goetghebeur 2001; Kenward 1998; Jansen et al. 2003). Another route for sensitivity analysis is to consider pattern-mixture models as a complement to selection models. Other approaches have been considered by Copas and Li (1997), Copas and Shi (2000), and Copas and Eguchi (2001).

More detail on some of these procedures can be found in Molenberghs and Verbeke (2005), and Little and Rubin (2002).

6. APPLICATION TO A DEMENTIA DISEASE STUDY

We apply the proposed method to the National Alzheimer's Coordinating Center (NACC) Uniform Data Set (UDS), which is an ongoing longitudinal database of subjects seen at one of the National Institute on Aging's 29 funded Alzheimer's Disease Centers (ADC) located throughout the USA.

Some studies have found amnestic MCI to be transient because future evaluations could yield a reversion to normal cognition (here we group normal and "impaired, not MCI" and denote by normal cognition for simplicity) as opposed to progression to dementia. In this section, we implement our proposed method to investigate the risk factors for transitions among normal cognition, MCI, dementia and death. There are 7932 subjects from 29 Alzheimer's Disease Centers included at the entry of this study. Follow-up visits for subjects are scheduled at approximately one-year intervals, with up to four clinical visits at present. There are 6722 subjects with complete data observed. The missing proportion is 13.27%.

In this analysis, we treat death as an absorbing state but allow transitions between all other states. Table 3 lists the total number of transitions between clinical diagnosis states at successive visits. Risk factor vector X_i includes: sex, congestive heart failure (CVCHF, yes/no), geriatric depression score (GDS), family history of dementia (fhdem, yes/no), diabetes (yes/no), hypertension (yes/no), education (years), MMSE, and age. Table 4 lists the baseline risk factors for the 7932 subjects.

Insert Tables 3 and 4 here

For simplicity, the four states normal cognition, MCI, dementia and death were coded

as 1, 2, 3 and 4, respectively. The multiplicative models for transition $\tilde{k}$ to k after the transformation are

$$q_{i\tilde{k}k}(t) = q_{0\tilde{k}k} \exp(X_i' \beta_{\tilde{k}k})$$

for $\tilde{k}, k = 1, 2, 3, 4$, $\tilde{k} \neq k$.

The missing data model in the EM algorithm is

$$\text{logit } \lambda_i(u_j) = \alpha_0 + X_i' \alpha_x + \alpha_1 I(Y_i(u_j) = 2) + \alpha_2 I(Y_i(u_j) = 3).$$

Tables 5 to 6 list the risk factors for the transitions among these four states. Here, we compare two methods: the complete case analysis and the proposed method. In this study we assume a power transformation of the form $h(u) = u\phi^u$. The estimates of ϕ are 1.033 with 95% confidence interval $(1.015, 1.051)$ for the complete case analysis and 1.046 with 95% confidence interval $(1.030, 1.062)$ for the proposed method. Both reveal that the process exhibits significant non-homogeneity, and the rate of evolution if the process is increasing as a function of time. For the assessment to the model fit of the time transformed process, we can compare goodness-of-fit of the time transformed model to goodness-of-fit for the homogeneous model and a piecewise constant model using the Bayesian Information Criterion (BIC) and a likelihood ratio test in the same spirit of Hubbard, Inous, and Fann (2007). Here we do not give much details about it, and interested people can refer Hubbard, Inous, and Fann (2007).

For the risk factors, complete case analysis and the proposed methods give different estimates. In the transitions between normal cognition and MCI, family history of dementia and age are significant, indicating that older people and people who have a family history of dementia have a higher risk of transition from normal cognition to MCI. In the transitions between normal cognition and dementia, age, GDS and hypertension affect the transition from normal to dementia (i.e., an older person, a person with higher GDS score or a person with hypertension is more likely to transition to dementia); and in the transitions from

normal cognition to death directly, GDS, diabetes and age are significant, indicating that a person has higher risk of transition to death if he/she has higher GDS score, diabetes, or is older. In the transitions between MCI and dementia, sex, GDS and age are significant, indicating that older people or people with higher GDS score have a higher risk of transition from MCI to dementia, and women have a lower risk of transition from MCI to dementia; in the direct transition from MCI to death, sex, GDS and age are significant, indicating that a person with higher GDS score or older age is more easy to death, also a man has a higher risk to death comparing to a woman; in the transition from dementia to death, sex, CVCHF, GDS, education and age are significant, indicating that a person has higher risk to death if he/she has a Congestive heart failure, higher GDS score, higher education or with an older age, and women has lower risk from dementia to death comparing to men.

