# Web-based Supplementary Materials for Bayesian variable selection for multistate Markov models with interval-censored data in an ecological momentary assessment of an attempt to quit smoking

by Matthew D. Koslovsky, Michael D. Swartz, Wenyaw Chan, Luis Leon-Novelo, Anna V. Wilkinson, Darla E. Kendzor, and Michael S. Businelle

## Web Appendix

*Web Appendix A: Convergence Stopping Rule*

To determine convergence of our proposed method, we set a stopping rule for the absolute value of the difference between the log-likelihood distribution evaluated at the current and next step of the algorithm (Wu, 1983). Formally, the algorithm stops if

$$\left|\ell\left(\boldsymbol{\Phi}^{(k+1)}|\boldsymbol{y}\right)-\ell\left(\boldsymbol{\Phi}^{(k)}|\boldsymbol{y}\right)\right|=\left|\left[Q\left[\boldsymbol{\Phi}^{(k+1)}|\boldsymbol{\Phi}^{(k)}\right]-Q\left[\boldsymbol{\Phi}^{(k)}|\boldsymbol{\Phi}^{(k)}\right]\right]-\left[R\left[\boldsymbol{\Phi}^{(k+1)}|\boldsymbol{\Phi}^{(k)}\right]-R\left[\boldsymbol{\Phi}^{(k)}|\boldsymbol{\Phi}^{(k)}\right]\right]\right|\leqslant\epsilon,$$

where $\ell$ is the log-likelihood function and $\epsilon$ is the stopping rule threshold. The Q-function is described in Eq. 4. To complete the derivation of our stopping rule, we define

$$R\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]=\sum_{\boldsymbol{\gamma}}\log\left(\pi(\boldsymbol{\gamma}|\boldsymbol{\Phi},\boldsymbol{y})\right)\times\pi\left(\boldsymbol{\gamma}|\boldsymbol{\Phi}^{(k)},\boldsymbol{y}\right)=E_{\boldsymbol{\gamma}|\cdot}\log(\pi(\boldsymbol{\gamma}|\boldsymbol{\Phi},\boldsymbol{y})).$$

Due to the hierarchical structure of this model, we assume that

$$\pi(\boldsymbol{\gamma}|\boldsymbol{\Phi},\boldsymbol{y})=\pi(\boldsymbol{\gamma}|\boldsymbol{\Phi})=\theta^{\sum_{r=1}^{p}(\gamma_{\lambda,r}+\gamma_{\mu,r})}(1-\theta)^{2p-\sum_{r=1}^{p}(\gamma_{\lambda,r}+\gamma_{\mu,r})}. \tag{A.1}$$

Taking the log and conditional expectation of A.1, we show

$$E_{\boldsymbol{\gamma}|\cdot}\log(\pi(\boldsymbol{\gamma}|\boldsymbol{\Phi}))=\sum_{r=1}^{p}\left[E_{\boldsymbol{\gamma}|\cdot}[\gamma_{\lambda,r}]+E_{\boldsymbol{\gamma}|\cdot}[\gamma_{\mu,r}]\right]\log(\theta)+\left(2p-\sum_{r=1}^{p}E_{\boldsymbol{\gamma}|\cdot}[\gamma_{\lambda,r}]+E_{\boldsymbol{\gamma}|\cdot}[\gamma_{\mu,r}]\right)\log(1-\theta).$$

Once the algorithm has converged, the final estimates, $\hat{\boldsymbol{\Phi}}$, maximize Eq. 4.

*Web Appendix B: Deterministic Annealing Variant*

The deterministic annealing EM algorithm (DAEM) redefines the EM algorithm's objective based on the principle of maximum entropy to reduce the algorithm's dependence on initial values. Instead of maximizing the Q-function, the new objective is to minimize the free energy function at gradually cooler temperatures. For EMVS, this corresponds to maximizing the negative free energy function

$$-\mathfrak{F}_t(\boldsymbol{\Phi}) = -\mathfrak{U}_t(\boldsymbol{\Phi}) + \frac{1}{t}\mathfrak{S}_t(\boldsymbol{\Phi}) = \frac{1}{t}\log\sum_{\boldsymbol{\gamma}}\left(\pi(\boldsymbol{\Phi},\boldsymbol{\gamma}|\boldsymbol{y})\right)^t,$$

where $\mathfrak{U}_t(\boldsymbol{\Phi})$ is the internal energy, $\mathfrak{S}_t(\boldsymbol{\Phi})$ is the entropy, and $\frac{1}{t}$ is interpreted as the temperature of the annealing process, with $0 < t < 1$. Equivalently, we can maximize $-\mathfrak{F}_t(\boldsymbol{\Phi})$ by iteratively maximizing the negative internal energy, conditioned on the current estimates of the unknown parameters, $\boldsymbol{\Phi}^{(k)}$ (Ueda and Nakano, 1998). The negative internal energy function is

$$-\mathfrak{U}_t(\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}) = \sum_{\boldsymbol{\gamma}}\log\left(\pi(\boldsymbol{\Phi},\boldsymbol{\gamma}|\boldsymbol{y})\right)\times\pi(\boldsymbol{\gamma}|\boldsymbol{\Phi}^{(k)})^t \tag{A.2}$$

In application, the annealing process starts with a high temperature (i.e., $t$ close to 0). When the temperature is high, the landscape of $-\mathfrak{F}_t(\boldsymbol{\Phi})$ is smooth, which prevents the EM algorithm from getting stuck in a local mode early in its iterative process. As the temperature iteratively cools, the effect of the inclusion posterior is strengthened. As a result, local modes begin to appear and the landscape of $-\mathfrak{F}_t(\boldsymbol{\Phi})$ progressively approaches the true incomplete posterior.

To formulate our ECM algorithm with a deterministic annealing variant, we introduce an annealing loop that regulates the influence of the inclusion posterior and replace Eq. 4 with equation Eq. A.2 in the E-step. The algorithm for this method is

**Step 1:** Set the initial $\boldsymbol{\Phi}^{(0)}$ and $t$.

**Step 2:** Carry out the ECM steps at the current temperature, $t$, until convergence:

(a) E-Step: Evaluate $\mathfrak{U}_t\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]$

(b) CM-steps: Perform steps 1-3 in Section 2.3, replacing Eq. 4 with Eq. A.2.

(c) Set $k \leftarrow k+1$

**Step 3:** Increase $t$.

**Step 4:** If $t < 1$, return to step 2 and use the final estimates given the previous $t$ to initiate at the current $t$. Else, stop.

Here, the conditional expectation is taken with respect to a tempered inclusion posterior distribution. The tempered probabilities of inclusion are calculated as

$$p^*_{r,t} = \frac{\left[\pi(\beta_r^{(k)}|\gamma_r = 1)P(\gamma_r = 1|\theta^{(k)})\right]^t}{\left[\pi(\beta_r^{(k)}|\gamma_r = 0)P(\gamma_r = 0|\theta^{(k)})\right]^t + \left[\pi(\beta_r^{(k)}|\gamma_r = 1)P(\gamma_r = 1|\theta^{(k)})\right]^t}.$$

At each cooling step, $t$, we find a global mode that is used to initiate the algorithm at the next temperature to find a new global mode. Assuming that the new global mode is close to the previous, the probability of converging at the true global mode is increased. Note that on the final annealing loop iteration ($t = 1$), Eq. 4 matches equation Eq. A.2. Thus, the parameter estimates that maximize the negative free-energy function are equivalent to the posterior mode estimates that maximize the log incomplete posterior. While convergence at the global mode is still not guaranteed, the variant removes the algorithm's dependence on initial parameter values and finds better estimates than the conventional EM algorithm (Ueda and Nakano, 1998).

