

# Estimation and Assessment of Markov Multistate Models with Intermittent Observations on Individuals

J. F. LAWLESS

*Department of Statistics and Actuarial Science,  
University of Waterloo, Waterloo, ON, N2L 3G1, Canada  
E-mail: [jlawless@uwaterloo.ca](mailto:jlawless@uwaterloo.ca)*

N. NAZERI RAD

*Department of Statistics and Actuarial Science,  
University of Waterloo, Waterloo, ON, N2L 3G1, Canada*

## Summary

Multistate models provide important methods of analysis for many life history processes, and this is an area where John Klein made numerous contributions. When individuals in a study group are observed continuously so that all transitions between states, and their times, are known, estimation and model checking is fairly straightforward. However, individuals in many studies are observed intermittently, and only the states occupied at the observation times are known. We review methods of estimation and assessment for Markov models in this situation. Numerical studies that show the effects of inter-observation times are provided, and new methods for assessing fit are given. An illustration involving viral load dynamics for HIV-positive persons is presented.

*Keywords:* Goodness of fit, Maximum likelihood, Panel data, Prevalence probability, Relative efficiency, Transition intensity, Transition probability

**The final publication:** “Lawless, J.F. and Nazeri Rad, N. (2015). Estimation and assessment of markov multistate models with intermittent observations on individuals. *Lifetime Data Analysis*, 21 (2), 160–179” is available at Springer via DOI: [10.1007/s10985-014-9310-z](https://doi.org/10.1007/s10985-014-9310-z).

## 1 INTRODUCTION

Multistate (MS) models are widely used to represent and analyze life history processes. In medicine, for example, they have been used to model disease progression in persons with psoriatic arthritis (O’Keeffe et al. 2011), the development of diabetic retinopathy (Cook and Lawless 2014), and patient outcomes following bone marrow transplantation (Klein and Shu 2002). In the MS framework individuals are in one of a set of  $K$  defined states at any point in time. We let  $Y(t)$  denote the state at time  $t$  for an individual, and consider processes  $\{Y(t), t \geq 0\}$ , where  $t = 0$  denotes a specified time origin. Process features of interest include transitions between states, sojourn times in states and the times to (first) entry to specific states, along with their relationship to covariates or other processes.

Continuous-time MS models can be specified under mild conditions in terms of transition intensity functions,

$$\lambda_{rs}(t|H(t)) = \lim_{\Delta t \rightarrow 0} \frac{P(Y(t + \Delta t) = s | Y(t^-) = r, H(t))}{\Delta t} \quad r \neq s \quad (1)$$

where  $H(t)$  denotes the history  $\{Y(s), 0 \leq s < t\}$  of the process and relevant covariates. Markov models have  $\lambda_{rs}(t|H(t)) = \alpha_{rs}(t)$  and semi-Markov models have  $\lambda_{rs}(t|H(t)) = \alpha_{rs}(B(t))$ , where  $B(t)$  is the elapsed time since entry to the current state ( $r$ ). These models and extensions involving covariates are readily fitted and assessed when continuous histories  $\{Y_i(t), 0 \leq t < b_i\}$  are observed for a random selection of individuals  $i = 1, \dots, n$  (Andersen et al. 1993; Cook and Lawless 2007; Aalen et al. 2008). However, in many studies data on individuals are collected intermittently, at visit times  $t_0 < t_1 < \dots < t_m$ , and only the states  $Y_i(t_j)$  occupied at these observation times are known; exact transition times, and in some cases the states visited between observation times  $t_{j-1}$  and  $t_j$ , are unknown. This is often termed panel data (e.g. Kalbfleisch and Lawless 1985; Hwang and Brookmeyer 2003).

Our objective in this paper is to consider the effects of intermittent observation on estimation and assessment of Markov models. Rather little has been written about this. Kalbfleisch and Lawless (1985) made general remarks and Hwang and Brookmeyer (2003) presented some numerical results on estimation for “progressive” models, in which the only transition permitted from state  $k$  is to state  $k + 1$  ( $k = 1, \dots, K - 1$ ). However, models with bi-directional transitions between certain states are needed in many applications. Examples include relapse-remitting conditions like repeated infections (Cook and Lawless 2007, Sect. 6.7.2), recurrent hospitalization spells (e.g. Kvist et al. 2010) and progression of diseases such as Type 1 diabetes (Cook and Lawless 2014) and multiple sclerosis (Mandel and Betensky 2008). Cook (2000) and Mehtala et al. (2011) have considered time homogeneous two-state Markov models. The former compares the efficiency of estimation of transition intensities based on intermittent observation relative to continuous observation, and considers study design for relapse-remitting conditions. Mehtala et al. (2011) focus mainly on estimation of transition intensities and the optimal choice of initial distribution for different ratios of the intensities for state 1 to 2 and state 2 to 1. They also consider two-phase designs, where a phase 1 study provides preliminary parameter estimates that can be used to choose an optimal design for a phase 2 study. In the present paper we consider both two- and three-state models in our numerical studies, and we compare uni-directional and bi-directional models. We consider efficiency of both transition intensity and transition probability estimates for Markov models, and also discuss goodness of fit procedures when observation times vary substantially between and within individuals. As we show, the effects of intermittent observation are much more severe for models allowing bi-directional transitions. Consider, for example, the two multi-state diagrams in Figure 1, which are considered in subsequent sections. In each case transitions in both directions can occur between states 1 and 2, and we might not even know the number of transitions between two consecutive observation times. Models (a) and (b) are compared in Section 3 with the corresponding three-state models in which transitions from state 2 to state 1 cannot occur.

In the following sections we review and extend methodology for Markov models. The extensions include a numerical study of the loss of information due to intermittent observation, improved methods for assessing goodness of fit when times between observations are variable, and a method for estimating power for likelihood ratio tests of fit. Section 2 reviews estimation of Markov model parameters based on intermittent observation of individuals and outlines a procedure for obtaining asymptotic covariance matrices. Section 3 presents a numerical study of the effects of inter-visit gap times on estimation in progressive and bi-directional models. Section 4 discusses model assessment and gives a method of estimating the power of likelihood ratio tests of fit. Section 5 applies the methodology here to the study

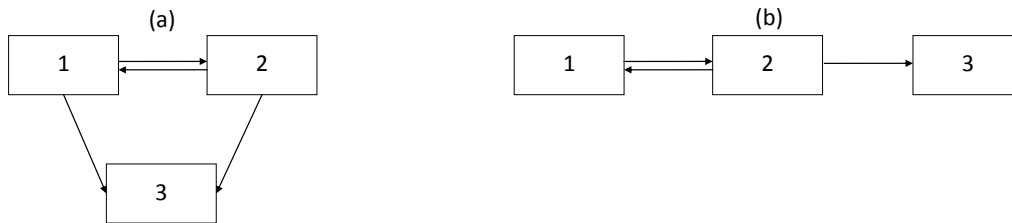


Figure 1: Two multi-state models

of viral load dynamics in a cohort of HIV-positive persons. We also describe and illustrate a method of obtaining empirical state occupancy probability (prevalence) estimates when observation times are related to an individual's previous process history. Section 6 contains some concluding remarks.

## 2 ESTIMATION AND EFFICIENCY WITH PANEL DATA

### 2.1 MAXIMUM LIKELIHOOD ESTIMATION

We consider a  $K$ -state Markov model with transition intensities (1) of the form  $q_{rs}(t)$ . The corresponding transition intensity matrix  $Q(t)$  has off-diagonal entries  $q_{rs}(t)$ ,  $r \neq s$  and diagonal entries  $q_{rr}(t) = -\sum_{s \neq r} q_{rs}(t)$ . Estimation of time-dependent intensities from panel data can be computationally challenging for many models (e.g. see Titman 2011; Cook and Lawless 2014) and in practice, models with piecewise-constant intensities are widely used. In this case it is assumed that for a set of specified cut-points  $0 = a_0 < a_1 < \dots < a_R = \infty$  we have

$$q_{rs}(t) = q_{rs}^{(l)} \quad a_{l-1} \leq t < a_l. \quad (2)$$

We will focus here on time-homogeneous models for which  $q_{rs}(t) = q_{rs}$  for some interval  $0 \leq t < T$ , and consider a cohort of individuals observed over  $(0, T)$  at common times  $0 = t_0 < t_1 < \dots < t_m = T$ . Piecewise-constant models are handled by summing log-likelihood functions and information matrices discussed below over the separate time intervals  $[a_{l-1}, a_l)$  in (2). In observational studies individuals  $i = 1, \dots, n$  may be seen at varying sets of times  $t_{ij}$  ( $j = 1, \dots, m_i$ ), and the norm is for the times between visits to vary substantially within and between individuals. We discuss this situation in Sections 4 and 5.

