# Web-based Supporting Materials for Fitting and Interpreting Continuous-Time Latent Markov Models for Panel Data

by Jane M. Lange and Vladimir N. Minin

## Appendix A: Complete data score and Hessian

Note: All vectors are assumed to be column vectors unless otherwise noted.

### CTMC parameters

The CTMC log-likelihood component is in the curved exponential family, with natural parameters $\log(\lambda_{ij})$ and $\sum_{i \neq j} \lambda_{ij}$ corresponding to sufficient statistics $n_T(i,j)$ and $d_T(i)$. Individual level baseline covariates $\mathbf{w}^h$ are added via $\log(\lambda_{ij}^h) = \boldsymbol{\beta}_{ij}^T \mathbf{w}^h$, where $h$ denotes the individual and $\mathbf{w}^h$ and $\boldsymbol{\beta}_{ij}$ are $p$-dimensional vectors corresponding to $p$ covariates. For convenience, we list the intensity parameters $\{log(\lambda_{ij}) : i, j \in S; i \neq j\}$ as a $q$-dimensional vector $\boldsymbol{\psi}$, indexing each $i, j$ pair in $\boldsymbol{\psi}$ by $u$. This allows us to derive the score and information for all intensity parameters simultaneously, which is particularly useful if one assumes the same covariate effect for more than one transition intensity. Using the notation $i[u]$ and $j[u]$ to yield the $i$ and $j$ corresponding to $u$, the $u$th entry of the vector score function for $\boldsymbol{\psi}$ is

$$\dot{l}(\boldsymbol{\psi})[u] = n_T\left(i[u], j[u]\right) - d_T\left(i[u]\right) exp\left(\psi[u]\right).$$

The Hessian matrix for $\boldsymbol{\psi}$ is diagonal with non-zero entries

$$\ddot{l}(\boldsymbol{\psi})[u,u] = -d_T\left(i[u]\right) exp\left(\psi[u]\right).$$

The score function when the rate matrix is parameterized with covariates $\mathbf{w}$ is given by

$$\dot{l}(\boldsymbol{\beta}|\mathbf{w}^h) = \nabla\boldsymbol{\psi}(\boldsymbol{\beta})^T \dot{l}\{\boldsymbol{\psi}(\boldsymbol{\beta})\},$$

where $\nabla\boldsymbol{\psi}(\boldsymbol{\beta})^T$ is the $p \times q$ matrix whose $m, u$ entry corresponds to $\frac{\partial\psi[u]}{\partial\beta[m]}$. The Hessian matrix in the presence of covariates is

$$\ddot{l}(\boldsymbol{\beta}|\mathbf{w}^h)[j,m] = -\sum_{u=1}^{q} \frac{\partial\psi[u]}{\partial\beta[j]} \frac{\partial\psi[u]}{\partial\beta[m]} d_T\left(i[u]\right) exp\left(\psi[u]\right).$$

In matrix form, this can be written as

$$\ddot{l}(\boldsymbol{\beta}|\mathbf{w}^h) = \nabla\boldsymbol{\psi}(\boldsymbol{\beta})^T (\nabla\boldsymbol{\psi}(\boldsymbol{\beta})) \circ \mathbf{D},$$

where $\mathbf{D}$ is a $q \times p$ matrix with each column consisting of column vector $\mathbf{v}$, such that entries $v[u] = -exp(\psi[u]) d_T\left(i[u]\right)$, and $\circ$ refers to the Hadamard (element-wise) product. Both the score and Hessian are additive across subjects, so the total score and Hessian are obtained by summing over corresponding subject-specific quantities.

### Initial and Emission distributions parameters

We limit our attention to the score and Hessian for the emission distribution, as the initial distribution is analogous. For a single subject,
$\mathbf{O}_T(i) = \{O_T(i,1), \ldots, O_T(i,r)\} \sim Multinomial\{\mathbf{e}_i, N(i)\}$, where $N(i) = \sum_{j=1}^{r} O_T(i,j)$ and $\boldsymbol{e}_i = \{e(i,1), \ldots, e(i,r)\}$. Sufficient statistics include the $r-1$ length vector
$\mathbf{o_{i[-1]}} = \{o_T(i,2), ..., o_T(i,r)\}$. The natural parameters are $\left\{\eta_{ij} = \log\left(\frac{e(i,j)}{e(i,1)}\right) : j = 2, ..., r\right\}$. In the absence of covariates, the score function for the parameters $\boldsymbol{\eta}_i = (\eta_{i2}, ..., \eta_{ir})$ is

$$\dot{l}(\boldsymbol{\eta}_i) = \mathbf{o_{i[-1]}} - N(i)\mathbf{e_{i[-1]}},$$

where $\mathbf{e}_{i[-1]} = \{e(i,2), ...e(i,r)\}$ is a $r-1$ length vector of emission probabilities written in terms of $\boldsymbol{\eta}_i$.