*Web Appendix C: Variance Estimation*

We present the closed-form expressions for $\partial^2 Q\left[\boldsymbol{\Phi}|\hat{\boldsymbol{\Phi}}\right]/\partial\boldsymbol{\Phi}\partial\boldsymbol{\Phi}'$ and $\text{var}\{\partial \log \pi(\boldsymbol{\Phi}, \boldsymbol{\gamma}|\boldsymbol{y})/\partial\boldsymbol{\Phi}|\hat{\boldsymbol{\Phi}}, \boldsymbol{y}\}$ from Eq. 5 in the main article. For simplicity, we redefine the Q-function as

$$\begin{aligned} Q\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right] = C &+ \sum_{i=1}^{m}\sum_{j=2}^{n_i}\left[\log P_{y_i(t_{i,j-1}),y_i(t_{i,j})}\left(\delta_{ij}|\boldsymbol{x}_{i,j-1}\right)\right] \\ &+ \sum_{r=0}^{p} -\frac{1}{2}\left\{\beta^2_{\lambda,r}E_{\boldsymbol{\gamma}|\cdot}\left[\frac{1}{v_0(1-\gamma_{\lambda,r}) + v_1\gamma_{\lambda,r}}\right] + \beta^2_{\mu,r}E_{\boldsymbol{\gamma}|\cdot}\left[\frac{1}{v_0(1-\gamma_{\mu,r}) + v_1\gamma_{\mu,r}}\right]\right\} \\ &+ \sum_{r=1}^{p} E_{\boldsymbol{\gamma}|\cdot}\left[\gamma_{\lambda,r} + \gamma_{\mu,r}\right]\log\left(\frac{\theta}{1-\theta}\right) + (a-1)\log\theta + (b+2p-1)\log(1-\theta), \end{aligned}$$

where $\boldsymbol{x}'_{ij} = (1, x_{ij1}, \ldots, x_{ijp})$, $\boldsymbol{\gamma}_\lambda = (\gamma_{\lambda,0}, \ldots, \gamma_{\lambda,p})$, $\boldsymbol{\gamma}_\mu = (\gamma_{\mu,0}, \ldots, \gamma_{\mu,p})$, $\boldsymbol{\beta}'_\lambda = (\lambda_0, \beta_{\lambda,1}, \ldots, \beta_{\lambda,p})$, and $\boldsymbol{\beta}'_\mu = (\mu_0, \beta_{\mu,1}, \ldots, \beta_{\mu,p})$. By setting $\gamma_{\lambda,0} \equiv 1$ and $\gamma_{\mu,0} \equiv 1$, $\lambda_0$ and $\mu_0$ are forced into the model. The second derivative of the Q-function can be partitioned into

$$\frac{\partial^2 Q\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]}{\partial\boldsymbol{\Phi}\partial\boldsymbol{\Phi}'} = \begin{bmatrix} \frac{\partial^2 Q\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}_\lambda'} & \frac{\partial^2 Q\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}_\mu'} & \mathbf{0}_{(p+1)\times 1} \\ \frac{\partial^2 Q\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]'}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}_\mu'} & \frac{\partial^2 Q\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]}{\partial\boldsymbol{\beta}_\mu\partial\boldsymbol{\beta}_\mu'} & \mathbf{0}_{(p+1)\times 1} \\ \mathbf{0}_{1\times(p+1)} & \mathbf{0}_{1\times(p+1)} & \frac{\partial^2 Q\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]}{\partial\boldsymbol{\theta}^2} \end{bmatrix}_{(2p+3)\times(2p+3)}.$$

Due to the complexity of the following derivations, we use multiple substitutions for clarity. Here, major derivations are presented first, and their components are defined after. For instance, second derivatives may be written as functions of first derivatives, which are defined thereafter. We let

$$\begin{aligned}\log P_{y_i(t_{i,j-1}),y_i(t_{i,j})}\left(\delta_{ij}|\boldsymbol{x}_{i,j-1}\right) = {} & \log\left(B+AC\right)I(N,N) + \log\left(A-AC\right)I(N,S) \\ & + \log\left(A+BC\right)I(S,S) + \log\left(B-BC\right)I(S,N) \\ & - \log(A+B),\end{aligned}$$

where the indicator function $I(y_i(t_{i,j-1}), y_i(t_{i,j}))$ represents the observed transition for subject $i$ from the assessment at time $t_{i,j-1}$ to $t_{i,j}$. Recall that $y_i(t_{i,j})$ is defined as $N$ if the subject is in a non-smoking state and $S$ if the subject is in a smoking state. Here, we set

$$A = \lambda^{i,j-1} = \exp(\boldsymbol{x}_{i,j-1}'\boldsymbol{\beta}_\lambda),$$

$$B = \mu^{i,j-1} = \exp(\boldsymbol{x}_{i,j-1}'\boldsymbol{\beta}_\mu),$$

and

$$C = \exp(-(A+B)\delta_{ij}).$$

It is important to note that in the following derivations, $A$, $B$, and $C$ depend on $i$ and $j$, but we suppress their subscripts for ease of reading.

Here, we provide the first derivatives of the Q-function with respect to $\boldsymbol{\Phi}$.

The first derivatives of the Q-function are

$$\begin{aligned}\frac{\partial Q\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]}{\partial \boldsymbol{\beta}_\lambda} &= \sum_{i=1}^{m}\sum_{j=2}^{n_i}\left[\frac{\frac{\partial(B+AC)}{\partial\boldsymbol{\beta}_\lambda}}{(B+AC)}I(N,N)+\frac{\frac{\partial(A-AC)}{\partial\boldsymbol{\beta}_\lambda}}{(A-AC)}I(N,S)\right.\\
&\left.+\frac{\frac{\partial(A+BC)}{\partial\boldsymbol{\beta}_\lambda}}{(A+BC)}I(S,S)+\frac{\frac{\partial(B-BC)}{\partial\boldsymbol{\beta}_\lambda}}{(B-BC)}I(S,N)-\frac{\frac{\partial(A+B)}{\partial\boldsymbol{\beta}_\lambda}}{(A+B)}\right]\\
&-\boldsymbol{\beta}_\lambda\circ\left[E_{\boldsymbol{\gamma}|\cdot}\left[\frac{1}{v_0(1-\boldsymbol{\gamma}_\lambda)+v_1\boldsymbol{\gamma}_\lambda}\right]\right],\end{aligned}$$

$$\begin{aligned}\frac{\partial Q\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]}{\partial \boldsymbol{\beta}_\mu} &= \sum_{i=1}^{m}\sum_{j=2}^{n_i}\left[\frac{\frac{\partial(B+AC)}{\partial\boldsymbol{\beta}_\mu}}{(B+AC)}I(N,N)+\frac{\frac{\partial(A-AC)}{\partial\boldsymbol{\beta}_\mu}}{(A-AC)}I(N,S)\right.\\
&\left.+\frac{\frac{\partial(A+BC)}{\partial\boldsymbol{\beta}_\mu}}{(A+BC)}I(S,S)+\frac{\frac{\partial(B-BC)}{\partial\boldsymbol{\beta}_\mu}}{(B-BC)}I(S,N)-\frac{\frac{\partial(A+B)}{\partial\boldsymbol{\beta}_\mu}}{(A+B)}\right]\\
&-\boldsymbol{\beta}_\mu\circ\left[E_{\boldsymbol{\gamma}|\cdot}\left[\frac{1}{v_0(1-\boldsymbol{\gamma}_\mu)+v_1\boldsymbol{\gamma}_\mu}\right]\right],\end{aligned}$$

and

$$\frac{\partial Q\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]}{\partial \theta} = \sum_{r=1}^{p}\left[E_{\boldsymbol{\gamma}|\cdot}\left[\gamma_{\lambda,r}+\gamma_{\mu,r}\right]\left[\frac{1}{\theta}+\frac{1}{1-\theta}\right]\right]+\frac{a-1}{\theta}-\frac{b+2p-1}{1-\theta},$$

where $\circ$ is the component-wise product of two vectors.

