Analysis of Interval-Censored Disease Progression Data via Multi-State Models under a Nonignorable Inspection Process

Baojiang Chen, Grace Y. Yi* and Richard J. Cook

Department of Statistics and Actuarial Science
University of Waterloo
200 University Avenue West
Waterloo, Ontario,
Canada N2L 3G1

*Corresponding author : Grace Y. Yi, phone: 519-888-4567 x35110; fax: 519-746-1875; email: yyi@uwaterloo.ca
Analysis of Interval-Censored Disease Progression Data via Multi-State Models under a Nonignorable Inspection Process

Summary

Progressive models provide a convenient framework for characterizing disease processes which arise, for example, when the state represents the degree of the irreversible damage incurred by the disease. In many settings, however, individuals are not seen in continuous time, but rather are only examined at periodic clinic visits. In this case, in standard analyses it is often assumed that the examination times are prespecified or are chosen completely independently of the disease process given covariates. If the life history and observation processes are not conditionally independent, the observation process contains information on the life history process, and more importantly, likelihoods based on the disease process alone are invalid (Grüger et al., 1991). With interval censored failure time data, joint models are nonidentifiable and data analysts must rely on sensitivity analyses to assess the effect of dependent observation times. However, this paper is concerned with the analysis of data from progressive multi-state disease processes in which individuals are seen at periodic, individual-specific, assessment times. We cast the problem in the convenient framework of incomplete longitudinal data problems. Maximum likelihood estimation, based on an EM algorithm, is advocated for parameter estimation, and a variance estimates are obtained using the Louis’s method (Louis, 1982). Simulation studies demonstrate that the proposed method works well under a variety of situations. Data from a cohort of patients with psoriatic arthritis are analyzed for illustration.

Keywords: EM Algorithm; Longitudinal Data; Maximum Likelihood; Progressive Models.
1 Introduction

Multi-state life history data arise in many research areas such as medicine, social sciences and public health, and multi-state models provide a convenient way to characterize the movement of individuals among distinct states. In health research, for instance, multi-state models are often employed to provide a comprehensive description of complex disease processes, including estimation of transition intensities and state occupancy probabilities for various states over time. With continuous time multi-state models, transition intensities are often of primary interest, and these are perhaps most widely modeled using Markov models. Their use can be traced back to Bartholomew (1983), Singer and Spilerman (1976a, b) and Wasserman (1980), among others. 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). With discrete time models, it is difficult to handle observation time which are not evenly spaced, but Kalbfleish and Lawless (1985, 1989) proposed an effective method for dealing with irregular spacing of observation times using time-homogeneous or power function transition intensities in Markov models. Gentleman et al. (1994) adapted that method to accommodate time nonhomogeneous models via piecewise constant transition intensities. When interest focuses on disease progression or models allowing transitions in only one direction, progressive models provide a convenient and attractive framework for modeling changes in responses over time. For progressive time-homogeneous Markov models, Satten (1999) provided a closed form of transition probabilities which are expressed in terms of transition intensities, and allows direct estimation via the likelihood function. Cook et al. (2004) described a conditional Markov model with multivariate random effects to handle clustered, conditionally Markov, progressive 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.

In most applications of multi-state models to data arising from intermittent inspection processes, it is assumed that the examination times are prespecified or are chosen completely independently of the response process. However, if the life history process and the follow-up process are not independent, the follow-up process may contain the information on the life history process so the processes must be modeled simultaneously to ensure the valid inferences (Grüger et al., 1991). Sun and Wei (2000) considered semiparametric regression methods for the analysis of panel count data when both the observation and censoring times may depend on covariates. Sun, Tong and He (2007) proposed an estimating equation approach for the analysis of panel count data with dependent observation times. Betensky and Finkelstein (2002) discussed test procedures for dependence between failure time and visit compliance for interval-censored data.
There remains, however, challenges when modeling various disease and follow-up processes. In cohort studies, clinical assessments may be scheduled at roughly equal intervals (e.g. annually), but patients may choose when they want to visit clinics for clinical examination according to their degree of disease activity. This creates somewhat akin to incomplete data in longitudinal studies when data may be missing at random (MAR) if missing status depends on the observed, typically past, responses, or missing not at random (MNAR), where the missing status may depend on the latent disease status. The latter situation is particularly difficult to deal with in general and in most settings analysts must rely on sensitivity analyses to examine the possible effect of this type of observation scheme. We consider, however, progressive models for chronic disease processes, which by their progressive nature, are convenient for jointly modeling the disease and observation processes. We provide a general method to handle this type of data. 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’ (1982) method.

The remainder of this paper is organized as follows. Section 2 reviews models and estimation for continuous time progressive models, and reports a numerical study of the naive analysis under dependent inspection schemes. Section 3 describes an EM algorithm for parameter estimation in joint models for the disease and inspection processes, and simulation studies are conducted to assess the performance of the proposed method. Data arising from the motivating psoriatic arthritis study (Gladman et al., 1995) are analyzed with the proposed method in Section 4. We conclude the paper with a general discussion in Section 5.

2 Continuous Time Progressive Multi-State Models

2.1 Notation and Model Formulation

Suppose there are $K$ states and the transition direction is irreversible. Figure 1 is an illustrative diagram of a $K$-state progressive transition model. Let $Y(t)$ represent the state occupied at time $t \geq 0$, and $\mathcal{H}(t) = \{Y(s), 0 \leq s < t\}$ denote the history of the response process which records the states occupied over the interval $[0, t)$. The transition probability function is written generally as $P(Y(s+t) = k|Y(s) = k', \mathcal{H}(s))$ for $s, t > 0$, and $k \geq k'$, but under a Markov model this simplifies to $P(Y(s+t) = k|Y(s) = k')$, which we denote compactly as $P_{k'k}(s, s+t)$, $k \geq k'$.

