

Semi-Markov Models with Phase-Type Sojourn Distributions by A.C. Titman, L.D. Sharples, 2010

Rahul Biswas

Department of Statistics, University of Washington Seattle, WA, 98195, USA

Abstract

1 Introduction

(VM's comment: Please do not modify my LaTeX template, your margins are too wide for some reason. Stick to my original style file)

Finite state continuous-time processes, with data being observed at irregularly spaced discrete time points, and no data on the trajectory of the process between these times (panel observed), are often encountered in practice. Examples are medical research on characterising the evolution of a disease within an individual (progressive disease like HIV (Longini & Clark, 1989), or, episodic disease with recurrent transitions like asthma (Saint-Pierre et al., 2003)), when disease statuses are recorded only at clinic visits, and exact transition times are unknown (Dean et al., 2009; Mandel, 2010). Time-homogeneous continuous time Markov chains (CTMCs) are the simplest and most tractable models in such scenarios (Kalbfleisch and Lawless, 1985). However homogeneous CTMCs have limiting restrictions of transition intensities constant over time and thereby exponential sojourn distributions, which is often unrealistic.

Inhomogeneous CTMCs (Kay, 1986; Hubbard, Inoue, and Fann, 2008; Titman, 2011) extend the setup to have transition intensities vary with respect to time since the process origin. But for diseases, ~~often~~ transition intensities **often** may depend on the time spent in the current state (sojourn time), not just external time. Semi-Markov models have such a property and are considered by Titman & Sharples (2010) (Cox & Miller 1965; McGilchrist & Hills 1991).

Although semi-Markov models are appealing, there are computational hurdles documented in fitting them to data. The likelihood is less tractable for observed panel data unless the model is assumed to be progressive, where a subject cannot reenter a

state once exited (Joly and Commenges, 1999; Foucher et al., 2010). In the presence of reversible transitions, semi-Markov models are tractable only under stringent restrictions of an evenly spaced two-state recurrent model (Rosychuk and Thompson; 2001), or, if at least one state has exponential sojourn distribution (Kang, Lagakos; 2007).

Titman and Sharples in their paper (2010) proposed a general approach to semi-Markov modeling and fitting to panel observed processes using a latent state CTMC. (VM's comment: **Don't hardcode your references! Learn how to use bibtex, please!**) Each state maps to multiple latent states, which are traversed according to an underlying CTMC. This framework yields rates of transitions between states that depend on the duration spent in that state, yet likelihoods are analytically tractable, even for disease processes with reversible transitions. The model is extended to incorporate misclassification error. Methods for inference of parameters while addressing non-identifiability concerns are discussed. The methods are applied to assess development of bronchitis obliterans syndrome in post-lung-transplantation patients, making and are compared with the standard popular method based on HMM. The quantities of scientific interest studied are the rate of disease onset, survival rates of patients before and after disease onset given survived for certain years after onset, and, extent of misclassification, which are one-dimensional functionals of the model parameters. (VM's comment: **Try to use less of the passive voice in your writing**)

In contrast to the restriction of exponentially distributed sojourn distributions in homogeneous CTMCs, the sojourn-time of observable states in such latent CTMC setup has a phase-type distribution, defined as the distribution of absorption time of a homogeneous CTMC with finitely many transient states and one absorbing state. An advantage is generality, as phase type distributions are dense in the class of all distributions with non-negative support, so any distribution with non-negative support can be approximated by a phase-type distribution (Neuts, 1974). Analytic tractability is also ensured with density, cumulative distribution function and failure rate being matrix exponentials. One disadvantage is that the model parameters may not be identifiable (Asmussen et al., 1996), which is a difficulty in frequentist estimation, but, typical scientifically meaningful functionals of sojourn distribution parameters are identifiable (Bladt et. al, 2003). Titman & Sharples (2010) constrained the latent CTMC parameters to yield a sojourn distribution from a subclass of phase type distributions called

Coxian phase-type distribution (details are in Section 2.3) which largely provides similar approximations to distributions as the general phase-type class while being faster for computation (Asmussen et al, 1996).