Then, the second derivatives of the Q-function are shown to be

$$\begin{aligned}\frac{\partial^2 Q\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]}{\partial \boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}'_\lambda} &= \sum_{i=1}^{m}\sum_{j=2}^{n_i}\left[\frac{\frac{\partial^2(A+BC)}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}'_\lambda}(A+BC)-\frac{\partial(A+BC)}{\partial\boldsymbol{\beta}_\lambda}\frac{\partial(A+BC)}{\partial\boldsymbol{\beta}'_\lambda}}{(A+BC)^2}I(S,S)\right.\\
&+\frac{\frac{\partial^2(B+AC)}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}'_\lambda}(B+AC)-\frac{\partial(B+AC)}{\partial\boldsymbol{\beta}_\lambda}\frac{\partial(B+AC)}{\partial\boldsymbol{\beta}'_\lambda}}{(B+AC)^2}I(N,N)\\
&+\frac{\frac{\partial^2(A-AC)}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}'_\lambda}(A-AC)-\frac{\partial(A-AC)}{\partial\boldsymbol{\beta}_\lambda}\frac{\partial(A-AC)}{\partial\boldsymbol{\beta}'_\lambda}}{(A-AC)^2}I(N,S)\\
&+\frac{\frac{\partial^2(B-BC)}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}'_\lambda}(B-BC)-\frac{\partial(B-BC)}{\partial\boldsymbol{\beta}_\lambda}\frac{\partial(B-BC)}{\partial\boldsymbol{\beta}'_\lambda}}{(B-BC)^2}I(S,N)\\
&\left.-\frac{\frac{\partial^2 A}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}'_\lambda}(A+B)-\frac{\partial A}{\partial\boldsymbol{\beta}_\lambda}\frac{\partial A}{\partial\boldsymbol{\beta}'_\lambda}}{(A+B)^2}\right]-\operatorname{diag}\left(E_{\boldsymbol{\gamma}|\cdot}\left[\frac{1}{v_0(1-\boldsymbol{\gamma}_\lambda)+v_1\boldsymbol{\gamma}_\lambda}\right]\right),\end{aligned}$$

$$\frac{\partial^2 Q\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]}{\partial\boldsymbol{\beta}_\mu\partial\boldsymbol{\beta}'_\mu} = \sum_{i=1}^{m}\sum_{j=2}^{n_i}\left[\frac{\frac{\partial^2(A+BC)}{\partial\boldsymbol{\beta}_\mu\partial\boldsymbol{\beta}'_\mu}(A+BC) - \frac{\partial(A+BC)}{\partial\boldsymbol{\beta}_\mu}\frac{\partial(A+BC)}{\partial\boldsymbol{\beta}'_\mu}}{(A+BC)^2}I(S,S)\right.$$
$$+\frac{\frac{\partial^2(B+AC)}{\partial\boldsymbol{\beta}_\mu\partial\boldsymbol{\beta}'_\mu}(B+AC) - \frac{\partial(B+AC)}{\partial\boldsymbol{\beta}_\mu}\frac{\partial(B+AC)}{\partial\boldsymbol{\beta}'_\mu}}{(B+AC)^2}I(N,N)$$
$$+\frac{\frac{\partial^2(A-AC)}{\partial\boldsymbol{\beta}_\mu\partial\boldsymbol{\beta}'_\mu}(A-AC) - \frac{\partial(A-AC)}{\partial\boldsymbol{\beta}_\mu}\frac{\partial(A-AC)}{\partial\boldsymbol{\beta}'_\mu}}{(A-AC)^2}I(N,S)$$
$$+\frac{\frac{\partial^2(B-BC)}{\partial\boldsymbol{\beta}_\mu\partial\boldsymbol{\beta}'_\mu}(B-BC) - \frac{\partial(B-BC)}{\partial\boldsymbol{\beta}_\mu}\frac{\partial(B-BC)}{\partial\boldsymbol{\beta}'_\mu}}{(B-BC)^2}I(S,N)$$
$$\left.-\frac{\frac{\partial^2 B}{\partial\boldsymbol{\beta}_\mu\partial\boldsymbol{\beta}'_\mu}(A+B) - \frac{\partial B}{\partial\boldsymbol{\beta}_\mu}\frac{\partial B}{\partial\boldsymbol{\beta}'_\mu}}{(A+B)^2}\right] - \text{diag}\left(E_{\boldsymbol{\gamma}|\cdot}\left[\frac{1}{v_0(1-\boldsymbol{\gamma}_\mu)+v_1\boldsymbol{\gamma}_\mu}\right]\right),$$

$$\frac{\partial^2 Q\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}'_\mu} = \sum_{i=1}^{m}\sum_{j=2}^{n_i}\left[\frac{\frac{\partial^2(A+BC)}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}'_\mu}(A+BC) - \frac{\partial(A+BC)}{\partial\boldsymbol{\beta}_\lambda}\frac{\partial(A+BC)}{\partial\boldsymbol{\beta}'_\mu}}{(A+BC)^2}I(S,S)\right.$$
$$+\frac{\frac{\partial^2(B+AC)}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}'_\mu}(B+AC) - \frac{\partial(B+AC)}{\partial\boldsymbol{\beta}_\lambda}\frac{\partial(B+AC)}{\partial\boldsymbol{\beta}'_\mu}}{(B+AC)^2}I(N,N)$$
$$+\frac{-\frac{\partial^2(B+AC)}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}'_\mu}(A-AC) - \frac{\partial(A-AC)}{\partial\boldsymbol{\beta}_\lambda}\frac{\partial(A-AC)}{\partial\boldsymbol{\beta}'_\mu}}{(A-AC)^2}I(N,S)$$
$$+\frac{-\frac{\partial^2(A+BC)}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}'_\mu}(B-BC) - \frac{\partial(B-BC)}{\partial\boldsymbol{\beta}_\lambda}\frac{\partial(B-BC)}{\partial\boldsymbol{\beta}'_\mu}}{(B-BC)^2}I(S,N)$$
$$\left.+\frac{\frac{\partial A}{\partial\boldsymbol{\beta}_\lambda}\frac{\partial B}{\partial\boldsymbol{\beta}'_\mu}}{(A+B)^2}\right],$$

and

$$\frac{\partial^2 Q\left[\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}\right]}{\partial\boldsymbol{\theta}^2} = \sum_{r=1}^{p}\left[E_{\boldsymbol{\gamma}|\cdot}\left[\gamma_{\lambda,r}+\gamma_{\mu,r}\right]\left[-\frac{1}{\theta^2}+\frac{1}{(1-\theta)^2}\right]\right] - \frac{a-1}{\theta^2} - \frac{b+2p-1}{(1-\theta)^2}.$$

where diag$(\cdot)$ represents a diagonal matrix. The second derivatives in these equations are shown below.