Let  $\theta$  be a parameter vector specifying the unknown transition intensities  $q_{rs}$  and any regression coefficients (see below). The likelihood function based on  $n$  independent individuals who are all seen at times  $0 = t_0 < t_1 < \dots < t_m = T$  is (Kalbfleisch and Lawless 1985)

$$L(\theta) = \prod_{i=1}^n \prod_{j=1}^m P_{y_i(t_{j-1})y_i(t_j)}(w_j; \theta), \quad (3)$$

where  $P_{rs}(w; \theta) = P(Y_i(t+w) = s | Y_i(t) = r; \theta)$  and  $w_j = t_j - t_{j-1}$  ( $j = 1, \dots, m$ ). In some special models (e.g. Hwang and Brookmeyer 2003; Jackson 2011) there are algebraic expressions for the transition probabilities  $P_{rs}(w; \theta)$  but in general they can be obtained numerically for specified  $w$  and  $\theta$  from the matrix exponential expression

$$P(w) = \exp(wQ). \quad (4)$$

This can be computed by methods described by Moler and Van Loan (2003); Kalbfleisch and Lawless (1985) and Jackson (2011) discuss algorithms for the computation and maximization of (3). The `msm` package in R (Jackson 2011) is one of several that will do this, and it also handles piecewise-constant models. Fixed covariates  $x$  may also be added via multiplicative specifications  $q_{rs}(x) = q_{rs0} \exp(\beta'_{rs}x)$ ,  $r \neq s$ , where  $x$  is a vector of covariates and  $\beta$  is a corresponding vector of regression coefficients. Covariates are also allowed to be time-varying in `msm`, provided they are piecewise-constant.

## 2.2 ASYMPTOTIC COVARIANCE MATRICES

The expected (Fisher) information matrix  $I(\theta)$  for  $\theta$  for a cohort of  $n$  independent individuals (Kalbfleisch and Lawless 1985, Sect. 3.1) has entries

$$\begin{aligned} I_{uv}(\theta) &= E\left(\frac{-\partial^2 \log L(\theta)}{\partial \theta_u \partial \theta_v}\right) \\ &= \sum_{j=1}^m \sum_{r,s=1}^K \frac{E_r(t_{j-1}; \theta)}{P_{rs}(w_j; \theta)} \frac{\partial P_{rs}(w_j; \theta)}{\partial \theta_u} \frac{\partial P_{rs}(w_j; \theta)}{\partial \theta_v}, \end{aligned} \quad (5)$$

where

$$E_r(t_{j-1}; \theta) = \sum_{i=1}^n P\{Y_i(t_{j-1}) = r | Y_i(t_0); \theta\} \quad (6)$$

is the expected number of individuals in state  $r$  at time  $t_{j-1}$ . This assumes that all individuals are seen at each of the times  $t_1, \dots, t_m$ . In many studies there is attrition (loss to followup, LTF) and in that case we multiply the  $i$ th term in (6) by  $\pi_i(t_{j-1})$ , the probability that individual  $i$  will not be LTF at time  $t_{j-1}$ . Kalbfleisch and Lawless (1985, Eq. 3.4) give an expression for the computation of derivatives in (5), but numerical differentiation is often simpler and accurate. This uses the approximation

$$\frac{\partial P_{rs}(w; \theta)}{\partial \theta_u} \approx \frac{P_{rs}(w; \theta + \Delta_u) - P_{rs}(w; \theta - \Delta_u)}{2\delta_u}, \quad (7)$$

where  $\Delta_u$  is a vector with a small value  $\delta_u$  in the position corresponding to  $\theta_u$  and 0 elsewhere.

The asymptotic covariance matrix  $V(\theta)$  for  $\sqrt{n}(\hat{\theta} - \theta)$  is the limit of  $nI(\theta)^{-1}$ , assuming some initial distribution of individuals at time  $t_0$  and assuming the Markov model is correct and that  $\theta$  is the true parameter vector. It is also the asymptotic covariance matrix when the model is misspecified, with  $\theta$  then representing the probability limit of  $\hat{\theta}$  as  $n \rightarrow \infty$  (White 1982). Our interest in  $V(\theta)$  is two-fold: (i) for studying the effect of the inter-visit times  $w_j$  on the information about various model parameters, and (ii) for study planning decisions concerning the length and frequency of followup. A third use of  $V(\theta)$  is for assessing when software such as `msm` may encounter convergence problems when trying to maximize  $L(\theta)$ ; this occurs frequently when models have numerous states or covariates. For some models the likelihood function may be nearly flat in certain directions when the  $w_j$  are too large.

## 3 NUMERICAL STUDIES

Hwang and Brookmeyer (2003) studied the effects of the number of visits and the distance between them on estimation of the transition intensities  $q_{k,k+1}$  in a progressive, time-homogeneous Markov model. Numerical results were given for a three-state model and they show that for a fixed study length, the loss of efficiency for  $q_{k,k+1}$  compared to continuous observation is small when the gap time  $w$  between visits is less than the average sojourn time  $q_{k,k+1}^{-1}$  in state  $k$ , and is moderate when  $w \leq 1.5 q_{k,k+1}^{-1}$ . However,

the situation is much different for models in which transitions in both directions between a pair of states are possible. Some insight into this can be gained from the two-state model with transition intensities  $q_{12}$  and  $q_{21}$ , for which the transition probability matrix  $P(t, t + w)$  is (Kalbfleisch and Lawless 1985, Sect. 4)

$$\begin{pmatrix} 1 - \pi(1 - \exp(-\alpha w)) & \pi(1 - \exp(-\alpha w)) \\ (1 - \pi)(1 - \exp(-\alpha w)) & \pi + (1 - \pi)\exp(-\alpha w) \end{pmatrix} \quad (8)$$

where  $\alpha = q_{12} + q_{21}$  and  $\pi = q_{12}\alpha^{-1}$ . As  $w$  increases, both rows of (8) approach the limiting distribution  $(1 - \pi, \pi)$ . This suggests that if the gap time  $w$  between visits is large, we will be able to estimate  $\pi$  well, but estimates of  $q_{12}$  and  $q_{21}$  will be very imprecise.

We consider below the efficiency of maximum likelihood estimates of transition intensities  $q_{rs}$  and of transition probabilities  $P_{rs}(t)$  based on panel data, relative to a study of the same size with continuous observation of individuals. The bi-directional two-state model was considered previously by Cook (2000), who showed that the relative efficiency for  $q_{12}$  and  $q_{21}$  dropped drastically as the gap between successive observation times increased. For a specific parameter  $\psi$  we define the relative efficiency of  $\hat{\psi}^{(w)}$  based on a panel study with individuals seen every  $w$  units of time as

$$reff(\hat{\psi}^{(w)}) = \frac{asvar(\hat{\psi}^{(0)})^{1/2}}{asvar(\hat{\psi}^{(w)})^{1/2}} \quad (9)$$

where  $\hat{\psi}^{(0)}$  is the estimate for a study in which observation is continuous (i.e.  $w = 0$ ), so that exact transition times are recorded. In (9) we use  $asvar(\hat{\psi}^{(w)})$ , the asymptotic variance of  $\hat{\psi}^{(w)}$ ; this is obtained from the asymptotic covariance matrix  $V(\theta)$  as described in Section 2.2. The ratios of asymptotic standard deviations in (9) reflect the relative lengths of confidence intervals for parameters. The asymptotic covariance matrix for the case of continuous observation is described for general counting processes by Andersen et al. (1993, Chap. 6). For a time-homogeneous Markov model the covariance matrix for the transition intensity estimates is diagonal, with

$$asvar(\sqrt{n}(\hat{q}_{rs}^{(0)} - q_{rs})) = \frac{q_{rs}}{E_r^{(0)}(T; \theta)}, \quad (10)$$

where

$$E_r^{(0)}(T; \theta) = \lim\left(\frac{1}{n}\sum_{i=1}^n \int_0^T P\{Y_i(t) = r | Y_i(0)\} dt\right) \quad (11)$$

is the average time spent in state  $r$  across all individuals over the time period  $(0, T)$ . The limit in (11) is taken by assigning some initial distribution of states at  $t = 0$ . As in Section 2.2, (11) can be adjusted for attrition by multiplying the integrands by  $\pi_i(t)$ , the probability that individual  $i$  is still under observation (not LTF) at time  $t$ .

