Supplementary Materials

Identification of physiological shock in intensive care units via Bayesian regime switching models

Contents

1 More Details on Electronic Health Record Data

1.1 Data Cleaning

There are 33 distinct response measurements for each of the 33,924 patients, and the length of each patient visit varies between 15 minutes and 50 days (the average length of stay is approximately 35 hours). Despite the large electronic health record (EHR) database available, not all of the information is necessary to analyze. To start, all 33 distinct response outcomes are not necessary to incorporate into the model because not all are relevant for the detection of hemorrhage. Based on clinical expertise, the relevant EHR measurements for hemorrhage detection are: hemoglobin concentration, heart rate, mean arterial pressure (MAP), lactate levels, red blood cell (RBC) transfusion order and administration times, and patient medications. In particular, hemoglobin, heart rate, MAP, and lactate serve as a four-dimensional longitudinal response, while the RBC transfusions and medication information serve as covariates. Given the response and covariate information, we then clean the training data based on the following criteria: (1) the patient did not die during their encounter; (2) the patient encounter lasts at least six hours; (3) if the patient encounter lasts more than 48 hours, only the first 48 hours are used; (4) less than 10% of heart rate and MAP measurements are missing; (5) patients do not have a pacemaker; (6) erroneous data points are removed (i.e., hemoglobin $\in [0, 20]$, heart rate $\in [20, 200]$, MAP $\in [25, 150]$); (7) at least 80% of the patient encounter is spent in intensive care unit (ICU) level-of-care. After adjusting for these criteria, the resulting dataset consists of 1,516 patient encounters. Due to the size of the data and model complexity, the compute time for training on all 1,516 subjects is computationally burdensome; hence, we use 500 subjects for training, and use five for testing. The reason for such a small test set is this allows us to have our clinical collaborators review these patient encounters closely and provide an in-depth commentary on the performance of our model with regards to uncovering their latent physiological changes, despite the data having no physiological-state labels.

1.2 Medications

The medication information most relevant for our use are: medication type, medication dose/strength, administration time, and administration type. With the help of clinical expertise, the medications can be separated based on the type of effect they are expected to have on heart rate and/or MAP. Similarly, we can separate medications based on if the administration type is *continuous* (e.g., intravenous) or *discrete* (e.g., oral pill). In total, there exists 44 distinct types of medications that are administered among all of the patients in our sample. Each medication can be separated based on how it is administered and how it affects heart rate and MAP, respectively. Hence, we have the following breakdown: 16 medications affect heart rate and are continuously administered, 18 medications affect heart rate and are discretely administered, 22 medications affect MAP and are continuously administered, and 28 medications affect MAP and are discretely administered. Thus, the 44 distinct medications leads to 84 medication effects to be learned in the model. Lastly, note that in the process of cleaning the medication data, any inconsistencies in a patient's medical record results in that patient being excluded from the training data. Section 3 in the manuscript details exactly how these medications are structured into the model.

2 Bayesian Computation Derivations

Full Conditional Distribution for $\text{vec}(\boldsymbol{\alpha}_*^{(i)})$

$$\begin{aligned}\pi(\text{vec}(\boldsymbol{\alpha}_*^{(i)}) \mid \text{rest}) &\propto \pi\Big(\text{vec}(\boldsymbol{\alpha}_*^{(i)}) \mid \text{vec}(\tilde{\boldsymbol{\alpha}}_*), \boldsymbol{\Upsilon}_\alpha\Big) \\ &\quad \times \prod_{k=2}^{n_i} f\Big(\boldsymbol{y}_k^{(i)} \mid \boldsymbol{y}_{k-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^k, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_{s_{i,k}}, \boldsymbol{R}\Big) \\ &\propto \text{N}_{16}\Big(\text{vec}(\boldsymbol{\alpha}_*^{(i)}) \mid \text{vec}(\tilde{\boldsymbol{\alpha}}_*),\ \boldsymbol{\Upsilon}_\alpha\Big) \times \prod_{k=2}^{n_i} \text{N}_4\Big(\boldsymbol{y}_k^{(i)} \mid \boldsymbol{\nu}_k^{(i)} + \mathbf{A}_{s_{i,k}}(\boldsymbol{y}_{k-1}^{(i)} - \boldsymbol{\nu}_{k-1}^{(i)}), \boldsymbol{R}\Big).\end{aligned}$$