The second derivatives of $(A+BC)$ are

$$\frac{\partial^2(A+BC)}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}'_\lambda} = \frac{\partial^2 A}{\partial\boldsymbol{\beta}_\lambda\partial\boldsymbol{\beta}'_\lambda}(1-BC\delta_{ij}) + BC\delta_{ij}^2\frac{\partial A}{\partial\boldsymbol{\beta}_\lambda}\frac{\partial A}{\partial\boldsymbol{\beta}'_\lambda},$$

$$\frac{\partial^2(A+BC)}{\partial\boldsymbol{\beta}_\mu\partial\boldsymbol{\beta}'_\mu} = C\frac{\partial^2 B}{\partial\boldsymbol{\beta}_\mu\partial\boldsymbol{\beta}'_\mu}(1-B\delta_{ij}) - \delta_{ij}\frac{\partial B}{\partial\boldsymbol{\beta}_\mu}\left(C\frac{\partial B}{\partial\boldsymbol{\beta}'_\mu}+\frac{\partial AB}{\partial\boldsymbol{\beta}'_\mu}\right),$$

and

$$\frac{\partial^2 (A+BC)}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}'_\mu} = -\delta_{ij} \frac{\partial A}{\partial \boldsymbol{\beta}_\lambda} \left[ B \frac{\partial C}{\partial \boldsymbol{\beta}'_\mu} + C \frac{\partial B}{\partial \boldsymbol{\beta}'_\mu} \right].$$

The second derivatives of $(B + AC)$ are

$$\frac{\partial^2 (B+AC)}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}'_\lambda} = C \frac{\partial^2 A}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}'_\lambda} (1 - A\delta_{ij}) - \delta_{ij} \frac{\partial A}{\partial \boldsymbol{\beta}_\lambda} \left( C \frac{\partial A}{\partial \boldsymbol{\beta}'_\lambda} + \frac{\partial BA}{\partial \boldsymbol{\beta}'_\lambda} \right),$$

$$\frac{\partial^2 (B+AC)}{\partial \boldsymbol{\beta}_\mu \partial \boldsymbol{\beta}'_\mu} = \frac{\partial^2 B}{\partial \boldsymbol{\beta}_\mu \partial \boldsymbol{\beta}'_\mu} (1 - AC\delta_{ij}) + AC\delta_{ij}^2 \frac{\partial B}{\partial \boldsymbol{\beta}_\mu} \frac{\partial B}{\partial \boldsymbol{\beta}'_\mu},$$

and

$$\frac{\partial^2 (B+AC)}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}'_\mu} = \frac{\partial A}{\partial \boldsymbol{\beta}_\lambda} \frac{\partial C}{\partial \boldsymbol{\beta}'_\mu} [1 - A\delta_{ij}].$$

The second derivatives of $(A - AC)$ are

$$\frac{\partial^2 (A-AC)}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}'_\lambda} = \frac{\partial^2 A}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}'_\lambda} - \frac{\partial^2 (B+AC)}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}'_\lambda},$$

$$\frac{\partial^2 (A-AC)}{\partial \boldsymbol{\beta}_\mu \partial \boldsymbol{\beta}'_\mu} = AC\delta_{ij} \left[ \frac{\partial^2 B}{\partial \boldsymbol{\beta}_\mu \partial \boldsymbol{\beta}'_\mu} - \delta_{ij} \frac{\partial B}{\partial \boldsymbol{\beta}_\mu} \frac{\partial B}{\partial \boldsymbol{\beta}'_\mu} \right],$$

and

$$\frac{\partial^2 (A-AC)}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}'_\mu} = -\frac{\partial^2 (B+AC)}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}'_\mu}.$$

The second derivatives of $(B - BC)$ are

$$\frac{\partial^2 (B-BC)}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}'_\lambda} = BC\delta_{ij} \left[ \frac{\partial^2 A}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}'_\lambda} - \delta_{ij} \frac{\partial A}{\partial \boldsymbol{\beta}_\lambda} \frac{\partial A}{\partial \boldsymbol{\beta}'_\lambda} \right],$$

$$\frac{\partial^2 (B-BC)}{\partial \boldsymbol{\beta}_\mu \partial \boldsymbol{\beta}'_\mu} = \frac{\partial^2 B}{\partial \boldsymbol{\beta}_\mu \partial \boldsymbol{\beta}'_\mu} - \frac{\partial^2 (A+BC)}{\partial \boldsymbol{\beta}_\mu \partial \boldsymbol{\beta}'_\mu},$$

and

$$\frac{\partial^2 (B-BC)}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}'_\mu} = -\frac{\partial^2 (A+BC)}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}'_\mu}.$$

Additionally,

$$\frac{\partial^2 A}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}'_\lambda} = A\boldsymbol{x}_{i,j-1}\boldsymbol{x}'_{i,j-1}$$

and

$$\frac{\partial^2 B}{\partial \boldsymbol{\beta}_\mu \partial \boldsymbol{\beta}'_\mu} = B\boldsymbol{x}_{i,j-1}\boldsymbol{x}'_{i,j-1}.$$

Here, we present the first derivatives necessary to compute the functions above.

The first derivatives of $(A + BC)$ are

$$\frac{\partial(A+BC)}{\partial\boldsymbol{\beta}_\lambda} = \frac{\partial A}{\partial\boldsymbol{\beta}_\lambda}\left(1 - BC\delta_{ij}\right)$$

and

$$\frac{\partial(A+BC)}{\partial\boldsymbol{\beta}_\mu} = C\frac{\partial B}{\partial\boldsymbol{\beta}_\mu}\left(1 - B\delta_{ij}\right).$$

The first derivatives of $(B + AC)$ are

$$\frac{\partial(B+AC)}{\partial\boldsymbol{\beta}_\lambda} = C\frac{\partial A}{\partial\boldsymbol{\beta}_\lambda}\left(1 - A\delta_{ij}\right)$$

and

$$\frac{\partial(B+AC)}{\partial\boldsymbol{\beta}_\mu} = \frac{\partial B}{\partial\boldsymbol{\beta}_\mu}\left(1 - AC\delta_{ij}\right).$$

The first derivatives of $(A - AC)$ are

$$\frac{\partial(A-AC)}{\partial\boldsymbol{\beta}_\lambda} = \frac{\partial A}{\partial\boldsymbol{\beta}_\lambda} - \frac{\partial(B+AC)}{\partial\boldsymbol{\beta}_\lambda}$$

and

$$\frac{\partial(A-AC)}{\partial\boldsymbol{\beta}_\mu} = AC\delta_{ij}\frac{\partial B}{\partial\boldsymbol{\beta}_\mu}.$$

The first derivatives of $(B - BC)$ are

$$\frac{\partial(B-BC)}{\partial\boldsymbol{\beta}_\lambda} = BC\delta_{ij}\frac{\partial A}{\partial\boldsymbol{\beta}_\lambda}$$

and

$$\frac{\partial(B-BC)}{\partial\boldsymbol{\beta}_\mu} = \frac{\partial B}{\partial\boldsymbol{\beta}_\mu} - \frac{\partial(A+BC)}{\partial\boldsymbol{\beta}_\mu}.$$

Additionally,

$$\frac{\partial A}{\partial\boldsymbol{\beta}_\lambda} = A\boldsymbol{x}_{i,j-1},$$

$$\frac{\partial A}{\partial\boldsymbol{\beta}_\mu} = \mathbf{0},$$

$$\frac{\partial B}{\partial\boldsymbol{\beta}_\lambda} = \mathbf{0},$$

$$\frac{\partial B}{\partial\boldsymbol{\beta}_\mu} = B\boldsymbol{x}_{i,j-1}$$

$$\frac{\partial C}{\partial \boldsymbol{\beta}_\lambda} = -C\delta_{ij}\frac{\partial A}{\partial \boldsymbol{\beta}_\lambda},$$

and

$$\frac{\partial C}{\partial \boldsymbol{\beta}_\mu} = -C\delta_{ij}\frac{\partial B}{\partial \boldsymbol{\beta}_\mu}.$$

