Biostatistics (2002), **3**, 3, *pp.* 407–420
Printed in Great Britain

A generalized mover–stayer model for panel data

RICHARD J. COOK, JOHN D. KALBFLEISCH, GRACE Y. YI

Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1
rjcook@vwaterloo.ca

SUMMARY

A generalized mover–stayer model is described for conditionally Markov processes under panel observation. Marginally the model represents a mixture of nested continuous-time Markov processes in which sub-models are defined by constraining some transition intensities to zero between two or more states of a full model. A Fisher scoring algorithm is described which facilitates maximum likelihood estimation based only on the first derivatives of the transition probability matrices. The model is fit to data from a smoking prevention study and is shown to provide a significant improvement in fit over a time-homogeneous Markov model. Extensions are developed which facilitate examination of covariate effects on both the transition intensities and the mover–stayer probabilities.

Keywords: Latent variables; Marginal likelihood; Markov model; Multi-state process; Time homogeneous intensity.

1. INTRODUCTION

Multi-state stochastic models provide a useful framework for the analysis of data from longitudinal studies when interest lies in dynamic aspects of the process under investigation. When subjects are observed continuously over a period of observation, transitions between states are observed and parametric, nonparametric, and semiparametric methods may be used (Andersen *et al.*, 1993). In contrast, when the subjects are seen at discrete time points, exact transition times are not observed and all that is known is the state occupied at each assessment. Such data are often referred to as panel data. Kalbfleisch and Lawless (1985, 1989) describe a Fisher scoring algorithm for maximum likelihood estimation of the transition intensities under a time-homogeneous Markov model in this setting. Applications of this methodology to problems in infectious disease (Gentleman *et al.*, 1994), rheumatology (Gladman *et al.*, 1995), and smoking prevention studies (Kalbfleisch and Lawless, 1985) highlight the scope of problems amenable to this type of analysis. Recently, however, there has been interest in considering more general models for panel data. Satten (1999) considers a mixed time-homogeneous Markov model for progressive disease processes. Cook (1999) considers a mixed time-homogeneous Markov model for the special case of a two-state alternating process under panel observation.

Here we present a generalized mover–stayer model in which, conditionally on latent mover–stayer variables, subjects follow a time-homogeneous Markov process. Marginally the model represents a mixture of nested continuous-time Markov processes in which sub-models are defined by constraining transition intensities out of one or more states in the full model to be zero. Thus, each individual can have one or more absorbing states and, once one of these states is entered, no further transitions can take place. The usual mover–stayer model constrains 'stayers' to remain in their initial state (see, for example,

Frydman, 1984). The model considered here is, therefore, more general in the sense that subjects may make transitions between a number of states before entering one of their 'stayer' states.

The remainder of the paper is organized as follows. In Section 2 we review the analysis of panel data under a time-homogeneous Markov model. In Section 3 the generalized mover–stayer model is introduced and extensions to accommodate regression analyses for both the transition intensities and the mover–stayer probabilities are described. The methods are applied to data from a smoking prevention study in Section 4, and general remarks are made in Section 5.

2. PANEL DATA FROM MARKOV PROCESSES

Suppose an individual makes transitions among K states according to a continuous-time Markov process. Let the states be identified by the integers $1, 2, \ldots, K$, and let $Y(t)$ represent the state occupied at time t for $t \geqslant 0$. Let $P(s, t)$ be the $K \times K$ transition probability matrix with (i, j) entry

$$P_{ij}(s, t) = \Pr\{Y(t) = j | Y(s) = i\}$$

for $0 \leqslant s \leqslant t$ and $i, j = 1, 2, \ldots, K$. The transition intensity from state i to j at time t is

$$\lambda_{ij}(t) = \lim_{\Delta t \to 0} \frac{P_{ij}(t, t + \Delta t)}{\Delta t}, \qquad i \neq j$$

and, by convention, we set $\lambda_{ii}(t) = -\sum_{j \neq i} \lambda_{ij}(t)$, $i, j = 1, \ldots, K$. Let $\Lambda(t)$ be the $K \times K$ transition intensity matrix with (i, j) entry $\lambda_{ij}(t)$, $i, j = 1, \ldots, K$.

In this paper, we confine our attention to time-homogeneous Markov models for which the transition intensities are independent of t. We therefore let $\lambda_{ij}(t) = \lambda_{ij}$, $i, j = 1, \ldots, K$, and write $\Lambda(t) = \Lambda$. It follows that $P(s, t) = P(0, t - s)$ which, for convenience, we write as $P(t - s)$. It can be seen that

$$P(t) = \exp(\Lambda t) = \sum_{i=0}^{\infty} \Lambda^i t^i / i!.$$

In most applications, we are interested in fitting models in which $\Lambda = (\lambda_{ij})$ is written as a function of a parameter vector, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_A)'$ say. For example, in a problem with no covariates and no structural zeros in Λ, we might take $A = k(k - 1)$ and define the elements of $\boldsymbol{\theta}$ as $\{\log \lambda_{ij}\}$. In many instances, we may wish to specify some structure which relates elements in the model. For example, in a progressive illness–death model with three states (1 = healthy, 2 = ill, 3 = dead), we may wish to relate the intensities for death in states 1 and 2 and specify $\lambda_{12} = \exp(\theta_1)$, $\lambda_{13} = \exp(\theta_1 + \theta_2)$, so that θ_2 measures the additional risk due to illness. When covariates are present we may wish to test whether covariate effects are significantly different for particular transition intensities, and therefore we need to fit models in which some covariate effects are constrained to be the same.

It is well known that simple formulas for the computation of $P(t)$ are available (see, for example, Cox and Miller, 1977, p. 151). If Λ is diagonalizable, then we can write $\Lambda = H\ D\ H^{-1}$ where $D = \text{diag}(d_1, d_2, \ldots, d_K)$ is a diagonal matrix of eigenvalues of Λ and H is a matrix whose columns are independent right eigenvectors. It is then easy to see that

$$P(t) = H \exp(Dt) H^{-1}. \tag{2.1}$$

Thus, once the eigenvalue decomposition of Λ is known, computation of the transition probability matrix is straightforward. One can also develop formulas to compute the matrix of first derivatives $\partial P(t)/\partial \theta_k$, $k = 1, \ldots, A$ (see, for example, Kalbfleisch and Lawless, 1985, 1989). We show how this is done in a more general context in the next section.

