Biometrics 66, 742–752
September 2010
DOI: 10.1111/j.1541-0420.2009.01339.x

Semi-Markov Models with Phase-Type Sojourn Distributions

Andrew C. Titman

Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, U.K.
email: a.titman@lancaster.ac.uk

and

Linda D. Sharples

MRC Biostatistics Unit, University Forvie Site, Robinson Way, Cambridge, CB2 0SR, U.K.
email: linda.sharples@mrc-bsu.cam.ac.uk

Summary. Continuous-time multistate models are widely used for categorical response data, particularly in the modeling of chronic diseases. However, inference is difficult when the process is only observed at discrete time points, with no information about the times or types of events between observation times, unless a Markov assumption is made. This assumption can be limiting as rates of transition between disease states might instead depend on the time since entry into the current state. Such a formulation results in a semi-Markov model. We show that the computational problems associated with fitting semi-Markov models to panel-observed data can be alleviated by considering a class of semi-Markov models with phase-type sojourn distributions. This allows methods for hidden Markov models to be applied. In addition, extensions to models where observed states are subject to classification error are given. The methodology is demonstrated on a dataset relating to development of bronchiolitis obliterans syndrome in post-lung-transplantation patients.

Key words: Bronchiolitis obliterans syndrome; Hidden Markov model; Multistate model; Panel observation; Phase-type distribution; Semi-Markov model.

1. Introduction

Continuous-time multistate models are widely used for categorical response data in the context of the natural history of chronic diseases. However, inference is difficult when the process is only observed at discrete time points, with no information about the times or types of events between observation times, unless a Markov assumption is made. Often it may be the case that the transition intensities of the process depend on time since entry into a state making the process *semi-Markov*.

If the observation scheme is sufficiently frequent to assume with validity that, although transition times are interval censored, all transitions are observed, then the likelihood for a semi-Markov model can be expressed with relative ease (Foucher et al., 2007). However, if the process is panel observed, such that multiple transitions may have occurred between observations, the likelihood is less tractable. In the case of progressive models, where a subject cannot reenter a state once it has been exited, the likelihood is expressible as a multidimensional integral. This integral is generally analytically intractable, but for progressive models with a small number of states, e.g., three or four, computation of the integral via numerical quadrature is feasible (Foucher et al., 2008). In the case of progressive models, where there is only one possible path of transitions, nonparametric estimation is possible via self-consistent estimators if a discrete-time assumption is made (Satten and Sternberg, 1999). Progressive models can also be fitted semiparametrically using penalized likelihood (Joly and Commenges, 1999).

Computation of the likelihood is more difficult for models that allow reverse transitions. For evenly spaced observations a minimum chi-square estimation approach has been used for a two-state recurrent model (Rosychuk and Thompson, 2001), avoiding computation of the likelihood. Stopping-time resampling (Chen, Xie, and Liu, 2005) has been proposed as a potential simulation based method of computation.

Kang and Lagakos (2007) noted that inference for panel observed semi-Markov models is significantly easier if at least one state in the model is Markov, i.e., it has an exponential sojourn time distribution. The likelihood for an individual in this case can then be factorized into sojourn times of departure from the Markov state. The components of the likelihood require solution of integral equations. These can be expressed as a sum over the number of unobserved transitions between states that occurred between observations. If a minimum staying time, or *guarantee time*, is assumed in the semi-Markov states then the number of possible unobserved transitions becomes finite.

Crespi, Cumberland, and Blower (2005) analyzed panel observed data from a two state recurrent disease process. They assumed the existence of a latent process that had the form of a time homogeneous birth–death process, with state space

$\{0, 1, 2, \ldots\}$, a subject was considered to be disease free if in state 0 and ill otherwise. Sojourns in the observable illness state are not therefore exponential and the observable process is semi-Markov. However, the latent Markov structure of the model allows the likelihood to be expressed as a hidden Markov model (HMM), for which computation is relatively straightforward. This method of analyzing non-Markov processes by the inclusion of latent states has a long history (Cox, 1955a).

In this article, a general approach to fitting semi-Markov models to panel-observed data using phase-type sojourn time distributions is developed, both as a means of appraising the fit of Markov models and a means of modeling in itself. In addition, the approach is extended to data where the observed states are subject to error.

The remainder of the article is structured as follows. Section 2 introduces notation, Sections 3 and 4 describe the general method and the extension to data with misclassified states. Sections 5 and 6 discuss issues relating to identifiability and methods for dealing with unknown initiation times. Section 7 applies the method to data from posttransplant patients. Finally the article concludes with a discussion. A modified likelihood ratio test for model assessment is provided in the Appendix.

2. Notation for General Markov and Hidden Markov Models

Throughout this section, to simplify the notation we describe the contribution to the likelihood for one subject. For a given individual data are realizations, $x_0, \ldots, x_n$, at times $t_0 = 0, t_1, \ldots, t_n$ of a continuous time, discrete state stochastic process $X(t)$, where the observation times t_i may both be irregularly spaced and differ between subjects. Throughout we assume that the sampling mechanism is noninformative for the underlying process (Gruger, Kay, and Schumacher, 1991), so that the sampling can be ignored in the construction of the likelihood.

If $X(t)$ is a continuous time Markov process, then it is governed by transition intensity matrix $Q(t)$ with (r, s) entry given by

$$\mu_{rs}(t) = \lim_{\delta t \downarrow 0} \frac{\mathbb{P}\{X(t + \delta t) = s \mid X(t) = r\}}{\delta t},$$

the instantaneous probability of making a transition between state r and s at time t, where t is measured relative to the origin of the process. Note that this does not depend on the history of the process due to the Markov assumption.

The transition probabilities, defined as

$$p_{rs}(u, t + u) = \mathbb{P}\{X(t + u) = s \mid X(u) = r\},$$

can be found by solving the Kolmogorov forward equations (Cox and Miller, 1965) and are usually complicated expressions of the transition intensities.

Thus for a Markov process the likelihood contribution for an individual with observations $x_0, \ldots, x_n$, at times $t_0 = 0, t_1, \ldots, t_n$ can be written as a product of transition probabilities

$$L = \prod_{i=1}^{n} p_{x_{i-1}x_i}(t_{i-1}, t_i).$$

Kalbfleisch and Lawless (1985) developed a Fisher scoring algorithm for computing the maximum likelihood estimate of the model parameters. For the special case of a time-homogeneous process the transition intensities are independent of time, $\mu_{rs}(t) = \mu_{rs}$, and the transition probabilities depend only on the time interval between successive observations.

