Lifetime Data Analysis (2014), **20**(3): 404-423
DOI: 10.1007/s10985-013-9271-7

A dynamic Mover-Stayer model for recurrent event processes subject to resolution

HUA SHEN

Department of Statistics and Actuarial Science,

University of Waterloo, Waterloo, ON, N2L 3G1, Canada

E-mail: hshen@uwaterloo.ca

RICHARD J. COOK

Department of Statistics and Actuarial Science,

University of Waterloo, Waterloo, ON, N2L 3G1, Canada

E-mail: rjcook@uwaterloo.ca

Summary

In studies of affective disorder, individuals are often observed to experience recurrent symptomatic exacerbations warranting hospitalization. Interest may lie in modeling the occurrence of such exacerbations over time and identifying associated risk factors. In some patients, recurrent exacerbations are temporally clustered following disease onset, but cease to occur after a period of time.We develop a dynamic Mover-Stayer model in which a canonical binary variable associated with each event indicates whether the underlying disease has resolved. An individual whose disease process has not resolved will experience events following a standard point process model governed by a latent intensity. When the disease process resolves, the complete data intensity becomes zero and no further event will occur. An expectation- maximization algorithm is described for parametric and semiparametric model fitting based on a discrete time dynamic Mover-Stayer model and a latent intensity-based model of the underlying point process.

Keywords: Dynamic Mover-Stayer model, EM algorithm, Recurrent event data, Semiparametric estimation

The final publication is available at Springer via http://dx.doi.org/10.1007/s10985-013-9271-7.

1 INTRODUCTION

1.1 RECURRENT EVENT PROCESSES

Recurrent data arise frequently in studies of chronic disease, actuarial science, industrial research and sociology. In health research, examples include exacerbations of symptoms in patients with respiratory disease (Grossman et al., 1998), seizures in individuals with epilepsy (Pledger et al., 1994), and recurrent episodes of bleeding in patients with thrombocytopenia (Heddle et al., 2003; Webert et al., 2006). There has been considerable statistical research in the last 20 years on methods for the analysis of recurrent event data. Models and methods can be broadly classified as intensity-based (Andersen et al., 1993), based on marginal mean or rate functions (Lawless and Nadeau, 1995), or based on random effect models (Lawless, 1987).

Frequently the recurrent event process ends upon on the occurrence of a terminal event. Graft rejection episodes in transplant recipients, for example, cease to occur upon total graft rejection (Cole et al., 1993), skeletal complications in patients with bone metastases end when a patient dies (Hortobagyi et al., 1998), and recurrent hospitalizations for cardiovascular events end upon death (Bourassa, Gurné, Bangdiwala et al., 1993). There has been considerable recent work on the development of statistical methods for the analysis of recurrent events in the presence of a terminal event. This phenomenon is naturally handled with intensity-based models (Andersen et al., 1993), but robust marginal methods have been developed (Cook and Lawless, 1997; Ghosh and Lin, 2000, 2002) as have models and methods incorporating random effects (Liu, Wolfe and Huang, 2004; Ye, Kalbfleisch and Schaubel, 2007).

We consider the setting in which recurrent events arise in a chronic disease processes but where some individuals have particularly long periods of time from their last event to a right-censoring time. This is motivated by the need to model recurrent event processes in which the recurrent events arise because of a transient underlying condition which can resolve. Unlike the case of a terminal event such as death, in this setting it is not known if and when the underlying condition has resolved. We handle this complication through use of a dynamic mover-stayer model. The model is comprised of an intensity function for event occurrence among individuals still experiencing the underling condition generating the events and a series of conditional probabilities for modeling the resolution of the underlying process.

Mixture models have been used extensively to model the presence of a so-called "cured fraction" in cancer studies featuring long-term survivors. Farewell (1982, 1986) proposed a parametric mixture model incorporating a logistic regression model for the latent cure status and a Weibull model for the survival times of those in the uncured group. Peng, Dear and Denham (1998) extended this approach to incorporate the generalized F failure time distribution and Taylor (1995) extended this further to enable nonparametric estimation of the survival distribution among susceptible individuals through a Kaplan-Meier type estimate. Kuk and Chen (1992) extended the cure rate model to accommodate a semiparametric proportional hazard model for the survival time and proposed estimation via an expectation-maximization (EM) algorithm. Peng and Dear (2000) further studied the semiparametric approach by allowing covariate effects on the cure rate. A zero-tail constraint was introduced by Sy and Taylor (2000) to deal with identifiability issues. Yamaguchi (1992) described a further interesting generalization of the notion of a cured fraction by introducing a latent failure time at which subjects became nonsusceptible to the event of interest. Asymptotic properties of maximum likelihood estimates from the cure rate model, including the existence, strong consistency and asymptotic normality, were studied by Fang, Li and Sun (2005); asymptotic variances were also derived to facilitate inferences using Wald-based pivotals.

Cure rate survival models are a special case of a more general class of mover-stayer models. In mover-stayer models the population is comprised of two sub-populations. In one sub-population, the so-called "mover" group, transitions among states are made according to a general multistate process. In the other sub-population individuals have a zero probability of moving from the initial state, and these individuals are called "stayers". Often Markov models are adopted for the multistate process for movers, but any multistate model can be specified in principle. Goodman (1961) proposed methods for consistent parameter estimation to address inconsistency of estimators developed by Blumen, Kogan and McCarthy (1955) in the discrete-time setting. Spilerman (1972) further generalized the mover-stayer model to allow the individual mobility rate to follow a continuous distribution. Frydman (1984) described how to obtain maximum likelihood estimates based on the observed likelihood, while Fuchs and Greenhouse (1988) used the EM algorithm with extensions to handle incomplete follow-up in the panel studies. Models incorporating dynamic mover-stayer indicators have received some attention including the multistate models by Heckman and Walker (1987), Yamaguchi (1994, 1998, 2003) and Cook, Kalbfleisch and Yi (2002).

The remainder of this paper is organized as follows. A brief description of the Danish study on the course of affective disorder is given in the next subsection. In Section **??** we introduce the notation and give the complete data likelihood for a general model. We next describe how to implement the EM algorithm, and give specific details on how to fit a semiparametric latent Markov model. The performance of the proposed algorithm for parametric and semiparametric models is examined empirically in Section **??**. Several models are fit to the motivating Danish study of affective disorder in Section **??** and concluding remarks and topics warranting further development are discussed in Section **??**.

1.2 Description of Danish Registry Data

A study of individuals with affective disorder was carried out in Denmark based on a registry of hospitalizations. For this study, a patient entered the cohort at the onset of affective disorder, defined by the first hospitalization for any mental disorder of inorganic etiology between 1994 and 1999. A total of 10,523 individuals satisfied this selection condition. Over the course of the study period there was an average of 1.618 admissions (SD=1.720), with a minimum of 1 and a maximum of 90.

Kvist et al. (2007) examined the impact of misspecification of the frailty distribution, using a non-parametric estimator for the joint gap times and a marginalized estimator for marginal gap times. Cook and Lawless (2012) investigated trends in this recurrent event process and discussed the tests for trends in detail.

The present goal is to describe a model for the pattern of event occurrence in which the events are the acute exacerbations of affective disorder and data feature individuals with unusually long periods of time without recurrence at the end of follow-up; see Figure **??**. This pattern prompted the development (Winokur, 1975) and examination (Kessing, Hansen, Andersen et al., 2004) of a theory that the disease process may "burn-out" for some affected individuals. This theory, in part, motivated the development of the dynamic mover-stayer model we describe in the section that follows.

2 Model Formulation and Algorithm for Estimation

2.1 Notation and Model Formulation

We suppose the process of interest begins with an initiating event representing the onset of disease. This could be, for example, the first seizure among individuals with epilepsy, the first acute exacerbation in persons with asthma, or the first hospitalization in individuals with affective disorder. We let $T_0 = 0$ denote the time of the initiating event and let T_j represent the time of the jth subsequent event, $j = 1, 2, \ldots$. The number of events over time period $(0, t]$ is denoted by $N(t) = \sum_{j=1}^{\infty} I(T_j \leq t)$, and $\{N(s), 0 \leq s\}$ denotes the corresponding counting process.