Suppose N individuals are under study and each individual independently follows a common continuous-time Markov process. Let $Y_\ell(t)$ denote the state occupied by individual ℓ at time $t \geqslant 0$, $\ell = 1, \ldots, N$. Let $t_{\ell 0}, t_{\ell 1}, \ldots, t_{\ell m_\ell}$ denote the $m_\ell + 1$ times at which individual ℓ is observed, and let $z_\ell = (i_{\ell 0}, i_{\ell 1}, \ldots, i_{\ell m_\ell})$ denote the states occupied at each observation time (i.e. $i_{\ell j} = Y_\ell(t_{\ell j})$, for $j = 0, 1, \ldots, m_\ell$). It is assumed that the observation times are independent of the process, although this can be relaxed to allow the observation status of the process at time t (i.e. whether it is observed or not observed) to depend on the *observed* history of the process up to time t^-. Finally, it is convenient to define the set

$$U_\ell = \{(i_{\ell,r-1}, i_{\ell r}, s_{\ell r}) : s_{\ell r} = t_{\ell r} - t_{\ell,r-1}, r = 1, \ldots, m_\ell\}, \tag{2.2}$$

the rth triple of which indicates the state occupied at $t_{\ell,r-1}$, the state occupied at $t_{\ell,r}$, and the time elapsed between $t_{\ell,r-1}$ and $t_{\ell,r}$.

Conditional on the initial state $Y_\ell(t_{\ell 0}) = i_{\ell 0}$, the contribution to the likelihood for $\boldsymbol{\theta}$ from individual ℓ is

$$L_\ell(\boldsymbol{\theta}) = \prod_{r=1}^{m_\ell} P_{i_{\ell,r-1} i_{\ell r}}(t_{\ell r} - t_{\ell,r-1}) = \prod_{(i,j,s) \in U_\ell} P_{ij}(s).$$

The full likelihood function for $\boldsymbol{\theta}$ is therefore

$$L(\boldsymbol{\theta}) = \prod_{\ell=1}^{N} L_\ell(\boldsymbol{\theta}). \tag{2.3}$$

It is straightforward to show that the observed information arising from (2.3) has (u, v) entry

$$\sum_{\ell=1}^{N} \sum_{(i,j,s) \in U_\ell} \left[\frac{1}{P_{ij}^2(s)} \frac{\partial P_{ij}(s)}{\partial \theta_u} \frac{\partial P_{ij}(s)}{\partial \theta_v} - \frac{1}{P_{ij}(s)} \frac{\partial^2 P_{ij}(s)}{\partial \theta_u \partial \theta_v} \right]. \tag{2.4}$$

We may estimate the contribution to the expected information from individual ℓ by taking the expectation with respect to the distribution of the state occupied by individual ℓ at time $t_{\ell r}$, conditional on the state individual ℓ occupied at $t_{\ell,r-1}$, for each $r = 1, 2, \ldots, m_\ell$. It can then be seen that the term involving the second derivatives in (2.4) has conditional, and hence marginal, expectation 0 for each ℓ. An estimate of the (u, v) entry of the expected information is then given by the sum of the conditional expectations of (2.4) given by

$$\sum_{\ell=1}^{N} \sum_{(i,s) \in U_\ell^*} \sum_{j=1}^{K} \frac{1}{P_{ij}(s)} \frac{\partial P_{ij}(s)}{\partial \theta_u} \frac{\partial P_{ij}(s)}{\partial \theta_v}, \tag{2.5}$$

where $U_\ell^* = \{(i_{\ell,r-1}, s_{\ell r}) \in U_\ell, r = 1, \ldots, m_\ell\}$. This form was suggested by Kalbfleisch and Lawless (1985, 1989) and Gentleman *et al.* (1994).

An alternative estimate is given by retaining only the first term in (2.4), which is justified because the expectation of the second term is zero. We then write

$$I_{u,v}(\boldsymbol{\theta}) = \sum_{\ell=1}^{N} \sum_{(i,j,s) \in U_\ell} \frac{1}{P_{ij}^2(s)} \frac{\partial P_{ij}(s)}{\partial \theta_u} \frac{\partial P_{ij}(s)}{\partial \theta_v}. \tag{2.6}$$

This approach proves to be the most convenient strategy for estimating the expected information in the next section.

3. THE GENERALIZED MOVER–STAYER MODEL

3.1 *Model formulation*

Frequently there is heterogeneity in panel data beyond that which is expected from an underlying Markov model. One particular type of heterogeneity arises when there are unusually long runs of observations in a particular state. When these runs occur in the initial state for many individuals, mover–stayer models provide a useful framework (see, for example, Frydman, 1984). In the usual mover–stayer model, a randomly selected individual ℓ either stays in its initial state, $i_{\ell 0} = i$ say, with probability π_i, or with complementary probability $1 - \pi_i$ 'moves' among the full set of states according to a common Markov process with transition intensity matrix Λ. In population studies of chronic degenerative disease processes, mover–stayer models are attractive because they accommodate the possibility that a substantial proportion of the population may be disease free over the course of observation and will therefore not experience degeneration.

In many contexts, long runs of observations in a particular state are observed after some initial transitions. This can happen, for example, in a behavioural study where there may be some experimental behaviour before individuals 'settle in' with a final choice. To describe such behaviour, we develop a different mover–stayer type model in which each individual ℓ is allowed to have a different set of absorbing states. If state k is an absorbing state for individual ℓ, once individual ℓ enters state k, no further transitions occur. Individual ℓ is then said to be a 'stayer' in state k. Thus, a typical individual will move among the states according to the underlying Markov process until it encounters one of its absorbing states, whereupon it is confined there. We term this model the 'generalized mover–stayer model'.

The generalized mover–stayer model is more formally specified as follows. Let $\boldsymbol{\alpha}_\ell = (\alpha_{\ell 1}, \ldots, \alpha_{\ell K})'$ be a vector of mover–stayer indicators for individual ℓ where $\alpha_{\ell k} = 0$ if the kth state is absorbing for individual ℓ and $\alpha_{\ell k} = 1$ otherwise. Conditional on $\boldsymbol{\alpha}_\ell$, we suppose that the process for this individual is timehomogeneous Markov with transition intensity matrix

$$\Lambda(\boldsymbol{\alpha}_\ell) = \begin{pmatrix} \alpha_{\ell 1}\lambda_{11} & \alpha_{\ell 1}\lambda_{12} & \cdots & \alpha_{\ell 1}\lambda_{1K} \\ \alpha_{\ell 2}\lambda_{21} & \alpha_{\ell 2}\lambda_{22} & \cdots & \alpha_{\ell 2}\lambda_{2K} \\ \vdots & \vdots & & \vdots \\ \alpha_{\ell K}\lambda_{K1} & \alpha_{\ell K}\lambda_{K2} & \cdots & \alpha_{\ell K}\lambda_{KK} \end{pmatrix}.$$