### 3.1 TWO-STATE MODELS

We compare the bi-directional two-state model with the progressive two-state model, which is equivalent to a survival model. We denote models M1a with  $q_{12} > 0$ ,  $q_{21} > 0$  and M1b with  $q_{12} > 0$ ,  $q_{21} = 0$ . For M1a we consider two cases, (i)  $q_{12} = 1$ ,  $q_{21} = 1$ , (ii)  $q_{12} = 1$ ,  $q_{21} = 4$ . In case (i) the states have the same average sojourn times whereas in case (ii) the sojourns in state 2 are much shorter, as

would occur when it represents a transient adverse condition. Two followup times  $T = 4$  and  $T = 10$  are considered, along with different fixed times  $w_j = \Delta t$  between visits. All individuals are in state 1 at  $t = 0$ . The ratios (9) of asymptotic standard deviations for estimators of transition intensities and probabilities are given in Table 1. They show that the relative precision of transition intensity estimators falls off rapidly with increasing inter-visit times for bi-directional models M1a, but much more slowly for transition probabilities, especially for larger  $t$ . Loss of efficiency is small for the progressive model M1b, until the inter-visit time is about twice the average sojourn time in state 1. Relative efficiencies for  $q_{12}$  and  $q_{21}$  are higher for model M1a(i) than for M1a(ii), but the opposite is true for the transition probabilities. For transition intensities the lower relative efficiency for M1a(ii) is due to the fact that on average more transitions occur over an inter-visit interval than for model M1a(i). Comparison of Table 1 with similar results when  $T = 10$  indicate no substantial change in relative efficiency for intensity estimators, but there can be significant gains for transition probability estimators; this is related to the speed with which the rows of (8) approach the limiting distribution as  $w$  increases. The results in Table 1 are consistent with earlier comments on Model M1a. They are also consistent with the results of Cook (2000) as far as the efficiency of  $q_{12}$  and  $q_{21}$  are concerned.

Table 1: Ratios of asymptotic standard deviations for estimators of transition intensities and probabilities in two-state Markov models with (1)  $q_{12} = q_{21} = 1$  (M1a(i)), (2)  $q_{12} = 1, q_{21} = 4$  (M1a(ii)), and (3)  $q_{12} = 1, q_{21} = 0$  (M1b)

	$\Delta t$	$q_{12}$	$q_{21}$	$P_{11}(1)$	$P_{11}(2)$	$P_{11}(4)$
M1a(i)	0.25	0.89	0.87	0.64	0.64	0.64
	0.5	0.73	0.71	0.62	0.64	0.64
	1	0.40	0.40	0.50	0.62	0.60
	2	0.08	0.09	0.13	0.44	0.51
M1a(ii)	0.25	0.63	0.64	0.97	0.97	0.97
	0.5	0.30	0.30	0.87	0.88	0.88
	1	0.04	0.04	0.64	0.70	0.70
	2	4.2e-04	4.4e-04	0.01	0.52	0.54
M1b	0.25	0.99	-	0.99	0.99	0.99
	0.5	0.99	-	0.99	0.99	0.99
	1	0.96	-	0.96	0.96	0.96
	2	0.86	-	0.86	0.86	0.86

Values of  $P_{11}(t)$  in M1a(i) are  $P_{11}(1) = 0.568, P_{11}(2) = 0.509, P_{11}(4) = 0.500$ , in M1a(ii)  $P_{11}(1) = 0.801, P_{11}(2) = 0.800, P_{11}(4) = 0.800$ , and in M1b  $P_{11}(1) = 0.368, P_{11}(2) = 0.135, P_{11}(4) = 0.018$ . The total followup time is 4 years.

### 3.2 THREE-STATE MODELS

We first consider the multistate diagram (b) in Figure 1, and each of the cases  $q_{21} > 0$  and  $q_{21} = 0$ ; we also consider a binary covariate  $x$ , with half of the individuals having  $x = 0$  and half  $x = 1$ . Models M2a and M2b have no covariates and transition intensities  $q_{12} = 1$ ,  $q_{21} = 2$ ,  $q_{23} = 1$  and  $q_{12} = 1$ ,  $q_{21} = 0$ ,  $q_{23} = 1$  respectively. Models M3a and M3b are the corresponding models with covariate  $x$ , and intensities  $q_{rs} \exp(\beta_{rs}x)$ , with  $\beta_{12} = -0.2$ ,  $\beta_{21} = -0.5$ , and  $\beta_{23} = -0.9$ . Finally, we investigated models M4a and M4b which are as in Figure 1(a), with transition intensities  $q_{12} = 1$ ,  $q_{21} = 2$ ,  $q_{13} = 0.25$  and  $q_{23} = 0.5$  for M4a, and  $q_{12} = 0.5$ ,  $q_{21} = 0$ ,  $q_{13} = 0.25$ ,  $q_{23} = 0.5$  for M4b. In each case, all individuals were in state 1 at  $t = 0$ .

Selected results are shown in Tables 2, 3, and 4. We investigated the same inter-visit and followup times as in Table 1. We once again give results only for scenarios with  $T = 4$ ; relative efficiencies with  $T = 10$  were broadly similar. Table 2 shows results for models M2a and M2b; as with the two-state models, for bi-directional model M2a there is substantial loss of efficiency for transition intensities as  $\Delta t$  increases, though as we would expect, this is much less for  $q_{23}$ . Except for  $P_{13}(1)$ , the loss of efficiency for transition probability estimators is negligible. Except when  $\Delta t = 2$ , the relative efficiency is near one for estimators in the uni-directional model M2b. Table 3 shows the relative efficiencies for estimators of baseline transition intensities and covariate effects for models M3a and M3b. For model M3a the loss of efficiency for covariate effects is slightly less than for the  $q_{rs}$  as  $\Delta t$  increases; the relative efficiencies for the  $q_{rs}$  are exactly the same as in Table 2. We note that there is almost no loss of efficiency for  $\beta_{23}$ . For model M3b there is negligible loss of efficiency for covariate effects, except for  $\beta_{12}$  when  $\Delta t = 2$ . We also compared results for transition probability estimators for individuals with  $x = 1$  for models M3a and M3b; results for those with  $x = 0$  are exactly as in Table 3. There is virtually no loss of efficiency relative to continuous observation for either the bi-directional or uni-directional cases.

Table 4 shows results for models M4a and M4b, based on Figure 1(a). The patterns seen here are qualitatively similar to those in Table 2: there are very substantial drops in relative efficiency for estimators of transition intensities in M4a, but aside from when  $\Delta t = 2$  the drops are small for transition probability estimators. As might be expected, estimators for transitions to the absorbing state 3 show smaller drops than for  $q_{12}$  and  $q_{21}$ . For the uni-directional model M4b there is a more substantial drop in relative efficiency for the transition intensity estimators than there was for model M2b in Table 2; this is because observed transitions from state 1 to state 3 over successive observation times have two possible paths.