In many clinical studies, the x_i may be measurements of a biomarker or screening test, which are subject to measurement error so that there is a nonzero probability that the state is misclassified. Instead of observing the x_i directly we observe $o_1, \ldots, o_n$. If it is assumed that the misclassification probabilities are given by $\mathbb{P}\{O(t) = s \mid X(t) = r\} = e_{rs}$, where e_{rs} remains constant through time, that $X(t)$ is a Markov process and conditional on the true underlying states, the observed states are independent, then the o_i are modeled by a HMM (Satten and Longini, 1996). In the presence of misclassification the contribution to the likelihood of each transition depends on the complete history of the process for that individual. For each individual, matrices $\mathbf{M_1}, \ldots, \mathbf{M_n}$ can be constructed, where $\mathbf{M_i}$ is an $R \times R$ matrix with (r, s) entry $e_{s,o_i} p_{rs}(t_{i-1}, t_i)$ with $t_0 = 0$; then the likelihood contribution for an individual can be written as a matrix product

$$L = \pi \mathbf{M_1} \mathbf{M_2} \ldots \mathbf{M_n} \mathbf{1}, \tag{1}$$

where π is the vector of initial state probabilities and $\mathbf{1}$ is a vector of ones of length R. This matrix representation is equivalent to the basis of the recursive forward algorithm for computing the likelihood of a HMM (Satten and Longini, 1996). The matrix of misclassification probabilities, with entries e_{rs}, may be known from previous studies but more usually is unknown. It can be estimated from the data simultaneously with the parameters governing $X(t)$, although further constraints may be needed to ensure the model is identifiable (see Section 5).

There are several approaches to computation and maximization of the likelihood for continuous time HMMs. One approach is to apply a continuous-time generalization of the forward–backward algorithm for discrete-time HMMs (Bureau, Shiboski, and Hughes, 2003). Alternatively, the likelihood can be computed directly and maximized using derivative free optimization techniques such as those provided in the `R` package `msm` (Jackson et al., 2003). These methods are available for general models where the underlying process is Markov and where there is conditional independence between the observed states. However, in practice additional assumptions may be necessary. For example, when using the software `msm` it is necessary to assume time homogeneity or piecewise time homogeneity or homogeneity conditional on covariates that may change over time in a piecewise constant manner. Additionally, there may be problems with identifiability due to the presence of both transition intensities and misclassification probabilities in the model, so that it is necessary to constrain some parameters, for example by putting an upper bound on misclassification probabilities. See Section 5 for further discussion of estimation and identifiability.

In some applications π, the vector of initial state probabilities may be unknown and can instead be included as unknown parameters (Bureau et al., 2003). Alternatively, in recurrent models, for tractability it is sometimes assumed that

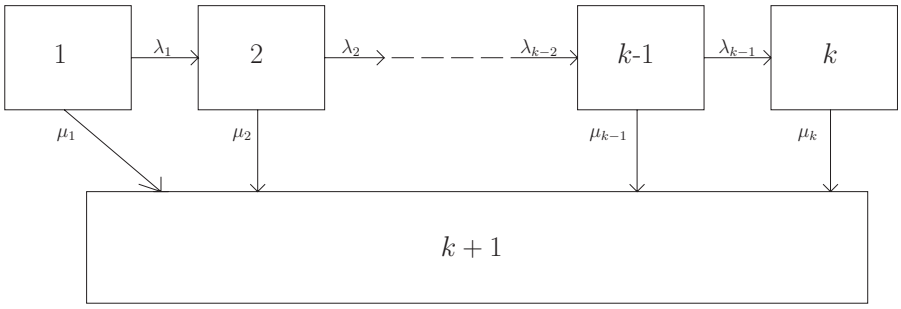

Figure 1. General Coxian phase-type distribution with k phases.

the Markov process is in equilibrium (Rosychuk and Thompson, 2003).

Covariates affecting the transition intensities can be modeled by $\mu_{rs}(t;\mathbf{z}) = \mu_{rs}(t)\exp(\beta_{rs}^T\mathbf{z})$, where $\mathbf{z}(\mathbf{t})$ is a vector of explanatory variables. Covariate effects may also be incorporated into the matrix of misclassification probabilities by assuming linearity on a *logit* scale $\text{logit}(e_{rs}) = \alpha_{rs}^T\mathbf{z}$.

3. Coxian Phase-Type Models

Here we begin by describing a Coxian phase-type distribution (Cox, 1955b) applied to a simple two state $(alive, dead)$ survival model, and then demonstrate how it can be applied to the sojourn distribution of each transient state of a general, multistate, semi-Markov model.

3.1 *Phase-Type Survival Model*

Consider a two state survival model $X(t)$ with states $\{1 = alive, 2 = dead\}$, for which the transition intensity from alive to dead is time inhomogeneous. For a Coxian phase-type model the sojourn time in the transient state is assumed to be governed by a latent Markov process $X^*(t)$ with k transient phases and one absorbing phase $k+1 (= dead)$. The latent process is progressive so that movement from transient phase $j \in \{1,\ldots,k\}$ is only possible to the absorbing state $k+1$ or to the adjacent phase $j+1$ (Figure 1). At time zero, the process is in phase 1. The parameters of this distribution are $(\lambda_1,\ldots,\lambda_{k-1})$, the transition intensities between transient phases, and $(\mu_1,\ldots,\mu_k)$, the transition intensities to the absorbing state. These parameters are constant with respect to time, but intensities vary between phases, which induces time inhomogeneity in the movement between the observable states $alive \rightarrow dead$. This formulation fits naturally into the HMM framework described in Section 2, so that derivation of the likelihood and implementation follows in an analogous way.

For many applications, a two-phase distribution may provide sufficient flexibility to describe the evolution of the process over time. Such a model allows transition intensities to be monotonically increasing (if $\mu_2 > \mu_1$), decreasing (if $\mu_2 < \mu_1$), or constant (if $\mu_2 = \mu_1$) with respect to time since entry into the state. In addition, the distribution defines the initial transition intensity on entry into the state and the limiting transition intensity, i.e., the hazard of transition as the time in the state tends to infinity. If $\mu_1 + \lambda_1 > \mu_2$ then the limiting intensity is μ_2 and otherwise it is given by some value between μ_1 and μ_2.

3.2 *Phase-Type Semi-Markov Model*

Consider a semi-Markov process $X(t)$ with state space $\mathcal{S} = \{1,\ldots,R\}$, where R is an absorbing state, and t represents time from entry into the initial state. To maintain clarity, we will assume there is exactly one absorbing state. However, the extension to the case of no absorbing state or multiple absorbing states is straightforward. In the setting of a multistate model the sojourn distribution of each nonabsorbing state r of $X(t)$ is assumed to be a k-phase Coxian phase-type distribution, with parameters $\lambda_{r_1},\ldots,\lambda_{r_{k-1}}$, the intensities for movement between phases of state r and $\mu_{r_1 s},\ldots,\mu_{r_k s}$, the intensities for movement out of state r to state s. For each of the observable states $r \in \mathcal{S}$ we assume there exists a latent process $X^*(t)$ with states $r_1,\ldots,r_k$ for which we observe only that the subject is in state r. It is also assumed that an individual enters transient state r in phase r_1, and passes through consecutive phases until the state is exited, which can occur from any phase.

To describe the process $X(t)$ in terms of the latent process $X^*(t)$ we let $\mathcal{S}^*$ consist of state phases $r_1,\ldots,r_k$ for each $r = 1,\ldots,R-1$, and state R, such that

$$\mathcal{S}^* = \{1_1, 1_2, \ldots, 1_k\} \cup \{2_1, 2_2, \ldots, 2_k\} \cup \ldots \cup \{(R-1)_1, \ldots, (R-1)_k\} \cup R,$$

so that it has dimension $\{k(R-1)+1\}$.

Note that it is not necessary for each observable state to have the same number of latent states. Indeed, in some cases, it might be appropriate to allow greater flexibility in states for which there is more information in the data.