Now, the components of the conditional covariance matrix, $\text{var}\left\{\frac{\partial \log \pi(\boldsymbol{\Phi},\boldsymbol{\gamma}|\boldsymbol{y})}{\partial \boldsymbol{\Phi}}\middle|\hat{\boldsymbol{\Phi}},\boldsymbol{y}\right\}$, are defined as:

$$\text{cov}\left[\frac{\partial \log \pi(\boldsymbol{\Phi},\boldsymbol{\gamma}|\boldsymbol{y})}{\partial \boldsymbol{\beta}}, \frac{\partial \log \pi(\boldsymbol{\Phi},\boldsymbol{\gamma}|\boldsymbol{y})}{\partial \boldsymbol{\beta}}\right] = \text{diag}\left(\hat{\boldsymbol{\beta}}^2 \circ \left[\frac{1}{v_0^2}(1-\boldsymbol{p}^*) + \frac{1}{v_1^2}\boldsymbol{p}^*\right] - \left[\hat{\boldsymbol{\beta}} \circ \left[\frac{1}{v_0}(1-\boldsymbol{p}^*) + \frac{1}{v_1}\boldsymbol{p}^*\right]\right]^2\right),$$

$$\text{cov}\left[\frac{\partial \log \pi(\boldsymbol{\Phi},\boldsymbol{\gamma}|\boldsymbol{y})}{\partial \boldsymbol{\beta}}, \frac{\partial \log \pi(\boldsymbol{\Phi},\boldsymbol{\gamma}|\boldsymbol{y})}{\partial \theta}\right] = -\left[\frac{1}{\hat{\theta}} + \frac{1}{1-\hat{\theta}}\right]\left[\frac{\hat{\boldsymbol{\beta}} \circ \boldsymbol{p}^*}{v_1} - (\hat{\boldsymbol{\beta}} \circ \boldsymbol{p}^*) \circ \left[\frac{\boldsymbol{p}^*}{v_1} + \frac{1-\boldsymbol{p}^*}{v_0}\right]\right],$$

and

$$\text{var}\left[\frac{\partial \log \pi(\boldsymbol{\Phi},\boldsymbol{\gamma}|\boldsymbol{y})}{\partial \theta}\right] = \left[\frac{1}{\hat{\theta}} + \frac{1}{1-\hat{\theta}}\right]^2 \text{var}\left[\sum_{r=1}^{p} \gamma_{\lambda,r} + \gamma_{\mu,r}\right],$$

where $\boldsymbol{\beta'} = (\boldsymbol{\beta}'_\lambda, \boldsymbol{\beta}'_\mu)$ and $\text{var}[\gamma_r] = \hat{\theta}(1-\hat{\theta})$. Additionally, $\boldsymbol{p}^{*'} = (\boldsymbol{p}^{*'}_\lambda, \boldsymbol{p}^{*'}_\mu)$, $\boldsymbol{p}^{*'}_\lambda = (p^*_{\lambda,0}, \ldots, p^*_{\lambda,p})$, and $\boldsymbol{p}^{*'}_\mu = (p^*_{\mu,0}, \ldots, p^*_{\mu,p})$. Recall that setting $p^*_{\lambda,0} \equiv 1$ and $p^*_{\mu,0} \equiv 1$ forces $\lambda_0$ and $\mu_0$ into the model.

*Web Appendix D: Data Generation*

To evaluate the performance of our method, we apply it to multiple simulated data sets in a variety of research scenarios. In all of the generated data sets, each individual is allowed to transition between smoking, S, and non-smoking, N, states over a period of time. First, each individual is randomly assigned to an initial state following a draw from a Bernoulli distribution with probability 0.50. Since there are two states in this process, an individual always transitions into the opposite state. True transition times follow an exponential distribution that is parametrized with transition rate $\lambda^{ij}$ if the individual is currently in an N state and $\mu^{ij}$ if he/she is in an S state. Here, we set the true transition rates to

$$\lambda^{ij} = \exp(0.5 - 0.5x_{ij1} + 0.5x_{ij2} - 0.5x_{ij5})$$

$$\mu^{ij} = \exp(0.5 + 0.5x_{ij1} + 0.5x_{ij5} - 0.5x_{ij6})$$

We restrict our analysis to weaker effects, under the assumption that performance will only improve with larger effects (Koslovsky et al., 2016). The full model contains 4 binary covariates $(x_{ij1}, \ldots, x_{ij4})$ and 16 continuous covariates $(x_{ij5}, \ldots, x_{ij20})$. Recall that since covariates can be associated with both transition rates, this creates $2^{2\times20} > 10^{12}$ possible models. Binary covariates are sampled from a Bernoulli distribution with probability 0.50. Binary covariates $x_{ij1}$ and $x_{ij3}$ are resampled from the same distribution after the individual has transitioned 10 times. Continuous covariates $(x_{ij6}, x_{ij8}, \ldots, x_{ij20})$ follow a multivariate normal distribution with mean 0, variance 1, and an exchangeable correlation structure, $\rho = 0$ or $\rho = 0.75$. Continuous covariates $x_{ij5}$ and $x_{ij7}$ are simulated from a random walk with the initial value sampled from a normal distribution with mean 0 and variance 1. Thereafter, the data are generated by $x_{i,j} = x_{i,j-1} + \epsilon_{i,j}$, where $\epsilon_{i,j}$ is sampled from a normal distribution with mean 0 and variance 0.001 at each transition.

The true transition process is iterated until an individual's cumulative transition time exceeds the period of time for which he/she is set to be followed (e.g., 30 or 70 units of time). In practice, we do not always observe an assessment exactly at each individual's transition time. We only observe their current state at each assessment time. For example, an individual may receive a post-quit random assessment at 8 a.m., slip at 8:05 a.m., and not perform another assessment until 8:20 a.m. Here, the exact transition time is unknown since we only observe N at 8 a.m. and S at 8:20 a.m. To analogue this, we parse through each individual's simulated transition data at equally or randomly spaced assessment times and identify the individual's state at the time the assessment is conducted. At each assessment time, we observe the individual's current smoking state and measured risk factors. Then all of the individuals' observations are compiled to create the data set for analysis.

*Web Appendix E: Simulations*

Here, we examine the performance of our method in various senarios. For scenarios 1-8, we simulate $m = 100$ and $m = 150$ individuals observed at $n_i = 30$ or $n_i = 70$ equally spaced assessments under both correlation structures. Equally spaced observations are a special case of a continuous-time Markov process which can be modeled discretely. Thus in these scenarios, we can model the data with a discrete-time transition model, as in (Diggle et al., 2013), for comparison:

$$logit\left(\omega_S\left(\boldsymbol{x}_{i,j-1}\right)|y_{i,j-1} = N\right) = 0.5 - 0.5x_{i,j-1,1} + 0.5x_{i,j-1,2} - 0.5x_{i,j-1,5} \tag{A.3}$$

and

$$logit\left(\omega_N\left(\boldsymbol{x}_{i,j-1}\right)|y_{i,j-1} = S\right) = 0.5 + 0.5x_{i,j-1,1} + 0.5x_{i,j-1,5} - 0.5x_{i,j-1,6} \tag{A.4}$$

where $\omega_S\left(\boldsymbol{x}_{i,j-1}\right)$ and $\omega_N\left(\boldsymbol{x}_{i,j-1}\right)$ are the probabilities that an individual is observed in a smoking or non-smoking state at $t_{ij}$, respectively. Note, we use the covariates observed in the current state to model the probability of transition in the next state, similar to Eq. 3 in the main article. Here, we can compare our method to any variable selection method designed for logistic regression models; we choose the LASSO (Tibshirani, 1996) and EMVS (Koslovsky et al., 2016). Note that EMVS for logistic regression models has already been shown to outperform stepwise selection procedures (Koslovsky et al., 2016).

For scenarios 9-16, we simulate $m = 100$ and $m = 150$ individuals observed at $n_i = 30$ or $n_i = 70$ randomly spaced assessments under both correlation structures. Since observation times are inconsistent, this data structure is more realistic for EMA data. Consequently, modeling the data with Eq. A.3 and A.4 for comparison is no longer valid.