Subject-level covariates $\mathbf{w}_i^h$ are added to the model via $\eta_{ij}^h = \boldsymbol{\gamma}_{ij}^T \mathbf{w}_i^h$, where $h$ indexes the individual. Let $\boldsymbol{\gamma}_i = (\boldsymbol{\gamma}_{i2}, \ldots, \boldsymbol{\gamma}_{ir})$ be the vector of all $p$ covariate parameters. The score is

$$\dot{l}(\boldsymbol{\gamma}_i|\mathbf{w}^h) = \nabla\boldsymbol{\eta}_i(\boldsymbol{\gamma}_i)^T\{\mathbf{o_{i[-1]}} - N_i\boldsymbol{e}_{i[-1])}\},$$

where $\nabla\boldsymbol{\eta}_i(\boldsymbol{\gamma}_i)^T$ is the $p \times (r-1)$ matrix of partial derivatives of $\boldsymbol{\eta}_i$ with respect to $\boldsymbol{\gamma}_i$ and $\boldsymbol{e}_{i[-1])}$ is written in terms of $\boldsymbol{\gamma}_i$. The Hessian matrix in the absence of covariates is given by

$$\ddot{l}(\boldsymbol{\eta}_i) = -\operatorname{Cov}(\mathbf{o_{i[-1]}}).$$

With covariates, the Hessian matrix is given by

$$\ddot{l}(\boldsymbol{\gamma}_i|\mathbf{w}^h) = -\left\{\nabla\boldsymbol{\eta}_i(\boldsymbol{\gamma}_i)^T \operatorname{Cov}(\mathbf{o_{i[-1]}})\nabla(\boldsymbol{\eta}_i(\boldsymbol{\gamma}_i)\right\}.$$

As before, the total score and Hessian are obtained by summing over the corresponding subject-specific quantitites.

## Appendix B: Recursions for hidden Markov models

Throughout, we abbreviate $x_1, \ldots, x_k$ by $\mathbf{x}_{1:k}$ and $o_1, \ldots o_k$ by $\mathbf{o}_{1:k}$.

### Forward and backward probabilities

Forward probabilities are defined as $\alpha_k(u) = \mathrm{P}(\mathbf{o}_{1:k}, X_k = u)$ and backward probabilities as $\beta_k(u) = \mathrm{P}(\mathbf{o}_{k+1:n}|X_k = u)$. When the last time coincides with the time of absorption, Y, the forward and backward probabilities are defined as before, with the exception that $\beta_k(u) = \frac{\partial}{\partial y}\mathrm{P}(\mathbf{o}_{k+1:n}, Y < y|X_k = u)$ and $\alpha_n(u) = \frac{\partial}{\partial y}\mathrm{P}(\mathbf{o}_{k+1:n}, Y < y)$, Forward and backward probabilities are calculated through Baum's recursive formulae [1].

## Filtering and conditional likelihood calculations

Filtering probabilities, $\mathrm{P}(X_k = j|\mathbf{o}_{1:k})$ and the conditional observed data likelihood $\mathrm{P}(O_k = o_k|\mathbf{o}_{1:k-1})$ are related to modified forward probabilities, $a_k(j) = \mathrm{P}(X_k = j, O_k = o_k|\mathbf{o}_{1:k-1})$. That is, $\mathrm{P}(O_k = o_k|\mathbf{o}_{1:k-1}) = \sum_{j\in S} a_k(j)$, and $\mathrm{P}(X_k = j|\mathbf{o}_{1:k}) = \frac{a_k(j)}{\sum_{l\in S} a_k(l)}$. The modified forward probabilities can be calculated recursively. Initialize

$$a_1(j) = \mathrm{P}(O_1 = o_1, X_1 = x_1) = e(x_1, o_1)\pi(x_1),$$

and the recursion is

$$a_{k+1}(j) = \sum_{i\in S} \frac{a_k(i)}{\sum_l a_k(l)} e(x_{k+1}, o_{k+1})\, \mathrm{P}_{ij}(t_{k+1} - t_k).$$

## Recursive smoothing for first moments of complete data sufficient statistics

First moment calculations define entries of $s_k(x_k, x_{k+1})$ as values of complete data sufficient statistics (section 2.4.1) on the interval $T_k = [t_k, t_{k+1}]$, conditional on $x_k$ and $x_{k+1}$. Thus, $s_k(x_k, x_{k+1})$ is defined as $\mathrm{E}[d_{T_k}|X_k = x_k, X_{k+1} = x_{k+1}]$ for entries corresponding to $d_T(i)$; as $\mathrm{E}[n_{T_k}(i,j)|X_k = x_k, X_{k+1} = x_{k+1}]$ for $n_T(i,j)$; 0 for $z_i$; and as $I(X_{k+1} = i, O_{k+1} = j)$ for $o_T(i,j)$. Initial values for the function $t_k(\mathbf{x}_{1:k})$ are set at $t_1(x_1) = 0$ for entries corresponding to $d_T(i)$ and $n_T(i,j)$; $I(X_1 = i)$ for $z_i$; and $I(X_1 = i, O_1 = j)$ for $o_T(i,j)$.