2 Methods

2.1 Likelihood for Continuous-time processes with panel data

Let $X(t)$ denote the continuous-time discrete state stochastic process on a finite state space $\{1, \dots, R\}$, and observed data for an individual subject consist of observed states x_0, \dots, x_n at times (can be irregular) t_0, \dots, t_n where time points and number of times n can be subject specific. The data is panel observed at discrete time points with no information about the trajectory of the process between observations. It is assumed throughout, that, the sampling mechanism from the continuous process is non-informative (Gruger, Kay, Schumacher, 1991).

General finite state continuous-time models may be defined according to the transition intensities from states r to s :

$$q_{rs}(t, \mathcal{F}_t) = \lim_{\delta t \downarrow 0} \frac{P(X(t + \delta t) = s | X(t) = r, \mathcal{F}_t)}{\delta t},$$

where \mathcal{F}_t is the filtration or past history of the process up to time t . (VM's comment: **Math equations are still parts of sentences and need punctuation.**)

Under a Markov assumption (Kalbfleisch and Lawless, 1985), the transition intensities are only a function of t and the likelihood for an individual is

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

where $p_{x_{i-1}, x_i}(t_{i-1}, t_i; \theta) = P(X(t_i) = x_i | X(t_{i-1}) = x_{i-1})$ are the transition probabilities. The transition probabilities can be expressed in terms of the transition intensities by solving the Kolmogorov forward equations (Cox & Miller; 1965). If time homogeneous, they **transition probabilities and intensities** are related through a matrix exponential.

2.2 Semi-Markov models

In a semi-Markov model the transition intensities depend on the time spent in the current state

$$q_{rs}(t, \mathcal{F}_t) = q_{rs}(u) = \lim_{\delta t \downarrow 0} \frac{P(X(t + \delta t) = s | X_i(t) = r, T^* = t - u)}{\delta t},$$

where T^* denotes the time since entry into the current state. In semi-Markov models the sojourn times in each state are not constrained to have exponential distributions in contrast to time homogeneous Markov models. Titman & Sharples specify the sojourn distribution of $X(t)$ to be of Coxian phase-type (See Section 2.3).

Computation of the likelihood for panel observed states is more difficult compared to the Markov case. In the case of progressive models, where the process cannot reenter a state once exited, there is a finite number of possible paths that an individual can take conditional on their observed states. Computation of the likelihood requires considering each path and integrating over the possible sojourn times in each state of the path (Foucher et al. 2010). Numerical quadrature methods can be applied to compute the likelihood but become unattractive for models with more than 3 or 4 states because of the increasing dimension of the integrals.

For more general models with reverse transitions, direct integration is not possible because the number of possible state visits is unbounded. Computation of the transition probabilities defined as $p_{r,s}(u, t) = P(X(t) = s | X(u) = r, T^* = u)$ requires solution to a system of integral equations (Howard 1964; De Dominics and Manca 1984). However, limitation for panel observed states is that the likelihood cannot be expressed as the product of transition probabilities, as the observation times will not correspond to the entry time into the observed state. Under restrictions of evenly-spaced two-state recurrent model (Rosychuk & Thompson, 2001), or, having one state with exponential sojourn distributions (Kang & Lagakos, 2007), computation could be made tractable.

2.3 Coxian Phase-type Models with misclassification error

Recall that $X(t)$ denotes a continuous time stochastic process with state space $\{1, \dots, R\}$, where t is time since process origin. Let R be the only absorbing state. Under-