Results for prevalence probabilities  $P_{1r}(t)$  have been shown in Tables 2, 3, and 4; they are often of main interest. Relative efficiencies for other transition probabilities are more similar to those for transition intensities. For example, for model M4a the relative efficiencies for  $P_{23}(2)$  are 0.86, 0.70, 0.38 and 0.03 for  $\Delta t = 0.25, 0.50, 1.0, 2.0$  respectively.

## 4 MODEL ASSESSMENT

Intermittent observation also affects our ability to check model assumptions. Kalbfleisch and Lawless (1985) discussed tests of fit for Markov models when inter-observation times are equi-spaced, but the assessment of models is much more challenging when observation times vary across individuals. Titman and Sharples (2010a) give a comprehensive review in this case and the R software `msm` provides procedures for models with piecewise-constant intensities. The methods in `msm` involve the comparison of model-based and empirical counts or estimates related to prevalence probabilities, defined as  $P_r(t) = P(Y(t) = r)$ , and transition probabilities  $P_{rs}(t_1, t_2)$ . For specified values of  $t, t_1$  and  $t_2$ , ‘‘em-

Table 2: Ratios of asymptotic standard deviations for estimators of transition intensities and probabilities in three-state Markov models with (1)  $q_{12} = 1, q_{21} = 2, q_{23} = 1$  (M2a) and (2)  $q_{12} = 1, q_{21} = 0, q_{23} = 1$  (M2b)

	$\Delta t$	$q_{12}$	$q_{21}$	$q_{23}$	$P_{11}(1)$	$P_{11}(2)$	$P_{11}(4)$	$P_{13}(1)$	$P_{13}(2)$	$P_{13}(4)$
M2a	0.25	0.80	0.74	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	0.5	0.56	0.49	0.98	0.97	1.0	1.0	1.0	1.0	1.0
	1	0.27	0.23	0.90	0.85	0.99	1.0	0.91	1.0	1.0
	2	0.13	0.11	0.72	0.58	0.89	1.0	0.67	0.99	1.0
M2b	0.25	0.99	-	1.0	0.99	0.99	0.99	1.0	1.0	1.0
	0.5	0.99	-	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	1	0.97	-	0.96	0.97	0.97	0.97	0.99	0.99	0.99
	2	0.87	-	0.84	0.87	0.87	0.87	0.95	0.95	0.95

Values of  $P_{11}(t)$  and  $P_{13}(t)$  in M2a are  $P_{11}(1) = 0.608, P_{11}(2) = 0.462, P_{11}(4) = 0.270, P_{13}(1) = 0.178, P_{13}(2) = 0.369, P_{13}(4) = 0.631$ , and in M2b  $P_{11}(1) = 0.368, P_{11}(2) = 0.135, P_{11}(4) = 0.018, P_{13}(1) = 0.264, P_{13}(2) = 0.594, P_{13}(4) = 0.908$ . The total followup time is 4 years. A table entry of 1.0 denotes a relative efficiency of 0.995 or greater.

Table 3: Ratios of asymptotic standard deviations for estimators of transition intensities and covariate effects in three-state Markov models with (1)  $q_{12} = 1, q_{21} = 2, q_{23} = 1, \beta_{12} = -0.2, \beta_{21} = -0.5, \beta_{23} = -0.9$  (M3a) and (2)  $q_{12} = 1, q_{21} = 0, q_{23} = 1, \beta_{12} = -0.2, \beta_{21} = 0, \beta_{23} = -0.9$  (M3b)

	$\Delta t$	$q_{12}$	$q_{21}$	$q_{23}$	$\beta_{12}$	$\beta_{21}$	$\beta_{23}$
M3a	0.25	0.80	0.74	1.0	0.85	0.80	1.0
	0.5	0.56	0.49	0.98	0.66	0.59	1.0
	1	0.27	0.23	0.90	0.35	0.30	1.0
	2	0.13	0.11	0.72	0.15	0.14	0.90
M3b	0.25	0.99	-	1.0	1.0	-	1.0
	0.5	0.99	-	0.99	0.99	-	1.0
	1	0.97	-	0.96	0.98	-	1.0
	2	0.87	-	0.84	0.90	-	0.99

The total followup time is 4 years. A table entry of 1.0 denotes a relative efficiency of 0.995 or greater.



Table 4: Ratios of asymptotic standard deviations for estimators of transition intensities and probabilities in three-state Markov models with (1)  $q_{12} = 1, q_{21} = 2, q_{13} = 0.25, q_{23} = 0.5$  (M4a) and (2)  $q_{12} = 0.5, q_{21} = 0, q_{13} = 0.25, q_{23} = 0.5$  (M4b)

	$\Delta t$	$q_{12}$	$q_{21}$	$q_{13}$	$q_{23}$	$P_{11}(1)$	$P_{11}(2)$	$P_{11}(4)$	$P_{13}(1)$	$P_{13}(2)$	$P_{13}(4)$
M4a	0.25	0.75	0.74	0.83	0.80	0.99	1.0	1.0	1.0	1.0	1.0
	0.5	0.48	0.48	0.66	0.62	0.97	1.0	1.0	0.99	1.0	1.0
	1	0.14	0.15	0.40	0.37	0.80	0.96	1.0	0.93	1.0	1.0
	2	7e-03	8e-03	0.04	0.05	0.08	0.80	0.97	0.56	0.99	1.0
M4b	0.25	0.97	-	0.94	0.98	1.0	1.0	1.0	0.98	1.0	1.0
	0.5	0.93	-	0.88	0.95	1.0	1.0	1.0	0.96	1.0	1.0
	1	0.83	-	0.75	0.90	0.99	0.99	0.99	0.90	0.99	1.0
	2	0.60	-	0.51	0.76	0.93	0.93	0.93	0.70	0.90	1.0

Values of  $P_{11}(t)$  and  $P_{13}(t)$  in M4a are  $P_{11}(1) = 0.515, P_{11}(2) = 0.364, P_{11}(4) = 0.188, P_{13}(1) = 0.263, P_{13}(2) = 0.469, P_{13}(4) = 0.725$  and in M4b are  $P_{11}(1) = 0.472, P_{11}(2) = 0.223, P_{11}(4) = 0.045, P_{13}(1) = 0.259, P_{13}(2) = 0.487, P_{13}(4) = 0.779$ . The total followup time is 4 years. A table entry of 1.0 denotes a relative efficiency of 0.995 or greater.

pirical” estimates based on the raw data require imputation, since we usually do not know what state an individual is in at a given time. Pearson-type tests of fit (e.g. Titman 2009, 2010; Titman and Sharples 2010a) are also included in msm.

There are several difficulties with methods that compare model-based and empirical counts or estimates. First, nonparametric Markov estimates of prevalence based on continuous observation are robust for non-Markov processes, assuming that loss to followup is not related to process history (Aalen et al. 2001; Datta and Satten 2001). It is plausible that flexible parametric models fitted to intermittently observed data would also be fairly robust in the case where the observation times are independent of process history. We examined this in a small simulation study based on the three-state model in Figure 1(b): we assumed the true model was semi-Markov with transition intensities of Weibull form, but fitted piecewise-constant Markov models. The study is reported in the University of Waterloo PhD thesis of the second author; briefly, Markov models with several pieces gave accurate estimates of the true prevalence probabilities  $P_{1r}(0, t)$ . Because of this robustness, tests of fit based on prevalence can detect departures from assumed parametric forms for Markov transition intensities, but they may not be able to detect non-Markov behavior.

A second problem is the need for imputation with empirical prevalence estimates, and the fact this is not accounted for in formal tests. In addition, prevalence or transition counts involve mixtures of multinomial variables, and limiting distributions of test statistics depend on unknown parameter values. To investigate these matters and develop procedures for obtaining honest p-values is difficult. Parametric bootstrap methods are exceedingly time-consuming except in very simple cases. Titman (2009) has obtained the limiting distribution of Pearson-type statistics based on transition counts as a linear combination of independent, single degree of freedom chi-squared variables, but this is complicated to apply and its adequacy for p-value calculation in finite samples was checked only for three and four state pro-

gressive models. A third problem is that the empirical estimates of prevalence or transition probabilities used by current methods may be biased when the times between successive visits for an individual are related to their process history. In the case of continuously observed processes, inverse probability of censoring weighted (IPCW) estimation can be introduced for Markov model estimation (e.g. Datta and Satten 2002; Cook et al. 2009) or empirical prevalence estimation. This does not handle dependent intermittent observation times. Nonparametric estimates that adjust for dependent visit times can be developed, as long as they depend only on observed process history up to the preceding visit. We describe such estimates of prevalence and illustrate their use in Section 5. However, this affects the distributions of test statistics and it is not currently clear how to obtain reliable p-values with a moderate computing burden if we want formal tests of fit.