Table 7 also lists the result for the missing data model. Significance of sex indicates that women are more likely to miss the observation than men; significance of fhdem and hypertension indicates that people with family history of dementia are less likely to attend this study, and people with hypertension are more likely to attend this study; significance of MMSE indicates that people with higher MMSE score are less likely to attend this study; significance of present state occupation indicate that the missing not at random mechanism is perhaps reasonable, and people are less likely to attend the study if they are in MCI or dementia comparing to normal cognition.

Insert Tables 5 to 7 here

7. DISCUSSION

In this paper we propose a likelihood-based method for the analysis of incomplete observations under the framework of non-homogeneous Markov processes using the time transformation model. This method is very appealing in that it can deal with all kinds of missing

data mechanisms and allow variation in transition intensities, even under the framework of non-homogeneity, and little work in the literature has been done to deal with this kind of data. Several methods are introduced for estimations with incomplete data under MAR or MNAR, including the Fisher-scoring method. The Fisher-scoring method is appealing since we do not need the second derivatives of the likelihood if the covariates are discrete. Applications to the NACC UDS demonstrate the usefulness of the proposed method.

Note that to obtain the consistent parameter estimates under MNAR, both the transition model and the model for the missing data process must be correctly specified. In practice, we aim to build a model which provides useful insight into the response process and observation process. Our strategy is therefore to build models that contain a large number of covariates, carry out tests of fit of nested models, and ultimately find a parsimonious model using standard procedures for model selection. The need for generalizations to deal with more complex models can be assessed by model expansion and the use of general model selection procedures via likelihood ratio tests.

Our method here assume the covariates are time independent, which is a limitation as most of other methods to deal with transition models. Relatively little work has been done on fitting regression models with time-dependent covariates. In the special case of a single interval-censored covariate that indicates the development of a particular condition, Goggins et al. (1999) develop methods for Cox regression for a right censored event time. Chen and Cook (2003) considered models and methods to deal with an interval-censored progressive covariate processs in recurrent event analyses. Cook, Zeng, and Lee (2008) consider an extension to the bivariate setting where both the covariate and failure times are interval-censored. The more general problem of interval-censored time varying covariates remains relatively open and worthy of future research.

ACKNOWLEDGEMENT

Dr. Xiao-Hua Zhou, Ph.D., is presently a Core Investigator and Biostatistics Unit Di-

rector at the Northwest HSR&D Center of Excellence, Department of Veterans Affairs Medical Center, Seattle, WA. Dr. Zhou's work was supported in part by U.S. Department of Veterans Affairs, Veterans Affairs Health Administration, HSR&D grants, and the National Science Foundation of China (NSFC 30728019). Both Drs. Zhou and Chen was support in part by National Institute on Aging grant U01AG016976. This paper presents the findings and conclusions of the authors. It does not necessarily represent those of VA HSR&D Service.

REFERENCES

Albert, P. S., and Waclawiw, M. A. (1998), "A Two-state Markov Chain for Heterogeneous Transitional Data: A Quasi-likelihood Approach, " *Statistics in Medicine*, 17, 1481–1493.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, New York, NY: Springer-Verlag.

Bartholomew, D. J. (1983), "Some Recent Developments in Social Statistics," *International Statistical Review*, 51, 1–9.

Chen, E. B., and Cook, R. J. (2003), "Regression Modeling with Recurrent Events and Time-dependent Interval-censored Marker Data," *Lifetime Data Analysis*, 9, 275–291.

Cook, R. J., Kalbfleisch, J. D., and Yi, G. Y. (2002), "A Generalized Mover-stayer Model for Panel Data," *Biostatistics*, 3, 407–420.