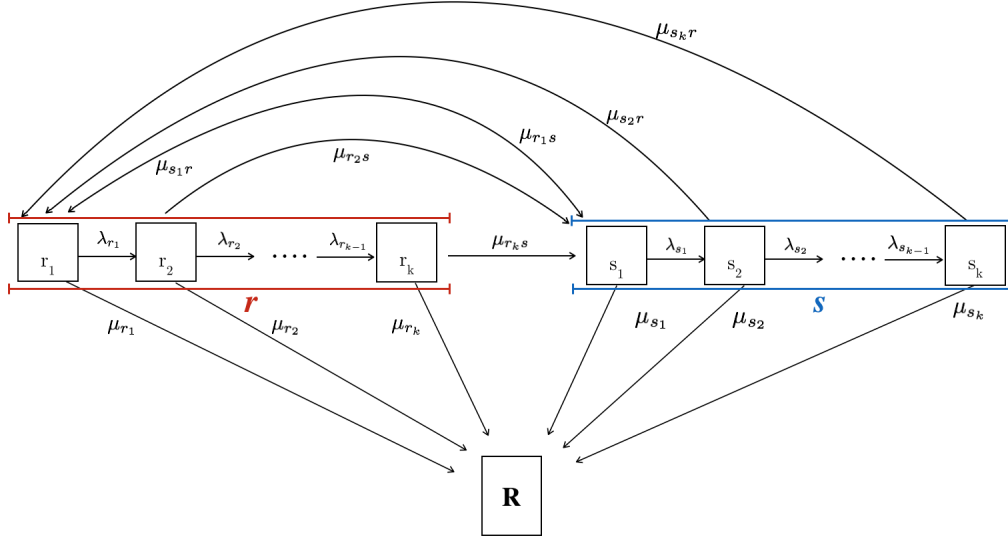


Figure 1: This figure summarizes the transition diagram of the underlying process $X^*(t)$. r, s are any transient state of $X(t)$ with underlying $X^*(t)$ states r_1, \dots, r_k and s_1, \dots, s_k respectively, and R is the only absorbing state.

lying $X(t)$ there is the latent homogeneous CTMC $X^*(t)$ with latent state space $\{1_1, \dots, 1_{s_1}\} \cup \dots \cup \{(R-1)_1, \dots, (R-1)_{s_{R-1}}\} \cup R$. $X(t) = r \Leftrightarrow X^*(t) \in \{r_1, \dots, r_{s_r}\}$ for $r < R$ and $X(t) = R \Leftrightarrow X^*(t) = R$. The sojourn distribution of each non-absorbing state r of $X(t)$ is assumed to be a k -phase Coxian phase-type distribution, defined to be the distribution of absorption time of a CTMC with k progressive states and one absorbing state. Thus for observable state r , the latent process $X^*(t)$ can have the states r_1, \dots, r_k , with parameters λ_{r_j} = the intensity for movement from r_j to r_{j+1} , for $j = 1, \dots, k-1$, with the assumption that these latent phases for each observable state is progressive thereby all other transitions between r_j 's have 0 intensity, and $\mu_{r_j s} =$, the intensity for movement out of a latent phase of r_j to state s with the assumption that any state s is entered only at latent phase s_1 , thereby all other intensities for transition from latent phases of r to latent phases of s are 0. (VM's comment: **This is an insanely long sentence and impossible to parse. Also, do not say $\lambda =$ and then put definition of it in words. Treat λ as a noun and use it in a sentence properly, using "is equal" to instead of $=$.**) Figure 1 summarizes the transition diagram of the underlying process $X^*(t)$. For greater estimability and interpretation of moving out of a state to another it is assumed that $\mu_{r_j s} = \tau_{r_j} \mu_{r_1 s} \forall s$,