If $\boldsymbol{\alpha}_\ell = \mathbf{1}$ the model of Section 2 is retrieved. If any components of $\boldsymbol{\alpha}_\ell$ are zero, however, the corresponding states in the chain become absorbing. The probability transition matrix corresponding to $\boldsymbol{\alpha}_\ell$ is

$$P(t|\boldsymbol{\alpha}_\ell) = \exp\{\Lambda(\boldsymbol{\alpha}_\ell)t\}. \tag{3.1}$$

The variables $\boldsymbol{\alpha}_\ell$ are unobserved and, to complete the model, we need to specify their distribution. We suppose that $\boldsymbol{\alpha}_\ell$, $\ell = 1, \ldots, N$ are independent and identically distributed. Further, we assume that $\alpha_{\ell k}$ is a Bernoulli random variable with $\Pr(\alpha_{\ell k} = 0) = \pi_k$, and $\Pr(\alpha_{\ell k} = 1) = 1 - \pi_k$, $k = 1, \ldots, K$, where $\alpha_{\ell k}$ and $\alpha_{\ell k'}$ are independently distributed. It is often useful to reparametrize these probabilities in terms of a basic parameter vector $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_B)'$. For example, one might set $B = K$ and take $\phi_k = \log(\pi_k/(1 - \pi_k))$, $k = 1, \ldots, K$ in the absence of covariates. Let $f(\boldsymbol{\alpha}_\ell; \boldsymbol{\phi}) = \prod_{k=1}^K \pi_k^{1-\alpha_{\ell k}}(1 - \pi_k)^{\alpha_{\ell k}}$ be the joint probability mass function of $\boldsymbol{\alpha}_\ell$. Note that given α_ℓ the process is Markov, but marginally the process $\{Y_\ell(t), t > 0\}$ does not satisfy the Markov property.

To write down the likelihood, we introduce some additional notation. We use the same definitions of the observation times $t_{\ell 0}, t_{\ell 1}, \ldots, t_{\ell m_\ell}$ and the states occupied, $z_\ell = (i_{\ell 0}, \ldots, i_{\ell m_\ell})$, as before, but suppose here that observation begins at $t_{\ell 0} = 0$, the time origin of the process.

The vector $\boldsymbol{z}_\ell$ often identifies some mover states for the ℓth individual and we define

$$\mathcal{M}_\ell = \{i : \ell \text{ is a known mover from state } i\}. \tag{3.2}$$

Note that $\mathcal{M}_\ell$ comprises all states which the ℓth individual is known to have exited. Let $\mathcal{S} = \{0, 1\}^K$ and

$$\mathcal{S}_\ell = \{\boldsymbol{\alpha} \in \mathcal{S} : \alpha_i = 1 \text{ for all } i \in \mathcal{M}_\ell\}. \tag{3.3}$$

Finally, we again use the set notation for U_ℓ as given in (2.3). The ℓth component of the likelihood can then be written as

$$L_\ell(\boldsymbol{\psi}) = \sum_{\boldsymbol{\alpha} \in \mathcal{S}_\ell} P_\ell(\boldsymbol{\alpha}, \boldsymbol{\theta}) f_\ell(\boldsymbol{\alpha}; \boldsymbol{\phi}) \tag{3.4}$$

where

$$\begin{aligned} P_\ell(\boldsymbol{\alpha}, \boldsymbol{\theta}) &= \prod_{(i,j,s) \in U_\ell} P_{ij}(s|\boldsymbol{\alpha}), \\ f_\ell(\boldsymbol{\alpha}; \boldsymbol{\phi}) &= \prod_{i \in \mathcal{M}_\ell} (1 - \pi_i) \cdot \prod_{i \notin \mathcal{M}_\ell} \pi_i^{1-\alpha_{\ell i}} (1 - \pi_i)^{\alpha_{\ell i}}, \end{aligned} \tag{3.5}$$

and $\boldsymbol{\psi} = (\boldsymbol{\theta}', \boldsymbol{\phi}')'$. The overall likelihood function is

$$L(\boldsymbol{\psi}) = \prod_{\ell=1}^{N} L_\ell(\boldsymbol{\psi}). \tag{3.6}$$

The expression (3.4) simplifies in a number of situations and sometimes, as in the ordinary homogeneous case, grouping across individuals leads to simplification. The formulation in (3.6), however, is fully general and serves to clearly illustrate the main ideas. A Fisher scoring algorithm for obtaining maximum likelihood estimates is described in Appendix A.

3.2 *Remarks on regression modelling*

In many applications there are covariates available for each individual under study and interest lies in the relationship between these covariates, the transition intensities and the mover–stayer probabilities. The discussion of the parametrization of the mover–stayer model has thus far been quite general and here we make specific remarks on the formulation of regression models.

Suppose that individual ℓ has an associated covariate vector $\boldsymbol{x}'_\ell = (x_{\ell 0}, x_{\ell 1}, \ldots, x_{\ell,d-1})$ where $x_{\ell 0} = 1$, $\ell = 1, \ldots, N$. Given $\boldsymbol{\alpha}_\ell$ and $\boldsymbol{x}_\ell$, individual ℓ is assumed to follow a time-homogeneous Markov model with the (i, j) entry of the conditional transition intensity matrix $\Lambda(\boldsymbol{\alpha}_\ell, \boldsymbol{x}_\ell)$ given by $\alpha_{\ell i}\lambda_{ij}(\boldsymbol{x}_\ell)$. Here $\alpha_{\ell i}$, $i = 1, \ldots, K$ are independent Bernoulli random variables with $\Pr(\alpha_{\ell i} = 0|\boldsymbol{x}_\ell) = \pi_i(\boldsymbol{x}_\ell)$. Given $\boldsymbol{x}_\ell$ and $\boldsymbol{x}_{\ell'}$, $\boldsymbol{\alpha}_\ell$ and $\boldsymbol{\alpha}_{\ell'}$ are assumed to be independent. Covariate effects may be naturally examined by specifying regression models of the form

$$\begin{aligned} \log \lambda_{ij}(\boldsymbol{x}_\ell) &= \boldsymbol{x}'_\ell \boldsymbol{\theta}_{ij}, \qquad i \neq j \\ \log(\pi_i(\boldsymbol{x}_\ell)/(1 - \pi_i(\boldsymbol{x}_\ell))) &= \boldsymbol{x}'_\ell \boldsymbol{\phi}_i, \end{aligned}$$

although other suitable link functions may be chosen. In these regression models, $\boldsymbol{\theta}_{ij} = (\theta_{ij0}, \theta_{ij1}, \ldots, \theta_{ij,d-1})'$ is a vector of d regression parameters relating the transition intensity from state i