Cook, R. J., Yi, G. Y., Lee, K. A., and Gladman, D. D. (2004), "A Conditional Markov Model for Clustered Progressive Multistate Processes under Incomplete Observation," *Biometrics*, 60, 436–443.

Cook, R. J., Zeng, L., and Lee, K. A. (2008), "A Multistate Model for Bivariate Interval-censored Failure Time Data," *Biometrics*, 64, 1100–1109.

Copas, J. B., and Eguchi, S. (2001), "Local Sensitivity Approximations for Selectivity Bias," *Journal of the Royal Statistical Society*, Ser. B, 63, 871–895.

Copas, J. B., and Li, H. G. (1997), "Inference for Non-Random Samples" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 59, 55–96.

Copas, J. B., and Shi, J. Q. (2000), "Meta-Analysis, Funnel Plots and Sensitivity Analysis," *Biometrics*, 1, 247–262.

Cox, D. R., and Miller, H. D. (1977), *The Theory of Stochastic Processes*, London: Chapman and Hall.

Dempster, A. P., Laird, N. M., and Rubins, D. B. (1977) "Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion)," *Journal of the Royal Statistical Society B*, 39, 1–38.

Gentleman, R. C., Lawless, J. F., Lindsey, J. C., and Yan, P. (1994), "Multi-state Markov Models for Analysing Incomplete Disease History Data with Illustrations for HIV Diease," *Statistics in Medicine*, 13, 805–821.

Goggins, W. B., Finkelstein, D. M., and Zaslavsky, A. M. (1999), "Applying the Cox Proportional Hazards Model When the Change Time of a Binary Time-varying Covariate is Interval-censored," *Biometrics*, 55, 445–451.

Grüger, J., Kay, R., and Schumacher, M. (1991), "The Validity of Inferences Based on Incomplete Observations in Disease State Models," *Biometrics*, 47, 595–605.

Harezlak, J., Gao, S., and Hui, S. (2003), "An Illness-death Stochastic Model in the Analysis of Longitudinal Dementia Data," *Statistics in Medicine*, 22, 1465–1475.

Hubbard, R. A., Inoue, L. Y. T., and Fann, J. R. (2008), " Modeling Nonhomogeneous Markov Processes via Time Transformation," *Biometrics*, 64, 843–850.

Ibrahim, J. G., Chen, M. H., and Lipsitz, S. R. (2001), "Missing Responses in Generalized Linear Mixed Models When the Missing Data Mechanism is Nonignorable," *Biometrika*, 88, 551–564.

Ibrahim, J. G., Chen, M-H, Lipsitz, S. R., and Herring, A. H. (2005), "Missing-Data Methods for Generalized Linear Methods: A Comparative Review," *Journal of the American Statistical Association*, 100, 332–346.

Jansen, I., Molenberghs, G., Aerts, M., Thijs, H., and Van Steen, K. (2003), "A Local Influence Approach Applied to Binary Data from a Psychiatric Study," *Biometrics*, 59, 409-418.

Kalbfleisch, J. D., and Lawless, J. F. (1985), "The Analysis of Panel Data under a Markov Assumption," *Journal of the American Statistical Association*, 80, 863–871.

Kalbfleish, J. D., and Lawless, J. F. (1989), "Some Statistical Methods for Panel Life History Data," *Proceedings of the Statistics Canada Symposium on Analysis of Data in Time*, 185–192. Ottawa, Ontario: Statistics Canada.

Kenward, M. G. (1998), "Selection Models for Repeated Measurements With Nonrandom Dropout: An Illustration of Sensitivity," *Statistics in Medicine*, 17, 2723–2732.

Laird, N. M. (1988), "Missing Data in Longitudinal Studies," *Statistics in Medicine*, 7, 305–315.

Little, R. J. A., and Rubin, D. B. (2002), *Statistical Analysis with Missing Data* (2st edn), John Wiley and Sons, Inc..

Louis, T. (1982), "Finding the Observed Information Matrix When Using the EM Algorithm," *Journal of the Royal Statistical Society B*, 44, 226–233.