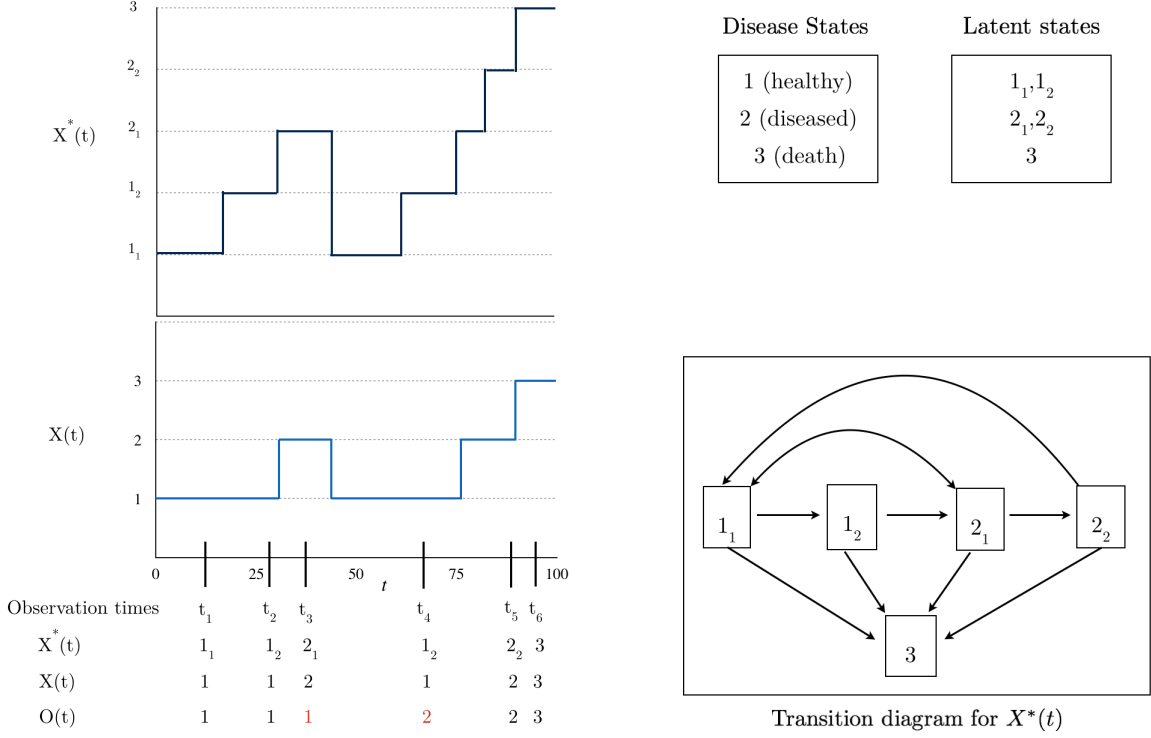


Figure 2: This is an illustration of a disease process with latent trajectory $X^*(t)$, corresponding disease trajectory $X(t)$, and misclassified trajectory $O(t)$. The three states of the disease process $X(t) : \{1=\text{healthy}, 2=\text{BOS}, 3=\text{death}\}$ with 3 the only absorbing state. $X(t)$ has underlying latent homogeneous CTMC $X^*(t)$ with state space $S^* = \{1_1, 1_2, 2_1, 2_2, 3\}$ with $\{3\}$ an absorbing state and rest transient. $O(t)$ is the observed process which is the misclassified version of $X(t)$. States which are misclassified in this example are marked in red.

$\tau_{r_1} = 1$. That is rates of exiting from latent phase r_j relative to r_1 change by the same factor irrespective of destination. It is noted that the resulting $X(t)$ is a semi-Markov process.

To incorporate that the states in the process are observed with misclassification error, it is considered that observed states are $O(t)$ which are related to $X(t)$ by taking a value with misclassification probability $e_{rs} = P(O(t) = s | X(t) = r) = P(O(t) = s | X^*(t) = r_j) = e_{r_j s}^*$ for $r, s = 1, \dots, R, j = 1, \dots, k$. It is noted that the process $(O(t), X^*(t))$ is a Hidden Markov Model (HMM). Thereby computational methods tailored for **statistical analysis of HMMs** can be used for inference. Figure 2 shows an example of a disease process with latent trajectory $X^*(t)$ and the corresponding dis-

ease trajectory $X(t)$ and misclassified trajectory $O(t)$ for a two-state reversible disease model.

2.4 Inference and Identifiability

Denoting o_0, \dots, o_n to be the sampled observations with misclassification at times t_0, \dots, t_n for an individual, the latent HMM structure of the model enables to write the likelihood contribution for the individual based on the forward-backward algorithm, as follows:

$$L = \pi \mathbf{M}_1^* \mathbf{M}_2^* \dots \mathbf{M}_n^* \mathbf{1}, \quad (1)$$

where \mathbf{M}_i^* is a $\{k(R-1)+1\} \times \{k(R-1)+1\}$ matrix with (r^*, s^*) entry $e_{s^* o_i}^* p_{r^* s^*}^*(t_i - t_{i-1})$ for $r^*, s^* \in S^*$, with $\mathbf{P}^*(t) = \{p_{r^* s^*}^*(t)\} = e^{\mathbf{A}t}$, and, π is the initial state distribution.