Insert Figure 1 here
The Markov transition intensity function at time \( t \) for transitions from state \( k \) to state \( k + 1 \), is
\[
\lambda_k(t) = \lim_{\Delta t \to 0} \frac{P(Y(t + \Delta t) = k + 1|Y(t) = k)}{\Delta t}, \quad k = 1, \ldots, K - 1,
\]
(Cox and Miller, 1977). A multi-state progressive model with state space \( \{1, 2, \ldots, K\} \) can then be described via the following transition intensity matrix, \( Q(t) \):
\[
Q(t) = \begin{pmatrix}
-\lambda_1(t) & \lambda_1(t) & 0 & \cdots & 0 & 0 \\
0 & -\lambda_2(t) & \lambda_2(t) & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & -\lambda_{K-1}(t) & \lambda_{K-1}(t) \\
0 & 0 & 0 & \cdots & 0 & 0
\end{pmatrix}.
\]

Under a time-homogeneous Markov model, let \( \lambda_{0k}(t) = \lambda_k, k = 1, \ldots, K - 1 \), and it follows from stationarity that \( P_{k'k}(s, s + t) = P_{k'k}(0, t) \), which we may now write simply as \( P_{k'k}(t) \). Let \( P(t) \) denote the \( K \times K \) matrix with \( (k', k) \) element \( P_{k'k}(t) \). We assume \( \lambda_1, \ldots, \lambda_{K-1} \) are distinct, and let \( \lambda = (\lambda_1, \ldots, \lambda_{K-1})' \). For a time-homogeneous model the transition probability from state \( k' \) to state \( k \) over \( [0, t] \) is given by
\[
P_{k'k}(t) = \begin{cases}
\sum_{j=k'}^{k} C(k', j, k; \lambda) \exp(-\lambda_j t), & \text{if } k' \leq k, \\
0, & \text{if } k' > k,
\end{cases}
\]
where the coefficients are given by
\[
C(k', j, k; \lambda) = \frac{\prod_{h=k}^{k-1} \lambda_h}{\prod_{h=k', h \neq j} (\lambda_h - \lambda_j)}
\]
for \( k' \leq j \leq k \), and \( C(j, j, j; \lambda) = 1, j = 1, 2, \ldots, K \) (Satten, 1999). In the simulations and application that follow we focus on time-homogeneous Markov models, but extensions which accommodate nonhomogeneous Markov models can be developed in the same spirit, and so we retain the dependence on \( t \) in the following remarks.

To model the dependence of the transition intensities on prognostic variables, we may incorporate covariates in the preceding formulation by expressing the transition intensities as functions of time (in the nonhomogeneous case) and the covariates. That is, let \( \lambda_k(t) = g_k(t, X) \) for some non-negative known function \( g_k(\cdot, \cdot), k = 1, \ldots, K - 1 \), where \( X \) represents the covariates vector. For a given individual \( i \), we often adopt models of the form
\[
\lambda_{ik}(t) = \lambda_{0k}(t) \exp(X_{ik} \beta_k), \quad k = 1, \ldots, K - 1,
\]
(2.1)
where the $\lambda_{0k}(t)$ are the baseline transition intensities which may or may not depend on $t$, and
$\beta_k$ is a vector of regression coefficients associated with the covariates of interest, $X_{ik}, k = 1, \ldots, K - 1$. This setup permits the baseline transition intensities and the regression coefficients to vary across the possible transitions. We let $X_i = (X_{i1}', \ldots, X_{iK-1}')'$ and $\beta = (\beta_1, \ldots, \beta_{K-1})'$ denote the vector of all covariates and regression coefficients.

With continuous time models and observation schemes, the response process $\{Y(t), t > 0\}$ may be observed at any time point $t$ over the period of observation. In practice, however, individuals are often observed at random, individual-specific times, which are not necessarily evenly spaced, and their states are determined at these visit times. Let $t_{i1} < t_{i2} < \cdots < t_{iJ_i}$ denote the assessment times for subject $i$, $H_{ij}^a = \{t_{ik}, k = 1, \ldots, j - 1\}$, $H_{ij}^y = \{Y_i(t_{ik}), k = 1, \ldots, j - 1\}$, and $H_{ij} = \{(t_{ik}, Y_i(t_{ik})), k = 1, \ldots, j - 1\}$. If we condition on the initial time of assessment and the initial state, the full observed data likelihood contribution from subject $i$, suppressing dependence on the covariates, is then

$$L_i = \prod_{j=2}^{J_i} P(t_{ij}, Y_i(t_{ij})|H_{ij}) = \prod_{j=2}^{J_i} P(Y_i(t_{ij})|H_{ij}) \prod_{j=2}^{J_i} P(t_{ij}|Y_i(t_{ij}), H_{ij}).$$

(2.2)

The model for the underlying stochastic process does not typically feature a dependence on the previous assessment times, and so $P(Y_i(t_{ij})|H_{ij}) = P(Y_i(t_{ij})|H_{ij}^y)$ is a quite natural assumption. Indeed we usually desire to base inferences strictly on the product of such terms. If $P(t_{ij}|Y_i(t_{ij}), H_{ij})$ does not depend on $Y_i(t_{ij})$ (i.e. the time of the assessment does not depend on the state of the underlying process) then we can treat $\prod_{j=2}^{J_i} P(Y_i(t_{ij})|H_{ij}^y)$ the same as if it were the probability of the observed states, conditional on the assessment times, and this is typically an implicit assumption in standard analyses. If, on the other hand, $P(t_{ij}|Y_i(t_{ij}), H_{ij})$ does depend on $Y_i(t_{ij})$, then we must consider the full likelihood based on (2.2). In this case, one needs to model the conditional distributions of the examination times (i.e. $P(t_{ij}|Y_i(t_{ij}), H_{ij})$) which can be challenging. In this paper we consider the problem in which subjects are scheduled to be examined at pre-specified assessment times denoted $a_1, a_2, \ldots, a_J$, and adopt the convenient framework commonly employed to handle incomplete longitudinal data.