Each scenario is simulated 250 times and evaluated in R (R Core Team, 2015). We initialize our method using estimates from the **msm** package (Jackson et al., 2011) for a two-state model including all possible covariates. Before selection, we standardize continuous covariates to mean 0 and variance 1. The sparsity parameter $\theta$ is initially set to 0.50. The variance of exclusion and inclusion are set similar to the method suggested by Koslovsky et al. (2016). Here, we claim a hazard ratio between (0.95, 1.05) to be clinically irrelevant, and we evaluate the local stability of regularization plots around our intuition with a 95% prior probability of inclusion for the hazards ratio that covers $(1/4, 4)$. Using this tuning procedure, we set the 95% prior probability of exclusion equivalent to a hazards ratio between $(0.95, 1.05)$, $v_0 = 0.0006$, for all models. For the comparisons in scenarios 1-8, EMVS for logistic regression models are tuned similarly. We use the **glmnet** package to compare our model with the LASSO for logistic regression (Friedman et al., 2009). Here, each LASSO model is tuned using 10-fold cross validation to identify the largest penalty value, $\lambda_{LASSO}$, that is within one standard error of the minimum mean square error (Breiman et al., 1984). For the annealing process, the inverse temperature is initially set to .2 and increased by .1 until $t = 1$, as described in Web Appendix B. Convergence at each temperature level and at each CM-step using the Newton-Raphson algorithm is determined with $\epsilon = 1e - 5$.

To evaluate the performance of our method, we calculate the average false positive (FP) and false negative (FN) rates with FPR = FP/(FP + TN) and FNR= FN/(FN+TP), where TP and TN are true positives and true negatives, respectively. Additionally, we assessed the bias (average of the posterior modes minus true values), the Monte Carlo error of the posterior modes (MCE), the square root of the average of the posterior variances estimated with Louis's method (SE), the coverage probability (CP) of the 95% equal-tail credible intervals, and the average mean squared error (MSE) of the steady-state probability of transition from a non-smoking to a smoking state.

## *Web Appendix F: Simulation Results*

The simulation results for scenarios 1-8 are found in Table 1. Here, our method produced relatively low FPR and FNR. The performance of our method improved with larger sample sizes and number of assessments observed and declined with higher correlation among covariates.

[Table 1 about here.]

Across all combinations of the number of individuals, number of assessments, and correlation structures, our method correctly included associated covariates in 99.4% of the simulations on average (min = 94.4% and max = 100%). Additionally, our method correctly excluded unassociated covariates in 99.2% of the simulations on average (min = 88.8% and max = 100%). Marginally, we observed that our method performed better for continuous covariates. It was least accurate for binary covariates whose values were allowed to change. With equally spaced assessment times, our method estimated unknown parameters with minimal bias, MCE very close to the SE, and CP around 92% on average (Table 2). The MSE for the steady-state probability of transition from a non-smoking to a smoking state was around 0.04 for scenarios 1-8. In terms of the MSE, our method performed better with higher correlation structures and was insensitive to the number of individuals and number of assessments (Table 3).

[Table 2 about here.]

[Table 3 about here.]

Our method outperformed or showed relatively equivalent performance to the LASSO in FPR and FNR in every setting (Table 1). In sensitivity analyses not shown, we found the LASSO tuned with a $\lambda_{LASSO}$ that was within one standard error of the minimum performed better than a $\lambda_{LASSO}$ that minimized the cross-validation error in our simulations. Additionally, the LASSO method was outperformed by EMVS for logistic regression models. EMVS for multistate models on data with equally spaced assessment times showed comparable performance to EMVS for logistic regression models. Here, EMVS for multistate models performed better in all of the 8 scenarios for average FPR. However, EMVS for logistic regression performed better in 4 of the 8 scenarios for average FNR. The differences in performance were small and could be attributed to tuning parameterization.

In scenarios 9-16, we found the performance of EMVS for multistate models on data with randomly spaced assessment times was marginally worse than that for data with for equally spaced assessment times (Table 4). This discrepancy could be due to smaller observation windows for some individuals, since their first and last assessments may be later or earlier than the equally spaced assessments, respectively.

[Table 4 about here.]

Across all combinations of the number of individuals, number of assessments, and correlation structures, the method correctly selected associated covariates in 98.7% of the simulations on average (min = 93.2% and max = 100%). Additionally, the method correctly excluded unassociated covariates in 98.7% of the simulations on average (min 87.6% and max 100%). Again, the method performed better for continuous covariates than binary covariates. With randomly spaced assessment times, our method estimated unknown parameters with minimal bias, MCE very close to the SE, and CP around 92% on average, similar to the results found with equally spaced assessment times (Table 5). Additionally, the MSE for the steady-state probability of transition from a non-smoking state to a smoking state with randomly spaced assessments was similar to the MSE observed with the equally spaced assessments (Table 3).

[Table 5 about here.]

[Figure 1 about here.]

## References

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees.* CRC press.

Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. (2013). *Analysis of longitudinal data.* OUP Oxford.

Friedman, J., Hastie, T., and Tibshirani, R. (2009). glmnet: Lasso and elastic-net regularized generalized linear models. *R Package Version* .

Jackson, C. H. et al. (2011). Multi-state models for panel data: the msm package for R. *Journal of Statistical Software* **38,** 1–29.

Koslovsky, M. D., Swartz, M. D., Leon-Novelo, L., Chan, W., and Wilkinson, A. V. (2016). Using the EM algorithm for Bayesian variable selection in logistic regression models with related covariates. *In Revisions* .

R Core Team (2015). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58,** 267–288.

Ueda, N. and Nakano, R. (1998). Deterministic annealing EM algorithm. *Neural Networks* **11,** 271–282.

Wu, C. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics* **11,** 95–103.

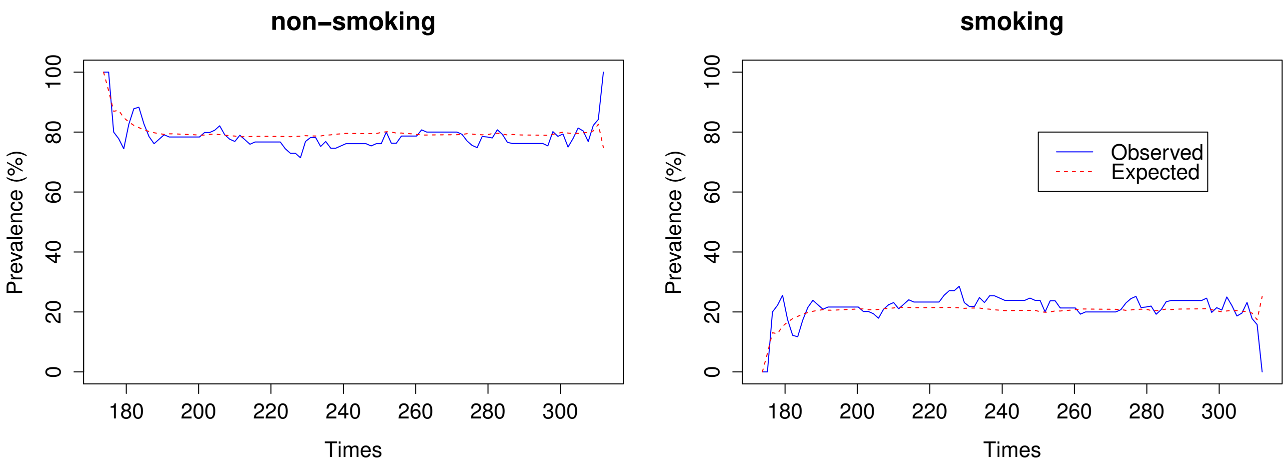


**Figure 1.** Estimated and observed smoking and non-smoking prevalence one week after the quit attempt in the PREVAIL Study. Time is in days.