Let $\mu_{r_1 s},\ldots,\mu_{r_k s}$ denote transition intensities from observable states r (from phases 1 to k respectively) to s, where

$$\mu_{r_j s} = \lim_{\delta t \downarrow 0} \frac{\mathbb{P}\{X^*(t+\delta t) = s_1 \mid X^*(t) = r_j\}}{\delta t}.$$

A general phase-type semi-Markov model would allow the competing transition intensities out of observable state r to vary independently. However, to allow greater estimability of the parameters, we make restrictions on this variation. We constrain $\mu_{r_j s} = \tau_{r_j}\mu_{r_1 s}$ for all s, where $\tau_{r_1},\ldots,\tau_{r_k}$ denote "phase effects" for state r and $\tau_{r_1} \equiv 1$. Hence in phase j, all transition intensities out of state r are τ_{r_j} times their respective values in phase 1. This constraint ensures that the probability of making an $r \rightarrow s$ transition, given that a transition out of r occurs, is constant through time (provided the $\mu_{r_j s}$ do not depend on time) and that the resulting sojourn

distributions are Coxian. In addition, the τ give a convenient parameterization, with $\tau = 1$ representing no time dependency with respect to entry into each state. The transition intensities for $X^*(t)$ are then as follows. The transition intensity between state phase r_j and s_1 for $j = 1, \ldots, k$ and $r = 1, \ldots, R-1,\ s = 1, \ldots, R-1,\ r \neq s$ is given by $\tau_{r_j}\mu_{r_1 s}$. Similarly the intensity between r_j and R is given by $\tau_{r_j}\mu_{r_1 R}$. In addition, for $r = 1, \ldots, R-1$, transitions from phase state r_j to r_{j+1} for $j = 1, \ldots, k-1$ have a transition intensity given by λ_{r_j}. All other transition intensities are zero. $X(t)$ relates to $X^*(t)$ in the following way. If the latent process at time t is $X^*(t) \in \{r_1, \ldots, r_k\}$ then the observable process is $X(t) = r$ for $r = 1, \ldots, R$.

The likelihood for such data can be calculated by directly applying methods for HMMs. In particular, the likelihood can be expressed as a matrix product as in equation (1), where for an individual the matrix $\mathbf{M_i}$ is now $\{k(R-1)+1\} \times \{k(R-1)+1\}$ with (r, s) entry $e_{s,x_i} p_{rs}(t_{(i-1)}, t_i)$, where $e_{s,x_i} = \mathbb{P}\{X(t) = x_i \mid X^*(t) = s\}$ takes the value 1 if s is a phase of the observed state x_i and 0 otherwise, for $s \in \mathcal{S}^*,\ i = 1, \ldots, n$. As the likelihood takes the form of a HMM, existing methods, such as derivative free numerical optimization, can be applied to compute the maximum likelihood estimate. Provided the maximum likelihood estimate lies within the interior of the parameter space and the dataset is sufficiently large, approximate standard error estimates can be found by inverting the Hessian matrix at the optimum. Where information is sparse bootstrapping may be required to obtain more accurate variance estimates.

4. Extension to Incorporate Misclassification Error (Hidden Semi-Markov Models)

If the states in the process are observed with misclassification error, the process is a hidden semi-Markov model (HSMM). Suppose the hidden semi-Markov process has observed states $O(t)$ related to the states of the underlying semi-Markov process $X(t)$ by misclassification probability matrix e. In the case of a general semi-Markov process $X(t)$, to calculate the likelihood contribution of a patient, it is necessary to sum over all possible sequences of true states that are compatible with the (misclassified) observed states $O(t)$. If reverse transitions are permitted or the sequence of observed states is long, the number of possible sequences is large and the calculations become unmanageable unless a recursive formula for the likelihood can be formulated.

A further advantage of the use of a phase-type semi-Markov process is that the extension to a HSMM is relatively straightforward. Suppose the hidden semi-Markov process has observed states $O(t)$ related to the states of the underlying semi-Markov process $X(t)$ by misclassification probability matrix e, such that $e_{rs} = P\{O(t) = s \mid X(t) = r\}$. As before, if we allow each state in $X(t)$ to have a phase-type distribution, we can introduce a further latent process, $X^*(t)$, such that if $X^*(t) \in \{r_1, \ldots, r_k\}$ then $X(t) = r$ for $r = 1, \ldots, R$. The stochastic relationship between $O(t)$ and $X(t)$ defined through e then means that we may express $O(t)$ as a hidden Markov process, with a $\{k(R-1)+1\} \times R$ misclassification probability matrix e^*, where

$$e^*_{r_j s} = e_{rs} \quad \text{for} \quad r, s = 1, \ldots, R \quad \text{and} \quad j = 1, \ldots, k,$$

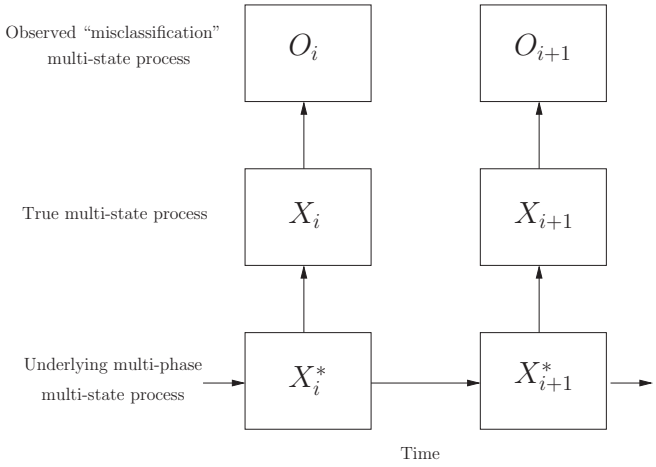

Figure 2. HSMM: The latent Markov process, $X^*(t)$, defines $X(t)$ deterministically. $O(t) \mid X(t)$ is then multinomial.

so that $P\{O(t) = s \mid X^*(t) = r_k\} = e_{rs}$ independent of k. Figure 2 illustrates this relationship.

This latent HMM structure ensures that the recursive forward algorithm can be applied and the likelihood contribution from an individual can be calculated as a matrix product,

$$L = \pi \mathbf{M_1^*} \mathbf{M_2^*} \ldots \mathbf{M_n^*} \mathbf{1},$$

where M_i^* is a $\{k(R-1)+1\} \times \{k(R-1)+1\}$ matrix with (r^*, s^*) entry $e^*_{s^* o_i} p_{r^* s^*}(t_{i-1}, t_i)$, for $r^*, s^* \in \mathcal{S}^*$.

Therefore, the only difference between the formulation of the likelihood in the hidden semi-Markov case compared to the semi-Markov case is that rather than being 0 or 1, the e_{rs} may lie between 0 and 1 and may be treated as unknown parameters.

5. Estimation and Identifiability

The incomplete nature of the observation scheme and the potential complexity of the proposed models mean that issues of parameter identifiability and estimability may arise. Precise characterization of when a model is identifiable for particular data is not straightforward. However, it is clear that we require the number of observations per patient to be sufficiently large so that the number of possible sequences of states exceeds the number of parameters to be estimated in the model.