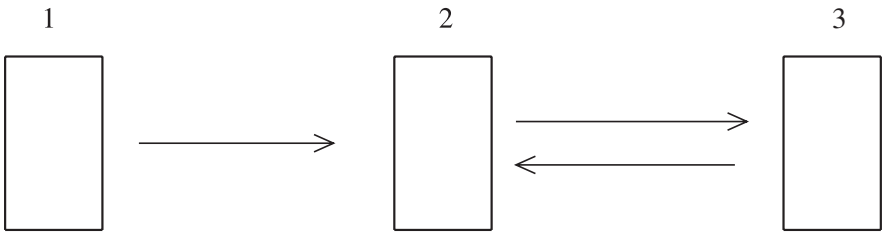

Fig. 1. Three-state diagram for smoking study.

to j to the covariates $\boldsymbol{x}_\ell$, and similarly $\boldsymbol{\phi}_i = (\phi_{i0}, \phi_{i1}, \ldots, \phi_{i,d-1})'$ is a vector of d regression parameters relating the mover–stayer probability from state i to the covariates $\boldsymbol{x}_\ell$. Of course the covariate vectors modulating the various transition intensities and mover–stayer probabilities do not need to be the same. The Fisher scoring algorithm may proceed in the same manner as described in Appendix A. We remark, however, that the number of parameters to be estimated increases rapidly with the introduction of new covariates since they may act on more than one model for each state of the process.

4. APPLICATION TO A SMOKING PREVENTION STUDY

The Waterloo Smoking Prevention Project is a randomized longitudinal study designed to investigate smoking behaviour among schoolchildren. A total of 6294 students from 100 schools in seven Ontario school boards participated in this study. Schools were randomized to receive either the regular health education programme, or one of four intensive anti-smoking programmes. The four intensive anti-smoking programmes involved identical teaching material, but differed in who delivered the material (teacher or public health nurse), and how that person was trained (by workshop or through the use of printed material). For the purpose of these analyses we pool the four treatment arms and define the treatment variable to be one for students from a school which was randomized to one of these four arms, and zero otherwise.

Questionaires regarding smoking attitudes and behaviour were administered annually from grade 6 to grade 12. To model the children's behaviour we define three states of interest. Children in state 1 have never smoked and are classified 'non-smokers'. State 2 represents children who are either regular smokers or are 'experimenting' with smoking, and children who have smoked but are currently not smoking are classified in state 3. The model is represented by the three-state diagram in Figure 1.

For illustration we have included sample data from children in two participating schools in Table 1. Inspection of the table reveals considerable variation in the spacing and number of assessments. The intermittently missing data arise due to absences from school and to students who moved to schools not participating in the study. From Table 1 we also note that some subjects exhibit long runs of visits in which they report being in the same state. These runs motivated the development of our generalized mover–stayer model since there was concern that an insufficient number of transitions were observed for a standard Markov model.

A primary objective of this study is to model smoking behaviour from the onset of smoking and so we restrict consideration to individuals who are known to have been non-smokers at the beginning of observation (in grade 6). This represents almost all of the sample. Our primary interest is to illustrate the application of the generalized mover–stayer model in this setting and to contrast it with the findings from a time-homogeneous Markov model. We also examine the effects of covariates representing treatment status and gender.

Table 1. *Sample data from two schools participating in the Waterloo smoking prevention project*

School A										School B									
			Assessment										Assessment						
ID	Group	Sex	1	2	3	4	5	6	7	ID	Group	Sex	1	2	3	4	5	6	7
1	Control	Male	1	2	2	2	2	2	2	1	Treatment	Male	1	–	–	–	–	–	–
2	Control	Female	1	1	–	–	3	3	3	2	Treatment	Female	1	1	1	1	1	3	3
3	Control	Female	1	2	–	–	–	–	–	3	Treatment	Male	1	1	1	2	2	2	2
4	Control	Male	1	1	1	–	2	2	2	5	Treatment	Male	1	–	1	1	1	1	1
5	Control	Male	1	1	1	1	1	1	1	5	Treatment	Female	1	1	1	1	1	1	1
6	Control	Female	1	3	3	3	2	2	2	6	Treatment	Female	1	1	1	3	2	2	2
7	Control	Male	1	3	3	3	3	–	–	7	Treatment	Male	1	1	1	–	2	2	2
8	Control	Male	1	1	1	–	1	1	1	8	Treatment	Female	1	1	1	–	2	2	2
9	Control	Male	1	3	3	2	–	–	–	9	Treatment	Female	1	–	1	1	1	1	1
10	Control	Male	1	1	1	1	1	1	3	10	Treatment	Male	1	1	1	1	1	1	2
11	Control	Male	1	–	–	–	2	3	3	11	Treatment	Male	1	1	2	3	3	3	3
12	Control	Female	1	1	1	1	1	1	–	12	Treatment	Female	1	2	2	–	2	2	2
13	Control	Male	1	1	1	–	1	1	1	13	Treatment	Female	1	1	1	1	1	1	1
14	Control	Male	1	1	–	–	1	1	1	14	Treatment	Male	1	1	1	1	1	1	3
15	Control	Female	1	1	1	–	2	–	–	15	Treatment	Female	1	1	1	1	1	2	2
16	Control	Male	1	1	1	–	–	–	–	16	Treatment	Male	1	3	2	2	2	2	2
17	Control	Male	1	1	1	2	2	2	2	17	Treatment	Female	1	1	1	2	2	2	2
18	Control	Female	1	1	2	2	–	2	2	18	Treatment	Male	1	1	1	2	2	2	2
19	Control	Male	1	1	1	1	1	1	1	19	Treatment	Male	1	1	1	1	1	3	3
20	Control	Male	1	1	1	1	1	1	1	20	Treatment	Female	1	1	1	1	1	2	2
21	Control	Male	1	2	2	2	2	–	–	21	Treatment	Female	1	1	–	–	–	–	–
22	Control	Female	1	1	1	–	1	–	–	22	Treatment	Female	1	1	1	–	–	–	–
23	Control	Female	1	1	1	–	2	2	2	23	Treatment	Female	1	1	1	1	1	1	1
24	Control	Female	1	1	2	–	3	2	2	24	Treatment	Male	1	1	1	3	2	2	2
25	Control	Female	1	1	1	2	2	2	3	25	Treatment	Female	1	1	1	1	2	–	–
26	Control	Female	1	–	–	–	–	–	–	26	Treatment	Male	1	1	1	1	1	3	2
27	Control	Female	1	1	3	2	2	2	2	27	Treatment	Female	1	1	1	1	1	1	1
28	Control	Male	1	1	1	1	1	1	3	28	Treatment	Female	1	1	1	2	2	2	2