The product of the probability density functions above leads to the following distribution:

$$\text{vec}(\boldsymbol{\alpha}_*^{(i)}) \mid \text{rest} \sim \text{N}_{16}(\boldsymbol{W}_i \boldsymbol{V}_i, \boldsymbol{W}_i)$$

where

$$\begin{aligned}\boldsymbol{W}_i &= \left\{\boldsymbol{\Upsilon}_\alpha^{-1} + \sum_{k=2}^{n_i}\Big(\boldsymbol{D}_{\alpha,k}^{(i)} - \mathbf{A}_{s_{i,k}}\boldsymbol{D}_{\alpha,k-1}^{(i)}\Big)^\top \boldsymbol{R}^{-1}\Big(\boldsymbol{D}_{\alpha,k}^{(i)} - \mathbf{A}_{s_{i,k}}\boldsymbol{D}_{\alpha,k-1}^{(i)}\Big)\right\}^{-1} \\ \boldsymbol{V}_i &= \boldsymbol{\Upsilon}_\alpha^{-1}\text{vec}(\tilde{\boldsymbol{\alpha}}_*) + \sum_{k=2}^{n_i}\Big(\boldsymbol{D}_{\alpha,k}^{(i)} - \mathbf{A}_{s_{i,k}}\boldsymbol{D}_{\alpha,k-1}^{(i)}\Big)^\top \boldsymbol{R}^{-1}\Big[\boldsymbol{y}_k^{(i)} - \mathbf{A}_{s_{i,k}}\boldsymbol{y}_{k-1}^{(i)} - (\boldsymbol{D}_{\omega,k}^{(i)} - \mathbf{A}_{s_{i,k}}\boldsymbol{D}_{\omega,k-1}^{(i)})\boldsymbol{\omega} \\ &\qquad - (\boldsymbol{X}_k^{(i)} - \mathbf{A}_{s_{i,k}}\boldsymbol{X}_{k-1}^{(i)})\boldsymbol{\beta} - (\mathbf{I} - \mathbf{A}_{s_{i,k}})\boldsymbol{\gamma}^{(i)}\Big]\end{aligned}$$

Full Conditional Distribution for $\boldsymbol{\gamma}^{(i)}$

$$\begin{aligned}\pi(\boldsymbol{\gamma}^{(i)} \mid \text{rest}) &\propto \pi\Big(\boldsymbol{\gamma}^{(i)} \mid \boldsymbol{y}_1^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{G}\Big) \cdot f\Big(\boldsymbol{y}_1^{(i)} \mid \boldsymbol{b}_1^{(i)} = s_{i,1}, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_{s_{i,1}}, \boldsymbol{R}\Big) \\ &\quad \times \prod_{k=2}^{n_i} f\Big(\boldsymbol{y}_k^{(i)} \mid \boldsymbol{y}_{k-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^k, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_{s_{i,k}}, \boldsymbol{R}\Big) \\ &\propto \text{N}_4(\boldsymbol{\gamma}^{(i)} \mid \boldsymbol{y}_1^{(i)} - \boldsymbol{D}_{\omega,1}^{(i)}\cdot\boldsymbol{\omega} - \boldsymbol{X}_1^{(i)}\cdot\boldsymbol{\beta},\ \boldsymbol{G}) \cdot \text{N}_4(\boldsymbol{y}_1^{(i)} \mid \boldsymbol{\gamma}^{(i)} + \boldsymbol{D}_{\omega,1}^{(i)}\cdot\boldsymbol{\omega} + \boldsymbol{X}_1^{(i)}\cdot\boldsymbol{\beta},\ \boldsymbol{\Gamma}_{s_{i,1}}) \\ &\quad \times \prod_{k=2}^{n_i} \text{N}_4\Big(\boldsymbol{y}_k^{(i)} \mid \boldsymbol{\nu}_k^{(i)} + \mathbf{A}_{s_{i,k}}\cdot(\boldsymbol{y}_{k-1}^{(i)} - \boldsymbol{\nu}_{k-1}^{(i)}), \boldsymbol{R}\Big).\end{aligned}$$