**Table 1**

***Simulation Scenarios 1-8 Results:*** *Comparison of selection performance between EMVS for multistate models and the LASSO and EMVS for logistic regression models when assessment times are equally spaced. The number of* ***Individuals****, number of* ***Assessments****, and correlation structure between covariates, $\rho$, is varied in each scenario. Here, the false positive (****FPR****) and false negative (****FNR****) rates are compared.*

| Individuals | Assessments | $\rho$ | FPR | | | FNR | | |
|---|---|---|---|---|---|---|---|---|
| | | | a | b | c | a | b | c |
| 100 | 30 | 0 | 0.010 | -0.026 | -0.011 | 0.017 | 0.017 | -0.023 |
| 100 | 30 | 0.75 | 0.030 | -0.038 | 0.005 | 0.020 | 0.019 | -0.056 |
| 100 | 70 | 0 | 0.002 | -0.037 | -0.028 | 0.000 | 0.000 | -0.002 |
| 100 | 70 | 0.75 | 0.002 | -0.030 | -0.026 | 0.000 | 0.000 | -0.001 |
| 150 | 30 | 0 | 0.004 | -0.031 | -0.017 | 0.002 | 0.002 | -0.013 |
| 150 | 30 | 0.75 | 0.011 | -0.036 | -0.014 | 0.003 | 0.003 | -0.019 |
| 150 | 70 | 0 | 0.000 | -0.041 | -0.031 | 0.000 | 0.000 | 0.000 |
| 150 | 70 | 0.75 | 0.000 | -0.029 | -0.021 | 0.000 | 0.000 | -0.001 |

[a] EMVS for multistate models
[b] EMVS for multistate models - EMVS for logistic models
[c] EMVS for multistate models - LASSO

**Table 2**

***Simulation Scenarios 1-8 Results:*** *Evaluation of* ***Bias*** *and Monte Carlo Error* ***(MCE)*** *of the estimated, unknown parameters, square root of the average Louis's method posterior variance estimates* ***(SE)****, and coverage probabilities* ***(CP)*** *of equal tailed 95% credible intervals at varying numbers of individuals, assessments, and correlation structures, $\rho$, for equally spaced assessment times.*

| | | 100 Individuals | | | | 150 Individuals | | | |
|---|---|---|---|---|---|---|---|---|---|
| | True Effect | Bias | MCE | SE | CP | Bias | MCE | SE | CP |
| 30 Assessments $\rho = 0$ | $\lambda_0$ | 0.029 | 0.127 | 0.108 | 0.900 | 0.022 | 0.091 | 0.088 | 0.928 |
| | $\beta_{\lambda,1}$ | 0.024 | 0.159 | 0.127 | 0.908 | 0.009 | 0.109 | 0.106 | 0.956 |
| | $\beta_{\lambda,2}$ | -0.026 | 0.134 | 0.108 | 0.932 | -0.010 | 0.085 | 0.088 | 0.960 |
| | $\beta_{\lambda,5}$ | 0.004 | 0.090 | 0.078 | 0.932 | -0.003 | 0.068 | 0.064 | 0.940 |
| | $\mu_0$ | 0.056 | 0.127 | 0.097 | 0.856 | 0.037 | 0.093 | 0.078 | 0.884 |
| | $\beta_{\mu,1}$ | -0.021 | 0.151 | 0.127 | 0.916 | -0.025 | 0.104 | 0.106 | 0.956 |
| | $\beta_{\mu,5}$ | 0.000 | 0.092 | 0.078 | 0.928 | 0.000 | 0.070 | 0.064 | 0.916 |
| | $\beta_{\mu,6}$ | 0.002 | 0.073 | 0.059 | 0.896 | 0.002 | 0.057 | 0.048 | 0.912 |
| 30 Assessments $\rho = 0.75$ | $\lambda_0$ | 0.042 | 0.126 | 0.102 | 0.868 | 0.029 | 0.097 | 0.083 | 0.900 |
| | $\beta_{\lambda,1}$ | -0.001 | 0.144 | 0.122 | 0.940 | 0.003 | 0.105 | 0.100 | 0.956 |
| | $\beta_{\lambda,2}$ | -0.017 | 0.138 | 0.101 | 0.916 | -0.011 | 0.095 | 0.084 | 0.928 |
| | $\beta_{\lambda,5}$ | 0.000 | 0.130 | 0.094 | 0.908 | -0.001 | 0.089 | 0.079 | 0.920 |
| | $\mu_0$ | 0.050 | 0.106 | 0.091 | 0.880 | 0.039 | 0.086 | 0.074 | 0.904 |
| | $\beta_{\mu,1}$ | -0.023 | 0.149 | 0.121 | 0.940 | -0.018 | 0.113 | 0.100 | 0.928 |
| | $\beta_{\mu,5}$ | -0.015 | 0.143 | 0.101 | 0.876 | -0.006 | 0.088 | 0.083 | 0.956 |
| | $\beta_{\mu,6}$ | 0.018 | 0.119 | 0.087 | 0.888 | 0.005 | 0.080 | 0.072 | 0.940 |
| 70 Assessments $\rho = 0$ | $\lambda_0$ | 0.013 | 0.083 | 0.071 | 0.896 | 0.010 | 0.066 | 0.058 | 0.904 |
| | $\beta_{\lambda,1}$ | -0.003 | 0.085 | 0.086 | 0.972 | 0.002 | 0.066 | 0.070 | 0.960 |
| | $\beta_{\lambda,2}$ | -0.006 | 0.072 | 0.072 | 0.944 | -0.007 | 0.057 | 0.058 | 0.956 |
| | $\beta_{\lambda,5}$ | 0.000 | 0.062 | 0.051 | 0.904 | -0.001 | 0.051 | 0.042 | 0.892 |
| | $\mu_0$ | 0.023 | 0.098 | 0.063 | 0.772 | 0.010 | 0.081 | 0.052 | 0.784 |
| | $\beta_{\mu,1}$ | -0.016 | 0.088 | 0.086 | 0.928 | -0.008 | 0.069 | 0.070 | 0.940 |
| | $\beta_{\mu,5}$ | -0.001 | 0.064 | 0.051 | 0.892 | 0.003 | 0.049 | 0.042 | 0.900 |
| | $\beta_{\mu,6}$ | 0.004 | 0.050 | 0.039 | 0.860 | 0.004 | 0.040 | 0.033 | 0.868 |
| 70 Assessments $\rho = 0.75$ | $\lambda_0$ | 0.002 | 0.080 | 0.067 | 0.900 | -0.001 | 0.071 | 0.055 | 0.896 |
| | $\beta_{\lambda,1}$ | 0.002 | 0.072 | 0.081 | 0.976 | 0.007 | 0.062 | 0.066 | 0.960 |
| | $\beta_{\lambda,2}$ | -0.005 | 0.064 | 0.067 | 0.960 | -0.004 | 0.056 | 0.055 | 0.940 |
| | $\beta_{\lambda,5}$ | -0.005 | 0.072 | 0.065 | 0.936 | 0.002 | 0.060 | 0.056 | 0.920 |
| | $\mu_0$ | 0.018 | 0.067 | 0.060 | 0.916 | 0.014 | 0.054 | 0.050 | 0.928 |
| | $\beta_{\mu,1}$ | -0.017 | 0.077 | 0.080 | 0.952 | -0.011 | 0.063 | 0.065 | 0.948 |
| | $\beta_{\mu,5}$ | -0.002 | 0.076 | 0.068 | 0.928 | 0.006 | 0.064 | 0.057 | 0.924 |
| | $\beta_{\mu,6}$ | 0.003 | 0.073 | 0.060 | 0.880 | 0.000 | 0.062 | 0.050 | 0.896 |