For the time-homogeneous Markov model the intensity matrix is given by

$$\Lambda = \begin{pmatrix} -\lambda_{12} & \lambda_{12} & 0 \\ 0 & -\lambda_{23} & \lambda_{23} \\ 0 & \lambda_{32} & -\lambda_{32} \end{pmatrix}.$$

The conditional intensity matrix for individual ℓ in the mover–stayer framework is given by

$$\Lambda(\boldsymbol{\alpha}_\ell) = \begin{pmatrix} -\alpha_{\ell 1}\lambda_{12} & \alpha_{\ell 1}\lambda_{12} & 0 \\ 0 & -\alpha_{\ell 2}\lambda_{23} & \alpha_{\ell 2}\lambda_{23} \\ 0 & \alpha_{\ell 3}\lambda_{32} & -\alpha_{\ell 3}\lambda_{32} \end{pmatrix},$$

where $\alpha_{\ell k}$ has a Bernoulli distribution with probability $\pi_k = P(\alpha_{\ell k} = 0)$, and $\alpha_{\ell k}$ are independent, $k = 1, 2, 3$, $\ell = 1, \ldots, N$. To estimate parameters $\boldsymbol{\lambda} = (\lambda_{12}, \lambda_{23}, \lambda_{32})'$ and $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)'$, we first reparametrize as $\theta_{ij} = \log \lambda_{ij}$ and $\phi_i = \log(\pi_i/(1 - \pi_i))$, $i, j = 1, 2, 3$ and let $\boldsymbol{\theta} = (\theta_{12}, \theta_{23}, \theta_{32})'$ and $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)'$.

We fit a time-homogeneous Markov model and a generalized mover–stayer model to these data. The maximum likelihood estimates, information-based standard errors, approximate 95% confidence intervals for the parameters, and the log-likelihoods are reported in Table 2. We report the estimates under the parametrizations used for optimization and under the more interpretable scales of intensities and probabilities. For robustness against model misspecification which may arise from school-to-school variation in the transition intensities and mover–stayer probabilities, we also report robust standard errors computed using the sandwich-type variance formula provided in Appendix B (Royall, 1986).

The inspection of the estimates for the Markov model and the generalized mover–stayer model suggests that there is a significant improvement in the fit to the data with the incorporation of the mover–stayer probabilities. The transition intensities for movers out of states 2 and 3 are considerably higher in the generalized model than the regular Markov model because we have accommodated the possibility that some individuals are 'stayers' in these states. The estimate of the 'stayer' probability in state 1, however, is extremely small which implies that there would essentially be no difference between the inferences drawn from an ordinary mover–stayer model and the Markov model. We next fit a reduced mover–stayer model with the constraint that $\pi_1 = 0$. This appears to be a reasonable constraint based on similarities of the log-likelihoods, estimates, and confidence intervals for the full and reduced models. Overall there is good agreement between the information-based and robust standard errors suggesting that there is little effect of intra-school correlation. This also suggests that likelihood ratio statistics can be used for inference here and would give results similar to those based on robust or information-based standard errors.

To illustrate an application involving covariates we fit two regression models involving treatment and gender. Based on the findings reported in Table 2 we constrain the stayer probability in state 1 to be zero in each of these models. Let $x_{\ell 0} = 1$, $x_{\ell 1} = 1$ if individual ℓ is in one of the treatment arms and $x_{\ell 1} = 0$ otherwise, $x_{\ell 2} = 1$ if individual ℓ is male and $x_{\ell 2} = 0$ otherwise, and finally let $\boldsymbol{x}_\ell = (1, x_{\ell 1}, x_{\ell 2})'$, $\ell = 1, \ldots, N$. The regression models involving both covariates take the form

$$\log \lambda_{ij}(x_\ell) = \boldsymbol{x}_\ell' \boldsymbol{\theta}_{ij}, \quad (i, j) = (1, 2), (2, 3), (3, 2)$$
$$\log(\pi_i(x_\ell)/(1 - \pi_i(x_\ell))) = \boldsymbol{x}_\ell' \boldsymbol{\phi}_i, \quad i = 2, 3,$$

where $\boldsymbol{\theta}_{ij} = (\theta_{ij0}, \theta_{ij1}, \theta_{ij2})'$, and $\boldsymbol{\phi}_i = (\phi_{i0}, \phi_{i1}, \phi_{i2})'$. The estimates arising from fitting regression models involving just treatment as well as treatment and sex, are reported in Table 3.

Model 1 includes only the treatment indicator and suggests that the treatment has no significant effect on either the transition intensities or the mover–stayer probabilities. Model 2 suggests that when we control for the assigned treatment, males have significantly lower transition intensities out of state 1 ($p = 0.048$). Specifically, among individuals in the same treatment group, the rate of transition out of state 1 for males is 92% that of females (95% CI (85.4, 99.9%)). In this study, there is some evidence that females are more likely to smoke at an earlier age than males.

5. DISCUSSION

The traditional mover–stayer model accommodates the possibility that some subjects may remain in their initial state indefinitely, but those who leave their initial state are assumed to follow a common Markov process. The generalized mover–stayer model presented here also allows some subjects to remain in their initial state. Others may make transitions through several states before entering a stayer state in which they will remain indefinitely. Under this model, once a subject has been observed to leave a particular state, that state will never be an absorbing state for them. A further generalization would allow states passed through to eventually become absorbing states. In this more general model, each time a new state is entered there is a probability that the state will, in this visit, be an absorbing, or stayer, state. Of

Table 2. *Estimates arising from Markov and conditionally Markov mover–stayer models to smoking prevention data*