The product of the probability density functions above leads to the following distribution:

$$\boldsymbol{\gamma}^{(i)} \mid \text{rest} \sim \text{N}_4(\boldsymbol{W}_i \boldsymbol{V}_i, \boldsymbol{W}_i)$$

where

$$\boldsymbol{W}_i = \left\{ \boldsymbol{G}^{-1} + \boldsymbol{\Gamma}_{s_{i,1}}^{-1} + \sum_{k=2}^{n_i} \left(\mathbf{I} - \mathbf{A}_{s_{i,k}}\right)^\top \boldsymbol{R}^{-1} \left(\mathbf{I} - \mathbf{A}_{s_{i,k}}\right) \right\}^{-1}$$

$$\boldsymbol{V}_i = \left(\boldsymbol{G}^{-1} + \boldsymbol{\Gamma}_{s_{i,1}}^{-1}\right)\left(\boldsymbol{y}_1^{(i)} - \boldsymbol{D}_{\omega,1}^{(i)} \cdot \boldsymbol{\omega} - \boldsymbol{X}_1^{(i)} \cdot \boldsymbol{\beta}\right) + \sum_{k=2}^{n_i} \left(\mathbf{I} - \mathbf{A}_{s_{i,k}}\right)^\top \boldsymbol{R}^{-1} \Big[\boldsymbol{y}_k^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{y}_{k-1}^{(i)} - \left(\boldsymbol{D}_{\alpha,k}^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{D}_{\alpha,k-1}^{(i)}\right) \mathrm{vec}(\boldsymbol{\alpha}_*^{(i)}) - (\boldsymbol{D}_{\omega,k}^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{D}_{\omega,k-1}^{(i)})\boldsymbol{\omega} - (\boldsymbol{X}_k^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{X}_{k-1}^{(i)})\boldsymbol{\beta}\Big]$$

Full Conditional Distribution for $\boldsymbol{\omega}$

$$\begin{aligned}\pi(\boldsymbol{\omega} \mid \text{rest}) &\propto \left\{ \prod_{i=1}^{N} \pi\left(\boldsymbol{\gamma}^{(i)} \mid \boldsymbol{y}_1^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{G}\right) \cdot f\left(\boldsymbol{y}_1^{(i)} \mid \boldsymbol{b}_1^{(i)} = s_{i,1}, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_{s_{i,1}}, \boldsymbol{R}\right) \right. \\ &\quad \left. \times \prod_{k=2}^{n_i} f\left(\boldsymbol{y}_k^{(i)} \mid \boldsymbol{y}_{k-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^k, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_{s_{i,k}}, \boldsymbol{R}\right) \right\} \cdot \pi(\boldsymbol{\omega} \mid \boldsymbol{\omega}_0, \boldsymbol{\Sigma}_\omega) \\ &\propto \left\{ \prod_{i=1}^{N} \mathrm{N}_4(\boldsymbol{\gamma}^{(i)} \mid \boldsymbol{y}_1^{(i)} - \boldsymbol{D}_{\omega,1}^{(i)} \cdot \boldsymbol{\omega} - \boldsymbol{X}_1^{(i)} \cdot \boldsymbol{\beta},\ \boldsymbol{G}) \cdot \mathrm{N}_4(\boldsymbol{y}_1^{(i)} \mid \boldsymbol{\gamma}^{(i)} + \boldsymbol{D}_{\omega,1}^{(i)} \cdot \boldsymbol{\omega} + \boldsymbol{X}_1^{(i)} \cdot \boldsymbol{\beta},\ \boldsymbol{\Gamma}_{s_{i,1}}) \right. \\ &\quad \left. \times \prod_{k=2}^{n_i} \mathrm{N}_4\left(\boldsymbol{y}_k^{(i)} \mid \boldsymbol{\nu}_k^{(i)} + \mathbf{A}_{s_{i,k}} \cdot (\boldsymbol{y}_{k-1}^{(i)} - \boldsymbol{\nu}_{k-1}^{(i)}), \boldsymbol{R}\right) \right\} \cdot \mathrm{N}_{84}(\boldsymbol{\omega} \mid \boldsymbol{\omega}_0, \boldsymbol{\Sigma}_\omega).\end{aligned}$$