Information on the nature of the event, individuals' characteristics at the event time, and any fixed covariates, are recorded in a $p \times 1$ covariate vector X_j observed upon the occurrence of the jth event. We let $\bar{X}(t) = \{X_0, \ldots, X_{N(t)}\}$ denote the history of this covariate vector when viewed in continuous time; because $N(t)$ is right-continuous this history includes X_j if $t = t_j$. Likewise we let $\bar{X}_j = \{X_0, \ldots, X_j\}$ denote the covariate history as a function of event count. To accommodate the possibility that the condition of interest is resolved upon the occurrence of the jth event, we let Z_j denote a time-dependent indicator variable such that $Z_j = 1$ if the individual remains at risk for future events following the jth event, and $Z_j = 0$ otherwise, $j = 0, 1, \ldots$. The indicator Z_j is a latent variable, but we learn that $Z_j = 1$ upon the occurrence of the $(j + 1)$st event, $j = 0, 1, \ldots$. As was done for the observed covariate vector, here we let $\bar{Z}(t) = \{Z_0, \ldots, Z_{N(t)}\}$ and $\bar{Z}_j = \{Z_0, \ldots, Z_j\}$.

The complete process history is denoted by $\mathcal{H}(t) = \{(N(s), X(s), Z(s)), 0 \leq s \leq t\}$, which includes the values of the latent variables realized over $[0, t]$, and the history excluding $\bar{Z}(t)$ is denoted

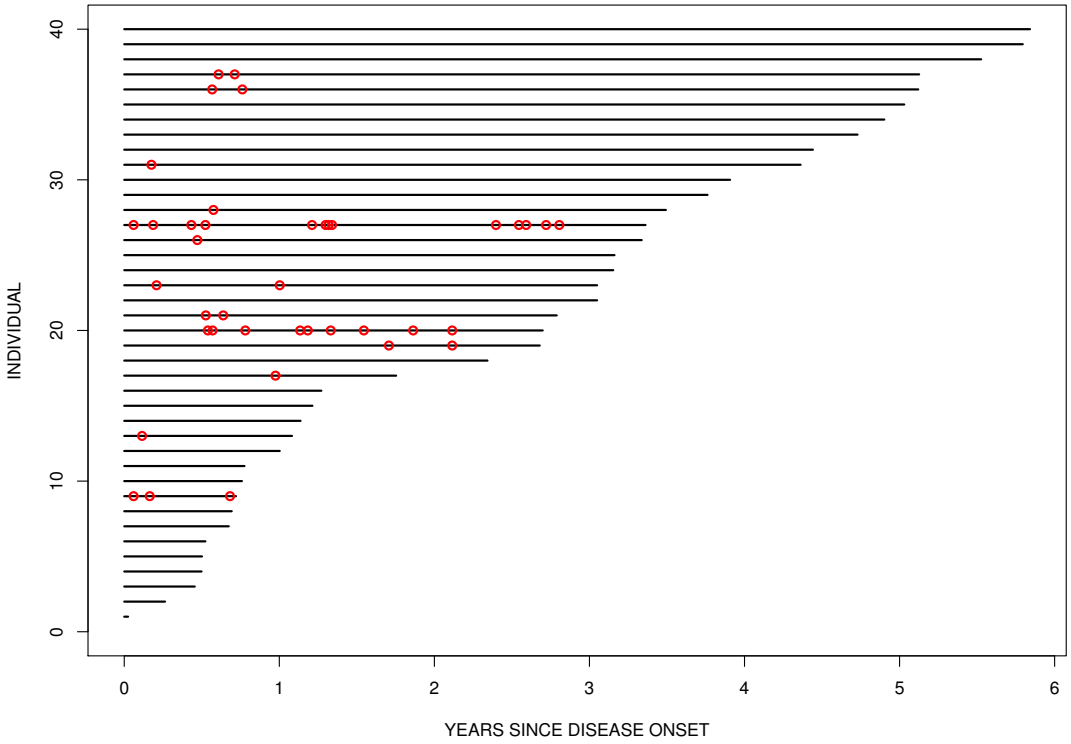

Figure 1: Timeline plots of recurrent acute episodes of affective disorder from time of disease onset for a selected sample of individuals

by $H(t) = \{(N(s), X(s)), 0 \leq s \leq t\}$. We let t^- denote an infinitesimal amount of time before t. Assuming two events cannot occur at the same time, the *complete data intensity function* is

$$\lambda(t|\mathcal{H}(t^-)) = \lim_{\Delta t \downarrow 0} \frac{P(\Delta N(t) = 1|\mathcal{H}(t^-))}{\Delta t} = Z_{N(t^-)}\lambda(t|H(t^-)) \ , \tag{1}$$

where $\Delta N(t) = N((t+\Delta t)^-) - N(t^-)$ denotes the number of the events over the interval $[t, t+\Delta t)$ and

$$\lambda(t|H(t^-)) = \lim_{\triangle t \downarrow 0} \frac{P(\triangle N(t) = 1|H(t^-))}{\triangle t} \tag{2}$$

is a *canonical event intensity function*. We use the term complete data intensity function for (**??**) because it contains the complete information over $[0, t)$ including information on the latent process; we use the term canonical intensity for (**??**) because it can be any intensity function useful for modeling recurrent event processes not subject to resolution. It may, for example, correspond to any point process model including modulated Markov models for which $\lambda(t|H(t^-)) = \lambda_0(t;\alpha)\exp(X'_{N(t^-)}\beta)$, or modulated semi-Markov models for which $\lambda(t|H(t^-)) = h_{N(t^-)}(B(t);\alpha)\exp(X'_{N(t^-)}\beta)$, where $h_j(w_j;\alpha)$ is the baseline hazard for the inter-arrival time $w_j = t_j - t_{j-1}$ and $B(t) = t - t_{N(t^-)}$ is the backwards recurrence time at $t > 0$ (Lawless, 1995). Mixed Markov and semi-Markov processes offer alternative frameworks (Cook and Lawless, 2007). The canonical intensity is not relevant alone for modeling the data, however, and the complete intensity is not useable since $\bar{Z}(t^-)$ is not observed. The *observed data intensity function* is obtained by marginalizing over the latent process and is of the form

$$E\{\lambda(t|\mathcal{H}(t^-))|H(t^-)\} = E(Z_{N(t^-)}|H(t^-)) \cdot \lambda(t|H(t^-)) \ .$$

As in models with fixed continuous frailty terms, here it is most convenient to adopt a latent variable approach to estimation and hence construct a complete data likelihood based on (**??**); we do so in the next section.

In general X_j can depend on the complete process history at t_j^- and the fact that an event occurred at t_j, so we denote the probability model by

$$P(X_j|\mathcal{H}(t_j^-), dN(t_j) = 1) = P(X_j|H(t_j^-), \bar{Z}_{j-1} = 1_{j-1}, dN(t_j) = 1) \tag{3}$$

where $\bar{Z}_{j-1} = 1_{j-1}$ is an $j\times 1$ vector of ones, and we somewhat informally let $dN(t) = \lim_{\Delta t \downarrow 0} \Delta N(t) = 1$ if an event occurs at time t and $dN(t) = 0$ otherwise.

The probability of remaining at risk following the jth event can depend upon $\mathcal{H}(t_j^-)$ and X_j, so at t_j we write this as

$$P(Z_j = 1|\mathcal{H}(t_j^-), dN(t_j) = 1, X_j) = P(Z_j = 1|H(t_j^-), \bar{Z}_{j-1} = 1_{j-1}, dN(t_j) = 1, X_j)\,. \tag{4}$$

This probability may therefore depend on the times of previous events and the history of the observable covariates over $[0, t_j]$ and is only relevant if $\bar{Z}_{j-1} = 1_{j-1}$. Discrete waiting time models are suitable for the resolution of the process and we may specify them based on logistic models. If $\dot{X}_j = (1, X_j')'$, a simple model is of the form

$$\text{logit}P(Z_j = 1|\mathcal{H}(t_j), dN(t_j) = 1, X_j) = \dot{X}_j'\eta_j \tag{5}$$

in which the odds the process does not resolve upon the occurrence of the jth event at t_j depends on the features X_j upon event occurrence. It is often convenient and reasonable to constrain $\eta_j = \eta$ and so there is one set of regression coefficients common across all logistic models.