Let $R_{ij}$ be the indicator random variable, which equals $1$ if response $Y_i(a_j)$ is observed and $0$ otherwise. Let $R_i = (R_{i1}, R_{i2}, \ldots, R_{ij})'$, and $r_i = (r_{i1}, r_{i2}, \ldots, r_{ij})'$ be a realization of $R_i$. Here we assume all the subjects are observed at the initial enrollment, i.e., $R_{i1} = 1$. Let

$$\lambda_{ij}^* = P(R_{ij} = 1|H_{ij}^r, Y_i, X_i)$$

be the conditional probability, where $H_{ij}^r$ denotes the history of the missing indicators until the $(j - 1)$st time point. When models with a first order dependence are of interest, we write

$$\lambda_{ij}^* = P(R_{ij} = 1|R_{ij-1}, Y_i, X_i).$$

A logistic regression model is commonly employed to postulate the conditional probability $\lambda_{ij}^*$, i.e.,

$$\logit(\lambda_{ij}^*) = Z_{ij}' \alpha,$$
where \( \alpha \) is a parameter vector, and \( Z_{ij} \) is a covariate vector featuring various missingness.

Let \( \theta = (\alpha', \beta')' \). Then the likelihood for the complete data is given by \( L(\theta) = \prod_{i=1}^{n} L_i(\theta; y_i) \), where

\[
L_i(\theta; y_i) = P(R_i = r_i | Y_i = y_i, X_i; \alpha)P(Y_i = y_i | X_i; \beta)
\]

(2.3)

or equivalently, its logarithm is

\[
\ell_i(\theta; y_i) = \sum_{j=2}^{J} \{r_{ij}\log \lambda_{ij}^{*} + (1 - r_{ij})\log(1 - \lambda_{ij}^{*})\}
\]

(2.4)

\[
+ \sum_{j=2}^{J} \sum_{k'=1}^{K} \sum_{k=k'}^{K} I(Y_i(a_{j-1}) = k', Y_i(a_j) = k)\log P_{ik'k}(a_j - a_{j-1}).
\]

### 2.2 Asymptotic Bias Under Dependent Inspection

In the presence of missing values, we may base inference about \( \theta \) on using (2.3) or (2.4). The detail will be presented in Section 3. Here we investigate the impact of ignoring missingness, or the fact of dependent inspection, through a simulation study. Specifically, we employ the available data analysis and the complete case analysis that are often used in practice due to their simplicity of implementation. We investigate this problem through application of the theory of misspecified models.

Let \( \ell^*(\beta^*) = \sum_{i=1}^{n} \log P(Y_i^{(*)} | X_i) \) be the naive likelihood function where \( Y_i^{(*)} \) represents the available data or the complete case data. Here \( \beta^* \) is used to stress that the associated parameter may be different from the parameter of interest \( \beta \). Solving

\[
S^*(\beta^*) = \frac{\partial \ell^*(\beta^*)}{\partial \beta^*} = 0
\]

leads to a naive estimate \( \hat{\beta}^* \).

White (1982) showed that \( \hat{\beta}^* \) converges to \( \beta^* \) almost surely, where \( \beta^* \) solves

\[
E_{Y,R,X}[S^*(\beta^*)] = 0.
\]

Here \( E_{Y,R,X} \) denotes the expectation taken with respect to the joint distribution (2.4) of \( (Y, R, X) \) which depends on \( \beta \) and \( \alpha \). In general, it is difficult to obtain an analytical expression for \( \beta^* \) by solving this equation. Instead, to understand the magnitude in bias, or difference \( \beta^* - \beta \), we proceed with a numerical evaluation.
We consider the case with $K = 3$ states and $J = 5$ time points, and assume that the intensity function for transitions from state $k$ to state $k + 1$ is

$$
\lambda_k = \lambda_0 e^{\beta_k x}, \quad k = 1, \ldots, K - 1,
$$

where the baseline function is modeled as $\lambda_0 = \lambda_0 e^{\gamma(k-1)}$, and $x$ is generated from $\text{Bin}(1, 0.5)$, representing a treatment indicator, for instance. The true values of the coefficients are taken to be $\beta_0 = 0.5$, $\gamma = 0.2$, and $\beta_k = 1/k, k = 1, \ldots, K - 1$. The study duration, $\tau$, is selected such that $P(T < \tau) = 0.9$ where $T$ denotes the time to entry of state $K$. The assessment time points are chosen as $a_j = (j - 1)/(J - 1) \cdot \tau, j = 1, \ldots, J$, equally cutting the interval $[0, \tau]$. We assume that all of the subjects are observed at the first assessment time, which is plausible in settings where the observation process begins upon entry to a clinic. The conditional probabilities $\lambda_{ij}^*$ are modeled as

$$
\logit(\lambda_{ij}^*) = \alpha_0 + \alpha_1 (1 - r_{i,j-1}) + \alpha_2 y_i(a_{j-1}) + \alpha_3 (y_i(a_j) - y_i(a_{j-1})) + \alpha_4 x_i.
$$

Note that $\alpha_2 \neq 0$ or $\alpha_3 \neq 0$ represents a nonignorable missing mechanism; if $\alpha_2 = 0$ and $\alpha_3 = 0$, this is a missing completely at random case (Little and Rubin, 1987; Laird, 1988).