	State		Std.[a]	Markov model			Mover–stayer			Reduced mover–stayer		
Parameter	i	j	error	MLE	SE	95% CI	MLE	SE	95% CI	MLE	SE	95% CI
θ_{ij}	1	2	Info	−1.662	0.020	(−1.703, −1.622)	−1.666	0.020	(−1.707, −1.626)	−1.666	0.020	(−1.707, −1.626)
			Robust		0.020	(−1.701, −1.622)		0.020	(−1.706, −1.628)		0.020	(−1.706, −1.627)
	2	3	Info	−0.638	0.035	(−0.706, −0.571)	1.518	0.163	(1.198, 1.838)	1.518	0.162	(1.201, 1.835)
			Robust		0.041	(−0.719, −0.557)		0.167	(1.191, 1.845)		0.167	(1.192, 1.845)
	3	2	Info	−0.164	0.042	(−0.247, −0.082)	2.078	0.160	(1.764, 2.391)	2.078	0.159	(1.766, 2.389)
			Robust		0.052	(−0.266, −0.062)		0.155	(1.774, 2.382)		0.155	(1.774, 2.381)
ϕ_i	1		Info				−12.149	6.413	(−24.719, 0.421)			
			Robust					3.271	(−18.560, −5.738)			
	2		Info				−1.189	0.087	(−1.358, −1.019)	−1.189	0.086	(−1.358, −1.019)
			Robust					0.092	(−1.369, −1.009)		0.092	(−1.369, −1.008)
	3		Info				−1.989	0.108	(−2.201, −1.777)	−1.989	0.108	(−2.201, −1.777)
			Robust					0.120	(−2.225, −1.755)		0.119	(−2.223, −1.755)
λ_{ij}	1	2	Info	0.190	–	(0.182, 0.198)	0.189	–	(0.181, 0.197)	0.189	–	(0.181, 0.197)
			Robust		–	(0.182, 0.197)		–	(0.182, 0.196)		–	(0.182, 0.196)
	2	3	Info	0.528	–	(0.494, 0.565)	4.563	–	(3.315, 6.283)	4.563	–	(3.323, 6.266)
			Robust		–	(0.487, 0.573)		–	(3.289, 6.330)		–	(3.292, 6.325)
	3	2	Info	0.849	–	(0.781, 0.922)	7.986	–	(5.835, 10.928)	7.985	–	(5.848, 10.904)
			Robust		–	(0.766, 0.940)		–	(5.896, 10.824)		–	(5.894, 10.818)
π_i	1		Info				5.23e−06	–	(1.84e−11, 0.604)			
			Robust					–	(8.70e−09, 3.21e−03)			
	2		Info				0.233	–	(0.205, 0.265)	0.234	–	(0.205, 0.265)
			Robust					–	(0.203, 0.267)		–	(0.203, 0.267)
	3		Info				0.120	–	(0.100, 0.145)	0.120	–	(0.100, 0.145)
			Robust					–	(0.098, 0.147)		–	(0.098, 0.147)
Log-likelihood				−11 245.841			−11 040.744			−11 040.744		

[a]The first row of each pair, denoted 'Info', contains standard errors and confidence intervals estimated based on the expected information matrix; the second row of each pair, denoted 'Robust' contains standard errors and associated confidence intervals based on the robust variance estimates.

Table 3. *Estimates arising from generalized mover–stayer Markov regression models*

	State			Model 1			Model 2		
Parameter	i	j	Covariate	MLE	SE[a]	SE[b]	MLE	SE[a]	SE[b]
θ_{ij}	1	2	Intercept	−1.685	0.046	0.045	−1.645	0.049	0.049
			Treatment	0.024	0.052	0.050	0.022	0.050	0.050
			Sex				−0.079	0.041	0.040
	2	3	Intercept	1.611	0.467	0.490	1.447	0.502	0.518
			Treatment	−0.107	0.524	0.521	−0.130	0.476	0.536
			Sex				0.448	0.405	0.377
	3	2	Intercept	2.158	0.440	0.459	1.981	0.479	0.484
			Treatment	−0.090	0.488	0.487	−0.119	0.455	0.503
			Sex				0.485	0.391	0.353
ϕ_i	1		Intercept						
			Treatment						
			Sex						
	2		Intercept	−1.249	0.203	0.200	−1.197	0.208	0.207
			Treatment	0.077	0.236	0.225	0.088	0.223	0.231
			Sex				−0.137	0.180	0.190
	3		Intercept	−2.346	0.316	0.333	−2.313	0.315	0.356
			Treatment	0.440	0.343	0.356	0.444	0.324	0.354
			Sex				−0.081	0.217	0.240
Log-likelihood				-11 039.307			-11 036.030		

[a] Standard errors based on the expected information matrix.
[b] Robust standard errors based on the sandwich variance formula given in Appendix B.

course, with panel data this sort of generalization is difficult since the complete path is unobserved and hence the number of times each state is entered is unknown.

The generalized mover–stayer model represents a discrete mixture of Markov processes in which each subject follows a specific sub-model defined by a reduced set of communicating states. For sub-models which share sub-classes, the transition intensities between states within these sub-classes are assumed to be the same. Multi-state models involving continuous mixing distributions have proven useful in many contexts (Aalen, 1987; Cook and Ng, 1997; Ng and Cook, 1997), but in general models for processes under panel observation have not been developed. This is in part due to the numerical challenges in obtaining the marginal likelihood function. Numerical methods such as Markov chain Monte Carlo may prove useful in this setting.

Another extension of interest involves the introduction of time-nonhomogeneous transition intensities. Gentleman *et al.* (1994) provide an example in which piecewise constant transition intensities are specified. Kalbfleisch and Lawless (1989) point out that a class of Weibull transition intensities may be specified and estimated under a time transformation. Of the two approaches, the piecewise constant formulation is the most flexible and in some contexts it would be worthwhile to combine such a model with a mover–stayer formulation.

The data from the Waterloo Smoking Prevention Project potentially involve both a longitudinal dependence arising from responses measured repeatedly over time, and a cross-sectional dependence arising from the clustering of students within schools. The focus of the transitional analyses reported here is on the longitudinal correlation, but to provide some degree of robustness against possible correlations

among students in the same school we adopted robust sandwich-type variance estimates. This is a reasonable strategy when the anticipated degree of between-cluster heterogeneity is at most mild, as was the case here. With more substantial clustering the effect of misspecification becomes more serious and biases in the resulting parameter estimates may become a central concern. If this is the case, it may be possible to obtain a suitable model either by the introduction of school-level random effects, or by the use of generalized estimating equations in the spirit of Albert and Waclawiw (1998).

Since the generalized mover–stayer model does not possess the Markov property, it is required that the processes be observed from the start. Frequently, it is possible to identify time origins corresponding to some initiating event. For example, studies of disease activity and progression often involve follow-up from disease onset, such as a known time of infection (e.g. time from transfusion-related HIV infection). In studies of labour force dynamics, information may be collected on employment status from the time of entry into the work-force. In some applications, however, it may be difficult to identify, or even conceptualize, a reasonable time origin. In rheumatologic diseases for example, disease activity may have been present for several years prior to clinical diagnosis. Retrospective ascertainment of the time origin of the disease should be attempted to provide information which can be used to deal with left-truncation in such settings.

ACKNOWLEDGEMENTS

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada to R. J. C. and J. D. K., and from the Canadian Institutes of Health Research (CIHR) to R. J. C. R. J. C. is a CIHR Investigator. We thank Professor K. S. Brown and the Health Behaviour Research Group (UW) for permission to use the data from the Waterloo Smoking Prevention Project which was funded by NHLBI (US) and NHRDP (Canada). We also thank Ms Ker-Ai Lee for programmeming assistance.