In view of these difficulties, we advocate formal model assessment within the Markov process family through likelihood ratio tests or local score tests. The latter have been considered by Stavola (1988); Kalbfleisch and Lawless (1989) and Gentleman et al. (1994); they have the advantage of not requiring the fitting of alternative models but for relative simplicity they require strong model assumptions. In the remainder of this section we consider likelihood ratio tests. These are easily implemented and are asymptotically equivalent to Pearson tests based on transition counts in cases where individuals are seen at a common set of visit times (Kalbfleisch and Lawless 1985; Titman 2009). Asymptotic chi-squared approximations for likelihood ratio statistics are reliable for obtaining p-values, provided samples are moderately large. We show below how the power of tests, and the effect of the gaps between visit times, can be assessed. The main remaining difficulty is in recognizing non-Markov behavior; simple nonparametric methods are unavailable and non-Markov models are difficult to fit in many settings involving intermittent observation, aside from certain simple uni-directional models (e.g. see Satten and Sternberg 1999). Titman and Sharples (2010b) have considered bi-directional semi-Markov models with phase-type state sojourn distributions, though they are unwieldy except in simple situations.

A MS model  $M_0$  can be tested against an expanded model  $M_1$  that includes it via a likelihood ratio statistic  $\Lambda = 2 \log L(\hat{\theta}_1) - 2 \log L(\hat{\theta}_0)$ , where  $\theta_0$  and  $\theta_1$  parameterize  $M_0$  and  $M_1$ , and  $L(\theta)$  is given by (3). Under the null model  $M_0$ ,  $\Lambda$  is asymptotically chi-squared under mild conditions, assuming that the observation times for individuals are independent of process history, conditional on observed history up to the immediately preceding observation time. In addition to considering the precision of estimation under specific models and observation plans, it is of interest when planning a study to consider the power for discriminating among different Markov models. We describe briefly how this may be done for likelihood ratio tests.

Under standard conditions,  $\Lambda$  is asymptotically  $\chi^2(d)$  under model  $M_0$ , where  $d$  is  $\dim(\theta_1) - \dim(\theta_0)$ . Under a sequence of alternative models  $M_1$  that approach the null model at rate  $1/\sqrt{n}$ ,  $\Lambda$  converges in distribution to non-central  $\chi^2(d, \delta)$  with degrees of freedom  $d$  and non-centrality parameter  $\delta$ . This can be used to approximate the power for a given alternative model by assuming that for a sample of  $n$  individuals, the likelihood ratio statistic  $\Lambda_n$  is approximately non-central  $\chi^2$  with non-centrality parameter  $\delta_n = n\gamma$ . We can estimate  $\gamma$  for a specific model  $M_1$  by using the fact that  $E(\Lambda_n) \doteq E(\chi^2(d, \delta_n)) = d + \delta_n$ . One way is to generate replicate samples with  $n$  individuals under  $M_1$ , and to obtain  $\Lambda$  for each, and then estimate  $E(\Lambda_n)$  by the mean  $\bar{\Lambda}_n$ . An alternative that involves less computation is to generate a single sample of large size  $N$  and obtain the observed value  $\Lambda_N$ . We can then estimate  $\gamma = \delta_N/N$  as  $\hat{\gamma} = (\Lambda_N - d)/N$ . Then,  $\delta_n$  for an arbitrary sample size is estimated by  $\hat{\delta}_n = n\hat{\gamma}$ .

Since the non-central chi-squared is for finite samples an approximation to the distribution of  $\Lambda$  based on asymptotic local power, our recommendation for study planning purposes when sample sizes are not very large is to use the procedure illustrated for preliminary estimates of power. Once approximate sample sizes and inter-visit times for a study have been identified, more accurate estimates (and confidence

intervals) for power can be obtained by simulating replicate samples.

## 5 APPLICATION: VIRAL LOAD VARIATION IN AN HIV COHORT

The Canadian observational cohort on HIV (CANOC) consists of HIV-positive individuals who initiated combination anti-retroviral therapy (cART) on or after January 1, 2000. Individuals in the cohort who achieved viral suppression (non-detectable presence of virus in the blood) were then seen at periodic visits that are nominally about 3 or 4 months apart, but can vary considerably between and within individuals (Raboud et al. 2010). At the visits, numerous measurements are taken, including biomarkers and clinical variables related to disease activity. The ability of patients to maintain a low viral load and avoid viral rebound is a key issue. Recent studies have examined factors related to viral rebounds (Grennan et al. 2012), and we address this here; the definition of viral rebound used is a viral load of over 1000 copies per mL. We examine variations in viral load and the occurrence of a first viral rebound by considering the multistate model shown in Figure 1(b), with states 1-3 defined as follows: state 1 is viral load (VL) 50 or less, state 2 is VL between 51 and 999, and state 3 is VL 1000 or more. State 3 is absorbing; when a VR occurs, treatment is typically adjusted so as to achieve viral suppression once again. It is assumed that the visit process is independent of the multistate process, conditional on observed data up to the previous visit time; this could be violated in some cases but is unavoidable without further assumptions that would be uncheckable.

There are systematic differences across the regions contributing to CANOC, and we focus here on the British Columbia cohort. Anti-retroviral therapy has evolved over time, and we consider the group of 1035 males who initiated cART in 2005 or later. Based on earlier work (Grennan et al. 2012), factors that will be considered as covariates include age at baseline, indication of injection drug use (IDU) at baseline, indication of men having sex with men (MSM) at baseline, and type of cART treatment (nucleoside reverse transcriptase inhibitor, NNRTI or protease inhibitor, PI). Table 5 summarizes features of the 1035 individuals and their followup visits; MSM or IDU status is missing for many individuals and we have chosen to treat missing (NA) as a separate category. We remark that there were 33 deaths among the 1035 individuals. Given the covariates considered, death is not significantly related to the viral load states, and for simplicity of presentation we do not include it in our model, except that the viral load process is terminated by death. If death were related to the viral load process, we could include it explicitly as a fourth state. We also note that for persons who did not die, end of followup terminated on February 28, 2010.

We fit Markov regression models with transition intensities  $q_{rs}(t|x) = q_{0rs}(t)exp(\beta_{rs}x)$ , for  $rs = 12, 21, 23$ . We consider piecewise-constant baseline intensities as in (2), and utilize the `msm` package in R (Jackson 2011). Table 6 shows estimates and standard errors for the baseline transition intensities and covariate effects for a two-piece model with a change in baseline intensities at  $t = 2$ ; the time units are years since initial viral suppression. The baseline intensities are for  $t < 2$ ; the Time regression coefficients represent the log ratios of the intensities for  $t \geq 2$  to those for  $t < 2$ . The covariate MSM is not included, and age is dichotomized, following initial fits and model assessment. Table 6 indicates that the  $2 \rightarrow 1$  transition intensity is much larger than the  $1 \rightarrow 2$  and  $2 \rightarrow 3$  intensities for all types of individuals. This is in part due to measurement error; a recorded measurement of less than 50 copies per mL usually means that no VL was detected, and measured viral loads over 50 may occur when no virus is actually present. In terms of covariate effects, older ages are associated with smaller  $2 \rightarrow 3$  transition intensity, and also with a smaller  $2 \rightarrow 1$  intensity. Injection drug users have a substantially higher intensity of transition to the viral rebound state than do non-users. Persons for whom IDU status is unknown are more similar to non-users. Those taking PI treatment have a higher  $1 \rightarrow 2$  intensity but a