In the study considered here we set $\alpha_0 = \log(4)$, $\alpha_1 = \log(0.75)$ and $\alpha_4 = \log(2)$. The parameters $\alpha_2$ and $\alpha_3$ are respectively considered to change from $\log(0.5)$ to $\log(2.0)$ to reflect varying degrees of the missing data proportion and the dependence of the missingness on the previous observation and the present observation. For example, as $\alpha_2$ and $\alpha_3$ increase, the missingness proportion reduces, and the dependence on unobserved data becomes weaker (if $\alpha_2 \leq 0$ and $\alpha_3 \leq 0$) or stronger (if $\alpha_2 \geq 0$ and $\alpha_3 \geq 0$). The results are displayed in Figures 2 and 3. Not surprisingly, the asymptotic biases produced by the complete case analysis are bigger than those obtained from the available data analysis, and as the absolute values of $\alpha_2$ and $\alpha_3$ decrease, the biases become smaller. Note that, for the baseline intensity function, the biases are all negative when $\alpha_2 < 0$ and $\alpha_3 < 0$, suggesting the naive baseline intensity estimates are underestimated. However, when $\alpha_2 > 0$ and $\alpha_3 > 0$, the biases of the baseline intensity function indicate overestimated results.

*Insert Figures 2 and 3 here*
3 Likelihood Methods for Dependent Inspection Schemes

3.1 An EM Algorithm

In this section we develop an EM algorithm for valid inference. In the E step, we construct the conditional expectation, at the $h$th iteration,

$$Q(\theta; \theta^{(h)}) = \sum_{i=1}^{n} Q_i(\theta; \theta^{(h)}),$$

where $Q_i(\theta; \theta^{(h)}) = E[\ell_i(\theta, y_i) | Y_i^{(o)}, \theta^{(h)}] = \sum_{y_i^{(m)}, w_i(y_i; \theta^{(h)})} \ell_i(\theta, y_i), y_i$ is written as $(y_i^{(m)}, y_i^{(o)})$ to explicitly indicate missing and observed components, and

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

facilitating the conditional probability of the missing data given the observed data. Here $L_i(\cdot)$ and $\ell_i(\cdot)$ are the complete data likelihood and log-likelihood given by (2.3) and (2.4), respectively. The dependence on the covariates is suppressed in the notation.

We note, by (2.4), that parameters $\alpha$ and $\beta$ can be separated in the conditional expectation $Q(\theta; \theta^{(h)})$ with the form:

$$Q(\theta; \theta^{(h)}) = Q_1(\alpha; \theta^{(h)}) + Q_2(\beta; \theta^{(h)}),$$

where

$$Q_1(\alpha; \theta^{(h)}) = \sum_{i=1}^{n} \sum_{j=2}^{J} \sum_{y_i^{(m)}} w_i(y_i; \theta^{(h)}) \cdot \{r_{ij}\log\lambda_{ij}^{*} + (1 - r_{ij})\log(1 - \lambda_{ij}^{*})\}$$

and

$$Q_2(\beta; \theta^{(h)}) = \sum_{i=1}^{n} \sum_{j=2}^{J} \sum_{k'=1}^{K} \sum_{k=1}^{K} \sum_{y_i^{(m)}} w_i(y_i; \theta^{(h)}) \cdot I(Y_i(a_{j-1}) = k', Y_i(a_j) = k) \cdot \log P_{ik'k}(a_j - a_{j-1}).$$

Consequently, in the M-step we can separately maximize the progressive model part $Q_2(\beta; \theta^{(h)})$ and the missing data process part $Q_1(\alpha; \theta^{(h)})$. Standard statistical software package such as SAS can be readily adapted to implement this step. Iterate through the E and M steps until $\theta^{(h)}$ converges. Denote the limit as $\hat{\theta}$.

To estimate variance for $\hat{\theta}$, we use the Louis's formula (Louis, 1982), given by

$$\Sigma(\hat{\theta}) = -\frac{\partial^2 Q(\hat{\theta}; \hat{\theta})}{\partial \theta \partial \theta'} - \sum_{i=1}^{n} \sum_{y_i^{(m)}} w_i(y_i; \hat{\theta}) S_i(\hat{\theta}) S_i'(\hat{\theta}) + \sum_{i=1}^{n} \left( \frac{\partial Q_i(\hat{\theta}; \hat{\theta})}{\partial \theta} \right)' \left( \frac{\partial Q_i(\hat{\theta}; \hat{\theta})}{\partial \theta} \right),$$

7
where $S_i(\hat{\theta}) = \partial\ell_i(\theta; y_i) / \partial\theta|_{\theta=\hat{\theta}}$. The estimate of the asymptotic covariance matrix of $\hat{\beta}$ is the lower $p_2 \times p_2$ block of $[\Sigma(\hat{\theta})]^{-1}$, where $p_2$ is the dimension of $\beta$.