The presence of misclassification and HMM setup increases possibility of non-identifiability (MacDonald & Zucchini, 1997). In fact for a HMM with two states and balanced ~~p~~observation times, it is not possible to simultaneously identify the misclassification probabilities and transition intensities or probabilities without putting constraints, for instance that misclassification probabilities are each < 0.5 (Rosychuk & Thompson, 2003).

In addition, the Coxian phase-type structure induces non-identifiability. For example, the sojourn time distribution for state r is exponential if $\tau_{r_j} = 1 \forall j = 1, \dots, k$. The parameters, $\lambda_{r_j}, j = 1, \dots, k-1$, are redundant and unidentifiable in this situation.

As ensuring identifiability is not straightforward, the authors suggest inspection of the likelihood function by evaluating the Hessian matrix to ensure the estimated parameters are at a maxima and performing optimization procedures from a wide range of starting values to ensure a strict global maxima has been attained, or, checking for unimodality of the likelihood. If non-identifiability is evident, reduction of the complexity of the model can be attempted. Else, the parameters can be constrained. The authors suggest performing the optimization by first maximizing the profile likelihood on a grid of possible fixed values of the λ_{r_j} . Other possible constraints suggested are having some λ_{r_j} as known constants, or constraining some of the sojourn distributions

to be exponential.

Due to identifiability problems for exponential sojourn distributions in the current model, the likelihood ratio test for Markov versus Coxian phase-type semi-Markov model may not have an asymptotic χ^2 distribution. Chen et. al (2001) proposed a penalized likelihood ratio tests for homogeneity in finite mixture models. Authors considered the same method but with an appropriate penalty for the current situation so that the λ_{r_j} 's are away from 0 and ∞ and identifiable irrespective of value of τ .

Titman & Sharples (2008) proposed a goodness of fit test applicable for Markov and Hidden Markov Models. The same is used for testing goodness of fit in current model which boils down to a HMM.

3 Application : Bronchiolitis obliterans syndrome

Bronchiolitis obliterans syndrome (BOS) is the irreversible progressive narrowing of bronchiolar lumens and airflow obstruction (Todd & Palmer, 2011), leading to impaired lung function. It is recorded to affect a majority of lung transplant recipients and is the principal factor limiting long-term transplant survival. It is measured by decline in forced expiratory volume in 1 second in liters (FEV_1) relative to a posttransplantation baseline measure, with a level of 80% of baseline or above deemed normal and less than 80% is the clinical marker for BOS onset (Estenne et al, 2002). The primary scientific objectives are inferring on the rate of BOS onset, survival rate of patients before BOS onset, and, after BOS onset given survived for certain years after onset, and the extent of misclassification. The dataset includes 364 post-lung-transplant patients who received transplants at Papworth Hospital between 1984 and 2006, with 242 heart-lung transplant and 122 both-lung transplant patients. Patients were scheduled to for visit at 9 months, 12 months, after that at 6 month intervals, but actual visit times were highly irregular, and can be of different number, motivating modelling the process by on the time continuum.

3.1 The methods applied to the BOS scenario

The model Titman and Sharples (2010) modelled the disease process on continuous time ($O(t)$, say) by an illness-death model with 3-states (of being well or in BOS or

dead) (Fig 2). It is assumed to be a Semi-Markov process. The available data is considered to be realizations sampled at discrete time points (may be irregular) for each patient. It is assumed throughout, that, the sampling mechanism from the continuous process is non-informative (Gruger, Kay, Schumacher, 1991).