Table 5: Characteristics of HIV cohort members

Age	18-45	> 45		MSM	No	Yes	NA
	576	459			318	292	425
cART	NNRTI	PI		IDU	No	Yes	NA
	402	633			323	287	425
Length of FU	$Q_{25}$	$Q_{50}$	$Q_{75}$	Gap times	$Q_{25}$	$Q_{50}$	$Q_{75}$
	0.986	1.863	2.973		0.112	0.189	0.255

FU stands for length of followup, in years; gap times are inter-visit times, in years;  $Q_p$  is the  $p$ th percentile. NA means not known.

smaller  $2 \rightarrow 3$  intensity. Finally, the  $1 \rightarrow 2$  transition intensity after two years is smaller than before two years; this may be a reflection of unobserved heterogeneity, with higher risk individuals tending to reach the VR state before two years. A likelihood ratio test of the piecewise constant model against the time-homogeneous model gives an observed log-likelihood difference of 28.4 on 3 degrees of freedom; most of the evidence for a difference is due to the  $1 \rightarrow 2$  intensity. Finally, the fitted model provides estimates of  $P_{13}(t|x)$ , the probability of a viral rebound by time  $t$ . The estimated probabilities vary considerably across covariate groups; for example, two-year probabilities for persons on NNRTI range from 0.11 for persons with age  $> 45$  and IDU = 0 to 0.31 for persons with age  $\leq 45$  and IDU = 1.

Table 6: The estimated regression coefficients and standard errors (in brackets) of piecewise constant Markov MS model

Transition	$\log(\hat{q}_{0rs})$	Age > 45 vs Age $\leq 45$	IDU=Yes vs No	IDU=NA vs No	cART=PI vs NNRTI	Time
1 $\rightarrow$ 2	-0.265 (0.214)	-0.085 (0.139)	0.003 (0.160)	0.112 (0.173)	0.426* (0.180)	-0.860* (0.180)
2 $\rightarrow$ 1	2.653 (0.242)	-0.338* (0.144)	-0.258 (0.177)	0.201 (0.171)	-0.068 (0.201)	-0.222 (0.173)
2 $\rightarrow$ 3	0.523 (0.257)	-0.555* (0.196)	0.808* (0.236)	-0.214 (0.264)	-0.781* (0.204)	0.137 (0.268)

\*  $|\frac{\hat{\beta}}{se(\hat{\beta})}| > 1.96$

The msm software package provides model checks based on transition counts and prevalence counts. The examination of state prevalence probabilities is widely used in assessing multistate models and in addition, the probability of being in state 3 is of special interest in the present context. We now discuss briefly the determination of empirical prevalence estimates that can be compared with model-based estimates, and consider the bias issue mentioned in Section 4. The msm software (function prevalence.msm, see also Titman and Sharples 2010a, Sect. 2.2) uses empirical estimates

$$\tilde{P}_{1r}(t_l) = \sum_{i=1}^n \delta_i(t_l) I(Y_i(t_l) = r) / \sum_{i=1}^n \delta_i(t_l), \quad (12)$$

for a specified set of times  $t_l$  ( $l = 1, \dots, m$ ), where  $C_i$  is a censoring time, assumed independent of the MS process, and  $\delta_i(t_l) = I(C_i \geq t_l)$  indicates whether a person is still under followup at  $t_l$ . The state

occupied at  $t_l$  is typically unknown when observation times are irregular, and then it is imputed according to the states occupied before and after  $t_l$ . In the following discussion, we assume the ‘‘mid-point’’ option is used; in this case the state is imputed when there are observation times before and after  $t_l$ , and the state is taken to be that occupied at the closest visit time. In addition, an individual who entered an absorbing state prior to  $t_l \leq C_i$  is taken to be in that state at  $t_l$ .

The estimator (12) is in general biased because of the imputation; the determination of bias is mathematically intractable but can be studied by simulation. If visit times  $t_{ij}$  are related to previous event history or covariates, then it seems likely that bias could be more severe. We consider briefly an alternative empirical approach that allows adjustment for process-related observation times; this will be developed and studied in more detail elsewhere. The approach employs so-called inverse intensity weighted (IIW) estimating functions introduced by Lin et al. (2004). Let the counting process  $\{N_i(t), t > 0\}$  represent visit times for person  $i$ , with  $N_i(t)$  equal to the number of times  $t_{ij} \leq t$ . Let  $H_i(t)$  represent the history of visits, the MS process and covariates up to time  $t^-$ , and  $\lambda_i(t|H(t))$  denote the visit intensity function. A crucial required assumption is that  $\lambda_i(t|H_i(t)) = \lambda_i(t|\tilde{H}_i(t))$ , where  $\tilde{H}_i(t)$  represents the observed history up to  $t^-$ . To estimate  $P_{1r}(t; \theta)$ , where  $\theta$  represents unknown parameters, we consider estimating function of the form (Lin et al. 2004)

$$U_w(\theta) = \sum_{i=1}^n \int_0^{\tau} \frac{g_i(t; \theta)}{\lambda_i(t|\tilde{H}_i(t))} \{Y_{ir}(t) - P_{1r}(t; \theta)\} dN_i(t), \quad (13)$$

where  $\tau$  is an upper limit on observation,  $Y_{ir}(t) = I(Y_i(t) = r)$  and  $g_i(t; \theta)$  is a specified function. It is readily seen that  $E\{U_w(\theta)\} = 0$  under the assumption that  $E\{Y_{ir}(t)\} = P_{1r}(t; \theta)$ . Moreover, if  $\lambda_i(t|\tilde{H}_i(t))$  is dropped from (13), the resulting estimating function does not have expectation zero in general and can be expected to produce biased parameter estimates. For model checking we would like nonparametric (empirical) prevalence estimates. If  $t_l$  ( $l = 1, \dots, m$ ) is actually the set of distinct observation times  $t_{ij}$  across individuals, then we can let  $\theta$  consist of elements  $\theta_l = P_{1r}(t_l)$ , and with  $g_i(t; \theta)$  in (13) an  $m \times 1$  unit vector, the estimating equation  $U_w(\theta) = 0$  gives estimates

$$\tilde{P}_{1r}(t_l) = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} I(t_{ij} = t_l) w_i(t_{ij}) Y_{ir}(t_{ij})}{\sum_{i=1}^n \sum_{j=1}^{m_i} I(t_{ij} = t_l) w_i(t_{ij})}, \quad (14)$$

where  $w_i(t_{ij}) = \lambda_i(t_{ij}|\tilde{H}_i(t_{ij}))^{-1}$  and the visit process is considered to be in discrete time. Since the visit intensity is unknown in general we have to estimate it, so that  $w_i(t_{ij})$  is replaced by an estimated value  $\hat{w}_i(t_{ij})$ . Consistency for the estimates  $\tilde{P}_{1r}(t_l)$  can be achieved if we have consistent estimation of the visit intensity; this is of theoretical interest but for practical purposes suggests that with a satisfactory model for the intensity the prevalence estimate should have little bias in large samples.

In many observational studies the number of distinct observation times  $t_{ij}$  is very large, and few (perhaps only one) times equal a given  $t_l$ . In this case the estimates (14) are highly discrete, so some type of smoothing is needed. Kernel methods are one option, and a simple approach that works quite well is to partition the time axis into  $m$  intervals  $I_l = (a_{l-1}, a_l]$  where  $a_0 = 0$  and  $a_m$  is an upper limit, and let  $t_l$  be the mid-point of  $I_l$ . We then use estimates

$$\tilde{P}_{1r}(t_l) = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} I(t_{ij} \in I_l) \hat{w}_i(t_{ij}) Y_{ir}(t_{ij})}{\sum_{i=1}^n \sum_{j=1}^{m_i} I(t_{ij} \in I_l) \hat{w}_i(t_{ij})} \quad (15)$$

where  $\hat{w}_i(t_{ij})^{-1}$  is the estimated conditional probability that  $t_{ij}$  is in  $I_l$ , given prior process and visit history, when  $Y_i(t_{ij})$  is a non-absorbing state. When an individual enters an absorbing state, it is assumed

that they are seen in all subsequent intervals up to  $C_i$ , and their component of the sum in the denominator of (15) is set equal to one for all subsequent  $I_l$ .