### 3.2 Simulation Study

In this section we report on a simulation study to assess the performance of the proposed method. We consider the case with $K = 3$ states and $J = 5$ time points, and a sample of $n = 500$ individuals. Two thousand samples are simulated for each parameter configuration. The intensity function and the missing data model are the same as (2.5) and (2.6) in Section 2.2, respectively. We set $\alpha_0 = \log(4)$, $\alpha_1 = \log(0.75)$, $\alpha_2 = \log(0.5)$, $\alpha_3 = \log(0.5)$ and $\alpha_4 = \log(2)$.

Here we consider the cases with no covariates (i.e., $\lambda_k = \lambda_{0k}$) and with one covariate (i.e., $\lambda_k = \lambda_{0k}e^{\beta_kx}$, $k = 1, 2$) included in the response model. We conduct three analyses – complete case analysis, available data analysis and the analysis using the proposed method. The results are reported in Tables 1 and 2, where SEL denotes the average standard error calculated based on the Louis’s formula, ASE is the average naive standard error calculated using the Hessian matrix, ESE is the empirical standard errors for the 2000 estimates, and CP represents the 95% coverage probability of the parameters. Table 1 displays the results for the case without covariates under two scenarios. In Scenario I the missingness proportions for $Y_2$ to $Y_5$ are about 20%, 28%, 30% and 30%, respectively, while in Scenario II the missingness proportions for $Y_2$ to $Y_5$ are about 48%, 60%, 65% and 65%, respectively. Similarly, Table 2 reports the results for the case with one covariate under two scenarios. In Scenario I the missingness proportions for $Y_2$ to $Y_5$ are about 25%, 28%, 30% and 30%, respectively, while in Scenario II the missingness proportions for $Y_2$ to $Y_5$ are about 45%, 60%, 60% and 62%, respectively.

Insert Tables 1 and 2 here

It can be seen that both the complete case analysis and the available data analysis produce biased estimates, whereas the proposed method yields satisfactory results with considerably smaller finite sample biases. As expected, as the proportion of missing observations increases the biases produced by the complete case and available data analyses become larger, but the proposed method retains small bias. Comparisons between the ASE and ESE suggest that the effect of missing data on variance estimation is not as striking as that on parameter estimation. Variance estimation based on SEL adjusts for missingness and the results agree with the empirical version (ESE) much better than the naive version of the ASE does. Furthermore, the coverage probabilities of the parameters obtained from the proposed method agree well with the nominal level 95%
under different settings, but the complete case and available data analyses yield coverage probabilities that are far away from the nominal value, and in some situations they may completely fail to capture the true values of the parameters.

In many settings, prevalence functions, such as the one giving the proportion of subjects in the absorbing state, are of interest. Graphical plots in Figure 4 reveal how the three methods differ in estimation of the survival function \( S(t) = 1 - P_{1K}(t) \) for the case without covariates. The estimates obtained from the proposed method are almost identical to the true survival functions, however, the survival probabilities estimated from the available data analysis and the complete case analysis are both above the true curve, revealing a positive bias in the survival probabilities. It is not surprising that the complete case analysis produces a curve that is farther apart from the true curve than the available data analysis does. The differences among the curves become more substantial as time increases.

Insert Figure 4 here

4 Application to the Psoriatic Arthritis Clinic Data

Psoriatic arthritis (PsA) is a progressive disease in the sense that without treatment, it can increase in severity causing disability through deformity and destruction of the joints. It is of interest to determine prognostic factors that relate to disease severity and rates of disease progression (Gladman et al., 1995, 1998). Upon entry to the clinic, a comprehensive list of demographic and clinical features are recorded. Covariates include duration of psoriasis at clinic entry (in years) (coded as PSORDUR), sex (coded as SEX, 0–Female, 1–Male), age at onset of PsA (in years) (coded as AGEPSA), family history of psoriasis (coded as FMPS, 0–No, 1–Yes), family history of PsA (coded as FMPSA, 0–No, 1–Yes) and erythrocyte sedimentation rate (ESR). Patients are then scheduled to be assessed annually and at each followup assessment the number of damaged joints, as determined by clinical examination, is recorded. Table 3 lists a sample data set. There are 703 subjects with complete covariates, in which 28 subjects have complete observations over the first 10 years of their participation in the clinic registry. That is, there are 675 subjects with missing observations at different assessment time points, leading to a missing proportion about 61.3%.

Insert Table 3 here
Here we consider a multi-state Markov model with four states defined by the number of damaged joints determined by clinical assessment, as used by Gladman et al. (1995, 1998). Specifically, 0, 1-4, 5-9 and 10 or more damaged joints correspond to states 1, 2, 3 and 4 representing no damage, mild, moderate and severe damage, respectively. The rationale behind this state structure is that a larger number of damaged joints corresponds to a more severe disease. Figure 5 displays the transitions among the four states.

**Insert Figure 5 here**

Let \( Y_i(a_j) \) denote the state subject \( i \) was in at time \( a_j, j = 0, \ldots, 10 \). The transition intensity functions are modeled as

\[
\lambda_{ik} = \lambda_{0k} \exp(\beta_1 \cdot \text{PSORDUR}_i + \beta_2 \cdot \text{AGEPSA}_i + \beta_3 \cdot \text{FMPS}_i \\
+ \beta_4 \cdot \text{FMPSA}_i + \beta_5 \cdot \text{ESR}_i + \beta_6 \cdot \text{SEX}_i), \quad k = 1, 2, 3,
\]

where the \( \lambda_{0k} \)'s are the baseline intensities. For the missing data process, we assume