Molenberghs, G., Kenward, M. G., and Goetghebeur, E. (2001), "Sensitivity Analysis for Incomplete Contingency Tables," *Applied Statistics*, 50, 15–29.

Molenberghs, G., and Verbeke, G. (2005), *Models for Discrete Longitudinal Data*, New York: Springer.

Ocana-Riola, R. (2005), "Non-homogeneous Markov Processes for Biomedical Data Analysis," *Biometrical Journal*, 47, 369–376.

Perez-Ocon, R., Ruiz-Castro, J. E., and Gamiz-Perez, M. L. (2001), "Non-homogeneous Markov Models in the Analysis of Survival After Breast Cancer," *Journal of the Royal Statistical Society Series C*, 50, 111–124.

Saint-Pierre, P., Combescure, C., Daures, J. P., and Godard, P. (2003), "The Analysis of Asthma Control under a Markov Assumption with Use of Covariates," *Statistics in Medicine*, 22, 3755–3770.

Singer, B., and Spilerman, S. (1976a), "The Representation of Social Processes by Markov Models," *American Journal of Sociology*, 82, 1–54.

Singer, B., and Spilerman, S. (1976b), "Some Methodoloical Issues in the Analysis of Longitudinal Surveys," *Annals of Economic and Sociological Measurement*, 5, 447–474.

Slazar, J., Schmitt, F., Yu, L., Mendiondo, M., and Kryscio, R. (2007), "Shared Radom Effects Analysis of Multi-state Markov Models: Application to a Longitudinal Study of Transitions to Dementia," *Statistics in Medicine*, 26, 568–580.

Van Steen, K., Molenberghs, G., and Thijs, H. (2001), "A Local Influence Approach to Sensitivity Analysis of Incomplete Longitudinal Ordinal Data," *Statistical Modelling: An International Journal*, 1, 125–142.

Verbeke, G., and Molenberghs, G. (2000), *Linear Mixed Models for Longitudinal Data*, New York: Springer.

Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., and Kenward, M. G. (2001), "Sensitivity Analysis for Non-random Dropout: A Local Influence Approach," *Biometrics*, 57, 43-50..

Wasserman, S. (1980), "Analyzing Social Networks as Stochastic Processes," *Journal of the American Statistical Association*, 75, 280–294.

White, H. (1982), "Maximum Likelihood Estimation under Misspecified Models," *Econometrica*, 50, 1–26.

Yesavage, J., O'Hara, R., Kraemer, H., Noda, A., Taylor, J., Ferris, S., Gely-Nargeot, M., Rosen, A., Friedman, L., Sheikh, J., and Derouesne C. (2002), "Moldeing the Prevalence and Incidence of Alzheimer's Disease and Mild Cognitive Impairment," *Journal of Psychiatric Research*, 36, 281–286.

Zhu, H. T., and Lee, S. Y. (2001), "Local Inference for Incomplete Data Models," *Journal of the Royal Statitical Society,* Ser. B, 63, 111–126.

Figure 1: Asymptotic percent relative bias of transition intensity with correctly specified linear predictor but the link is misspecified as probit (true link is logit): q_{12} and q_{13} are the transition intensities from state 1 to state 2 and 3 respectively, and ϕ is the time transformation parameter