## Recursive smoothing for second moments of complete data sufficient statistics

The recursive smoothing method to obtain second and cross moments of complete data sufficient statistics conditional on the entirety of a subject's observed data, $\mathbf{o}$, proceeds with a similar framework and terminology as for first moments (Section 3.2.3.) First, we recursively define a functional that corresponds to $\mathrm{E}[\mathbf{S}[t_1, t_k]\mathbf{S}[t_1, t_k]^T|\mathbf{x}_{1:k}]$, the second moments of complete sufficient statistics on the interval $[t_1, t_k]$, conditional on $\mathbf{x}_{1:k}$. Next, we define the recursive updates of the auxiliary function, $\tau_k(x_k)$. Finally, we compute the auxiliary function updates for $t_1, \ldots, t_n$, enabling us to calculate the target quantity $\mathrm{E}[\mathbf{S}[t_1, t_n]\mathbf{S}[t_1, t_n]^T|\mathbf{o}_{1:n}]$.

The recursive definition of $\mathrm{E}[\mathbf{S}[t_1, t_{k+1}]\mathbf{S}[t_1, t_{k+1}]^T|\mathbf{x}_{1:k+1}]$ involves not only $\mathrm{E}[\mathbf{S}[t_1, t_k]\mathbf{S}[t_1, t_k]^T|\mathbf{x}_{1:k}]$ but also the first moment, $\mathrm{E}[\mathbf{S}[t_1, t_k]|\mathbf{x}_{1:k}]$. Thus it makes sense to consider jointly the first and second moments of complete data sufficient statistics conditional on $\boldsymbol{x}_{1:k}$. We define the joint recursive function of latent states as

$$\mathbf{t}(\mathbf{x}_{1:k+1}) = \left\{ t^{(1)}(\mathbf{x}_{1:k+1}), t^{(2)}(\mathbf{x}_{1:k+1}) \right\},$$

where

$$\begin{aligned} t^{(1)}(\mathbf{x}_{1:k+1}) &= \mathrm{E}[\mathbf{S}[t_1, t_{k+1}]|\mathbf{x}_{1:k+1}] \\ &= t^{(1)}(\mathbf{x}_{1:k}) + \mathrm{E}[\mathbf{S}[t_k, t_{k+1}]|x_k, x_{k+1}] \end{aligned}$$

and

$$\begin{aligned} t^{(2)}(\mathbf{x}_{1:k+1}) &= \mathrm{E}[\mathbf{S}[t_1, t_{k+1}]\mathbf{S}[t_1, t_{k+1}]^T|\mathbf{x}_{1:k+1}] \\ &= t^{(2)}(\mathbf{x}_{1:k}) + \mathrm{E}[\mathbf{S}[t_k, t_{k+1}]|x_k, x_{k+1}]t^{(1)}(\mathbf{x}_{1:k})^T + t^{(1)}(\mathbf{x}_{1:k})\,\mathrm{E}[\mathbf{S}[t_k, t_{k+1}]|x_k, x_{k+1}]^T \\ &+ \mathrm{E}[\mathbf{S}[t_k, t_{k+1}]\mathbf{S}[t_k, t_{k+1}]^T|x_k, x_{k+1}]. \end{aligned}$$

The first component is identical to first moment recursive function (eq. (3) in the main text); the second corresponds to second and cross moments of complete data sufficient statistics conditional on latent states $\mathbf{x}_{1:k}$. The calculation of $t^{(2)}(\mathbf{x}_{1:k+1})$ follows from the conditional independence of $\mathbf{S}[t_l, t_{l+1}]$ and $\mathbf{S}[t_j, t_{j+1}]$ given the endpoints $x_l, x_{l+1}, x_j, x_{j+1}$ and the fact that $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$ if X and Y are independent. We assign the function

$$\begin{aligned} \mathbf{s}_k(x_k, x_{k+1}) &= \left\{ \boldsymbol{s}_k^{(1)}(x_k, x_{k+1}), \boldsymbol{s}_k^{(2)}(x_k, x_{k+1}) \right\} \\ &= \left\{ E\left[\mathbf{S}[t_k, t_{k+1}]|x_k, x_{k+1}\right], \mathrm{E}\left[\mathbf{S}[t_k, t_{k+1}]\mathbf{S}[t_k, t_{k+1}]^T|x_k, x_{k+1}\right] \right\}. \end{aligned}$$

The specific values of $t_1^{(1)}(x_1)$ and $s_k^{(1)}(x_k, x_{k+1})$ for latent CTMC sufficient statistics were provided previously. Appendix Table 1 summarizes specific details of $s_k^{(2)}(x_k, x_{k+1})$ and $t_1^{(2)}(x_1)$ for all pairs of latent CTMC complete data sufficient statistics.