\[
\logit(\lambda^*_{ij}) = \alpha_0 + \alpha_1 \cdot \text{PSORDUR}_i + \alpha_2 \cdot \text{AGEPSA}_i + \alpha_3 \cdot \text{FMPS}_i \\
+ \alpha_4 \cdot \text{FMPSA}_i + \alpha_5 \cdot \text{ESR}_i + \alpha_6 \cdot \text{SEX}_i + \alpha_7 \cdot Z_{ij12} + \alpha_8 \cdot Z_{ij13} \\
+ \alpha_9 \cdot Z_{ij14} + \alpha_{10} \cdot Z_{ij23} + \alpha_{11} \cdot Z_{ij24} + \alpha_{12} \cdot Z_{ij34} \\
+ \alpha_{13} \cdot I(Y_i(a_{j-1}) = 2) + \alpha_{14} \cdot I(Y_i(a_{j-1}) = 3) + \alpha_{15} \cdot I(Y_i(a_{j-1}) = 4) \\
+ \alpha_{16} \cdot r_{i,j-1},
\]

where \( Z_{ijk'k} = I(Y_i(a_{j-1}) = k', Y_i(a_j) = k) \) is the indicator featuring the transitions from states \( k' \) to \( k \) at time \( a_j, k' < k \).

**Insert Table 4 here**

Table 4 reports the results obtained from the proposed method as well as from the complete case and available data analyses that ignore the missing data mechanism. The duration of psoriasis at clinic entry has a significant effect on the rate of PsA progression (\( \hat{\beta}_1 = 0.057 \) with p-value<0.001); that is, the relative rate of progression increases 5.9% for each additional year since diagnosis, controlling for other factors. The age at onset of PsA is also significantly associated with the rate of transition (\( \hat{\beta}_2 = -0.073; \) p-value<0.001); that is, the older the age at onset the slower the rate of progression (the risk decreases about 7.0% for each additional year of age.
at onset of PsA, when controlling other factors). A family history of psoriasis or PsA were not significantly related to the rate of progression ($\hat{\beta}_3 = -0.064$; p-value=0.560 and $\hat{\beta}_4 = -0.103$, p-value=0.423 respectively), but ESR level has an effect on PsA progression ($\hat{\beta}_5 = 0.013$; p-value<0.001) such that those with a higher ESR value have rates of damage (the relative risk increases about 1.3% for one unit of ESR increasing when controlling other factors). The effect of SEX is significant ($\hat{\beta}_6 = 0.177$; p-value=0.030), indicating that males have higher rates of progression than females ($RR = 1.194$).

In Figure 6 we plot the transition probabilities starting from state 1 to other possible states to show the differences of the three analyses. It is seen that the available data analysis tends to yield less different curves from those obtained from the proposed method than the complete case analysis does. The probabilities staying in state 1 decrease as time goes by, while transition probabilities $P_{13}(t)$ and $P_{14}(t)$ have increasing trends with time. However, transition probabilities $P_{12}(t)$ obtained from the three methods are quite different. The proposed method produces a first-increasing-then-decreasing curve, the available data analysis yields a first-increasing-then-stable curve, but the complete case analysis leads to a fairly straight, increasing curve.

Insert Figure 6 here

For the missing data process model, we find that the $\alpha_j$ coefficients with $j = 7, 8, 9, 10, 12, 13, 14$ and 15 are all significant, suggesting that nonignorable missing mechanisms are perhaps reasonable. In particular, we report that $\hat{\alpha}_{13} = 0.403$ with p-value< 0.001, $\hat{\alpha}_{14} = 0.757$ with p-value< 0.001 and $\hat{\alpha}_{15} = 0.779$ with p-value< 0.001. It suggests that the more severe the disease at the previous assessment, the more likely he or she would appear for the present assessment. This seems to make intuitive sense since patients may be more likely to attend a clinic when the disease becomes more severe. If subjects are missing at an assessment, they are less likely to be observed at the next assessment because the estimate of $\alpha_{16}$ is 1.865 with p-value< 0.001. As for the covariates, only the duration of initial psoriasis is significant ($\hat{\alpha}_1 = -0.026$; p-value< 0.001), indicating the shorter the duration, the more likely for a patient to appear for assessment.

5 Discussion

In this paper we develop a full likelihood analysis for jointly modeling disease progression via a multistate model, and the observation process. It is well known that identifiability is often an issue associated with nonignorable mechanisms. However, unlike the case with less structured
response models, the progressive nature of this response model means that nonignorable inspection processes can be easily handled. This is achieved because there are a small number of states that could have been visited between any two time points and for any particular set of states, the number of transitions is known exactly.

Here we adopt a convenient framework that is often used to deal with longitudinal missing data problems. Specifically, we considered a setting with a discrete time observation scheme, which is appropriate when visits occur at pre-specified times. This facilitates use of indicator variables to indicate availability of data at these prescheduled times. More generally, however, in the continuous time observation setting, one can specify a stochastic model for the observation process which characterizes the intensity function for the assessment times. This would require stronger assumptions about the observation scheme, and some loss of robustness. The proposed method is easy to implement and can accommodate various missing data mechanisms. The simulation studies demonstrate that the proposed method performs well under various situations.

In this paper we focus the discussion on incomplete response data, but in practice data often feature missing covariates. In principle, the proposed method can be adapted to accommodate missing covariate data, or missing covariate and response data. The joint likelihood of the two types of missing data indicators, the response and the covariates that may be missing, may be formulated for complete data, and an EM algorithm can be used again for estimation in the spirit discussed here.

A number of important questions can be posed using the covariate information provided at clinic entry. However, in other settings interest may lie in the effect of time-varying covariates. Relatively little work has been done on fitting regression models with interval censored 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. Cook et al. (2008) consider an extension to the bivariate setting where both the covariate and failure times are interval censored.