In the case where the visit process is completely independent of the MS process, we could use all weights equal to one in (15), and it and (12) then give alternative prevalence estimators. We reiterate that individuals must not be censored when they enter an absorbing state; rather, they should be retained up to their independent censoring times  $C_i$ . This must be done with all of the estimators (12), (14) and (15), otherwise state-dependent censoring is introduced. (We thank an anonymous referee for pointing this out for (12) and for noting that `prevalence.msm` has an option for doing this.) In the case of (14) and (15), we need either to model the visit intensity following entry to an absorbing state or to assume that after entry an individual is seen at all subsequent times  $t_l$  up to their censoring time, and to set the corresponding weights equal to one. In the CANOC illustration, the former is challenging because an individual's treatment is adjusted following a viral rebound, and the viral load process usually begins again in state 1. Thus we have adopted the second approach in describing (15).

We now return to the CANOC data, and our three-state model. The administrative censoring times  $C_i$  for each person are known, since followup in our data ended on February 28, 2010. The visit process and prevalence estimates take the end of followup to be the minimum of  $C_i$  and a person's time to death. We based weights  $\hat{w}_i(t_{ij})$  on a Cox model fitted to the gap times between successive observations. The key covariates in the model are age, injection drug use, type of cART, and state at the last visit. Older persons, injection drug use, and being in state 1 at the last visit are associated with longer times between visits. We chose intervals  $I_l$  using values  $a_l = 0, 0.25, 0.50, \dots, 4.0$ ; this gave a substantial number of observation times in each interval without being too severe a discretization of the visit process. Figure 2 shows marginal prevalence estimates for states 1, 2, and 3 obtained (i) using the fitted Markov multistate model ("Multistate Model Prevalence"), (ii) using the estimates (15) ("IIW Prevalence"), and (iii) using the R `msm.prevalence` function ("Empirical-msm Prevalence"). We also computed estimates using (15) with all weights set equal to one; they were very close to the estimates given by (12). We see from Figure 2 that the empirical msm estimates and the IIW estimates are both close to the Markov model estimates, although they lie on opposite sides of the Markov estimates for states 1 and 3. The fact that the msm and IIW estimates are similar likely reflects the fact that although visit times are significantly related to the previous state and to covariates, the estimated weights do not vary a great deal, with most between 1.0 and 1.5. Simulations conducted to assess the variability of the estimators, which will be reported elsewhere, do not suggest lack of fit.

Finally, we provide a brief assessment of the effect of inter-visit times on estimation efficiency and on the power to detect non-homogeneity of the transition intensities. Individuals are nominally seen every three months, and we first examine the effect of visits every 6 months versus every 3 months. Many individuals do not adhere closely to the scheduled visits, but this will indicate the approximate effect if median inter-visit times were closer to 6 months than to three. We consider the estimated model in Table 6 as the true process, except that for simplicity we assume there is no change in transition intensities after two years. The total followup time for each individual was adjusted to the nearest quarter-year value. We used the procedures in Section 2.2 to get the asymptotic covariance matrix for baseline intensities and covariate effects, and Table 7 shows the results. The standard errors for transition intensities and covariate effects are substantially larger when  $\Delta t = 0.5$  years than when it is 0.25 years. The median inter-visit time in the actual data (see Table 5) is a little over 2 months, and the standard errors in Table 6 are a little smaller than those in Table 7 for  $\Delta t = 0.25$  in most cases. Estimates of baseline prevalence probabilities are essentially the same in each case, but estimates of other transition probabilities such as  $P_{23}(1)$  are much less precise when  $\Delta t = 0.5$ . Aside from issues of patient care, insights into viral load dynamics would be reduced if a 6 month visit schedule were used.

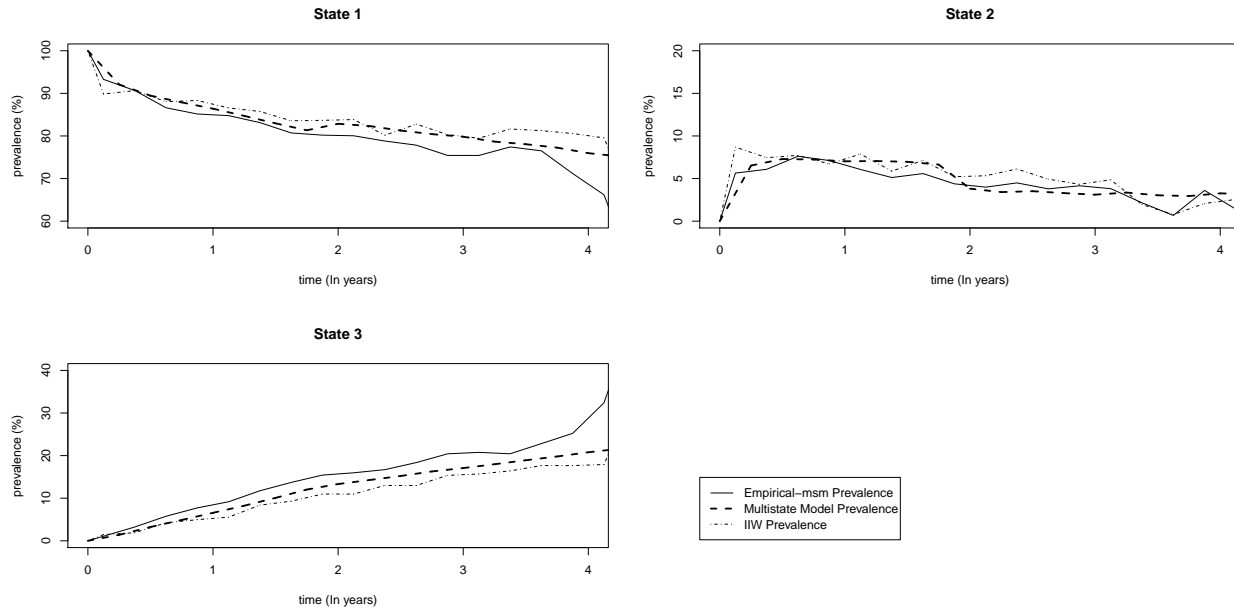


Figure 2: Comparison of MSM, empirical-msm, and IIW prevalences across all individuals.

We also considered the power of the likelihood ratio test to detect non-homogeneity under the two observation schemes. To reduce computation time we rounded up the total followup time for each person to the nearest year. We took the fitted model in Table 6 as the alternative model  $M_1$  and generated sample paths for 10,000 individuals under  $M_1$  and the distribution of total followup times. We then estimated the power of likelihood ratio tests of the time homogeneous null model  $M_0$  against  $M_1$ , when inter-visit times were  $\Delta t = 0.25$  and  $0.5$  years. With  $n = 1035$  (the sample size for the cohort) the power for a test of size  $0.05$  was estimated as described at the end of Section 4 to be  $0.91$  and  $0.90$  when  $\Delta t$  was  $0.25$  and  $0.5$  respectively. If visits are one year apart, the power is reduced to  $0.66$ .

## 6 CONCLUDING REMARKS

Transition intensities in a bi-directional Markov model cannot be estimated precisely when the observation times at which state occupancy is observed are too far apart. Therefore, when followup studies are planned the relative importance of intensity estimation should be considered; the methods in this paper can be used to assess precision and power to detect lack of fit, given a proposed set of followup times. In observational cohorts where times between visits vary widely, the possible dependence of visit times on previous observed state history should also be assessed; such dependence can bias some model assessment procedures. We have introduced a form of inverse intensity weighted estimation for state prevalence functions in Section 5; these procedures are analogous to, but distinct from, inverse probability of censoring (IPC) weighted estimation that has been used to deal with state-dependent censoring when individuals are observed continuously over time (e.g. see Cook et al. 2009 and references therein). A more thorough development and investigation of the IIW estimators will be given elsewhere, along with a comparison with unweighted estimators.