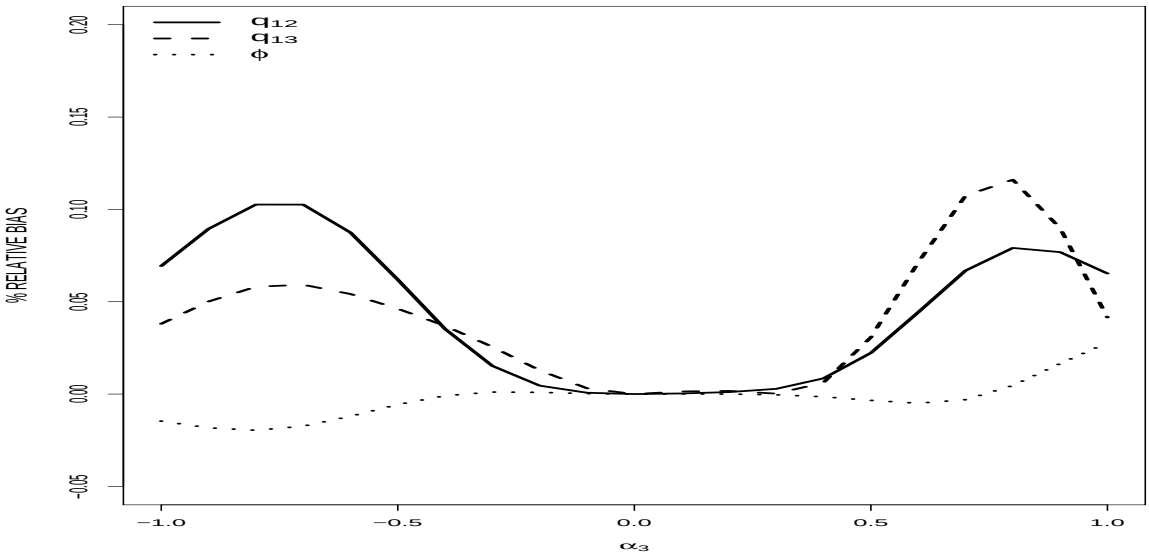

Figure 2: Asymptotic percent relative bias of transition intensity with correctly specified link but the linear predictor is misspecified ($\alpha_3 \neq 0$ so data are MNAR but are assumed MAR): q_{12} and q_{13} are the transition intensities from state 1 to state 2 and 3 respectively, and ϕ is the time transformation parameter

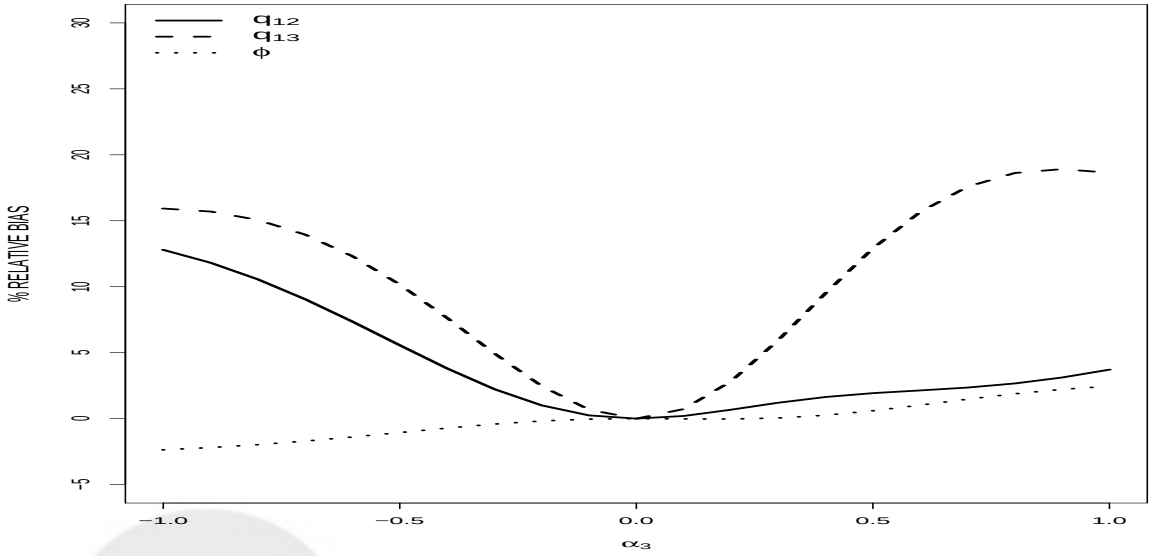

Table 1: Empirical performance of the proposed method and naive method for simulation studies with sample size $n = 200$: $\alpha_3 \neq 0$ leads to a missing not at random mechanism; $\alpha_3 = 0$ leads to a missing at random mechanism; the missing proportions are about 25%, 35% and 40% with $\alpha_3 = \log(2.0)$, $\log(1.5)$ and $\log(1.0)$