The presence of misclassification increases the possibility of identifiability problems. Relabeling of the underlying states in HMMs can often produce the same likelihood function (MacDonald and Zucchini, 1997). It is also known that for the HMM with two states and balanced observation times, it is not possible to simultaneously identify the misclassification probabilities and the transition intensities (or probabilities) without making some additional constraints, for instance that the misclassification probabilities are each less than 0.5 (Rosychuk and Thompson, 2003).

As ensuring identifiability is not straightforward, inspection of the likelihood function is required. This might involve evaluation of the Hessian matrix to ensure the estimated parameters are at a maximum and performing optimization procedures from a wide range of starting values to ensure a global

maximum has been attained. In cases where identifiability seems a problem, it would be necessary to reduce the complexity of the model. Possible constraints would be to fix some of the λ_{r_j} to be known constants, or constrain some of the sojourn distributions to be exponential. Alternatively, one could consider performing the optimization by first maximizing the profile likelihood on a grid of possible fixed values of the λ_{r_j}.

In addition to potential general identifiability problems, there are obvious problems of identifiability in the phase-type model when the sojourn time distribution in one or more of the states is exactly exponential. The sojourn time distribution for state r will be exponential if $\tau_{r_j} = 1$ for all $j = 1, \ldots, k$, meaning transition intensities out of the state are independent of the state phase occupied. The parameters, λ_{r_j}, $j = 1, \ldots, k-1$, are redundant and unidentifiable in this situation. In addition, if the estimates for λ_{r_j} are close to the boundaries of the parameter space, i.e., $\lambda_{r_j} = 0$ or $\lambda_{r_j} \to \infty$, some of the μ and τ parameters can become redundant. Similarly, for panel observed data, boundary estimates, where some of the $\mu_{r_i s}$ are zero, can arise if estimates of λ_{r_j} become large. For instance, if the process is Markov or close to Markov, a semi-Markov model where the intensities to death from state r are zero, but there is a high intensity to death during the first phase in state $r+1$, can have the same likelihood as the correct Markov model.

A consequence of the identifiability problems for exponential sojourn distributions is that likelihood ratio tests of Markov versus phase-type semi-Markov models may not have standard asymptotic χ^2 distributions. Davies (1977, 1987) considered this problem for models with a single nuisance parameter that is unidentifiable under the null. He proposed a method of obtaining an approximate upper bound for the p-value of a test statistic. However, it does not seem straightforward, either theoretically or practically, to extend those results to tests for phase-type semi-Markov models with multiple nuisance parameters. Moreover, this approach does not address the computational problems of finding the global maximum.

Chen, Chen, and Kalbfleisch (2001) considered likelihood ratio tests for homogeneity in finite mixture models, where there is a similar problem of an unidentifiable alternative if the null model is true. They proposed a modified likelihood ratio test by introducing a penalty term into the likelihood function. A similar idea may be of use for phase-type semi-Markov models. Details of this development are given in the Appendix.

6. Unknown Initiation Times

In many contexts the initiation time of the process is unknown. This is not problematic for Markov models for which only the current state and current time are required, but it does have an effect on the likelihood calculations for semi-Markov models because the time of entry into the current state is needed. In the context of recurrent processes, most authors assume the process is in equilibrium (Rosychuk and Thompson, 2001; Crespi et al., 2005). For models with absorbing states this is not appropriate because at equilibrium all subjects are in the absorbing state. Satten and Sternberg (1999) in the context of nonparametric modeling of unidirectional processes assumed the unknown times between initiation and entry into a state were independent of subsequent transition times. They treat the time from initial observation until the first transition as separate nuisance functions to be estimated. This leads to some information loss because the time between the first observation and the first transition is related to the sojourn time in the initially observed state. However, this formulation means that methods applicable to unidirectional models when the times of initiation are known can be easily adapted. Kang and Lagakos (2007), in a parametric setting, only dealt with the case of known initiation times, but suggested methods similar to Satten and Sternberg's might be applicable.

For phase-type semi-Markov models the time since the initiation of the process is not needed if the underlying latent Markov state at the first observation time is known. One approach is therefore to assume that the probability of occupancy of latent state r_j at the first observation time is given by π_{r_j}, the r_jth entry of π, where π is a $\{k(R-1)\}$ vector summing to 1 and k is the number of phases. The entries of π would need to be estimated from the data, leading to $\{k(R-1)-1\}$ additional parameters. If it is reasonable to assume the times from initiation to first observation are identically distributed, then this method is most suitable if the subjects are assumed to be homogeneous.

However, we might expect the initial state occupancy probabilities to depend on covariates. Dependence of the initial state occupancy vector on the covariates could be incorporated, though this would greatly increase the number of unknown parameters to be estimated and is unlikely to be feasible in practice. Often it may be necessary to adopt a more naive model for initiation in order to maintain identifiability.

7. Example: Bronchiolitis Obliterans

Bronchiolitis obliterans is the irreversible, progressive airway obstruction leading to impairment of lung function. It is the major limiting factor to long-term survival among lung transplant recipients. Bronchiolitis obliterans can only be reliably assessed histologically. In practice, however, bronchiolitis obliterans syndrome (BOS) is defined as decline in forced expiratory volume in 1 second in liters (FEV_1) and this is used as a surrogate measure. The dataset includes 364 post-lung-transplant patients who received transplants at Papworth Hospital between 1984 and 2006. A total of 242 of the patients were heart-lung transplant patients and the remaining 122 were double-lung transplant patients. The rate of BOS onset, survival before and after BOS, onset, and the extent of misclassification are of primary interest. As BOS is known to increase the rate of death of transplant recipients, and for some patients onset of BOS may be associated with rapid decline in function, the underlying process is likely to be best modeled by a semi-Markov process.

The standard definition for BOS is expressed in terms of decline in FEV_1 relative to a posttransplantation baseline measure. A level of 80% of baseline or above is considered normal function whereas less than 80% is the clinical marker for BOS onset (Estenne et al., 2002). Because there is measurement error associated with FEV_1, we focus on an illness–death model for BOS with misclassification. BOS is not defined until at least 6 months after transplant. A patient's baseline FEV_1 measure is established during the first 6 months. Time for

Table 1
Observed transitions for the BOS data. States are 1 = *well*, 2 = *BOS*, 3 = *dead*, *C* = *observed final state mortality censored, HL = heart-lung transplantation, DL = double-lung transplantation*

		To state			
		1	2	3	*C*
HL	1	1190	198	36	55
From state	2	47	773	113	38
DL	1	198	68	23	51
From state	2	20	160	21	27

the process is therefore measured from 6 months after transplant. Although most patients should be in state 1 at this time, a better fitting model is achieved if the disease status is assumed unknown at 6 months and is to be estimated from the data.

Patients were scheduled to make clinic visits at 9 months, 12 months, and then at 6-month intervals thereafter. However, actual visit times were highly irregular. In particular, many patients suffered acute events. During these periods their lung function would be measured much more frequently. To ensure noninformative observation times, a necessary condition for these analyses, we choose to work on a reduced dataset. For each patient, we take the nearest observation to a scheduled visit time and exclude all other observations. In addition to measurements of lung function, time of death or administrative censoring was recorded. Date of death was taken to be known exactly. The dataset contains 2654 assessments of lung function, 193 deaths, and 171 administrative censoring times. Observed transitions are shown in Table 1.