Example 1 Suppose the canonical intensity is Markov with $\lambda(t|H(t^-)) = \lambda\alpha(\lambda t)^{\alpha-1}$, and a logistic model is used for the latent indicator with (**??**) taking the form $\text{logit}P(Z_j = 1|\mathcal{H}(t_j), dN(t_j) = 1, X_j) = \eta_0 + \eta_1 j + \eta_2 X$ where $X_j = (j, X)'$ with X being an indicator of a treatment $(X = 1)$ or control $(X = 0)$ condition. In this case $\exp(\eta_1)$ is the relative odds, given X, that the process remains unresolved at the jth event compared to at the previous event; the parameter η_1 therefore reflects the tendency for the process to remain unresolved upon the occurrence of each event, regardless of the times of the events. The coefficient η_2 reflects the possible effect of treatment on the odds the process remains unresolved after a given number of events.

The mean function gives the expected number of events as a function of the time and so is defined by $E\{N(t)|X\} = \sum_{n=0}^{\infty} nP(N(t) = n|X)$. To compute this, note that

$$\begin{aligned} P(N(t) = n|X) &= P(N(t) = n|\bar{Z}_n = 1_n, X)P(\bar{Z}_n = 1_n|X) \\ &\quad + P(T_n \le t|\bar{Z}_{n-1} = 1_{n-1}, X)P(\bar{Z}_{n-1} = 1_{n-1}|X)\,, \end{aligned}$$

where

$$P(N(t) = n|\bar{Z}_n = 1_n, X) = \Lambda(t|X)^n\, e^{-\Lambda(t|X)}/n!\,,$$

with $\Lambda(t|X) = \int_0^t \lambda(s|X)ds$ and

$$P(T_n \le t|\bar{Z}_{n-1} = 1_{n-1}, X) = 1 - \sum_{r=0}^{n-1} P(N(t) = r|\bar{Z}_{r-1} = 1_{r-1}, X)\,,$$

since the latent process is a Poisson process, and

$$P(\bar{Z}_n = 1_n|X) = P(Z_0 = 1|X)\prod_{j=1}^{n} P(Z_j = 1|\bar{Z}_{j-1} = 1_{j-1}, X)\,.$$

Figure **??** contains plots of the mean function based on the canonical intensity, and the mean functions for the marginal (observed) processes discussed here for the treatment ($X = 1$) and control ($X = 0$) groups; we set $\lambda = 36$, $\alpha = 0.50$, $\eta_1 = \log 0.95$, $\eta_2 = \log 0.75$ and determined η_0 to give $E(N(1)) = 0.75$ (left panel) or $E(N(1)) = 3$ (right panel). As expected there is a large difference in the expected number of events between the canonical and marginal models since the latter incorporate the chance that the process resolves during follow-up. The covariate effect on the mover-stayer process leads to two marginal mean functions (under the proposed model) with the difference between them reflecting magnitude of the effect of treatment on the mover-stayer indicator.

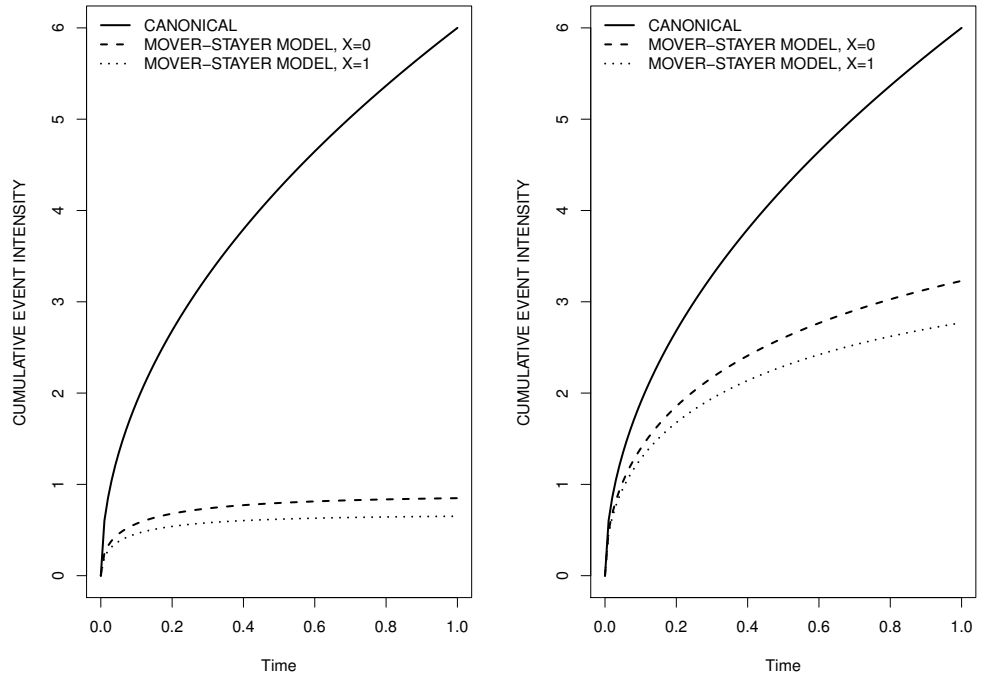

Figure 2: Plots of the cumulative canonical intensity $(\lambda t)^\alpha$ and mean functions for the treatment ($X = 1$) and control ($X = 0$) group in the dynamic mover-stayer model; $\lambda = 36$, $\alpha = 0.5$, $\eta_1 = \log 0.95$, $\eta_2 = \log 0.75$, η_0 is obtained to give $E(N(1)) = 0.75$ (left panel) and 3.0 (right panel)

2.2 Parameter Estimation and Statistical Inference

2.2.1 An EM Algorithm for Parametric Modeling

To describe the algorithm for estimation we return to the general case with a canonical Markov intensity of an unspecified form. Let θ_1 denote the parameter indexing the canonical intensity in (**??**), θ_2 parameterize (**??**), and θ_3 parameterize (**??**).

If the latent process were observable over an interval $[0, C]$, the complete data likelihood would be proportional to the probability of observing $\{(t_j, X_j, Z_j), j = 0, 1, \ldots, n\}$ over $[0, C]$ and is given

by $L_C \propto L_{C1}(\theta_1) \cdot L_{C2}(\theta_2) \cdot L_{C3}(\theta_3)$ where

$$\begin{aligned}
L_{C1}(\theta_1) &\propto \prod_{j=1}^{n} \left\{ \lambda(t_j|\mathcal{H}(t_j^-)) \exp\left(-\int_{t_{j-1}}^{t_j} \lambda(u|\mathcal{H}(u^-))du \right) \right\} \exp\left(-\int_{t_n}^{C} \lambda(u|\mathcal{H}(u^-))du \right) \\
L_{C2}(\theta_2) &\propto P(Z_0|\mathcal{H}(0^-), dN(0) = 1, X_0) \prod_{j=1}^{n} P(Z_j|\mathcal{H}(t_j^-), dN(t_j) = 1, X_j) \\
L_{C3}(\theta_3) &\propto P(X_0|\mathcal{H}(0^-), dN(0) = 1) \prod_{j=1}^{n} P(X_j|\mathcal{H}(t_j^-), dN(t_j) = 1)
\end{aligned}$$

and $\mathcal{H}(0^-) = \emptyset$. Terms involving the probability model for the observed covariates can be omitted if the covariate process is non-informative (i.e. the parameters indexing the distribution of the covariates are not functionally related to the parameters of the processes of interest). In this case we use the partial complete data likelihood

$$L_C(\theta) \propto L_{C1}(\theta_1) \cdot L_{C2}(\theta_2)\,, \tag{6}$$

where

$$L_{C1}(\theta_1) \propto \prod_{j=1}^{n} \left\{ \lambda(t_j|H(t_j^-)) \right\} \exp\left(-\sum_{k=0}^{n} \int_{t_k}^{t_{k+1}} Z_k^{I(k=n)} \lambda(u|H(u^-))du \right) \tag{7}$$

is the contribution pertaining to θ_1, with $t_0 = 0$ and $t_{n+1} = C$, and

$$\begin{aligned}
L_{C2}(\theta_2) &\propto P(Z_0|H(0^-), dN(0) = 1, X_0) \cdot \prod_{j=1}^{n} P(Z_j|H(t_j^-), \bar{Z}_{j-1} = 1_{j-1}, dN(t_j) = 1, X_j) \\
&\propto P(Z_0|H(0^-)) \cdot \prod_{j=1}^{n} P(Z_j|H(t_j^-), \bar{Z}_{j-1} = 1_{j-1}) \qquad (8)
\end{aligned}$$

is the contribution related to the latent process, where $H(0^-) = \emptyset$, and $\theta = (\theta_1', \theta_2')'$. The missing variable in the above complete data likelihood is Z_n, the indicator of whether the process continues following the occurrence of the last observed event.