		Proposed Method				Complete Case			
α_3	Parameters	BIAS%	ASE	ESE	CP%	BIAS%	ASE	ESE	CP%
$\log(2.0)$	q_{12}	-1.7	0.043	0.044	94.2	247.4	0.126	0.124	9.7
	q_{13}	-1.6	0.025	0.024	95.2	314.0	0.058	0.061	1.8
	q_{21}	-1.1	0.081	0.080	94.8	144.6	0.225	0.222	66.1
	q_{23}	-0.8	0.040	0.040	94.7	196.9	0.115	0.115	57.6
	ϕ	1.2	0.094	0.094	94.4	-43.8	0.247	0.249	52.7
$\log(1.5)$	q_{12}	1.8	0.081	0.078	94.7	342.3	0.154	0.157	12.7
	q_{13}	1.3	0.042	0.039	95.1	443.2	0.069	0.066	3.2
	q_{21}	3.7	0.143	0.143	94.4	259.7	0.258	0.261	52.4
	q_{23}	0.8	0.098	0.097	95.1	291.8	0.130	0.130	49.2
	ϕ	1.3	0.136	0.137	94.8	-45.6	0.298	0.304	57.1
$\log(1.0)$	q_{12}	-1.0	0.100	0.101	94.7	476.7	0.263	0.268	20.3
	q_{13}	-1.3	0.040	0.039	95.5	861.8	0.114	0.111	4.3
	q_{21}	1.7	0.184	0.185	94.6	718.2	0.399	0.394	27.9
	q_{23}	0.7	0.141	0.139	95.7	613.4	0.143	0.145	36.4
	ϕ	1.2	0.131	0.131	94.7	-26.2	0.492	0.488	34.8

The true values of the parameters are $q_{12} = 0.2$, $q_{13} = 0.1$, $q_{21} = 0.2$, $q_{23} = 0.1$, and $\phi = 1.0$.

Table 2: Empirical performance of the proposed method and naive method for simulation studies with sample size $n = 1000$: $\alpha_3 \neq 0$ leads to a missing not at random mechanism; $\alpha_3 = 0$ leads to a missing at random mechanism; the missing proportions are about 25%, 35% and 40% with $\alpha_3 = \log(2.0)$, $\log(1.5)$ and $\log(1.0)$

		Proposed Method				Complete Case			
α_3	Parameters	BIAS%	ASE	ESE	CP%	BIAS%	ASE	ESE	CP%
$\log(2.0)$	q_{12}	-1.0	0.030	0.030	0.951	171.1	0.105	0.106	10.6
	q_{13}	-0.8	0.015	0.014	0.954	139.4	0.046	0.046	11.4
	q_{21}	-1.0	0.050	0.050	0.947	124.4	0.184	0.191	42.5
	q_{23}	-0.5	0.032	0.032	0.950	81.1	0.095	0.100	23.5
	ϕ	0.7	0.056	0.055	0.949	-54.8	0.219	0.217	62.2
$\log(1.5)$	q_{12}	0.3	0.036	0.034	0.955	245.2	0.134	0.136	14.0
	q_{13}	0.3	0.018	0.017	0.951	182.8	0.059	0.061	15.5
	q_{21}	0.5	0.062	0.061	0.948	164.5	0.243	0.260	38.6
	q_{23}	-0.6	0.038	0.037	0.944	128.4	0.126	0.133	22.8
	ϕ	0.1	0.067	0.066	0.945	-61.2	0.238	0.242	58.5
$\log(1.0)$	q_{12}	-1.0	0.035	0.034	0.954	113.0	0.402	0.405	13.8
	q_{13}	-1.0	0.018	0.017	0.956	234.9	0.143	0.139	20.6
	q_{21}	-0.5	0.060	0.060	0.950	138.9	0.534	0.540	19.2
	q_{23}	-1.6	0.038	0.039	0.945	95.8	0.292	0.291	26.9
	ϕ	0.4	0.067	0.064	0.956	-43.1	0.422	0.429	34.6

The true values of the parameters are $q_{12} = 0.2$, $q_{13} = 0.1$, $q_{21} = 0.2$, $q_{23} = 0.1$, and $\phi = 1.0$.