Initially a range of HMM were fitted to the dataset. Biologically, the process is believed to be irreversible. However, initial exploration of the dataset suggested that the fit of a progressive misclassification HMM, where it is assumed recovery in the underlying process is not possible, is poor and some improvement was achieved by fitting a reversible misclassification HMM, that allowed actual recovery from BOS onset and misclassification. The likelihood ratio statistic testing $\mu_{21} = 0$ against $\mu_{21} > 0$ is 18.68. Note that, because this is a test at the boundary of the parameter space, the correct asymptotic null distribution is a 50:50 mixture of χ^2_1 and a point mass at 0. However, applying a Pearson-type goodness-of-fit statistic (Titman and Sharples, 2008) suggests a poor fit in both cases (100.2 on 25 degrees of freedom and 98.09 on 24 degrees of freedom for the progressive and reversible models respectively). Thus a HMM is not a good representation of the process.

One reason for the lack of fit of the HMM may be the lack of time homogeneity with respect to both time from start of the process (transplant + 6 months) and onset of BOS. Thus HSMMs may provide a better fit to the data and they can be implemented using the phase-type methodology. For simplicity we set $k = 2$, which allows for monotonically increasing or decreasing intensities for state exit. This gives an underlying five-state time-homogeneous Markov model as shown in Figure 3.

We emphasize here that although the latent state-phase Markov model is time homogeneous, the model for movement among the three states of well, BOS, and death is time inhomogeneous with respect to both time in the well state and time in the BOS state. The time dependence is governed by parameters τ_{1_2} and τ_{2_2}, which give the factor by which the intensities change between phases so that $\mu_{1_2r} = \mu_{1_1r}\tau_{1_2}$ for $r = 2, 3$ and $\mu_{2_2r} = \mu_{2_1r}\tau_{2_2}$ for $r = 1, 2$.

To better ensure that the global maximum likelihood estimate was found, the maximum likelihood estimate was computed conditional on the values of λ_1 and λ_2, on a 5×5 grid of possible values. Full optimization was then performed starting at the best value on the grid. We also inspected the estimated covariance matrix and calculated pairwise profile likelihood plots for pairs of parameters that were highly correlated. These further checks suggested the estimate found was a global maximum and that the likelihood surface, although quite flat in some directions, appeared unimodal.

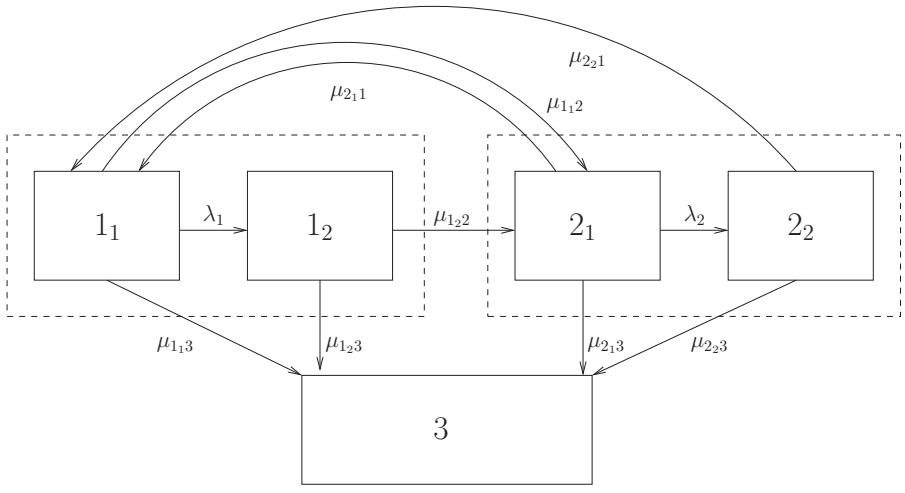

Figure 3. Latent Markov model underlying the phase-type semi-Markov model for the BOS dataset. The model has two 2-phase transient states and one absorbing state.

The HSMM, using 13 parameters, represented a significant improvement in fit compared to the HMM, which used 9 parameters. Applying the modified likelihood ratio test proposed in the Appendix, with $C_r = 10$ and $\alpha_r = 1$ for $r = 1, 2$, gave a test statistic of 24.48, which can be approximately compared to a χ^2_2 distribution. Note that the modified test means that parameters λ_1 and λ_2 do not contribute to the number of degrees of freedom. The result of the test shows that survival depends on time since entry to the BOS state as well as time from transplantation.

The possibility of unknown initiation times in the HSMM, as discussed in Section 6, was considered by allowing the state occupancies in latent states $1_1, 1_2, 2_1, 2_2$, to be unknown. However, the estimates for the proportion in state phases 1_2 and 2_2 were not significantly different from zero. Therefore, for parsimony, we assumed that at 6 months, a patient was in state 1_1 with probability $1 - \pi_2$ or 2_1 with probability π_2. Both the HMM and HSMM include the parameter, π_2. Transplantation type had a significant effect on the probability of initially being in state 2, π_2, with double-lung patients having a higher probability. This was represented in the model by $\text{logit}(\pi_2) = \text{logit}(\pi_2^0) + \theta_2^{DL}$. In addition, transplant type also affected e_{12}, the probability of being misclassified to state 2 given true state 1, with double-lung-transplant patients having a greater probability of misclassification. This was represented in the model by $\text{logit}(e_{12}) = \text{logit}(e_{12}^0) + \upsilon_{12}^{DL}$. However, transplant type did not have a significant effect on BOS onset rates. These results may indicate that double-lung-transplant patients in this cohort took longer to reach a stable baseline FEV_1 and the value established by 6 months may be less reliable.

Parameter estimates for the two final models are given in Table 2. In both the HMM and HSMM, $\mu_{1_1 3} < \mu_{2_1 3}$; hence as expected the mortality rate is substantially higher for patients with BOS than those without.

The HSMM estimates higher mortality in the first few years following transplantation and matches the Kaplan–Meier survival estimate more closely than the HMM, although there remains some lack of fit for the double-lung-transplant recipients (Figure 4).

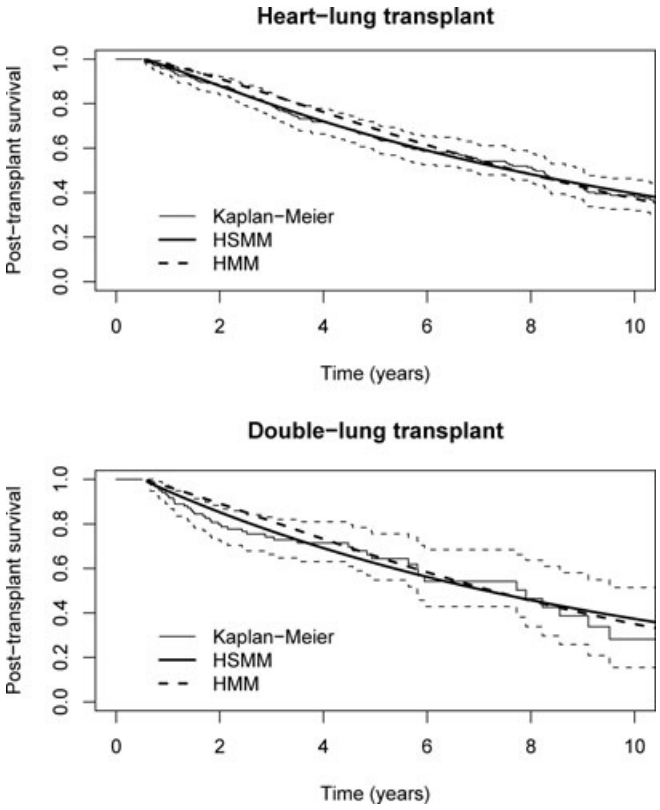

Figure 4. Estimates of posttransplant survival for heart-lung and double-lung transplantation patients—comparing fit based on Kaplan–Meier estimate, HMM, and HSMM.