The auxiliary functions likewise have two components corresponding to first and second moments: $\boldsymbol{\tau}(x_k) = \left\{\tau^{(1)}(x_k), \tau^{(2)}(x_k)\right\}$. The updates for $\boldsymbol{\tau}(x_k)$ follow from a multivariate version of eq. (4) in the main text. The $\tau^{(1)}(x_k)$ component is defined as in eq. (4). The $\tau^{(2)}(x_k)$ component is defined recursively as

$$\begin{aligned} \tau_{k+1}^{(2)}(x_{k+1}) = \mathrm{P}(o_{k+1}|\mathbf{o}_{1:k})^{-1} \Bigg\{ \sum_{x_k} [\tau^{(2)}(x_k) + \tau_k^{(1)}(x_k) s_k^{(1)}(x_k, x_{k+1})^T \\ + s_k^{(1)}(x_k, x_{k+1})\tau_k^{(1)}(x_k)^T + \mathrm{P}(X_k = x_k|\mathbf{o}_{1:k})\,\mathrm{E}[\mathbf{S}[t_k, t_{k+1}]\mathbf{S}[t_k, t_{k+1}]^T|x_k, x_{k+1}]] \\ \times e(x_{k+1}, o_{k+1})\,\mathrm{P}_{x_k x_{k+1}}(t_{k+1} - t_k) \Bigg\}. \end{aligned}$$

The final recursion allows us to calculate $\mathrm{E}[t_n^{(2)}(\mathbf{x}_{1:n})|\mathbf{o}_{1:k}] = \sum_{x_n \in X} \tau_n^{(2)}(x_n)$, giving us the expected value of second moments of complete data sufficient statistics conditional on the observed data.

Table 1: Definition of $s_k^{(2)}(x_k, x_{k+1})$ and $t_1^{(2)}(x_1)$ for second moment calculations.

| Statistics | $s^{(2)}(x_k, x_{k+1})$ | $t_1^{(2)}(x_1)$ |
|---|---|---|
| $d_T(i), d_T(j)$ | $\mathrm{E}[d_{T_k}(i)d_{T_k}(j)\|x_k, x_{k+1}]$ | 0 |
| $d_T(i), n_T(j, m)$ | $\mathrm{E}[d_{T_k}(i)n_{T_k}(j, m)\|x_k, x_{k+1}]$ | 0 |
| $d_T(i), o_T(j, m)$ | $\mathrm{E}[d_{T_k}(i)I(X_{k+1} = j, O_{k+1} = m)\|x_k, x_{k+1}]$ | 0 |
| $n_T(i, l), n_T(j, m)$ | $\mathrm{E}[n_{T_k}(i, l)n_{T_k}(j, m)\|x_k, x_{k+1}]$ | 0 |
| $o_T(j, m), o_T(l, r)$ | $I(X_{k+1} = j, O_{k+1} = m, X_{k+1} = l, O_{k+1} = r)$ | $I(X_1 = j, O_1 = m, X_1 = l, O_1 = r)$ |
| $n_T(i, l), o_T(l, r)$ | $\mathrm{E}[n_{T_k}(i, l)I(X_{k+1} = l, O_{k+1} = r)\|x_k, x_{k+1}]$ | |
| $z_i, z_m$ | 0 | $I(X_1 = i)I(X_1 = m)$ |
| $z_i, o_T(j, m)$ | 0 | $I(X_1 = i)I(X_1 = j, O_1 = m)$ |
| $n_T(j, m), z_i$ | 0 | 0 |
| $d_T(j), z_i$ | 0 | 0 |

## Appendix C: Differentiated joint moments of transitions and state occupancy durations with known absorption times

We assume that the CTMC has one absorbing state $g$. Differentiated joint moments in the presence of known absorption times rely on the fact that if an individual is absorbed at time $t$, transitions to $g$ occur only once and no time is spent in $g$. These joint moments formulae use the joint moments defined in Section 3.2.1, which we refer to as $M_{ij}(t)[a,b] = \mathrm{E}[n_t(i,j)I(X_0 = a)|X_t = b]$; $H_i(t)[a,b] = \mathrm{E}[d_t(i)I(X_t = b)|X_0 = a]$; $U_{ijlm}(t)[a,c] = \mathrm{E}[n_t(i,j)n_t(l,m)I(X_t = c)|X_0 = a]$; $W_{ij}(t)[a,c] = \mathrm{E}[d_t(i)d_t(j)I(X_t = c)|X_0 = a]$; and $V_{ilm}(t)[a,c] = \mathrm{E}[d_t(i)n_t(l,m))I(X_t = c)|X_0 = a]$.