The expectation-maximization (EM) algorithm of Dempster et al. (1979) offers a convenient way of maximizing the observed data likelihood. To do this we define

$$Q(\theta; \widehat{\theta}) = Q_1(\theta_1; \widehat{\theta}) + Q_2(\theta_2; \widehat{\theta}) \tag{9}$$

where $Q_1(\theta_1; \widehat{\theta}) = E(\log L_{C1}(\theta_1)|H(C); \widehat{\theta})$ and $Q_2(\theta_2; \widehat{\theta}) = E(\log L_{C2}(\theta_2)|H(C); \widehat{\theta})$. Since $\log L_{C1}(\theta_1)$ and $\log L_{C2}(\theta_2)$ are linear in Z_n, only

$$\zeta(\widehat{\theta}) = P(Z_n = 1|H(C); \widehat{\theta}), \tag{10}$$

given by

$$\frac{P(Z_n = 1|H(t_n^-), \bar{Z}_{n-1} = 1_{n-1}; \widehat{\theta}_2) \exp(-\int_{t_n}^{C} \lambda(u|H(u^-); \widehat{\theta}_1)\, du)}{P(Z_n = 1|H(t_n^-), \bar{Z}_{n-1} = 1_{n-1}; \widehat{\theta}_2) \exp(-\int_{t_n}^{C} \lambda(u|H(u^-); \widehat{\theta}_1)\, du) + P(Z_n = 0|H(t_n^-), \bar{Z}_{n-1} = 1_{n-1}; \widehat{\theta}_2)}\,,$$

is required at the E-step to compute (**??**). The maximum likelihood estimator is obtained by iteratively maximizing (**??**) as follows. If $\widehat{\theta}^r$ denotes the estimate of θ at the rth iteration, we maximize $Q(\theta; \widehat{\theta}^r)$ with respect to θ to obtain $\widehat{\theta}^{r+1}$. This process is repeated iteratively until $\|\widehat{\theta}^{r+1} - \widehat{\theta}^r\| \leq \epsilon$ where ϵ is a pre-specified tolerance, at which point we let the final value be the maximized likelihood estimate. Variance estimation can be carried out using the method of Louis (1982); see Appendix A for details.

The model formulation in this section is quite general. In the next section we consider a special model with a Markov canonical intensity with a proportional latent rate function and consider semi-parametric modeling of the canonical Markov intensity. To do this we introduce subscripts to index individuals and adopt counting process notation.

2.2.2 An EM Algorithm For Semiparametric Modeling of a Markov Process

Let m be the number of subjects in the study, n_i be the number of events for subject i, $[0, C_i]$ denote the period of observation for subject i and let $Y_i(u) = I(u \leq C_i)$ indicate whether they are under observation at time u. Let $Z_i(u) = Z_{iN_i(u^-)}$ denote the latent variable expressed as a continuous time varying indicator. Under a Markov latent intensity $\lambda(t|H(t^-)) = \lambda_0(t)\exp(X\beta)$, where $\lambda_0(t) = d\Lambda_0(t)/dt$ is the baseline latent intensity for an individual with $X = 0$. We also let $\Lambda_0(s,t) = \int_s^t d\Lambda_0(u)$. In counting process notation the complete data likelihood for the recurrent event process (Cook and Lawless, 2007) is

$$L_{C1}(\lambda_0(\cdot),\beta) = \prod_{i=1}^{m}\left[\prod_{j=1}^{n_i}[Y_i(u)d\Lambda(u|X_i)]^{Y_i(u)dN_i(u)}\exp\left(-\int_0^\infty Z_i(u)Y_i(u)d\Lambda(u|X_i)\right)\right]$$

and $L_{C2}(\theta_2)$ is the same as in (**??**). The complete log-likelihood is then $\ell_C(\theta) = \ell_{C1}(\lambda_0(\cdot),\beta) + \ell_{C2}(\theta_2)$ where

$$\ell_{C1}(\lambda_0(\cdot),\beta) = \sum_{i=1}^{m}\left\{\int_0^\infty Y_i(u)dN_i(u)(\log d\Lambda_0(u) + X_i\beta) - \int_0^\infty Z_i(u)Y_i(u)d\Lambda_0(u)\exp(X_i\beta)\right\},$$

and

$$\ell_{C2}(\theta_2) = \sum_{i=1}^{m}\Big[\sum_{j=0}^{n_i-1}\log P(Z_{ij}|H(t_{ij}^-),\bar{Z}_{j-1} = 1_{j-1}) + \log P(Z_{in_i}|H(t_{in_i}^-),\bar{Z}_{n_i-1} = 1_{n_i-1})\Big],$$

where we define $\bar{Z}_{-1}$ as the null set. Here $\theta = (\lambda_0(\cdot),\beta',\theta_2')$ where $\lambda_0(\cdot)$ is the latent baseline rate function, β is the covariate effect on the intensity of the latent process, and for the particular model discussed in Section **??**, for example, $\theta_2 = (\eta_0,\eta_1,\eta_2)$ is the parameter vector for the mover-stayer probability model. Then (**??**) becomes $Q(\theta;\widehat{\theta}) = Q_1(\lambda_0(\cdot),\beta;\widehat{\theta}) + Q_2(\theta_2;\widehat{\theta})$, and

$$Q_1(\lambda_0(\cdot),\beta;\widehat{\theta}) = \sum_{i=1}^{m}\left\{\int_0^\infty Y_i(u)dN_i(u)(\log d\Lambda_0(u) + X_i\beta) - \int_0^\infty \zeta_i(u;\widehat{\theta})Y_i(u)d\Lambda_0(u)\exp(X_i\beta)\right\}$$

where if $u < T_{in_i}$, $\zeta_i(u;\widehat{\theta}) = 1$; and if $T_{in_i} \leq u \leq C_i$, $\zeta_i(u;\widehat{\theta}) = E(Z_i(u)|H_i(C_i);\widehat{\theta})$ is given by (**??**) with $\exp(-\int_{t_n}^{C}\lambda(u|H(u^-);\widehat{\theta}_1)du)$ reduced to $\exp(-\Lambda(t_{in_i},C_i|X_i;\widehat{\theta}_1))$, where $\Lambda(s,t|X_i;\widehat{\theta}_1) = \int_s^t\lambda(u|X_i;\widehat{\theta}_1)du$ and $\theta_1 = (\lambda_0(\cdot),\beta)'$. The argument u in $\zeta_i(u;\widehat{\theta})$ is therefore introduced to facilitate writing a general expression for this expectation.

When maximizing $Q_1(\lambda_0(\cdot),\beta;\widehat{\theta})$ with respect to $\lambda_0(\cdot)$ and β, we obtain the two equations

$$\sum_{i=1}^{m}\Big[Y_i(u)dN_i(u) - \zeta_i(u;\widehat{\theta})Y_i(u)\exp(X_i\beta)d\Lambda_0(u)\Big] = 0,\ 0 < u \tag{11}$$

$$\sum_{i=1}^{m}\left[\int_0^\infty Y_i(u)dN_i(u)X_i - \int_0^\infty \zeta_i(u;\widehat{\theta})Y_i(u)d\Lambda_0(u)\exp(X_i\beta)X_i\right] = 0\,. \tag{12}$$

For a given β, we obtain the "profile" estimate

$$d\widehat{\Lambda}_0(u;\beta) = \frac{\sum_{i=1}^m Y_i(u)dN_i(u)}{\sum_{i=1}^m Y_i(u)\zeta_i(u;\widehat{\theta})\exp(X_i\beta)},$$

and substitute this into (**??**) to obtain the equation

$$\sum_{i=1}^{m}\int_0^\infty Y_i(u)dN_i(u)\left[X_i - \frac{\sum_{i=1}^m Y_i(u)\zeta_i(u;\widehat{\theta})\exp(X_i\beta)X_i}{\sum_{i=1}^m Y_i(u)\zeta_i(u;\widehat{\theta})\exp(X_i\beta)}\right].$$

This looks very much like the usual Cox partial likelihood score equation with offsets. For each subject i we can construct a pseudo-dataset with $n_i + 1$ lines: first n_i lines correspond to the period from 0 to t_{in_i} and have an offset of zero; the last line corresponds to the period from t_{in_i} to C_i and has an offset of $\log \zeta_i(u; \widehat{\theta})$. Existing software can therefore be used to obtain updated estimates of $\lambda_0(\cdot)$ and β.