The effect of the semi-Markov nature of the model is apparent in the estimated probabilities of survival post-BOS onset. For the HMM, the estimated future survival is fixed

Table 2
Parameter estimates for hidden Markov and phase-type HSMMs for BOS dataset

	HMM		HSMM	
Parameter	Estimate	CI	Estimate	CI
λ_1			0.207	$(0.047, 0.913)$
$\mu_{1_1 2}$	0.197	$(0.166, 0.234)$	0.339	$(0.240, 0.478)$
$\mu_{1_1 3}$	0.029	$(0.017, 0.047)$	0.016	$(0.002, 0.120)$
$\log(\tau_{1_2})$			-1.093	$(-1.694, -0.492)$
λ_2			1.369	$(0.286, 6.550)$
$\mu_{2_1 1}$	0.035	$(0.017, 0.073)$	0.113	$(0.038, 0.335)$
$\mu_{2_1 3}$	0.195	$(0.164, 0.232)$	0.500	$(0.214, 1.172)$
$\log(\tau_{2_2})$			-1.158	$(-1.914, -0.401)$
e_{21}	0.007	$(0.002, 0.023)$	0.006	$(0.001, 0.026)$
e_{12}^0	0.028	$(0.018, 0.045)$	0.025	$(0.015, 0.042)$
υ_{12}^{DL}	1.306	$(0.614, 1.998)$	1.314	$(0.561, 2.068)$
π_2^0	0.094	$(0.053, 0.094)$	0.079	$(0.037, 0.161)$
θ_2^{DL}	0.937	$(0.081, 1.794)$	1.019	$(0.004, 2.034)$
$-2\times$ LL	3005.06		2976.51	
No. pars	9		13	

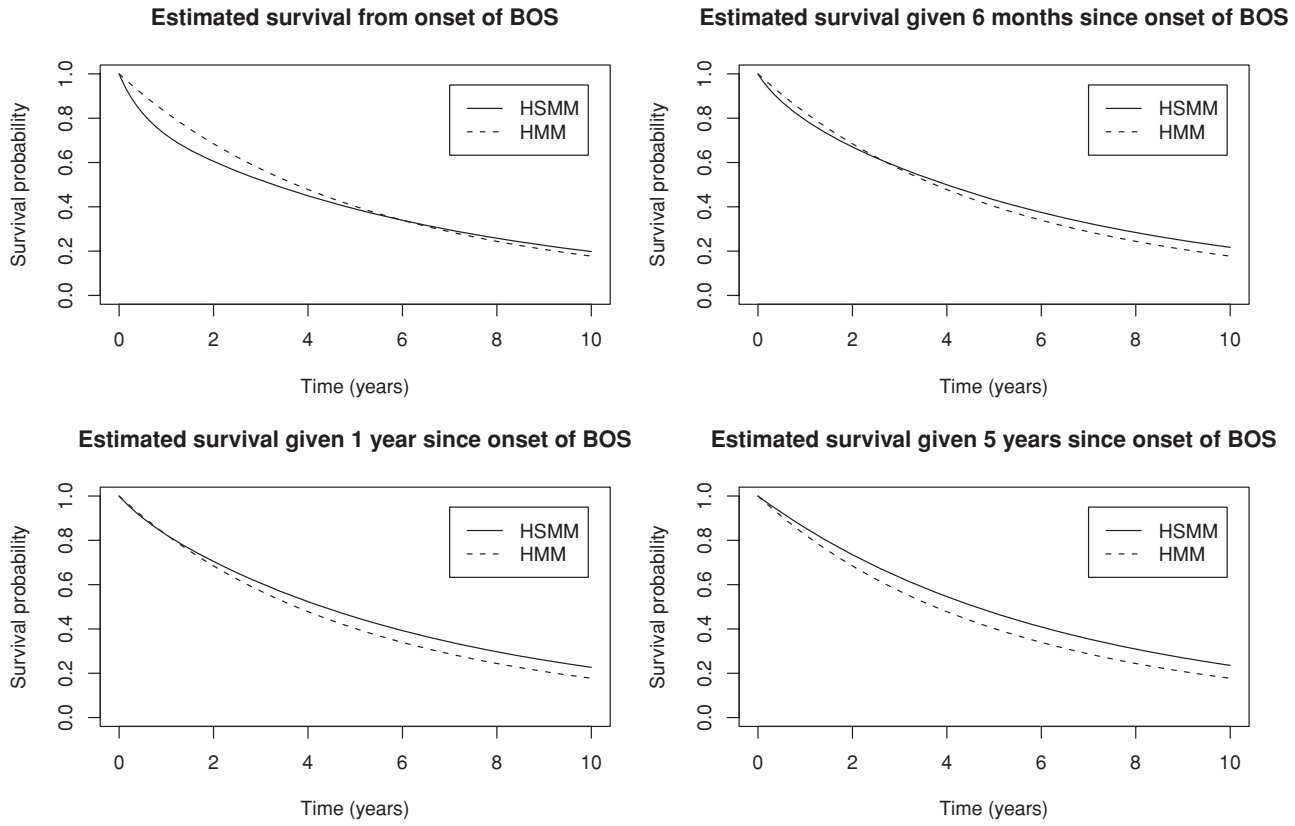

Figure 5. Estimates of survival given differing lengths of times since BOS onset—comparing estimates based on the HMM and HSMM.

irrespective of time spent in the BOS state. In contrast, the HSMM estimates lower short-term survival for patients who have only just developed BOS, but improved survival for patients who have already had BOS for a year or more (Figure 5).

The HSMM estimated a significantly decreasing hazard of making a transition (with respect to time of entry into the state) for both state 1 and state 2 ($\log(\tau_{r_2}) < 0, r = 1, 2$). Jackson et al. (2002) found that post-lung-transplant patients generally exhibited two possible types of decline in lung function, a rapid acute BOS onset, or a slower chronic onset. This suggests a random effects or frailty model may be appropriate. However, there are strong similarities between the data observed from a random effects model and the fitted semi-Markov model. A patient who has spent a long time in state 1 has a lower hazard of progressing to state 2; this is consistent with a random-effects model, where a long time in state 1 would suggest a low frailty. It is also possible that, despite observations being 6 months or more apart, the assumption of independent misclassification conditional on the underlying process is not appropriate. Part of the underlying semi-Markov process may therefore be modeling these shorter-term fluctuations, rather than the true BOS process. This may also account for the allowance of $2 \to 1$ transitions in the underlying process.