$O(t)$ denotes the disease process with the three states : $S = \{1=\text{healthy}, 2= \text{BOS}, 3=\text{death}\}$ with 3 the only absorbing state. Titman and Sharples considered $O(t)$ to have an underlying latent homogeneous CTMC with state space $S^* = \{1_1, 1_2, 2_1, 2_2, 3\}$ with $\{3\}$ an absorbing state and rest transient, and rate matrix $\mathbf{\Lambda}$ constant with time, thus leading to a 2-phase Coxian phase-type sojourn distribution. If the latent process $X^*(t) \in r_1, \dots, r_k$ then an accurately observed process $X(t) = r, r = 1, 2, k = 1, 2$, if $X^*(t) = 3, X(t) = 3$. It is assumed that an individual enters transient state r in phase r_1 and passes progressively through consecutive phases until the state is exited which can occur through any phase. Thus $\lambda_{r_j s_1}, r, s = 1, 2, j = 1, 2, \lambda_{r_j, 3}$ denote the underlying transition intensities of going from observable state r to s . For greater estimability and practical interpretation of the intensities of moving out of a state to another, it is assumed that $\lambda_{1_1 2_2} = \tau_1 \lambda_{1_1 2_1}, \lambda_{1_2 3} = \tau_1 \lambda_{1_1 3}, \lambda_{2_2 1_1} = \tau_2 \lambda_{2_1 1_1}, \lambda_{2_2 3} = \tau_2 \lambda_{2_1 3}$. That is rates of exiting from state 1_2 compared to 1_1 change by the same factor irrespective of destination and similar for 2_2 compared to 2_1 . In addition to these, transition intensities from phase 1_1 to 1_2 and from 2_1 to 2_2 are kept as parameters, and all other transition intensities of the latent process are 0. To incorporate misclassification, we consider $O(t)$ takes a value with misclassification probability $e_{rs} = P(O(t) = s | X(t) = r) = P(O(t) = s | X^*(t) = r_j)$ for $r, s = 1, 2, j = 1, 2$. It is noted that the process $(O(t), X^*(t))$ fits into the Hidden Markov Model framework (See also eqn. (1); so also called Hidden Semi Markov Model, HSMM). A better fitting model is recorded if the disease status is assumed unknown at the first observation time, and incorporating probability of initial states π in the model, which is estimated from data.

After a first analysis, π_{1_2} and π_{2_2} were estimated to be nearly 0, so in the model a priori they were taken to be 0, and $\pi_{1_1} = 1 - \pi_2$ (say), $\pi_{2_1} = \pi_2$. Also, transplantation type was seen to have a significant effect on probability of initial state being 2, and on the probability of a state being misclassified as 2, given true state is 1. So they were modelled as $\text{logit}(\pi_2) = a_0 + a_1 1(DL)$, and $\text{logit}(e_{12}) = b_0 + b_1 1(DL)$, where $1(DL)$ is

the indicator of double lung transplant.

It is noted that the final quantities of interest, first passage cumulative distribution function, transition probabilities, hazard rates and survival probabilities are one-dimensional functionals of the parameter vector of the latent CTMC, and point and interval estimates of the former can be obtained from those of the later using Delta method.

Parameter estimation and Computation The authors kept only the nearest observation to a scheduled visit time and excluded others from analysis, for each patient, as an attempt to ensure non-informative sampling. Denoting w_0, \dots, w_n to be the sampled observations at times t_0, \dots, t_n for an individual, the latent HMM structure of the model enables to write the likelihood contribution for the individual based on the forward-backward algorithm, as follows.

$$L = \pi \mathbf{M}_1^* \mathbf{M}_2^* \dots \mathbf{M}_n^* \mathbf{1} \quad (2)$$

where \mathbf{M}_i^* is a 5×5 matrix with (r^*, s^*) entry $e_{s^* w_i}^* p_{r^* s^*}^*(t_i - t_{i-1})$ for $r^*, s^* \in S^*$, with $P^*(t) = ((p_{r^* s^*}^*)) = e^{\mathbf{A}t}$, and, $\pi = (1 - \pi_2, 0, \pi_2, 0, 0)$ taken as the initial state probabilities.

The 13 parameters are estimated by a maximum likelihood estimate (MLE) considering all patients. To ensure global optimum, the authors computed the MLE conditional on the values of $\lambda_{1_1 1_2}$ and $\lambda_{2_1 2_2}$ on a 5×5 grid of possible values. Full optimization was then performed starting at the best value of the grid. Standard optimization algorithms like Nelder-Mead or Baum-Welch are mentioned. The authors inspected the estimated covariance matrix and for pairs of highly correlated parameter estimates, pairwise profile likelihood plots were plotted. A unimodal appearing likelihood surface while being quite flat in some directions is mentioned. The authors recorded these as checks for global maxima. It is noted that we require number of observations to be sufficiently large to have number of possible sequences of states exceeding the number of parameters to be estimated in the model.