When the complete-data statistic of interest is $S = d_t(i)$, the differentiated joint moment is given by

$$\frac{\partial}{\partial y}\,\mathrm{E}[d_t(i)I(Y < t)|X_0 = a] = I(i \neq g)\sum_{c\neq g} H_i(t)[a,c]\lambda_{cg}.$$

When $S = d_t(i)d_t(j)$, the differentiated joint expectation is identical, except $I(i \neq g)$ is replaced by $I(i,j \neq g)$, and $H_i(t)[a,c]$ is replaced by the duration cross moment $W_{ij}(t)[a,c]$.

For $S = n_t(i,j)$, the differentiated joint expectation is

$$\frac{\partial}{\partial y}\,\mathrm{E}[n_t(i,j)I(Y < y)|X_0 = a] = I(i,j \neq k)\sum_{c\neq k} M_{ij}(t)[a,c]\lambda_{ck} + I(i \neq k, j = k)\,\mathrm{P}_{ai}(t)\lambda_{ik}.$$

For $S = n_t(i,j)n_t(l,m)$ the differentiated joint expectation is given by

$$\begin{aligned}\frac{\partial}{\partial y}\,\mathrm{E}[n_t(i,j)n_t(l,m)I(Y < y)|X_0 = a] &= I(i,j,l,m \neq g)\sum_{c\neq g} U_{ijlm}(t)[a,c]\lambda_{cg}\\ &+ I(i,l,m \neq g, j = g)M_{lm}(t)[a,i]\lambda_{ig} + I(i,j,l \neq g, m = g)M_{ij}(t)[a,l]\lambda_{lg}\\ &+ I(i,l \neq g, i = l, j = m = g)\,\mathrm{P}_{ai}(t)\lambda_{ig}.\end{aligned}$$

For $S = n_t(l,m)d_t(i)$, the differentiated joint expectation is given by

$$\frac{\partial}{\partial y}\,\mathrm{E}[d_t(i)n_t(l,m)I(Y < y)|X_0 = a] = I(i,j,l,\neq g)\sum_{c\neq g} V_{ilm}(t)[a,c]\lambda_{cg} + I(i,l \neq g, m = g)H_i(t)[a,l]\lambda_{lm}.$$

## Appendix D: Delta method standard errors of disease process functionals

Suppose $\boldsymbol{\psi}$ is a $p\times 1$ vector of latent model parameters with MLE $\hat{\boldsymbol{\psi}}$, and $F(\boldsymbol{\psi},t)$ is a one-dimensional functional. Let $\nabla F(\hat{\boldsymbol{\psi}},t)$ be the $p\times 1$ gradient of $F(\boldsymbol{\psi},t)$ with respect to $\boldsymbol{\psi}$ evaluated at $\hat{\boldsymbol{\psi}}$. The asymptotic distribution of the functional estimates $F(\hat{\boldsymbol{\psi}},t)$ is normal with mean $F(\boldsymbol{\psi},t)$ and an approximate covariance matrix given by

$$\mathrm{Cov}(F(\hat{\boldsymbol{\psi}},t)) = \nabla F(\hat{\boldsymbol{\psi}},t)^T\,\mathrm{Cov}(\hat{\boldsymbol{\psi}},t)\nabla F(\hat{\boldsymbol{\psi}},t).$$

Functionals such as CDFs, hazard functions, and transition probabilities involve the matrix exponential; thus we require the derivative of $\exp(\mathbf{\Lambda}(\boldsymbol{\psi})t)$ with respect to entries of $\boldsymbol{\psi}$. These derivative involve similar integrals as first moments of occupancy durations and transition counts (Section 3.2.1) and are computed with similar methods [2]. For example, consider the functional $\mathrm{P}_{ij}(t, \boldsymbol{\psi}) = \exp(\mathbf{\Lambda}(\boldsymbol{\psi})t)$. Then $\frac{\partial \mathrm{P}_{ij}(t,\boldsymbol{\psi})}{\partial \psi[k]}$ is the $i, j$ entry of the matrix given by $\int_0^t e^{\mathbf{\Lambda}(\boldsymbol{\psi})\tau} \mathbf{B}_{\psi[k]} e^{\mathbf{\Lambda}(\boldsymbol{\psi})(t-\tau)} d\tau$, where $\mathbf{B}_{\psi[k]} = \{B_{\psi[k]}(i,j)\}$ and $B_{\psi[k]}(i,j) = \frac{\partial \lambda_{ij}(\boldsymbol{\psi})}{\partial_{\psi[k]}}$.