We have focussed on the time-homogeneous Markov model in this paper. This assumption can be easily relaxed to increase the flexibility of the model. Weakly parametric (e.g. piecewise constant intensities) models may be adopted to model \( \lambda_{0k}(t) \) in model (2.1) along the lines of Gentleman et al. (1994). Alternatively, one can use splines to obtain smoother estimates of transition intensities if desired (Staniswalis et al., 1997), or local likelihood methods (Loader, 1996, 1999). Nonparametric methods such as those of Turnbull (1976) can in principle be adapted for the setting of dependent observation schemes when models are progressive. Interval censored recurrent event data (e.g., Thall and Lachin, 1988; Wellner and Zhang, 2000) arise from progressive models, and further work in this area is warranted.
Acknowledgements

The authors are grateful to Dr. Dafna Gladman for permission to use the data from the psoriatic arthritis study. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (G. Y. Yi and R. J. Cook) and the Canadian Institutes for Health Research (R. J. Cook). R. J. Cook is Canada Research Chair in Statistical Methods for Health Research.

References


Figure 1: A diagram of K-state progressive process
Figure 2: Asymptotic biases resulted from the complete case analysis with a three-state model under a dependent inspection process
Figure 3: Asymptotic biases resulted from the available data analysis with a three-state model under a dependent inspection process.
Figure 4: Comparisons of the estimated survival functions obtained from the three analyses with the true curve for the case without covariates ($K = 3$ and $J = 5$).
Figure 5: Four-state progression diagram for psoriatic arthritis data
Figure 6: Transition probabilities for the analysis of the psoriatic arthritis data
Table 1: Simulation results for the case without covariates

<table>
<thead>
<tr>
<th>Parameters†</th>
<th>EM</th>
<th>Complete Case</th>
<th>Available Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>SEL</td>
<td>ASE</td>
</tr>
<tr>
<td>Scenario I: Transition Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.004</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.001</td>
<td>0.035</td>
<td>0.033</td>
</tr>
<tr>
<td>Missing Data Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.006</td>
<td>0.134</td>
<td>0.115</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.008</td>
<td>0.124</td>
<td>0.114</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>-0.002</td>
<td>0.058</td>
<td>0.052</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>-0.003</td>
<td>0.051</td>
<td>0.048</td>
</tr>
<tr>
<td>Scenario II: Transition Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.001</td>
<td>0.026</td>
<td>0.023</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.003</td>
<td>0.041</td>
<td>0.040</td>
</tr>
<tr>
<td>Missing Data Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.006</td>
<td>0.105</td>
<td>0.092</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.002</td>
<td>0.103</td>
<td>0.090</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.004</td>
<td>0.052</td>
<td>0.049</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>0.002</td>
<td>0.051</td>
<td>0.049</td>
</tr>
</tbody>
</table>

† $\lambda_1 = 0.500$ and $\lambda_2 = 0.611$

Scenario I: $\alpha_0 = \log(40), \alpha_1 = \log(0.75), \alpha_2 = \log(0.85), \alpha_3 = \log(0.95)$

Scenario II: $\alpha_0 = \log(40), \alpha_1 = \log(0.75), \alpha_2 = \log(0.05), \alpha_3 = \log(0.05)$
Table 2: Simulation Results for the case with covariates

<table>
<thead>
<tr>
<th>Parameters†</th>
<th>EM</th>
<th>Complete Case</th>
<th>Available Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>SEL</td>
<td>ASE</td>
</tr>
<tr>
<td>Scenario I:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transition Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda_{01} )</td>
<td>0.001</td>
<td>0.033</td>
<td>0.031</td>
</tr>
<tr>
<td>( \lambda_{02} )</td>
<td>0.005</td>
<td>0.051</td>
<td>0.046</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.013</td>
<td>0.108</td>
<td>0.098</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>0.001</td>
<td>0.117</td>
<td>0.102</td>
</tr>
<tr>
<td>Missing Data Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_0 )</td>
<td>0.006</td>
<td>0.159</td>
<td>0.137</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.004</td>
<td>0.125</td>
<td>0.110</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>-0.003</td>
<td>0.063</td>
<td>0.056</td>
</tr>
<tr>
<td>( \alpha_3 )</td>
<td>-0.001</td>
<td>0.063</td>
<td>0.059</td>
</tr>
<tr>
<td>( \alpha_4 )</td>
<td>0.006</td>
<td>0.106</td>
<td>0.091</td>
</tr>
<tr>
<td>Scenario II:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transition Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda_{01} )</td>
<td>0.002</td>
<td>0.037</td>
<td>0.033</td>
</tr>
<tr>
<td>( \lambda_{02} )</td>
<td>0.005</td>
<td>0.057</td>
<td>0.053</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.003</td>
<td>0.116</td>
<td>0.105</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-0.003</td>
<td>0.126</td>
<td>0.117</td>
</tr>
<tr>
<td>Missing Data Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_0 )</td>
<td>0.007</td>
<td>0.128</td>
<td>0.118</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>-0.001</td>
<td>0.105</td>
<td>0.098</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>-0.003</td>
<td>0.060</td>
<td>0.053</td>
</tr>
<tr>
<td>( \alpha_3 )</td>
<td>-0.006</td>
<td>0.060</td>
<td>0.056</td>
</tr>
<tr>
<td>( \alpha_4 )</td>
<td>0.004</td>
<td>0.099</td>
<td>0.090</td>
</tr>
</tbody>
</table>