The second term is

$$\begin{aligned} Q_2(\theta_2; \widehat{\theta}) \;=\; & \sum_{i=1}^{m} \Big[\sum_{j=0}^{n_i-1} \log P(Z_{ij} = 1 | H(t_{ij}^-), \bar{Z}_{j-1} = 1_{j-1}) \\ & + \zeta_i(\widehat{\theta}) \log P(Z_{in_i} = 1 | H(t_{in_i}^-), \bar{Z}_{n_i-1} = 1_{n_i-1}) \\ & + (1 - \zeta_i(\widehat{\theta})) \log P(Z_{in_i} = 0 | H(t_{in_i}^-), \bar{Z}_{n_i-1} = 1_{n_i-1}) \Big] \,. \end{aligned}$$

where $\zeta_i(\widehat{\theta}) = \zeta_i(u; \widehat{\theta})$ for $T_{in_i} \leq u$. Maximization of $Q_2(\theta_2; \widehat{\theta})$ with respect to θ_2 can be done by fitting logistic regression to pseudo-datasets, which contains $n_i + 2$ lines for each subject i: the first n_i lines correspond to $Z_{i0} = 1, \ldots, Z_{i,n_i-1} = 1$ and have weight 1; the next line corresponds to the possibility that $Z_{in_i} = 1$ and has weight $\zeta_i(\widehat{\theta})$; the final line corresponds to the other possibility that $Z_{in_i} = 0$ and has associated weight $1 - \zeta_i(\widehat{\theta})$.

3 Simulation Studies

Here we conduct simulation studies to evaluate the performance of the EM algorithm in fitting the dynamic mover-stayer model with a latent Markov process. We first generate a treatment indicator X as a Bernoulli random variable with $P(X = 1) = 1 - P(X = 0) = 0.5$. The Z_j are generated according to model (**??**) with a common η vector with $\eta_1 = \log 0.95$ and $\eta_2 = \log 0.75$ so that for given X, the probability of remaining a mover decreases with each event, and for each value of j the odds of remaining a mover are 25% lower in the treatment group with $X = 1$. For the baseline intensity of the latent Markov process of the form $\lambda\alpha(\lambda t)^{\alpha-1}$ we fix $\alpha = 1$ to correspond to a time-homogeneous latent process, and $\alpha = 0.50$ to correspond to a time-nonhomogeneous latent process; we set $\beta = \log 0.75$ to correspond to a 25% reduction in the rate of events among individuals at risk of events. For a given α and β, λ is determined so that the expected number of events over $(0, C]$ is specified at the particular value six among individuals who remain movers throughout the interval $(0, C]$. We then solve for η_0 so that the marginal expectation satisfies $E[N(C)] = 0.75$, 1.5, or 3.0. Five hundred datasets of $m = 500$ individuals were simulated for each parameter configuration. Parametric analysis and semiparametric analysis were carried out for each simulated dataset. Standard errors (SEs) were obtained using the method of Louis (1982) and the performance of the estimators was assessed in terms of empirical bias, empirical and model-based standard errors, and empirical coverage probability. The empirical bias (EBIAS), empirical standard error (ESE), average model-based standard error (ASE) computed according to Louis (1982), and empirical coverage probability expressed as a percentage (ECP) are given in Table **??** for the parametric analyses; the empirical coverage probability is defined as the fraction of simulations for which the sample confidence interval contained the true parameter value. The empirical bias and empirical SEs are also reported for the semiparametric analyses.

The empirical biases are generally small and decrease with increasing expected numbers of events. There is also good agreement between the empirical and average model-based SEs and the empirical coverage probability is compatible with the nominal level of 95%. The results are roughly comparable for the parametric and semiparametric analyses and the methods perform well when there is a trend in the latent rate function.

Table 1: Empirical results for maximum likelihood estimates obtained by the EM algorithm for parametric and semiparametric models with $\lambda(t|H(t^-)) = \lambda\alpha(\lambda t)^{\alpha-1}\exp(\beta X)$ and $P(Z_j = 1|\bar{Z}_{j-1} = 1_{j-1}, X) = \text{expit}(\eta_0 + \eta_1 j + \eta_2 X)$; $m = 500$, $nsim = 500$

		$E(N(C)) = 0.75, \eta_0 = -0.085$						$E(N(C)) = 1.5, \eta_0 = 0.709$						$E(N(C)) = 3, \eta_0 = 1.733$					
		Parametric				Semiparametric		Parametric				Semiparametric		Parametric				Semiparametric	
	VALUE	EBIAS	ESE	ASE	ECP	EBIAS	ESE	EBIAS	ESE	ASE	ECP	EBIAS	ESE	EBIAS	ESE	ASE	ECP	EBIAS	ESE
Time Homogeneous Rate																			
η_0	$---$	0.002	0.107	0.109	95.2	-0.001	0.107	-0.000	0.104	0.103	94.8	-0.003	0.104	0.003	0.115	0.117	95.4	0.000	0.116
η_1	-0.051	-0.007	0.073	0.072	95.6	0.020	0.082	-0.002	0.045	0.044	94.8	0.007	0.052	-0.000	0.036	0.035	94.6	0.004	0.041
η_2	-0.288	-0.006	0.144	0.141	94.2	0.013	0.146	-0.001	0.125	0.125	96.0	0.005	0.128	0.003	0.135	0.135	94.4	0.007	0.137
λ	6.857	0.045	0.509	0.499	95.4			0.010	0.365	0.359	94.0			0.010	0.275	0.262	92.8		
β	-0.288	-0.012	0.116	0.116	96.0	-0.026	0.125	0.002	0.086	0.084	94.8	-0.001	0.087	0.001	0.066	0.061	92.0	0.001	0.067
$\Lambda_0(C)$	6.857					-0.491	1.146					-0.069	0.666					-0.016	0.335
Time Non-homogeneous Rate																			
η_0	$---$	0.002	0.107	0.109	95.0	-0.001	0.107	-0.001	0.104	0.103	95.1	-0.004	0.104	0.003	0.115	0.117	95.6	0.000	0.116
η_1	-0.051	-0.007	0.074	0.072	95.2	0.020	0.082	-0.002	0.045	0.044	94.9	0.007	0.051	0.000	0.037	0.036	94.8	0.004	0.041
η_2	-0.288	-0.006	0.144	0.142	94.4	0.013	0.146	0.000	0.125	0.125	95.5	0.006	0.128	0.004	0.135	0.135	94.4	0.007	0.137
λ	47.020	0.899	7.859	7.840	95.8			0.292	6.219	6.055	93.5			0.454	5.085	4.976	94.0		
α	0.500	0.001	0.022	0.023	97.2			0.001	0.017	0.017	95.5			-0.000	0.013	0.013	95.4		
β	-0.288	-0.012	0.116	0.116	95.8	-0.026	0.125	0.001	0.086	0.084	95.1	-0.001	0.087	0.001	0.067	0.061	92.0	0.001	0.067
$\Lambda_0(C)$	6.857					-0.491	1.146					-0.070	0.667					-0.016	0.335

4 APPLICATION TO A COHORT STUDY OF INDIVIDUALS WITH AFFECTIVE DISORDER

We consider the cohort of 10,523 individuals with a first episode of affective disorder between January 1, 1994 and December 31, 1999. Among these individuals, 3,802 (36.1%) are male and 6,721 (63.9%) are female. A total of 17,021 hospitalizations are made over this window of calendar time giving a mean of 1.618 visits per individual (SD=1.720). A total of 1,106 (10.5%) of these individuals were bipolar at the time of the first admission; among the 9,417 (89.5%) patients who were unipolar at the study entry, 9,228 remain as unipolar, and 189 become bipolar by the end of follow-up. We consider a dataset comprised of 9417 patients who are unipolar at the first admission and who had a total of 14,497 admissions (mean=1.539 and SD=1.272). Follow-up of these individuals is censored at the end of the observation period, upon the diagnosis of bipolar disorder, schizophrenia, or an organic disorder, or at the time of death. There are 3,298 (35.0%) male individuals with total of 4,860 visits (mean=1.474 and SD=1.105) and 6,119 (65.0%) female patients with a total of 9,637 visits (mean=1.575 and SD=1.352).