APPENDIX A

A Fisher scoring algorithm for maximum likelihood estimation

From (3.6), we obtain the score vector with uth element

$$S_u(\psi) = \frac{\partial \log L(\psi)}{\partial \psi_u} = \sum_{\ell=1}^{N} \frac{1}{L_\ell} \frac{\partial L_\ell}{\partial \psi_u}, \tag{A.1}$$

where $u = 1, \ldots, A + B$. The negative of the matrix of second derivatives has (u, v) entry

$$-\frac{\partial^2 \log L}{\partial \psi_u \partial \psi_v} = \sum_{\ell=1}^{N} \frac{1}{L_\ell^2} \frac{\partial L_\ell}{\partial \psi_u} \frac{\partial L_\ell}{\partial \psi_v} - \sum_{\ell=1}^{N} \frac{1}{L_\ell} \frac{\partial^2 L_\ell}{\partial \psi_u \psi_v}, \tag{A.2}$$

$u, v = 1, \ldots, A + B$. The Fisher information is the expectation of (A.2). By considering the observed z_ℓ as one outcome in a multinomial trial for the ℓth individual, it can be seen that the expectation of the second term in the right-hand side of (A.2) is zero. As a consequence, an estimate of the (u, v) element of the Fisher information is provided by

$$I_{uv}(\psi) = \sum_{\ell=1}^{N} \frac{1}{L_\ell^2} \frac{\partial L_\ell}{\partial \psi_u} \frac{\partial L_\ell}{\partial \psi_v}. \tag{A.3}$$

Note that the simplifications leading to (2.6) for the homogeneous Markov model and (A.3) both appeal to the multinomial form of the likelihood function but these are quite different multinomial

distributions. In deriving (2.6) we make use of the multinomial distribution of the state occupied at subsequent visits conditional on the previous state occupied, and therefore there are at most K outcomes in the multinomial distribution. In deriving (A.3), the entire observed sequence of states for individual ℓ is viewed as a realization from a multinomial distribution with $m_\ell \cdot K$ possible outcomes. The size of this sample space suggests that (A.3) is much more convenient to work with than the estimate analogous to (2.5).

We now describe the algorithm. Let $\hat{\psi}_{(0)} = (\hat{\boldsymbol{\theta}}'_{(0)}, \hat{\boldsymbol{\phi}}'_{(0)})'$ be an initial estimate of $\boldsymbol{\psi}$, $S(\boldsymbol{\psi})$ be the $(A+B) \times 1$ score vector with components given by (A.1) and $I(\boldsymbol{\psi})$ be the $(A+B) \times (A+B)$ matrix with entries given by (A.3). The maximum likelihood estimate of $\boldsymbol{\psi}$ can often be obtained by recursive application of the equation

$$\hat{\psi}_{(r)} = \hat{\psi}_{(r-1)} + I^{-1}(\hat{\psi}_{(r-1)}) S(\hat{\psi}_{(r-1)})$$

until convergence is achieved, where it is assumed that $I(\hat{\psi}_{(r)})$ is nonsingular, $r = 1, 2, \ldots$.

This procedure needs only the first derivatives of the likelihood components L_ℓ and these are obtained using only the first derivatives of the transition matrices $P(s|\alpha)$. In particular, we find that

$$\frac{\partial L_\ell}{\partial \psi_u} = \sum_{\boldsymbol{\alpha} \in \mathcal{S}_\ell} P_\ell(\boldsymbol{\alpha}, \boldsymbol{\theta}) f_\ell(\boldsymbol{\alpha}; \boldsymbol{\phi}) \sum_{(i,j,s) \in U_\ell} \frac{1}{P_{ij}(s|\boldsymbol{\alpha})} \frac{\partial P_{ij}(s|\boldsymbol{\alpha})}{\partial \theta_u}$$

for $u = 1, \ldots, A$, and

$$\frac{\partial L_\ell}{\partial \psi_u} = \sum_{\boldsymbol{\alpha} \in \mathcal{S}_\ell} P_\ell(\boldsymbol{\alpha}, \boldsymbol{\theta}) f_\ell(\boldsymbol{\alpha}; \boldsymbol{\phi}) \left\{ - \sum_{i \in \mathcal{M}_\ell} \frac{1}{1-\pi_i} \frac{\partial \pi_i}{\partial \phi_{u-A}} + \sum_{i \notin \mathcal{M}_\ell} \left(\frac{1-\alpha_{\ell i}}{\pi_i} - \frac{\alpha_{\ell i}}{1-\pi_i} \right) \frac{\partial \pi_i}{\partial \phi_{u-A}} \right\}$$

for $u = A+1, \ldots, A+B$.

To compute $P_{ij}(s|\boldsymbol{\alpha})$, we suppose that $\Lambda(\boldsymbol{\alpha})$ is diagonalizable and utilize the decomposition $\Lambda(\boldsymbol{\alpha}) = H(\boldsymbol{\alpha}) D(\boldsymbol{\alpha}) H^{-1}(\boldsymbol{\alpha})$ which can be easily computed for any given $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$. In this, $D(\boldsymbol{\alpha})$ is the diagonal matrix of eigenvalues $d_1(\boldsymbol{\alpha}), \ldots, d_K(\boldsymbol{\alpha})$, and $H(\boldsymbol{\alpha})$ is the $K \times K$ matrix whose ith column is a right eigenvector corresponding to $d_i(\boldsymbol{\alpha})$. Recall that the ith row of $\Lambda(\boldsymbol{\alpha})$ is $(\lambda_{i1}, \ldots, \lambda_{iK})$ if $i \in \mathcal{M}_\ell$; when $i \notin \mathcal{M}_\ell$, the ith row of $\Lambda(\boldsymbol{\alpha})$ may be either $(0, \ldots, 0)$ or $(\lambda_{i1}, \ldots, \lambda_{iK})$ depending on whether $\alpha_{\ell i} = 0$ or 1.