†\( \lambda_1 = 0.500, \lambda_2 = 0.611, \beta_1 = 1.000 \) and \( \beta_2 = 0.500 \)

Scenario I: \( \alpha_0 = \log(40), \alpha_1 = \log(0.75), \alpha_2 = \log(0.85), \alpha_3 = \log(0.95), \alpha_4 = \log(20) \)

Scenario II: \( \alpha_0 = \log(40), \alpha_1 = \log(0.75), \alpha_2 = \log(0.85), \alpha_3 = \log(0.95), \alpha_4 = \log(20) \)
Table 3: Sample data of the psoriatic arthritis study

<table>
<thead>
<tr>
<th>ID</th>
<th>PSORDUR</th>
<th>AGEPSA</th>
<th>FMPS</th>
<th>FMPSA</th>
<th>ESR</th>
<th>SEX</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21.5</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>38.3</td>
<td>40</td>
<td>1</td>
<td>0</td>
<td>36</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>3</td>
<td>15.1</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>23.0</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>7.1</td>
<td>34</td>
<td>0</td>
<td>0</td>
<td>83</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>11.7</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>.</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>7.4</td>
<td>28</td>
<td>1</td>
<td>1</td>
<td>16</td>
<td>1</td>
<td>.</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>4.2</td>
<td>23</td>
<td>1</td>
<td>1</td>
<td>34</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>2</td>
<td>.</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>9</td>
<td>11.8</td>
<td>49</td>
<td>1</td>
<td>0</td>
<td>23</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>56.7</td>
<td>31</td>
<td>1</td>
<td>0</td>
<td>47</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>41.1</td>
<td>32</td>
<td>1</td>
<td>0</td>
<td>65</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>10.1</td>
<td>25</td>
<td>1</td>
<td>0</td>
<td>26</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>4</td>
<td>4</td>
<td>.</td>
<td>4</td>
<td>.</td>
<td>4</td>
<td>.</td>
<td>.</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>31.6</td>
<td>70</td>
<td>1</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>14</td>
<td>33.0</td>
<td>40</td>
<td>1</td>
<td>0</td>
<td>17</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>1</td>
<td>.</td>
<td>1</td>
<td>.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>35.9</td>
<td>51</td>
<td>0</td>
<td>0</td>
<td>57</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>1</td>
<td>.</td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>2</td>
<td>.</td>
<td>.</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>10.1</td>
<td>36</td>
<td>1</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>17</td>
<td>23.6</td>
<td>43</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>4</td>
<td>4</td>
<td>.</td>
<td>4</td>
<td>.</td>
<td>4</td>
<td>.</td>
<td>.</td>
<td>4</td>
</tr>
<tr>
<td>18</td>
<td>16.8</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>74</td>
<td>0</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>0.2</td>
<td>43</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>0</td>
<td>1</td>
<td>.</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>19.9</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>0</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>29.9</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>99</td>
<td>0</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>4</td>
</tr>
<tr>
<td>22</td>
<td>11.9</td>
<td>21</td>
<td>1</td>
<td>1</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>12.3</td>
<td>23</td>
<td>1</td>
<td>0</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>31.0</td>
<td>36</td>
<td>1</td>
<td>1</td>
<td>73</td>
<td>0</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>4</td>
</tr>
<tr>
<td>25</td>
<td>10.6</td>
<td>36</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>26</td>
<td>9.1</td>
<td>50</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>30.2</td>
<td>36</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1</td>
</tr>
</tbody>
</table>

PSORDUR: duration of psoriasis at time of clinic entry (years); AGEPSA: age at onset of psoriatic arthritis (years);
FMPS: family history of psoriasis (0–No, 1–Yes); FMPSA: family history of psoriatic arthritis (0–No, 1–Yes);
ESR: erythrocyte sedimentation rate; SEX: 0–Female, 1–Male.
Table 4: Analyses of the psoriatic arthritis data

<table>
<thead>
<tr>
<th>Transition Model</th>
<th>Proposed Method</th>
<th>Complete Case</th>
<th>Available Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline Intensities</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{01}$</td>
<td>0.064</td>
<td>0.016</td>
<td>0.013</td>
</tr>
<tr>
<td>$\lambda_{02}$</td>
<td>0.116</td>
<td>0.030</td>
<td>0.013</td>
</tr>
<tr>
<td>$\lambda_{03}$</td>
<td>0.150</td>
<td>0.042</td>
<td>0.030</td>
</tr>
<tr>
<td><strong>Covariate Effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSORDUR</td>
<td>0.057</td>
<td>0.005</td>
<td>0.044</td>
</tr>
<tr>
<td>AGEPSA</td>
<td>-0.073</td>
<td>0.007</td>
<td>0.025</td>
</tr>
<tr>
<td>FMPS</td>
<td>-0.064</td>
<td>0.110</td>
<td>0.579</td>
</tr>
<tr>
<td>FMPSA</td>
<td>-0.103</td>
<td>0.130</td>
<td>0.101</td>
</tr>
<tr>
<td>ESR</td>
<td>0.013</td>
<td>0.002</td>
<td>0.012</td>
</tr>
<tr>
<td>SEX</td>
<td>0.177</td>
<td>0.083</td>
<td>0.298</td>
</tr>
</tbody>
</table>

PSORDUR: duration of psoriasis at time of clinic entry (years); AGEPSA: age at onset of psoriatic arthritis (years); FMPS: family history of psoriasis (0–No, 1–Yes); FMPSA: family history of psoriatic arthritis (0–No, 1–Yes); ESR: erythrocyte sedimentation rate; SEX: 0–Female, 1–Male.