For both the HMM and HSMM the estimate of e_{12} was greater than that of e_{21}, suggesting that short-term fluctuations are more likely to result in overdiagnosis of BOS.

Because the phase-type HSMM is a class of HMM, it is possible to apply the Pearson-type goodness-of-fit test for HMMs proposed by Titman and Sharples (2008) to assess the overall fit of the model for the BOS data. This test imputes values for the next scheduled observation time for patients who died in order to calculate the expected transition probabilities, which results in noninteger values for observed numbers of deaths in particular time interval groups. Table 3 shows the resulting contingency table of observed and expected transitions, stratified by transition type (including transitions to death and censoring for which observation times are known exactly), length of time interval, observation number in an individual's series, and transplant type. The statistic had a value of 37.1. The null distribution of the statistic cannot be accurately calculated without bootstrapping, which would be time consuming in this case. However, $\chi^2_D(0.95)$, where D is the number of independent cells, gives an upper bound to the 95% point of the statistic. In this case $D = 32$ giving an upper bound of 46.2; hence based on this conservative test there is no apparent evidence of poor fit.

8. Discussion

Phase-type sojourn distributions allow semi-Markov and HSMMs to be fitted to panel observed multistate data with relative ease. The likelihood for such models can be expressed as a particular type of HMM. Standard methods and freely available software for fitting HMMs can therefore be used. Through these methods we were able to better

Table 3

Goodness-of-fit table for phase-type HSMM for BOS data. States are 1 = *well*, 2 = *BOS*, 3 = *dead*, *C* = *observed final state mortality censored. Transitions are stratified by: TT = transplant type, ON = observation number in an individual's series, TE = time elapsed since last observation, HL = heart-lung transplantation, DL = double-lung transplantation.*

TT	ON	TE		$1 \to 1$	$1 \to 2$	$1 \to 3$	$1 \to C$	$2 \to 1$	$2 \to 2$	$2 \to 3$	$2 \to C$
			Obs	381	46	3.9	10	4	52	4.6	1
		$\leqslant 0.44$	Exp	374.0	52.8	4.2	10.0	5.7	50.3	4.5	1.0
	$\leqslant 6$		Dev	0.14	0.87	0.57	0.00	0.52	0.10	0.48	0.00
			Obs	303	65	19.1	23	18	80	20.4	8
		>0.44	Exp	305.1	62.9	19.1	23.0	10.1	86.7	21.8	7.9
HL			Dev	0.02	0.07	0.13	0.00	6.25	0.54	0.20	0.00
			Obs	243	30	2.6	12	10	343	19.4	16
		$\leqslant 0.51$	Exp	245.5	27.5	2.5	12.0	15.5	328.5	28.8	15.6
	>6		Dev	0.03	0.22	0.45	0.00	1.20	0.67	3.49	0.01
			Obs	263	57	10.4	10	15	298	68.6	13
		>0.51	Exp	271.0	46.9	12.6	9.9	19.2	305.6	56.3	13.5
			Dev	0.24	2.17	0.47	0.00	0.92	0.22	2.87	0.02
			Obs	91	30	3.3	14	5	18	2.2	9
		$\leqslant 0.44$	Exp	93.2	28.8	2.1	14.1	4.6	18.4	2.0	9.04
	$\leqslant 6$		Dev	0.06	0.05	1.22	0.00	0.04	0.03	0.50	0.01
			Obs	70	30	16.7	28	12	52	6.8	7
		>0.44	Exp	74.1	29.6	11.9	29.0	9.7	51.8	9.5	6.7
DL			Dev	0.23	0.01	2.0	0.04	0.53	0.01	0.86	0.01
			Obs	9	3	0.6	6	2	48	3.3	10
		$\leqslant 0.51$	Exp	9.9	2.3	0.2	6.1	2.4	46.8	4.3	9.8
	>6		Dev	0.10	0.20	2.55	0.01	0.10	0.06	0.74	0.01
			Obs	28	5	2.4	3	1	42	8.7	1
		>0.51	Exp	25.1	8.0	2.3	3.0	3.2	41.9	6.6	1.0
			Dev	0.34	1.13	0.10	0.00	1.51	0.04	0.97	0.00

characterize the natural history of lung function decline after thoracic transplantation. In particular we can confirm our previous observations, based on less sophisticated models, that BOS has an aggressive form with acute onset and rapid subsequent decline. However, if this decline can be halted soon after BOS onset, possibly with more aggressive treatment, subsequent mortality rates will decrease. This feature was not apparent in HMMs. It was also clear that our current screening test has good sensitivity and specificity, because the chance of a false diagnosis of BOS is less than 3% and the chance of a false negative test is even lower. Fitting these semi-Markov and HSMMs provides a specific test of a time homogeneous Markov against semi-Markov assumption. Although a standard likelihood ratio test cannot be constructed due to the lack of identifiability of the semi-Markov model if the Markov model is true, a modified likelihood ratio test can be constructed with an approximate χ^2 distribution. According to this test a HSMM had superior fit in our BOS example and a goodness-of-fit test confirmed the overall adequacy of the HSMM.

Phase-type semi-Markov models may also have a wider applicability to more accurately describe the process of interest. However, a general problem with panel observed data from semi-Markov models is the lack of information in the data. In particular, the transition intensities at times soon after entry into a state may be difficult to estimate because of the lack of consecutive observations in short time intervals. This can lead to boundary maximum likelihood estimates. Similarly, when there is little evidence against a Markov model in the data, standard optimization algorithms such as Nelder–Mead or Baum–Welch may fail to find the global maximum. Further, in our example we fitted two phases to each state to allow monotonically increasing or decreasing intensities out of the state. It is theoretically possible to increase the number of phases to allow greater flexibility. If there is no misclassification and the model is progressive allowing $k = 3$ or more may be beneficial, particularly for the sojourn distribution for the first state, if there were sufficient observations. For the BOS dataset, increasing the number of phases was not viable and this is likely to be the case in general for models with reverse transitions or misclassification.

Some of the problems of estimation might be alleviated by adopting a penalized likelihood approach, similar to that used in the construction of the modified likelihood ratio test in the Appendix. Choosing smoothing parameters based on cross-validation would be computationally difficult. However, using an approximate crossvalidation score, such as that proposed in Joly et al. (2002), may be effective.

A further extension to the model would be to allow time dependence with respect to calendar time. Existing approaches to fitting time inhomogeneous Markov models based on parametric assumptions such as piecewise constant intensities can, in principle, be applied to the underlying latent Markov process. However, as with other possible model extensions, incorporating time inhomogeneity will increase the possibility of nonidentifiability.

References

Bureau, A., Shiboski, S., and Hughes, J. P. (2003). Applications of continuous time hidden Markov models to the study of misclassified disease outcomes. *Statistics in Medicine* **22,** 441–462.

Chen, H., Chen, J., and Kalbfleisch, J. D. (2001). A modified likelihood ratio test for homogeneity in finite mixture models. *Journal of the Royal Statistical Society, Series B* **63,** 19–29.

Chen, Y., Xie, J., and Liu, J. S. (2005). Stopping-time resampling for sequential Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* **67,** 199–217.