Table 3: Number of transitions made between states at successive clinic visits

	Normal	MCI	Dementia	Death
Normal	4410 (87.9)	393 (7.8)	65 (1.3)	149 (3.0)
MCI	289 (13.5)	1342 (62.7)	397 (18.6)	111 (5.2)
Dementia	42 (1.3)	62 (1.9)	2702 (82.2)	481 (14.6)

Table 4: Baseline risk factors for the 7962 subjects

Risk factor	Summary
Female, %	55.9
Congestive heart failure, %	4.0
GDS	2.0 ± 2.5
Family history of dementia, %	27.9
Diabetes, %	12.6
Hypertension, %	54.5
Years of education	15.7 ± 8.5
MMSE	26.3 ± 8.1
Age	75.5 ± 9.3

Table 5: Comparisons of two methods for the multiplicative effects on the transition intensities in the studies of Alzheimer's disease: hazard ratios and 95% confidence intervals

	Complete Case			EM Algorithm		
Parameter	HR	95%LCL	95%UCL	HR	95%LCL	95%UCL
Normal → MCI:						
SEX(F)	1.055	0.469	2.376	1.018	0.834	1.243
CVCHF	1.003	0.579	1.736	1.000	0.600	1.668
GDS	0.994	0.757	1.305	0.990	0.731	1.342
fhdem	1.388	1.106	1.743	1.269	1.029	1.565
diabete	1.007	0.196	5.174	1.197	0.885	1.619
hypert	0.965	0.365	2.550	0.928	0.274	3.146
EDUC	1.009	0.989	1.029	1.011	0.974	1.049
MMSE	1.024	0.968	1.084	1.025	0.950	1.107
AGE	1.029	1.017	1.042	1.025	1.013	1.037
Normal → Dementia:						
SEX(F)	0.846	0.614	1.167	0.919	0.702	1.205
CVCHF	0.252	0.034	1.854	0.237	0.043	1.299
GDS	1.151	1.050	1.263	1.124	1.054	1.199
fhdem	1.383	1.018	1.877	1.206	0.924	1.574
diabete	0.927	0.197	4.360	1.003	0.335	3.000
hypert	1.714	1.048	2.804	1.565	1.185	2.067
EDUC	1.004	0.965	1.044	1.004	0.981	1.028
MMSE	1.010	0.958	1.065	1.010	0.964	1.059
AGE	1.082	1.041	1.125	1.076	1.045	1.108
Normal → Death:						
SEX(F)	0.854	0.659	1.106	0.928	0.781	1.102
CVCHF	2.171	0.819	5.750	1.868	0.538	6.486
GDS	1.089	1.003	1.182	1.087	1.033	1.143
fhdem	0.724	0.470	1.114	0.748	0.503	1.111
diabete	1.699	1.066	2.709	1.933	1.273	2.934
hypert	1.197	0.758	1.890	1.138	0.926	1.398
EDUC	1.003	0.978	1.029	1.007	0.991	1.023
MMSE	1.017	0.984	1.052	1.018	0.989	1.049
AGE	1.100	1.059	1.141	1.099	1.071	1.127
MCI → Normal:						
SEX(F)	0.776	0.463	1.299	0.819	0.535	1.253
CVCHF	1.629	0.917	2.893	1.797	0.975	3.312
GDS	1.011	0.883	1.157	1.014	0.879	1.170
fhdem	1.001	0.520	1.926	0.882	0.426	1.826
diabete	1.110	0.219	5.623	1.001	0.288	3.475
hypert	1.022	0.491	2.128	1.036	0.626	1.714
EDUC	1.005	0.953	1.060	1.008	0.983	1.034
MMSE	1.025	0.986	1.066	1.026	0.987	1.067
AGE	0.984	0.932	1.040	0.980	0.905	1.062

Table 6: Comparisons of two methods for the multiplicative effects on the transition intensities in the studies of Alzheimer's disease: hazard ratios and 95% confidence intervals (Continued)