Following the results in Section 2, it follows that the transition probability matrix can be expressed as

$$P(s|\boldsymbol{\alpha}) = H(\boldsymbol{\alpha}) \exp(D(\boldsymbol{\alpha}) s) H^{-1}(\boldsymbol{\alpha}).$$

The first derivatives with respect to the components of $\boldsymbol{\theta}$ are

$$\frac{\partial P(s|\boldsymbol{\alpha})}{\partial \psi_u} = \frac{\partial P(s|\boldsymbol{\alpha})}{\partial \theta_u} = H(\boldsymbol{\alpha}) V^{(u)}(\boldsymbol{\alpha}) H^{-1}(\boldsymbol{\alpha}), \quad u = 1, \ldots, A,$$

where $V^{(u)}(\boldsymbol{\alpha})$ is the $K \times K$ matrix with (i,k) entry given by

$$\frac{g_{ik}^{(u)} (\exp\{d_i(\boldsymbol{\alpha}) s\} - \exp\{d_k(\boldsymbol{\alpha}) s\})}{d_i(\boldsymbol{\alpha}) - d_k(\boldsymbol{\alpha})},$$

if $d_i(\boldsymbol{\alpha}) \neq d_k(\boldsymbol{\alpha})$. If $d_i(\boldsymbol{\alpha}) = d_k(\boldsymbol{\alpha})$ or, in particular, if $i = k$ the (i,k) element is

$$g_{ik}^{(u)} s \exp\{d_i(\boldsymbol{\alpha}) s\}.$$

In these expressions, $g_{ik}^{(u)}$ is the (i, k) entry in $G^{(u)} = H^{-1}(\boldsymbol{\alpha})(\partial\Lambda(\boldsymbol{\alpha})\partial\theta_u)H(\boldsymbol{\alpha})$. A derivation of this result for the case of distinct eigenvalues appears in Kalbfleisch and Lawless (1985).

Note that the assumption that $\Lambda(\boldsymbol{\alpha})$ is diagonalizable does not imply that all eigenvalues are distinct. One could also develop more general decompositions that would apply when $\Lambda(\boldsymbol{\alpha})$ is not diagonalizable, but for practical purposes, the diagonalizable case is sufficient.

APPENDIX B

Robust variance estimation

To protect against possible residual correlation in the latent mover–stayer indicators and the transition times, in addition to information-based standard errors we report robust standard errors based on the sandwich-type variance formula (Royall, 1986). We generalize the notation of Appendix A and let $\boldsymbol{S}^{(h)}(\boldsymbol{\psi})$ denote the score vector given by (A.1) but constructed based only on students from school h, $h = 1, 2, \ldots, H$, where H denotes the total number of schools. Then

$$\boldsymbol{S}(\boldsymbol{\psi}) = \sum_{h=1}^{H} \boldsymbol{S}^{(h)}(\boldsymbol{\psi}).$$

From White (1982) we know that if $L(\boldsymbol{\psi})$ is constructed from a misspecified model, then $\hat{\boldsymbol{\psi}}$ converges to $\boldsymbol{\psi}^*$ almost surely, where $\boldsymbol{\psi}^*$ solves $E_T(\boldsymbol{S}(\boldsymbol{\psi}))$ where E_T denotes an expectation taken with respect to the true distribution. Moreover,

$$\sqrt{H}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}^*) \longrightarrow N(0, C(\boldsymbol{\psi}^*))$$

almost surely where $C(\boldsymbol{\psi}) = A^{-1}(\boldsymbol{\psi})B(\boldsymbol{\psi})A^{-1}(\boldsymbol{\psi})$ and

$$A(\boldsymbol{\psi}) = E_T(\partial\boldsymbol{S}(\boldsymbol{\psi})/\partial\boldsymbol{\psi})$$
$$B(\boldsymbol{\psi}) = E_T(\boldsymbol{S}(\boldsymbol{\psi})\boldsymbol{S}'(\boldsymbol{\psi})).$$

We estimate the matrix $C(\boldsymbol{\psi})$ evaluated at $\boldsymbol{\psi}^*$ with $\hat{A}^{-1}(\hat{\boldsymbol{\psi}})\hat{B}(\hat{\boldsymbol{\psi}})\hat{A}^{-1}(\hat{\boldsymbol{\psi}})$ where

$$\hat{A}(\hat{\boldsymbol{\psi}}) = H^{-1}\sum_{h=1}^{H} \partial S^{(h)}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}}$$
$$\hat{B}(\hat{\boldsymbol{\psi}}) = H^{-1}\sum_{h=1}^{H} S^{(h)}(\boldsymbol{\psi})[S^{(h)}(\boldsymbol{\psi})]'|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}}.$$

REFERENCES

AALEN, O. O. (1987). Mixing distributions on a Markov chain. *Scandinavian Journal of Statistics* **14**, 281–289.

ALBERT, P. S. AND WACLAWIW, M. A. (1998). A two-state Markov chain for heterogeneous transitional data: a quasi-likelihood approach. *Statistics in Medicine* **17**, 1481–1493.

ANDERSEN, P. K., BORGAN, O., GILL, R. D. AND KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.

COOK, R. J. (1999). A mixed model for Markov processes under panel observation. *Biometrics* **55**, 178–183.

COOK, R. J. AND NG, E. T. M. (1997). A logistic-bivariate normal model for overdispersed two-state Markov processes. *Biometrics* **53**, 358–364.

COX, D. R. AND MILLER, H. D. (1977). *The Theory of Stochastic Processes*. London: Chapman and Hall.

FRYDMAN, H. (1984). Maximum likelihood estimation in the model. *Journal of the American Statistical Association* **79**, 632–638.

GENTLEMAN, R. G., LAWLESS, J. F., LINDSEY, J. AND YAN, P. (1994). Multistate Markov models for analysing incomplete disease history data, with illustrations for HIV disease. *Statistics in Medicine* **13**, 805–821.

GLADMAN, D. D., FAREWELL, V. T. AND NADEAU, C. (1995). Clinical indicators of progression in psoriatic arthritis (PSA): multivariate relative risk model. *Journal of Rheumatology* **22**, 675–679.

KALBFLEISCH, J. D. AND LAWLESS, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* **80**, 863–871.

KALBFLEISCH, J. D. AND LAWLESS, J. F. (1989). Some statistical methods for panel life history data. *Proceedings of the Statistics Canada Symposium on Analysis of Data in Time*.

LAWLESS, J. F. AND FONG, D. (1999). State duration models in clinical and observational studies. *Statistics in Medicine* **18**, 2365–2376.

LAWLESS, J. F. AND YAN, P. (1993). Some statistical methods for follow-up studies of disease with intermittent monitoring. In Hoppe, F. M. (ed.), *Multiple Comparisons, Selection, and Applications in Biometry*, New York: Marcel Dekker.

NG, E. T. M. AND COOK, R. J. (1997). Modeling two-state disease processes with random effects. *Lifetime Data Analysis* **3**, 315–335.

ROYALL, R. M. (1986). Model robust confidence intervals using maximum likelihood estimators. *International Statistical Review* **54**, 221–226.

SATTEN, G. A. (1999). Estimating the extent of tracking in interval-censored chain-of-events data. *Biometrics* **55**, 1228–1231.

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1–25.

[*Received December 28, 2000*; *revised August 30, 2001*; *accepted for publication September 17, 2001*]