There are other issues, not addressed here, that deserve study. One is the assessment of models with time-varying covariates. A second is the potential presence of unobserved heterogeneity in the processes

Table 7: Standard errors for estimators of baseline transition intensities, covariate effects, and transition probabilities if all gap times were 3 months ( $\Delta t = 0.25$ ), or 6 months ( $\Delta t = 0.5$ ).

$\Delta t$	$\log(q_{012})$	$\log(q_{021})$	$\log(q_{023})$	$\beta_{12.age}$	$\beta_{21.age}$	$\beta_{23.age}$
0.25	0.248	0.281	0.210	0.185	0.205	0.171
0.5	0.548	0.623	0.226	0.430	0.479	0.179
$\Delta t$	$\beta_{12.cART}$	$\beta_{21.cART}$	$\beta_{23.cART}$	$\beta_{12.IDU1}$	$\beta_{21.IDU1}$	$\beta_{23.IDU1}$
0.25	0.208	0.236	0.174	0.199	0.224	0.199
0.5	0.442	0.505	0.189	0.507	0.578	0.211
$\Delta t$	$\beta_{12.IDU2}$	$\beta_{21.IDU2}$	$\beta_{23.IDU2}$	$P_{11}(1)$	$P_{11}(2)$	$P_{13}(1)$
0.25	0.286	0.302	0.231	0.012	0.021	0.011
0.5	0.709	0.758	0.238	0.013	0.020	0.011
$\Delta t$	$P_{13}(2)$	$P_{23}(1)$				
0.25	0.021	0.024				
0.5	0.020	0.037				

of individuals, and the effect of inter-visit times on its detection. A third issue concerns settings where certain transition times can be observed; this occurs, in particular, when death is one of the states. The broad conclusions here should still apply for transitions whose exact times cannot be observed, but some investigation would be useful. Finally, various authors have noted that models for which the covariates affect the transition intensities in simple ways (such as the multiplicative models used in this paper) do not provide simple interpretations of covariate effects on transition or prevalence probabilities. Andersen and Klein (2007) consider “pseudo-observation” methods for prevalence probabilities; we note that another direct method would be to use marginal models for prevalence functions in conjunction with estimating functions like (13), and a comparison of approaches would be worthwhile.

## ACKNOWLEDGEMENTS

This research was supported in part by a grant to J.F. Lawless from the Natural Sciences and Engineering Research Council of Canada. The authors thank Dr. Janet Raboud and the CANOC project, which was supported by funding from Emerging Team Grant 53444 (Dr. Robert Hogg, PI) from the Canadian Institutes of Health Research, for the HIV viral load data and for valuable discussion. They also thank Ker-ai Lee for assistance in managing the data, and Richard Cook and Brian Tom for helpful comments. Finally, the authors are grateful for comments by two anonymous reviewers which led to substantial improvements in the paper.

## REFERENCES

Aalen, O.O., Borgan, O., Fekjaer, H. (2001). Covariate adjustment of event histories estimated from Markov chains: the additive approach. *Biometrics*, **57**: 993-1001.



- Aalen, O.O., Borgan, O., Gjessing, H.K. (2008). *Survival and Event History Analysis: A Process Point of View*. New York: Springer Science + Business Media, LLC.
- Andersen, P.K., Borgan, O., Gill, RD., Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.
- Anderson, P.K., Klein, J.P. (2007). Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scandinavian Journal of Statistics*, **34**: 3-16.
- Cook, R.J. (2000). Information and efficiency considerations in planning studies based on two-state Markov process. *Journal of Statistical Research*, **34**: 161-178.
- Cook, R.J., Lawless, J.F. (2007). *Statistical Analysis of Recurrent Events*. New York: Springer Science+ Business Media, LLC.
- Cook, R.J., Lawless, J.F. (2014). Statistical issues in modeling chronic disease in cohort studies. *Statistics in Biosciences*, **6**: 127-161.
- Cook, R.J., Lawless, J.F., Lakhal-Chaieb, L., Lee, K.A. (2009). Robust estimation of mean functions and treatment effects for recurrent events under event-dependent censoring and termination: application to skeletal complications in cancer metastatic to bone. *Journal of the American Statistical Association*, **104**: 60-75.
- Datta, S., Satten, G.A. (2001). Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson-Aalen estimators of integrated transition hazards for non-Markov models. *Statistics and Probability Letters*, **55**: 403-411.
- Datta, S., Satten, G.A. (2002). Estimation of integrated transition hazards and stage occupancy probabilities for non-Markov systems under dependent censoring. *Biometrics*, **58**: 792-802.
- de Stavola, B.L. (1988). Testing departures from time homogeneity in multistate Markov processes. *Applied Statistics*, **37**: 242-250.
- Gentleman, R.C., Lawless, J.F., Lindsey, J.C., Yan, P. (1994). Multi-state Markov models for analysing incomplete disease history data with illustrations for HIV disease. *Statistics in Medicine*, **13**: 805-821.
- Grennan, J.T. et al. (2012). Magnitude of virologic blips is associated with a higher risk for virologic rebound in HIV-infected individuals: a recurrent events analysis. *Journal of Infectious Diseases*, **205**: 1230-1238.
- Hwang, W., Brookmeyer, R. (2003). Design of panel studies for disease progression with multiple stages. *Lifetime Data Analysis*, **9**: 261-274.
- Jackson, C.H. (2011). Multi-state models for panel data: the msm package for R. *Journal of Statistical Software*, **38** (8): 1-28.
- Kalbfleisch, J.D., 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, Lawless, J.F. (1989). Some statistical methods for panel life history data. *Proceedings of the Statistics Canada Symposium on the Analysis of Data in Time*, Ottawa, Statistics Canada: 185-192.
- Klein, J.P., Shu, Y. (2002). Multi-state models for bone transplantation studies. *Statistical Methods in Medical Research*, **11**: 117-139.
- Kvist, K., Andersen, P.K., Angst, J., and Kessing, L.V. (2010). Event dependent sampling of recurrent events. *Lifetime Data Analysis*, **16**, 580-598.
- Lin, H., Scharfstein, D.O., Rosenheck, R.A. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society, Series B*, **66**: 791-813.
- Mandel, M., Betensky, R.A. (2008). Estimating time-to-event from longitudinal ordinal data using random effects Markov models: application to multiple sclerosis progression. *Biostatistics*, **9**: 750-764.
- Mehtala, J., Auranen, K., Kulathinal, S. (2011). Optimal designs for epidemiologic longitudinal studies with binary outcomes. *Statistical methods in medical research*, DOI: 10.1177/0962280211430663.
- Moler, C., Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, **45**: 3-49.
- O’Keeffe, A.G., Tom, B.D.M., Farewell, V.T. (2011). A case-study in the clinical epidemiology of Psoriatic arthritis: Multistate models and causal arguments. *Journal of the Royal Statistical Society, Series C*, **60**: 675-699.
- Raboud, J., Loutfy, M.R., Su, D., Bayoumi, A.M., Klein, M.B., Cooper, C., Machouf, N., Rouke, S., Walmsley, S., Rachlis, A., Harrigan, R., Smieja, M., Tsoukas, C., Montaner, J.S., Hogg, R.S. (2010). Regional differences in rates of HIV-1 viral load monitoring in Canada: Insights and implications for antiretroviral care in high income countries. *BMC Infectious Diseases*, **10**: 40
- Satten, G.A., Sternberg, M.R. (1999). Fitting semi-Markov models to interval-censored data with unknown initiation times. *Biometrics*, **55**: 507-513.
- Titman, A.C. (2009). Computation of the asymptotic null distribution of goodness-of-fit tests for multi-state models. *Lifetime Data Analysis*, **15**: 519-533.
- Titman, A.C., Sharples, L.D. (2010a). Model diagnostics for multi-state models. *Statistical Methods in Medical Research*, **19**: 621-651.
- Titman, A.C., Sharples, L.D. (2010b). Semi-Markov models with phase-type sojourn distributions. *Biometrics*, **66**: 742-752.
- Titman, A.C. (2011). Flexible nonhomogeneous Markov models for panel observed data. *Biometrics*, **67**: 780-787.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**: 1-26.