	Complete Case			EM Algorithm		
Parameter	HR	95%LCL	95%UCL	HR	95%LCL	95%UCL
MCI → Dementia:						
SEX(F)	0.710	0.591	0.852	0.710	0.596	0.845
CVCHF	0.542	0.265	1.106	0.515	0.264	1.002
GDS	1.084	1.045	1.126	1.076	1.036	1.117
fhdem	1.134	0.932	1.380	1.177	0.981	1.412
diabete	1.000	0.747	1.339	1.029	0.779	1.360
hypert	1.013	0.841	1.220	1.067	0.895	1.273
EDUC	0.986	0.963	1.010	0.984	0.967	1.002
MMSE	1.000	0.977	1.024	0.999	0.978	1.021
AGE	1.021	1.009	1.033	1.024	1.012	1.036
MCI → Death:						
SEX(F)	0.708	0.665	0.754	0.592	0.549	0.637
CVCHF	2.221	1.649	2.992	1.998	1.591	2.508
GDS	1.166	1.144	1.190	1.158	1.134	1.184
fhdem	0.579	0.344	0.976	0.552	0.518	0.587
diabete	0.647	0.310	1.353	0.878	0.486	1.587
hypert	1.385	0.895	2.145	1.432	0.954	2.148
EDUC	1.006	0.998	1.014	1.004	0.986	1.022
MMSE	1.012	0.996	1.028	1.013	0.996	1.028
AGE	1.082	1.076	1.089	1.083	1.077	1.090
Dementia → Normal:						
SEX(F)	0.683	0.346	1.349	0.719	0.388	1.333
CVCHF	0.000	0.000	Inf	0.000	0.000	Inf
GDS	1.172	1.075	1.278	1.206	1.135	1.281
fhdem	1.003	0.734	1.370	1.132	0.862	1.487
diabete	1.000	0.221	4.523	1.004	0.321	3.142
hypert	0.946	0.578	1.546	0.871	0.643	1.180
EDUC	1.006	0.979	1.034	0.995	0.972	1.019
MMSE	0.994	0.973	1.016	0.995	0.978	1.013
AGE	0.995	0.959	1.033	0.999	0.972	1.027
Dementia → MCI:						
SEX	0.535	0.284	1.008	0.569	0.290	1.114
CVCHF	0.904	0.415	1.968	0.751	0.374	1.505
GDS	1.190	1.133	1.250	1.195	1.142	1.250
fhdem	0.986	0.760	1.280	1.067	0.850	1.340
diabete	1.202	0.841	1.717	1.439	0.730	2.835
hypert	1.052	0.825	1.342	0.999	0.799	1.249
EDUC	0.949	0.876	1.029	0.997	0.980	1.015
MMSE	0.991	0.968	1.015	0.986	0.965	1.008
AGE	0.987	0.972	1.003	0.992	0.979	1.006
Dementia → Death:						
SEX	0.671	0.627	0.719	0.649	0.591	0.713
CVCHF	1.735	1.573	1.914	1.779	1.551	2.041
GDS	1.140	1.131	1.149	1.130	1.117	1.143
fhdem	1.058	0.972	1.151	1.121	0.925	1.358
diabete	0.823	0.586	1.155	0.808	0.600	1.089
hypert	1.009	0.938	1.085	0.972	0.883	1.070
EDUC	1.008	0.999	1.015	1.009	1.000	1.016
MMSE	0.882	0.879	0.886	0.879	0.874	0.884
AGE	1.050	1.046	1.054	1.046	1.042	1.050

Table 7: Missing data model in the analysis of Alzheimer's disease

Parameter	Estimate	SE	p-value
Intercept	3.172	0.271	<0.001
SEX(F)	-0.225	0.057	<0.001
CVCHF	0.161	0.153	0.292
GDS	-0.005	0.011	0.689
fhdem	-1.355	0.060	<0.001
diabete	-0.147	0.084	0.081
hypert	0.156	0.058	0.007
EDUC	0.005	0.004	0.213
MMSE	-0.009	0.003	0.005
AGE	-0.004	0.003	0.204
I(MCI)	-1.205	0.069	<0.001
I(Dementia)	-0.676	0.070	<0.001