## Appendix E: Detailed discussion of simulation results

**Figure 3A in the main text shows the mean of point estimates of the Weibull (1.5,1) and Weibull (.75,10) hazard and CDF functions from the different latent CTMC models. Figure 3B in the main text shows the bias of the same point estimates. The x-axis is sojourn time, and x-limits were chosen to zoom in on the early portion of the sojourn time period. The bias in approximations reflects the closeness of the data-generating distribution to that of the latent CTMC model as well as the functional to be estimated. CDFs were generally less biased than hazard functions. Latent CTMC hazard estimates may be quite biased at the times corresponding to the distribution's tail, when latent CTMC hazards are asymptotically constant. Accordingly, the bias of the latent CTMC approximation of the Weibull (1.5,1) hazard is small until $t = 1$, but increases considerably at times thereafter. In contrast, the Weibull (.75,10) hazard function decreases and flattens out over time, as does the bias of the latent CTMC estimates.**

**Interestingly, discrete sampling schemes may also lead to increased bias in estimates of both hazard functions and CDFs. Both model II and IV characterize sojourn distributions by 2 transient latent states. We expected that the mean of model II and IV estimates of hazard and CDFs would be similar given that they assume the same latent CTMC model for each disease state sojourn time, although estimates would be more variable due to missing information. This was in fact true for estimates of Weibull(1.5,1) hazard functions and CDFs. However, bias of estimates of Weibull(.75, 10) hazards and CDFs depended on whether the data were survival data or discretely observed. In particular, Model IV was poor at estimating the early portion of the hazard function. We suspect this bias is related to the frequency of the sampling scheme relative to the rate of change of the hazard function and that the bias would be mitigated by more closely spaced observations.**

**We expected there would be a bias-variance tradeoff to adding more latent states to the latent CTMC model. In fact, model III (with 3 latent states), did have less biased estimates of hazard and CDF relative to model II. Model III estimates did have somewhat higher variance (not shown); and overall, the RMSE of the estimates (Figure 3C in the main text) from model II and III were quite similar. The one exception was for the tail end of the Weibull(1.5, 1) hazard function, when model III's estimates were considerably less biased. Overall, on the basis of the RMSE of point estimates, there is little to recommend model III over model II. We expect that adding more states**

to the model (e.g., 4 versus 3) would yield more variable estimates, and RMSE would favor models with III states. This was borne out by limited investigations with such models (results not shown).

Our investigations of delta-method standard errors on average represented 92% of the true variability of the estimates, but performance varied by model, functional, time, and data generating distribution (Appendix Figure A-1). Generally, delta method standard errors from model III better reflected estimate variability than model II. Coverage of 95% confidence intervals based on delta-method standard errors is shown in Figure 3D of the main text. Again, performance was quite mixed. Nominal coverage was attained when the bias was small and the delta-method standard errors provided good approximations of the true variability of the estimates. Poor coverage resulted when point estimates were quite biased (Weibull(1.5,1) hazards for $t > 1.5$), or when the delta-method standard errors underestimated the true variability of the estimates (Appendix Figure A-1), as in Weibull(75,10) CDF and hazard functions. Coverage of model IV estimates for small t was also poor for Weibull(1.5,1) functionals at t near 0, which appeared to be due to skewness in the estimates' distributions at this boundary.

## Appendix F: Additional models fit to BOS data

In addition to the model presented in the main text, we fit two additional models to the BOS data. The first alternative model (model 1) had $\lambda_{2_1 1_1}$ and $\lambda_{2_2 1_1}$ set to zero. Alternative model 2 had three latent healthy states, $1_1, 1_2, 1_3$, and no reversible transitions. Other aspects of the model 1 and 2 were identical to the original model. After fitting these models, model 1 had a maximum log-likelihood of -1251.757 and model 2, a maximum likelihood of -1251.389. Figure A-2 shows the estimated disease rate and first passage distribution to BOS onset under the original and two alternative models. While the point estimates for the hazard rates differ slightly between models, the first passage distributions for BOS onset are virtually indistinguishable. There was also evidence that the intensity parameters in model 2 were not uniquely identifiable: the information matrix was not positive definite at the maximum log-likelihood, and two separate runs of the algorithm from different starting values generated distinct sets of intensity rate estimates at log-likelihoods that differed by $< .3$.