**Table 3**
***Simulation scenarios 1-16:*** *Evaluation of the steady-state probability* ***MSE*** *at varying numbers of individuals, assessments, and correlation structures, $\rho$, for equal spaced and randomly spaced assessment times.*

| Individuals | Assessments | $\rho$ | Equally Spaced | Randomly Spaced |
|---|---|---|---|---|
| 100 | 30 | 0 | 0.0581 | 0.0583 |
| 100 | 30 | 0.75 | 0.0349 | 0.0345 |
| 100 | 70 | 0 | 0.0586 | 0.0590 |
| 100 | 70 | 0.75 | 0.0348 | 0.0352 |
| 150 | 30 | 0 | 0.0583 | 0.0575 |
| 150 | 30 | 0.75 | 0.0347 | 0.0345 |
| 150 | 70 | 0 | 0.0582 | 0.0583 |
| 150 | 70 | 0.75 | 0.0347 | 0.0349 |

**Table 4**

***Simulation Scenarios 9-16 Results:*** *Selection performance using EMVS for multistate models when assessment times are randomly spaced. The number of **Individuals**, number of **Assessments**, and correlation structure between covariates, $\rho$, is varied in each scenario. Here, the false positive (**FPR**) and false negative (**FNR**) rates are presented.*

| Individuals | Assessments | $\rho$ | FPR | FNR |
|---|---|---|---|---|
| 100 | 30 | 0 | 0.016 | 0.024 |
| 100 | 30 | 0.75 | 0.049 | 0.039 |
| 100 | 70 | 0 | 0.003 | 0.001 |
| 100 | 70 | 0.75 | 0.005 | 0.001 |
| 150 | 30 | 0 | 0.007 | 0.010 |
| 150 | 30 | 0.75 | 0.017 | 0.010 |
| 150 | 70 | 0 | 0.001 | 0.000 |
| 150 | 70 | 0.75 | 0.001 | 0.000 |

**Table 5**
***Simulation Scenarios 9-16 Results:*** *Evaluation of* ***Bias*** *and Monte Carlo error (****MCE****) of the estimated, unknown parameters, square root of the average Louis's method posterior variance estimates (****SE****), and coverage probabilities (****CP****) of equal tailed 95% credible intervals at varying numbers of individuals, assessments, and correlation structures, ρ, for randomly spaced assessment times.*

| | | 100 Individuals | | | | 150 Individuals | | | |
|---|---|---|---|---|---|---|---|---|---|
| | True Effect | Bias | MCE | SE | CP | Bias | MCE | SE | CP |
| 30 Assessments $\rho = 0$ | $\lambda_0$ | 0.008 | 0.153 | 0.121 | 0.860 | -0.025 | 0.112 | 0.099 | 0.892 |
| | $\beta_{\lambda,1}$ | -0.008 | 0.157 | 0.141 | 0.916 | 0.004 | 0.129 | 0.118 | 0.936 |
| | $\beta_{\lambda,2}$ | -0.052 | 0.154 | 0.119 | 0.908 | -0.035 | 0.116 | 0.098 | 0.936 |
| | $\beta_{\lambda,5}$ | 0.005 | 0.103 | 0.085 | 0.904 | 0.003 | 0.076 | 0.070 | 0.944 |
| | $\mu_0$ | 0.000 | 0.146 | 0.108 | 0.844 | -0.019 | 0.118 | 0.088 | 0.856 |
| | $\beta_{\mu,1}$ | -0.071 | 0.173 | 0.140 | 0.896 | -0.059 | 0.144 | 0.117 | 0.924 |
| | $\beta_{\mu,5}$ | 0.001 | 0.103 | 0.086 | 0.896 | -0.003 | 0.077 | 0.070 | 0.912 |
| | $\beta_{\mu,6}$ | 0.007 | 0.082 | 0.064 | 0.896 | 0.008 | 0.063 | 0.052 | 0.896 |
| 30 Assessments $\rho = 0.75$ | $\lambda_0$ | -0.031 | 0.150 | 0.114 | 0.868 | -0.044 | 0.116 | 0.094 | 0.872 |
| | $\beta_{\lambda,1}$ | 0.020 | 0.182 | 0.134 | 0.892 | 0.019 | 0.128 | 0.112 | 0.948 |
| | $\beta_{\lambda,2}$ | -0.037 | 0.141 | 0.114 | 0.936 | -0.028 | 0.112 | 0.093 | 0.920 |
| | $\beta_{\lambda,5}$ | 0.015 | 0.169 | 0.103 | 0.860 | -0.002 | 0.101 | 0.086 | 0.916 |
| | $\mu_0$ | -0.016 | 0.120 | 0.102 | 0.896 | -0.024 | 0.095 | 0.083 | 0.908 |
| | $\beta_{\mu,1}$ | -0.046 | 0.166 | 0.133 | 0.908 | -0.039 | 0.121 | 0.112 | 0.940 |
| | $\beta_{\mu,5}$ | -0.016 | 0.168 | 0.111 | 0.864 | -0.010 | 0.105 | 0.092 | 0.924 |
| | $\beta_{\mu,6}$ | 0.033 | 0.143 | 0.098 | 0.888 | 0.018 | 0.092 | 0.080 | 0.940 |
| 70 Assessments $\rho = 0$ | $\lambda_0$ | -0.001 | 0.094 | 0.081 | 0.924 | -0.013 | 0.079 | 0.067 | 0.904 |
| | $\beta_{\lambda,1}$ | -0.002 | 0.093 | 0.098 | 0.952 | -0.002 | 0.074 | 0.080 | 0.964 |
| | $\beta_{\lambda,2}$ | -0.009 | 0.091 | 0.081 | 0.940 | -0.006 | 0.068 | 0.066 | 0.920 |
| | $\beta_{\lambda,5}$ | 0.000 | 0.065 | 0.057 | 0.896 | 0.000 | 0.052 | 0.046 | 0.916 |
| | $\mu_0$ | -0.001 | 0.100 | 0.072 | 0.860 | -0.006 | 0.082 | 0.059 | 0.836 |
| | $\beta_{\mu,1}$ | -0.033 | 0.096 | 0.097 | 0.948 | -0.028 | 0.078 | 0.080 | 0.952 |
| | $\beta_{\mu,5}$ | 0.000 | 0.067 | 0.057 | 0.888 | 0.002 | 0.053 | 0.047 | 0.912 |
| | $\beta_{\mu,6}$ | 0.000 | 0.055 | 0.043 | 0.888 | 0.000 | 0.046 | 0.036 | 0.856 |
| 70 Assessments $\rho = 0.75$ | $\lambda_0$ | -0.020 | 0.088 | 0.077 | 0.896 | -0.025 | 0.076 | 0.063 | 0.872 |
| | $\beta_{\lambda,1}$ | 0.013 | 0.084 | 0.092 | 0.940 | 0.009 | 0.076 | 0.075 | 0.924 |
| | $\beta_{\lambda,2}$ | -0.006 | 0.073 | 0.076 | 0.964 | -0.002 | 0.060 | 0.062 | 0.964 |
| | $\beta_{\lambda,5}$ | -0.001 | 0.075 | 0.072 | 0.940 | -0.002 | 0.066 | 0.061 | 0.936 |
| | $\mu_0$ | -0.013 | 0.076 | 0.068 | 0.920 | -0.012 | 0.061 | 0.056 | 0.932 |
| | $\beta_{\mu,1}$ | -0.014 | 0.088 | 0.092 | 0.948 | -0.012 | 0.069 | 0.075 | 0.964 |
| | $\beta_{\mu,5}$ | -0.003 | 0.079 | 0.076 | 0.936 | 0.002 | 0.062 | 0.064 | 0.960 |
| | $\beta_{\mu,6}$ | 0.011 | 0.077 | 0.066 | 0.932 | 0.007 | 0.056 | 0.055 | 0.928 |