We fit parametric and semiparametric (Andersen and Gill, 1982) Poisson regression models for the recurrence of acute episodes, with a single covariate indicating gender ($X = 1$ for females, $X = 0$ for males). These results are reported in the first three columns of Table **??**. Dynamic mover-stayer models are also fitted for which the latent variable model controls for the cumulative number of events (j) and gender; we denote the vector of covariates by $\dot{X}_j = (1, j, X)'$. A reduced dynamic mover-stayer model is also fitted with $\dot{X}_j = (1, j)'$ which simply controls for the cumulative number of acute episodes. The canonical event intensity model in these dynamic mover-stayer models also controls for gender. Both parametric (top half) and semiparametric (bottom half) event intensity models are reported in Table **??**.

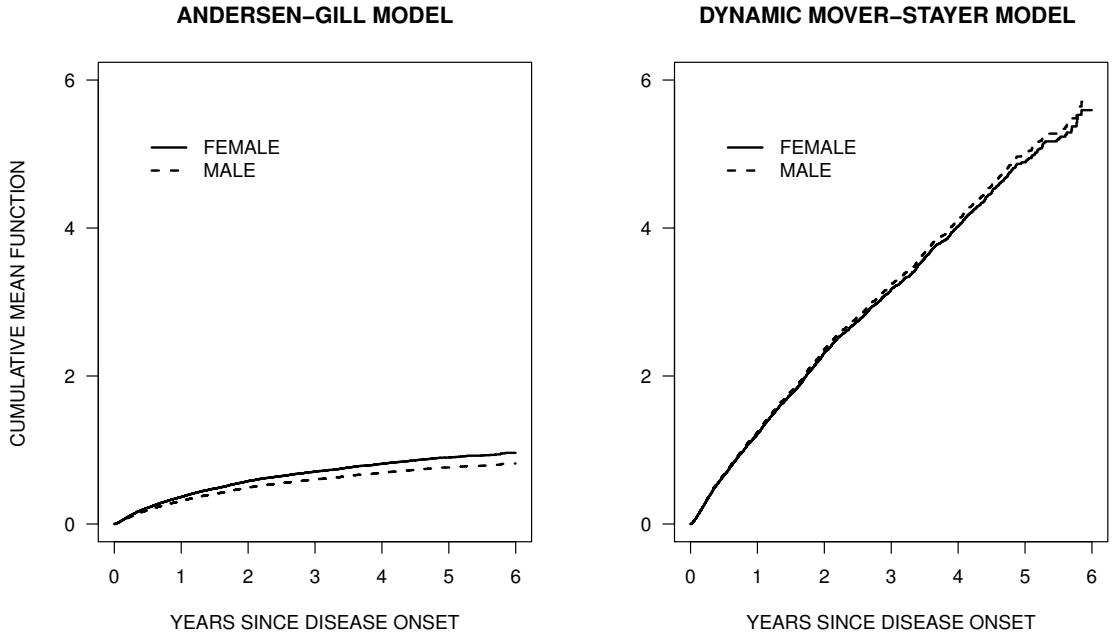

Figure 3: Plots of the estimated cumulative intensities for females and males with affective disorder; the left panel gives the cumulative mean function estimates based on the Andersen-Gill model and the right panel gives the cumulative canonical event intensity based on the dynamic mover-stayer model with covariate $\dot{X}_j = (1, j, X)'$ in the mover-stayer component and gender (X) in the canonical intensity model.

We focus the following discussion on the results of the analyses based on the semiparametric intensity model. The estimated regression coefficient for gender from the semiparametric Andersen-Gill model suggests women have a 17.6% increased rate of recurrence compared to men ($RR = 1.176$, 95% CI $(1.109, 1.247)$, $p < 0.001$). The estimates of the cumulative mean functions based on the fitted Andersen-Gill model are given in the left panel of Figure **??** and reveal a small absolute difference between genders in the cumulative expected number of episodes over time. The first semiparametric dynamic mover-stayer model reveals an insignificant association between gender on the latent intensity of recurrence ($RR = 0.980$, 95% CI $(0.882, 1.089)$, $p = 0.709$), but women have a significantly higher odds of remaining at risk of recurrence based on the mover-stayer component ($OR = 1.184$, 95% CI $(1.083, 1.293)$, $p < 0.001$). The dynamic mover-stayer model therefore suggests that the higher expected number of episodes for women may arise from a lower tendency for women to experience resolution of the disease. The right panel of Figure **??** gives the semiparametric estimate of the cumulative canonical event intensity for males and females. These estimates are much higher than those of the left panel since they correspond to the canonical process which does not accommodate resolution. Moreover the two estimates are very similar, reflecting the insignificant gender effect seen in this model.

Upon the removal of gender from the mover-stayer component (see the last three columns of Table **??**) the effect of gender on the latent rate remains insignificant ($RR = 1.034$, 95% CI $(0.934, 1.146)$, $p = 0.516$). The findings from the parametric and semiparametric analyses are in broad agreement.

5 DISCUSSION

We have described a dynamic mover-stayer model for the analysis of recurrent event data which is useful when there is a substantial fraction of individuals with an unduly long final gap time. This formulation is most appropriate when the underlying condition leading to the recurrent events can resolve but this resolution is not observable. There are a number of other medical conditions where this scenario can arise, and it is particularly relevant for registry studies where limited information is collected on individuals between records of events of interest. In the motivating example, the reasons for any resolution could include the identification of a suitable dose or type of medication or a change in a stressful environment leading to exacerbations of symptoms. Details on these and other possible explanations are often unavailable in the settings of registry studies but accommodation of such eventualities is often sensible in model formulation.

The formulation in Section **??** is quite flexible given the general form of the latent intensity. We have emphasized simple latent Markov models in our derivations and simulations. Natural extensions include the use of baseline rates which stratify on the cumulative number of events, latent semi-Markov models, or models with hybrid time scales. The expectation-maximization algorithm was described for parametric and semiparametric baseline rates within the latent Markov family of models, but adaptations to these other intensities are relatively straightforward. The introduction of random effects to offer a further avenue for explaining heterogeneity, while possible, may require large sample sizes to ensure convergence. Price and Manatunga (2001) illustrate the interplay between cure rate models and frailty models and Yu (2008) describes a mixture cure model with the latent mover-stayer and frailty variables realized at the time origin. Aalen (1992) discusses the use of a compound Poisson random effect distribution as a means of accommodating a fraction of nonsusceptible individuals as well as heterogeneity in risk among susceptible individuals. More general dynamic mover-stayer models can be specified by building upon these static latent variable models. Issues of estimability arise and become more challenging the more flexible the model components become and examination of profile likelihood contours can be instructive when investigating reasons for convergence problems.