Hypothesis tests of significance of different parameters were performed by using the idea of a penalised likelihood ratio test to account for non-identifiability having an approximate χ^2 distribution (Chen et al., 2001), in the current setup. Since the

Table 1: Parameter estimates for hidden Markov and phase-type HSMMs for BOS dataset

Parameter	HMM Estimate	HMM CI	HSMM Estimate	HSMM CI
λ_1			0.21	(0.05,0.95)
μ_{112}	0.12	(0.09,0.15)	0.36	(0.26,0.50)
μ_{113}	0.04	(0.03,0.07)	0.00	(0.00,0.002)
τ_{12}			0.37	(0.21,0.65)
λ_2			29.00	(0.00,82.17)
μ_{211}	0.03	(0.01,0.13)	1.00	(0.18,5.72)
μ_{213}	0.25	(0.20,0.30)	7.02	(0.79,1.94)
τ_{22}			0.01	(0.00,0.27)
e_{21}	0.10	(0.07,0.15)	0.00	(0.00,0.03)
e_{12}^{HL}	0	(0,1)	0.02	(0.02,0.04)
e_{12}^{DL}	0	(0,1)	0.56	(0.42,0.68)
π_{HL}	0	(0,1)	0.09	(0.05,0.14)
π_{DL}	0	(0,1)	0.76	(0.19,0.72)
$-LL$	-2533.91		-1617.978	
No. pars	9		13	

setup becomes that of a HMM, the authors used a Pearson-type goodness of fit test for HMMs proposed by them (2008) to assess the overall fit of the data.

4 Results

Table 1 shows MLEs and 95% confidence intervals for the parameters of the model. The MLEs were unique on the basis of numerical investigations with different starting values.

The first passage CDF for BOS development is reported in Fig 3a. The probability of an individual remaining BOS free at 5 years post transplant given initially healthy is 28%, with a 95% confidence interval of (22%, 38%). This is in agreement with existing literature where existing estimates of a 5 year disease free probability were between 15% and 37% (Chan, Allen; 2004). The model predicts that the rate of entry into the

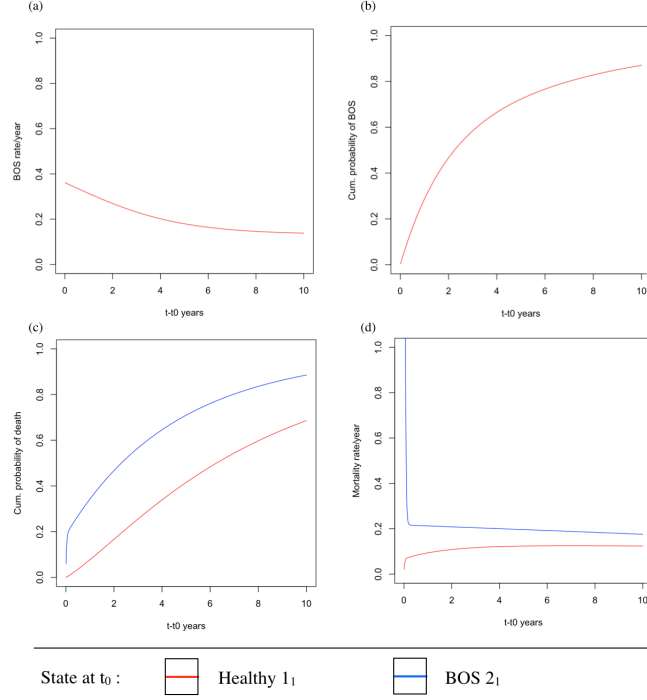
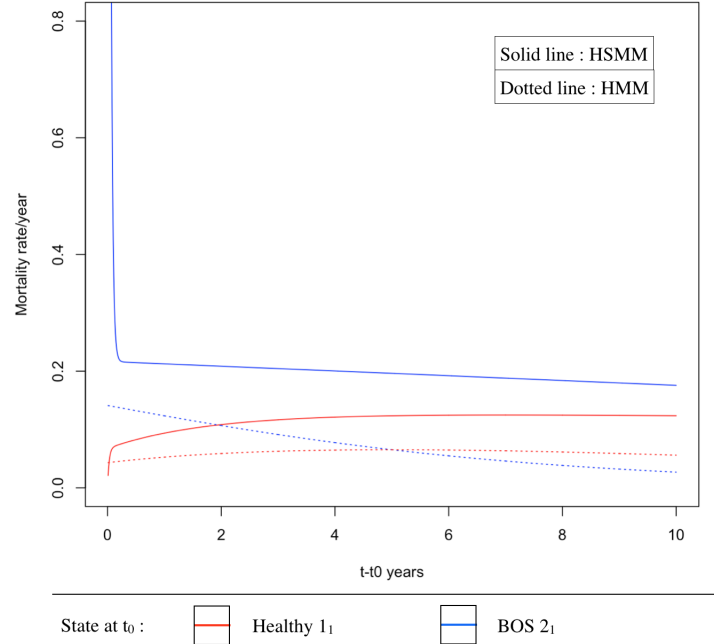