We compared model 1 to the original model via a likelihood ratio test, as they are nested. The null hypothesis (the transition between BOS to healthy states has a rate of zero) is on the boundary of the parameter space; thus the distribution of the likelihood ratio test statistic is not a standard chi-square distribution, but rather a 50:50 mixture of a chi-square with 1 df and a point mass at zero [3]. The p-value for the LR test statistic was .006, which supports the hypothesis that the reversible transition improves the fit. On the basis of the Bayesian information criterion (BIC), the original model edges ahead of model 1 and 2, although the difference in BIC between the original model and model 1's is $< .5$ (table 3). This suggests that it is possible that the original model with reversible transitions may be over-fitting the data.

**The evidence that a model with reversible Healthy to BOS transitions apparently offered a modest but statistically significant improvement in fit may seem unappealing given the biology of the disease. However, both misclassification and disease progression are reflected in the transitions observed in the data. It is possible that our model has not correctly specified the misclassification process. BOS is diagnosed with FEV1, a continuous measure with inherent variability. Our model assumes that misclassification probabilities are constant over the course of the disease. Misclassified disease outcomes are more likely to occur in individuals who have recently developed BOS, since their FEV1 may be near the diagnostic cutoff. The model that allows for BOS $\rightarrow$ Healthy transitions may reflect non-constant misclassification probabilities. Examining this scenario via simulation would allow one to assess this hypothesis in more depth.**

## References


[1] Baum L, Petrie T, Soules G, Weiss N. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Annals of Mathematical Statistics 1970; **41**(1):164–171.

[2] Najfeld I, Havel TF. Derivatives of the matrix exponential and their computation. Advances in Applied Mathematics 1994; **16**:321–375.

[3] Self SG, Liang KY. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of the American Statistical Association 1987; **82**(398):605–610, doi:10.1080/01621459.1987.10478472.

Table 2: Maximum likelihood estimates of BOS model intensity rates, emission probabilities, and initial probabilities.

| | Transition | | | | |
|---|---|---|---|---|---|
| Intensity rates | i | j | Point estimate | 95% CI | |
| | $1_1$ | $1_2$ | 0.39 | 0.11 | 1.42 |
| | $1_1$ | $2_1$ | 0.39 | 0.27 | 0.56 |
| | $1_1$ | 3 | 0.01 | 0 | 0.29 |
| | $1_2$ | $2_1$ | 0.14 | 0.09 | 0.23 |
| | $1_2$ | 3 | 0.004 | 0.00017 | 0.11 |
| | $2_1$ | $1_1$ | 0.06 | 0.01 | 0.31 |
| | $2_1$ | $2_2$ | 3.12 | 0.97 | 9.99 |
| | $2_1$ | 3 | 0.73 | 0.27 | 1.94 |
| | $2_2$ | $1_1$ | 0.02 | 0.004 | 0.06 |
| | $2_2$ | 3 | 0.19 | 0.15 | 0.23 |
| Emission | e(Healthy,BOS) | Double lung | 0.076 | 0.042 | 0.133 |
| | | Heart lung | 0.018 | 0.01 | 0.031 |
| | e(BOS,Healthy) | | 0.011 | 0.004 | 0.028 |
| Initial Distribution | $\pi(BOS_1)$ | Heart-lung | 0.061 | 0.035 | 0.103 |
| | | Double lung | 0.043 | 0.014 | 0.124 |

Table 3: **Summary of models fit to BOS data with and without BOS → Healthy transitions**

| | BOS → Healthy transition | No. Healthy latent states | No. parameters | Log likelihood | BIC |
|---|---|---|---|---|---|
| Original model | Yes | 2 | 13 | -1248.602 | 2572.03 |
| Model 1 | No | 2 | 12 | -1251.757 | 2572.58 |
| Model 2 | No | 3 | 11 | -1251.389 | 2583.36 |