Model assessment is challenging in settings with latent variables and this is particularly true of mixture models of this type. A particular issue of concern is the fact that there may be multiple con-

Table 2: Results of fitting Poisson model and dynamic mover-stayer model† to study of affective disorder with parametric and semiparametric models; Markov model is a parametric Poisson model or Anderson-Gill (1982) semiparametric model, $m = 9,417$

				Dynamic Mover-Stayer Models					
	Poisson Model			$\dot{X}_j = (1, j, X)$			$\dot{X}_j = (1, j)$		
	EST	SE	p-value	EST	SE	p-value	EST	SE	p-value
Parametric Models									
Mover-Stayer Model									
η_0	-	-	-	-0.6344	0.0376		-0.5219	0.0257	
η_1	-	-	-	0.5184	0.0232	< 0.0001	0.5210	0.0233	< 0.0001
η_2	-	-	-	0.1682	0.0433	0.0001			
Recurrent Event Model									
λ	0.1555	0.0058		1.2170	0.0548		1.1729	0.0548	
α	0.6970	0.0087		0.9574	0.0140		0.9570	0.0140	
β	0.1573	0.0299	< 0.0001	-0.0268	0.0515	0.6023	0.0222	0.0505	0.6600
Semiparametric Models									
Mover-Stayer Model									
η_0	-	-	-	-0.6418	0.0387		-0.5289	0.0264	
η_1	-	-	-	0.5760	0.0338	< 0.0001	0.5796	0.0340	< 0.0001
η_2	-	-	-	0.1685	0.0451	0.0002			
Recurrent Event Model									
β	0.1620	0.0299	< 0.0001	-0.0201	0.0539	0.7088	0.0338	0.0521	0.5158

p-values are based on Wald statistics
† SEs for estimates from parametric models obtained by Louis' method (1982) and by nonparametric bootstrap (200 bootstrap samples) for fitted semiparametric models

figurations of the baseline intensity and the mover-stayer model which render similar mean functions. Clear ideas regarding which component of the model covariates are to be placed can help circumvent this challenging problem.

In many settings with recurrent events, the events are not observed but only known to occur between to assessment times. In cohort studies of patients with osteoporosis for example, asymptomatic fractures may be detected upon periodic radiographic examination. Establishment of suitable medications or other changes in lifestyle and diet may minimize risk of further fractures, but it can be difficult to determine if these changes have taken place. The dynamic mover-stayer model offers a way of describing this phenomenon but adaptations to enable model fitting with interval-censored data are required. Cook, Kalbfleisch and Yi (2002) offer one such approach in the content of parametric Markov models.

We conclude by noting that mortality adds a further complication. In the motivating study we found a mortality rate of about 5% over the course of the six year period but simply censored individuals at the time of death. A multistate model offers a suitable framework for analysis which can be readily adapted to deal with the added complication of mortality. This is a topic of ongoing research.

ACKNOWLEDGEMENTS

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (RGPIN 155849) and the Canadian Institutes forHealth Research (FRN 13887). Hua Shen is supported by a Grant from the Division of High Impact Clinical Trials of the Ontario Institute for Cancer Research. Richard Cook is a Canada Research Chair in Statistical Methods for Health Research. The data on the psychiatric hospitalizations were kindly provided by Professor Lars Vedel Kessing and Professor Per Kragh Andersen who have used the data in prior publications on recurrent events.

A APPENDIX

A.1 APPENDIX 1: IMPLEMENTATION OF THE EM ALGORITHM

For an individual with n events observed at times $t_1 < t_2 < \cdots < t_n < C$, the only missing quantity is Z_n. If we have a single covariate X, the dataframe used at the rth step of the EM algorithm to maximize $Q_1(\theta_1; \widehat{\theta}^r)$ has the usual counting process form with the addition of a weight which is 1 for all lines except the last one

ID(i)	enum(j)	estart	estop	estatus	weight	rtrunc	tstatus	X
1	0	0	t_1	1	1	NA	1	X_1
1	1	t_1	t_2	1	1	NA	2	X_1
1	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	X_1
1	$n-1$	t_{n-1}	t_n	1	1	NA	2	X_1
1	n	t_n	C	0	$\widehat{\zeta}^r$	NA	2	X_1

In a parametric analysis with a baseline rate for the latent process of the form $\lambda_0(t; \lambda, \alpha) = \lambda\alpha(\lambda t)^{\alpha-1}$, $Q_1(\theta_1; \widehat{\theta}^r)$ is maximized to give $\widehat{\theta}^{r+1}$ by the Splus command

$$\begin{aligned}&\text{censorReg}(\text{censor}(\text{estop}, \text{estatus}) \sim X, \text{truncation} = \text{censor}(\text{estart}, \text{rtrunc}, \text{tstatus}),\\ &\quad \text{weights} = \text{weight}, \text{distribution} = \text{“weibull”})\,.\end{aligned}$$

and in the semiparametric analysis by the call

$$\text{coxph}(\text{Surv}(\text{estart}, \text{estop}, \text{estatus}) \sim \text{X} + \text{offset}(\log(\text{weight})), \text{method} = \text{“breslow”}) .$$

The data used to maximized $Q_2(\theta_2; \widehat{\theta}^r)$ has the form

ID(i)	enum(j)	Z	X	weight
1	0	1	X_1	1
1	1	1	X_1	1
1	2	1	X_1	1
1	$\vdots$	$\vdots$	X_1	$\vdots$
1	$n-1$	1	X_1	1
1	n	1	X_1	$\widehat{\zeta}^r$
1	n	0	X_1	$1-\widehat{\zeta}^r$

A simple logistic regression call

$$glm(Z \sim enum + X, \text{weights} = \text{weight}, \text{family} = \text{binomial}(\text{link} = \text{logit})) ,$$

yields $\widehat{\theta}_2^{r+1}$. New dataframes are then created with $\widehat{\zeta}^r$ replaced with $\widehat{\zeta}^{r+1}$ and the procedure is repeated until $\|\widehat{\theta}^{r+1} - \widehat{\theta}^r\| < \epsilon$ for some specified value of ϵ.

A.2 Appendix 2: Variance Estimation via Louis (1982)

Let $S_C(\theta) = \partial \log L_C(\theta)/\partial\theta$ and $I_C(\theta) = -\partial S_C(\theta)/\partial\theta$ where $L_C(\theta)$ is the complete data likelihood for which Z_n is treated as known, given by (**??**). If $L(\theta)$, $S(\theta) = \partial \log L(\theta)/\partial\theta$ and $I(\theta) = -\partial S(\theta)/\partial\theta$ are the observed data likelihood, score and information matrix, then

$$I(\theta) = E_{Z_n}\left[I_C(\theta)|H(C)\right] - E_{Z_n}\left[S_C(\theta)S_C'(\theta)|H(C)\right] \tag{13}$$

where $S_C(\theta) = (S_{C1}'(\theta_1), S_{C2}'(\theta_2))'$ and $S_{Ck}(\theta_k) = \partial \log L_{Ck}(\theta_k)/\partial\theta_k$, $k = 1, 2$, and

$$I_C(\theta) = \begin{bmatrix} I_{C1}(\theta_1) & 0 \\ 0 & I_{C2}(\theta_2) \end{bmatrix} ,$$

where $I_{Ck} = -\partial S_{Ck}(\theta_k)/\partial\theta_k$, $k = 1, 2$. We estimate $I(\widehat{\theta})$ in (**??**) by running the EM algorithm to the point of convergence and using the expression in (**??**) evaluated at the MLE $\widehat{\theta}$ to take the required expectation. Standard software can be readily exploited to do this in both the parametric and semiparametric settings.

The first matrix on the right hand side of (**??**) is obtained by extracting the values stored in the information matrices produced at the final M-step. Each individual contributes to the complete data likelihood and complete data score, so we can compute their contributions to $S_{C1}(\theta_1)$ and $S_{C2}(\theta_2)$, stack them and then take a weighted average to estimate the second term in (**??**).