Figure 3: (a). Disease rate conditional on being in healthy state 1_1 at t_0 . (b). Cumulative probability of having transitioned to BOS state at least once, conditional on being in 1_1 at t_0 . (c). Cumulative probability of death. (d). Mortality rate per year, as a function of state at t_0 .

diseased states decrease with time since transplant; Initially disease rates are 35%40%, and decrease to 18% per year after 5 years (Figure 3b). Heterogeneity in the lung transplant patient population in terms of progression to BOS might have led to the nonconstant disease hazard of BOS. Decreasing BOS rates are agreeing with the fact that risk for experiencing infections or acute rejection episodes, which may trigger BOS development, is high at the initial period right after transplant (Jackson et al., 2002).

The cumulative probabilities of death given initial state is healthy versus BOS are plotted in Fig 3c. By 2 years post transplant, about 17% of those healthy at the beginning of the study are estimated to have died. After BOS initiation, about 66% remain alive for 1 year, 43% for 3 years, 29% for 5 years. These estimates are in agreement with the literature estimates of survival after bilateral lung transplant of 74%, 46%, and 26% at 1, 3, and 5 years after the onset of BOS, respectively (Finlen et al., 2010).



In Fig. 3d, we note that till before BOS onset, mortality rates are very low. After having BOS onset, mortality rates increase drastically to 50% per year, and then decrease to 20% after 2 years, followed by a gradual decline in rate. This trend is reminiscent of the identification of distinct BOS patient populations: those with acute onset and rapidly deteriorating lung function, and those with more gradual onset and slowly progressing disease (Jackson et. al, 2002; Lama et. al, 2007).

5 Discussion

References

- Alred, G. J., Brusaw, C. T., and Oliu, W. E. (2003 (seventh edition)). Handbook of Technical Writing. St. Martin's, New York.
- Goossens, M., Mittelbach, F., and Rahtz, S. (1997). The LaTeX Companion. Addison-Wesley, Reading, Mass.
- Gull, S. F. (1989). Developments in maximum-entropy data analysis. In Skilling, J., editor, Maximum Entropy and Bayesian Methods, pages 53–71. Kluwer Academic, Dordrecht.

- Hanson, K. M. (1993). Introduction to Bayesian image analysis. In Loew, M. H., editor, Medical Imaging: Image Processing, volume 1898 of Proc. SPIE, pages 716–731.
- Lamport, L. (1994). LaTeX: A Document Preparation System. Addison-Wesley, Reading, Mass.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machine. J. Chem. Phys., 21:1087–1091.
- Mittelbach, F., Goossens, M., Braams, J., and Carlisle, D. (2004 (second edition)). The LaTeX Companion. Addison-Wesley, Reading, Mass.
- Perelman, L. C., Paradis, J., and Barrett, E. (1997). Mayfield Handbook of Technical and Scientific Writing. Mountain View, Mayfield. <http://mit.imoat.net/handbook/>.