Cox, D. R. (1955a). The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. *Proceedings of the Cambridge Philosophical Society* **51,** 33–41.

Cox, D. R. (1955b). A use of complex probabilities in the theory of stochastic processes. *Proceedings of the Cambridge Philosophical Society* **51,** 313–319.

Cox, D. R. and Miller, H. D. (1965). *The Theory of Stochastic Processes*. London: Chapman and Hall.

Crespi, C. M., Cumberland, W. G., and Blower, S. (2005). A queuing model for chronic recurrent conditions under panel observation. *Biometrics* **61,** 193–198.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64,** 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74,** 33–43.

Estenne, M., Maurer, J. R., Boehler, A., Egan, J. J., Frost, A., Hertz, M., Mallory, G. B., Snell, G. I., and Yousem, S. (2002). Bronchiolitis obliterans syndrome 2001: An update of the diagnostic criteria. *Journal of Heart and Lung Transplantation* **21,** 297–310.

Faddy, M. J. (2002). Penalised maximum likelihood estimation of the parameters in a Coxian phase-type distribution. In *Matrix-Analytic Methods: Theory and Applications*, G. Latouche and P. Taylor (eds), 107–113. Hackensack, New Jersey: World Scientific.

Foucher, Y., Giral, M., Soulillou, J.-P., and Daures, J.-P. (2007). A semi-Markov model for multistate and interval-censored data with multiple terminal events. Application in renal transplantation. *Statistics in Medicine* **26,** 5381–5393.

Foucher, Y., Giral, M., Soulillou, J.-P., and Daures, J.-P. (2008). A flexible semi-Markov model for interval-censored data and goodness-of-fit testing. *Statistical Methods in Medical Research* doi:10.1177/0962280208093889.

Gruger, J., Kay, R., and Schumacher, M. (1991). The validity of inferences based on incomplete observations in disease state models. *Biometrics* **47,** 595–605.

Jackson, C. H., Sharples, L. D., McNeil, K., Stewart, S., and Wallwork, J. (2002). Acute and chronic onset of bronchiolitis obliterans syndrome (BOS): Are they different entities? *Journal of Heart and Lung Transplantation* **21,** 658–666.

Jackson, C. H., Sharples, L. D., Thompson, S. G., Duffy, S. W., and Couto, E. (2003). Multistate Markov models for disease progression with classification error. *The Statistician* **52,** 193–209.

Joly, P. and Commenges, D. (1999). A penalized likelihood approach for a progressive three-state model with censored and truncated data: Application to AIDS. *Biometrics* **55,** 887–890.

Joly, P., Commenges, D., Helmer, C., and Letenneur, L. (2002). A penalized likelihood approach for an illness-death model with interval-censored data: Application to age-specific incidence of dementia. *Biostatistics* **3,** 433–443.

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.

Kang, M. and Lagakos, S. W. (2007). Statistical methods for panel data from a semi-Markov process, with application to HPV. *Biostatistics* **8,** 252–264.

MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-Valued Time Series*. New York: Chapman and Hall.

Rosychuk, R. J. and Thompson, M. E. (2001). A semi-Markov model for binary longitudinal responses subject to misclassification. *Canadian Journal of Statistics* **29,** 395–404.

Rosychuk, R. J. and Thompson, M. E. (2003). Bias correction of two-state latent Markov process parameter estimates under misclassification. *Statistics in Medicine* **22,** 2035–2055.

Satten, G. A. and Longini, I. M. (1996). Markov chains with measurement error: Estimating the "true" course of a marker of the progression of human immunodeficiency virus disease. *Applied Statistics* **45,** 265–309.

Satten, G. A. and Sternberg, M. R. (1999). Fitting semi-Markov models to interval-censored data with unknown initiation times. *Biometrics* **55,** 507–513.

Titman, A. C. and Sharples, L. D. (2008). A general goodness-of-fit test for Markov and hidden Markov models. *Statistics in Medicine* **27,** 2177–2195.

Received October 2008. *Revised July* 2009. *Accepted August* 2009.

Appendix: Modified Likelihood Ratio Test

We propose to maximize a modified log-likelihood function

$$pl(\theta) = l(\theta) + \sum_j \sum_r (C_{r_j} \log \lambda_{r_j} - C_{r_j} \lambda_{r_j} \alpha_{r_j}), \quad \text{(A1)}$$

where $l(\theta)$ is the standard log-likelihood function and C_{r_j} and α_{r_j} are constants to be chosen.

This can be thought of as the log-likelihood function that would result if an additional auxiliary experiment had been performed where the precise lengths of the phases for each state were observed. We say that C_{r_j} sojourns of length α_{r_j} were observed in phase j of state r.

The aim of the modification is to ensure estimates of λ_{r_j} are away from $\lambda_{r_j} = 0$ and $\lambda_{r_j} = \infty$. In addition, λ_{r_j} becomes identifiable regardless of the value of τ. Choice of C_{r_j} determines the degree of penalization and α_{r_j}, the default length of phase j for state r.

We note that Faddy (2002), in the context of fitting Coxian phase-type distributions to fully observed length-of-stay data, proposed a penalized log-likelihood function with an alternative penalty. He considered $\zeta_j = \lambda_j + \mu_j, j = 1, \ldots, k$, where it is assumed $\lambda_k = 0$, then took $C \sum_{j=1}^{k} (\zeta_j - \bar{\zeta})^2/(k-1)$ as the penalty, where $\bar{\zeta} = \sum_{j=1}^{k} \zeta_j/k$. However, this is not suitable for our application because it does not exclude the case $\lambda_j = 0$ for $j = 1, \ldots, k-1$. Moreover, the penalty is dependent on τ (through μ_j) under this formulation.

We can construct a modified likelihood ratio test for $\tau_r = 1,\ r = 1, \ldots, R-1$, by considering

$$T(X) = 2\{pl(\hat{\theta}) - pl(\widetilde{\mu}, \tau = \mathbf{1}, \lambda = \lambda_0)\},$$

where $\lambda_{r_{0j}} = (\alpha_{r_j})^{-1}$. Note that $\widetilde{\mu}$ corresponds to the Markov model maximum likelihood estimate.

The null distribution of T is nonstandard and depends on the choice of C. However, we can write T as $T(X) = T_{\lambda_0}(X) + A(\hat{\lambda}, \lambda_0)$ where $T_{\lambda_0}(X)$ is the standard likelihood ratio test statistic, based on fixing $\lambda = \lambda_0$ under both

the null and alternative. $A(\hat{\lambda}, \lambda_0)$ is nonnegative and its distribution depends on C. The mean of A decreases as C is increased. Note that one could base a likelihood ratio test on $T_{\lambda_0}(X)$, which has an asymptotic χ^2_{R-1} distribution. However, this would have lower power at detecting a semi-Markov process with true values of λ away from λ_0. Potentially, a choice of C can be obtained to ensure reasonable power of the test for a wide range of λ and allow χ^2_{R-1} to be a reasonable approximation to the null distribution. Limited simulations in the case of fully observed phase-type data suggest that a value of $C \approx 10$ may be appropriate. In principle, a parametric bootstrap could also be used to approximate the null distribution. Such a bootstrap is more feasible with the modified likelihood because identifiability allows convergence of the optimization algorithm to be more rapid.