In the semiparametric setting, let $u_1 < \cdots < u_R$ denote the R unique event times over the entire sample, let $d\Lambda_0 = (d\Lambda_0(u_1), \ldots, d\Lambda_0(u_R))'$, and let $\theta_1 = (d\Lambda_0', \beta)'$, then let $S_{C1}(\theta_1) = (S_{C11}'(\theta_1), S_{C12}'(\theta_1))'$, where $S_{C11}(\theta_1) = (S_{C11u_1}(\theta_1), \ldots, S_{C11u_R}(\theta_1))'$ and

$$\begin{aligned} S_{C11u}(\theta_1) = S_{C11}(\lambda_0(u)) &= Y(u)\left\{dN(u) - Z(u)d\Lambda_0(u)\exp(\beta X)\right\}, \quad 0 < u \\ S_{C12}(\theta_1) = S_{C12}(\beta) &= \int_0^\infty Y(u)X\left\{dN(u) - Z(u)d\Lambda_0(u)\exp(\beta X)\right\} . \end{aligned}$$

Then

$$I_{C1}(\theta_1) = \begin{bmatrix} -\partial S_{C11}(\theta_1)/\partial\theta_1' \\ -\partial S_{C12}(\theta_1)/\partial\theta_1' \end{bmatrix} = -\begin{bmatrix} \partial S_{C11}(\theta_1)/\partial d\Lambda_0' & \partial S_{C11}(\theta_1)/\partial\beta \\ \partial S_{C12}(\theta_1)/\partial d\Lambda_0' & \partial S_{C12}(\theta_1)/\partial\beta \end{bmatrix},$$

where

$$\begin{aligned}
\frac{\partial S_{C11}(\theta_1)}{\partial d\Lambda_0(u)} &= -Y(u)Z(u)\exp(\beta X) \\
\frac{\partial S_{C12}(\theta_1)}{\partial d\Lambda_0(u)} &= -Y(u)XZ(u)\exp(\beta X) \\
\frac{\partial S_{C11}(\theta_1)}{\partial \beta} &= -Y(u)Z(u)d\Lambda_0(u)\exp(\beta X)X \\
\frac{\partial S_{C12}(\theta_1)}{\partial \beta} &= -\int_0^\infty Y(u)Z(u)d\Lambda_0(u)\exp(\beta X)X^2 .
\end{aligned}$$

Then we can obtain $S_C(\theta)$ and $I_C(\theta)$ and proceed as in the parametric setting.

REFERENCES

Aalen OO (1992) Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability* 2: 951–972

Andersen PK, Gill RD (1982) Cox's regression model for counting processes: A large sample study. *Annals of Statistics* 10: 1100-1120

Andersen P, Borgan O, Gill R, Keiding N (1993) *Statistical Models Based on Counting Process*. Springer-Verlag

Blumen IM, Kogan M, McCarthy PJ (1955) The industrial mobility of labor as a probability process. *Cornell Studies of Industrial and Labor Relations, Ithaca, New York: Cornell University Press* 6

Bourassa MG, Gurné O, Bangdiwala SI, Ghali JK, Young JB, Rousseau M, Johnstone DE, Yusuf S (1993) Natural history and patterns of current practice in heart failure. *Journal of the American College of Cardiology* 22: 4, 14A-19A

Cole EH, Farewell VT, Aprile M, Cattran DC, Pei YP, Fenton SS, Zaltzman J, Cardella CJ (1995) Renal transplantation in older patients: the University of Toronto experience. *Geriatric Nephrology and Urology* 5: 85-92

Cook RJ, Kalbflesch JD, Yi GY (2002) A generalized mover-stayer model for panel data. *Biostatistics* 3: 3, 407-420

Cook RJ, Lawless JF (1997) Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine* 16: 911-924

Cook RJ, Lawless JF (2007) *The Statistical Analysis of Recurrent Events*. Springer, New York

Cook RJ, Lawless JF (2012) Concepts and tests for trend in recurrent event processes. *Journal of Iranian Statistical Society* in press

Dempster AP, Laird NM, Rubin DB (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39: 1, 1-38

Fang HB, Li G, Sun J (2005) Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. *Scandinavian Journal of Statistics* 32: 59-75

Farewell VT (1982) The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* 38: 1041-1046

Farewell VT (1986) Mixture models in survival analysis: Are they worth the risk? *Canadian Journal of Statistics* 14: 257-262

Frydman H (1984) Maximum likelihood estimation in the mover-stayer model. *Journal of the American Statistical Association* 79: 632-638

Fuchs C, Greenhouse JB (1988) The EM algorithm for maximum likelihood estimation in the mover-stayer model. *Biometrics* 44: 605-613

Ghosh D, Lin DY (2000) Nonparametric analysis of recurrent events and death. *Biometrics* 56: 554-562

Ghosh D, Lin DY (2002) Marginal regression models for recurrent and terminal events. *Statistica Sinica* 12: 663-688

Goodman LA (1961) Statistical methods for the mover-stayer model. *Journal of the American Statistical Association* 56: 841-868

Grossman R, Mukherjee J, Vaughan D, Eastwood C, Cook R, LaForge J, Lampron N (1998) A 1-year community-based health economic study of ciprofloxacin vs usual antibiotic treatment in acute exacerbations of chronic bronchitis: the Canadian Ciprofloxacin Health Economic Study Group. *Chest* 113: 131-141

Heckman JJ, Walker JR (1987) Using goodness of fit and other criteria to choose among competing duration models: a case study of Hutterite data. *Sociological Methodology* 17: 247-307

Heddle NM, Cook RJ, Webert KE, Sigouin C, Rebulla P (2003) Methodologic issues in the use of bleeding as an outcome in transfusion medicine studies. *Transfusion* 43: 742-752

Hortobagyi GN, Theriault RL, Lipton A, Porter L, Blayney D, Sinoff C, Wheeler H, Simeone JF, Seaman JJ, Knight RD, Heffernan M, Mellars K, Reitsma DJ (1998) Long-term prevention of skeletal complications of metastatic breast cancer with pamidronate. Protocol 19 Aredia Breast Cancer Study Group. *Journal of Clinical Oncology* 16: 2038-2044

Kessing LV, Hansen MG, Andersen PK, Angst J (2004) The predictive effect of episodes on the risk of recurrence in depressive and bipolar disorders–a life-long perspective. *Acta Psychiatrica Scandinavica* 109: 339-344

Kuk AYC, Chen C-H (1992) A mixture model combining logistic regression with proportional hazards regression. *Biometrika* 79: 531-541

Kvist K, Andersen PK, Angst J, Kessing LV (2010) Event dependent sampling of recurrent events. *Lifetime Data Analysis* 16: 580-598

Lawless JF (1987) Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* 15: 209-225

Lawless JF (1995) The analysis of recurrent events for multiple subjects. *Journal of the Royal Statistical Society. Series C* 44: 487–498

Lawless JF, Nadeau JC (1995) Some simple robust methods for the analysis of recurrent events. *Technometrics* 37: 158-168

Liu L, Wolfe RA, Huang X (2004) Shared frailty models for recurrent events and a terminal event. *Biometrics* 60: 747-756

Louis TA (1982) Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society* 44: 2, 226-233

Peng YW, Dear KBG, Denham JW (1998) A generalized F mixture model for cure rate estimation. *Statistics in Medicine* 17: 813-830

Peng YW, Dear KBG (2000) A nonparametric mixture model for cure rate estimation. *Biometrics* 56: 237-243

Pledger GW, Sackellares JC, Treiman DM, Pellock JM, Wright FS, Mikati M, Sahlroot JT, Tsay JY, Drake ME, Olson L, et al (1994) Flunarizine for treatment of partial seizures: results of a concentration-controlled trial. *Neurology* 44: 1830-1836

Price DL, Manatunga AK (2001) Modelling survival data with a cured fraction using frailty models. *Statistics in Medicine* 20: 1515–1527.

Spilerman S (1972) Extensions of the mover-stayer model. *American Journal of Sociology* 78: 3, 599-626

Sy JP, Taylor JMG (2000) Estimation in a Cox proportional hazards cure model. *Biometrics* 56: 227-236

Taylor JMG (1995) Semi-parametric estimation in failure time mixture models. *Biometrics* 51: 899-907

Webert KE, Cook RJ, Sigouin CS, Rebulla P, Heddle NM (2006) The risk of bleeding in thrombocytopenic patients with acute myeloid leukemia. *Haematologica* 91: 1530-1537

Winokur G (1975) The Iowa 500: heterogeneity and course in manic-depressive illness (bipolar). *Comprehensive Psychiatry* 16: 125-131

Yamaguchi K (1992) Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of "permanent employment" in Japan. *Journal of the American Statistical Association* 87: 284-292

Yamaguchi K (1994) Some accelerated failure-time regression models derived from diffusion process models: an application to a network diffusion analysis. *Sociological Methodology* 24: 267-300

Yamaguchi K (1998) Mover-Stayer models for analyzing event nonoccurrence and event timing with time-dependent covariates: an application to an analysis of remarriage. *Sociological Methodology* 28: 1, 327-361

Yamaguchi K (2003) Accelerated failure-time mover-stayer regression models for the analysis of last episode data. *Sociological Methodology* 33: 81-110

Ye Y, Kalbfleisch JD, Schaubel DE (2007) Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* 63: 78-87

Yu B (2008) A frailty mixture cure model with application to hospital readmission data. *The Biometrical Journal* 50: 386–394.