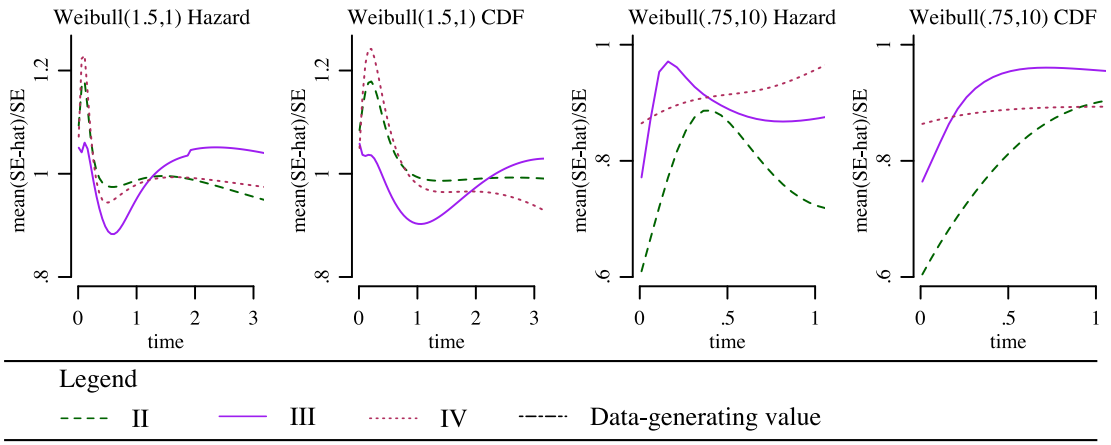


Figure A-1: Ratio of average delta-method standard errors to the empirical standard errors of the estimates from simulated data. Models II and III fit survival data with Coxian PH models with 2 and 3 transient states, respectively; Model IV fits discretely observed data from a 2-state reversible model assuming sojourn distributions analogous to model II.

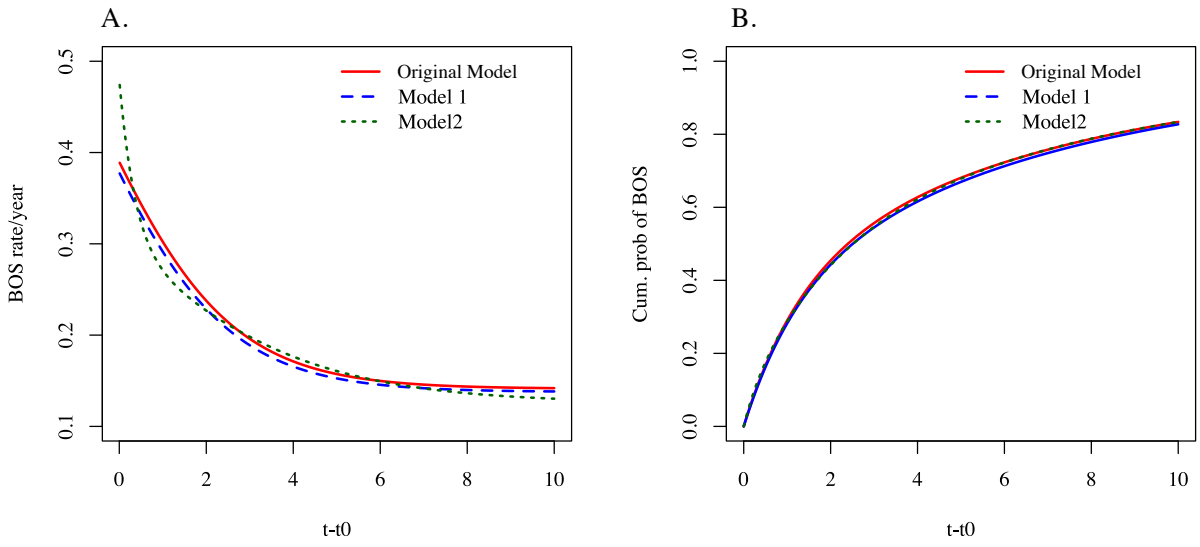


Figure A-2: **Point estimates for the BOS onset hazard rate (A) and CDF (B) with the original model and models without reverse BOS→ healthy transitions. The first alternative model (model 1) had $\lambda_{2_1 1_1}$ and $\lambda_{2_2 1_1}$ set to zero. Alternative model 2 had three latent healthy states, $1_1, 1_2, 1_3$, and no reversible transitions.**

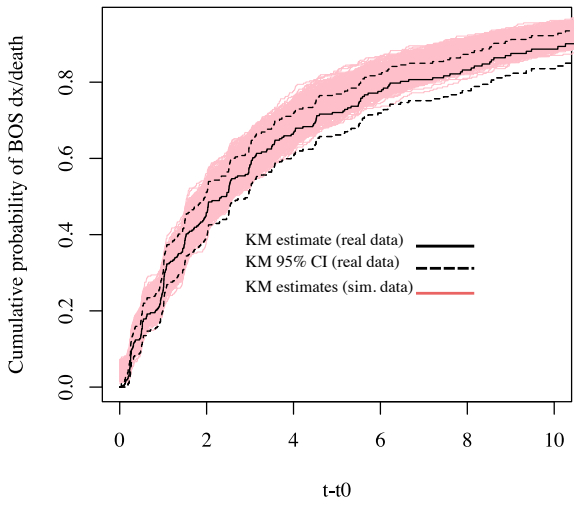


Figure A-3: **Kaplan-Meier estimates of time to first diagnosis of BOS or death from both the real data set and data simulated from the fitted model. Dashed lines are the 95% confidence intervals of the real data estimates.**