The product of the probability density functions above leads to the following distribution:

$$\boldsymbol{\omega} \mid \text{rest} \sim \mathrm{N}_{84}(\boldsymbol{W}\boldsymbol{V}, \boldsymbol{W})$$

where

$$\boldsymbol{W} = \left\{ \boldsymbol{\Sigma}_\omega^{-1} + \sum_{i=1}^{N} \boldsymbol{D}_{\omega,1}^{(i)\top} \left(\boldsymbol{G}^{-1} + \boldsymbol{\Gamma}_{s_{i,1}}^{-1}\right) \boldsymbol{D}_{\omega,1}^{(i)} + \sum_{k=2}^{n_i} \left(\boldsymbol{D}_{\omega,k}^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{D}_{\omega,k-1}^{(i)}\right)^\top \boldsymbol{R}^{-1} \left(\boldsymbol{D}_{\omega,k}^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{D}_{\omega,k-1}^{(i)}\right) \right\}^{-1}$$

$$\begin{aligned}\boldsymbol{V} &= \boldsymbol{\Sigma}_\omega^{-1} \boldsymbol{\omega}_0 + \sum_{i=1}^{N} \boldsymbol{D}_{\omega,1}^{(i)\top} \left(\boldsymbol{G}^{-1} + \boldsymbol{\Gamma}_{s_{i,1}}^{-1}\right) \left(\boldsymbol{y}_1^{(i)} - \boldsymbol{\gamma}^{(i)} - \boldsymbol{X}_1^{(i)} \cdot \boldsymbol{\beta}\right) \\ &\quad + \sum_{k=2}^{n_i} \left(\boldsymbol{D}_{\omega,k}^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{D}_{\omega,k-1}^{(i)}\right)^\top \boldsymbol{R}^{-1} \Big[\boldsymbol{y}_k^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{y}_{k-1}^{(i)} - (\boldsymbol{D}_{\alpha,k}^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{D}_{\alpha,k-1}^{(i)}) \mathrm{vec}(\boldsymbol{\alpha}_*^{(i)}) \\ &\quad - (\boldsymbol{X}_k^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{X}_{k-1}^{(i)})\boldsymbol{\beta} - (\mathbf{I} - \mathbf{A}_{s_{i,k}})\boldsymbol{\gamma}^{(i)}\Big]\end{aligned}$$

Full Conditional Distribution for $\boldsymbol{\beta}$

$$\begin{aligned}\pi(\boldsymbol{\beta} \mid \text{rest}) &\propto \left\{ \prod_{i=1}^{N} \pi\left(\boldsymbol{\gamma}^{(i)} \mid \boldsymbol{y}_1^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{G}\right) \cdot f\left(\boldsymbol{y}_1^{(i)} \mid \boldsymbol{b}_1^{(i)} = s_{i,1}, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_{s_{i,1}}, \boldsymbol{R}\right) \right. \\ &\quad \left. \times \prod_{k=2}^{n_i} f\left(\boldsymbol{y}_k^{(i)} \mid \boldsymbol{y}_{k-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^k, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_{s_{i,k}}, \boldsymbol{R}\right) \right\} \cdot \pi(\boldsymbol{\beta} \mid \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_\beta) \\ &\propto \left\{ \prod_{i=1}^{N} \mathrm{N}_4(\boldsymbol{\gamma}^{(i)} \mid \boldsymbol{y}_1^{(i)} - \boldsymbol{D}_{\omega,1}^{(i)} \cdot \boldsymbol{\omega} - \boldsymbol{X}_1^{(i)} \cdot \boldsymbol{\beta},\ \boldsymbol{G}) \cdot \mathrm{N}_4(\boldsymbol{y}_1^{(i)} \mid \boldsymbol{\gamma}^{(i)} + \boldsymbol{D}_{\omega,1}^{(i)} \cdot \boldsymbol{\omega} + \boldsymbol{X}_1^{(i)} \cdot \boldsymbol{\beta},\ \boldsymbol{\Gamma}_{s_{i,1}}) \right. \\ &\quad \left. \times \prod_{k=2}^{n_i} \mathrm{N}_4\left(\boldsymbol{y}_k^{(i)} \mid \boldsymbol{\nu}_k^{(i)} + \mathbf{A}_{s_{i,k}} \cdot (\boldsymbol{y}_{k-1}^{(i)} - \boldsymbol{\nu}_{k-1}^{(i)}), \boldsymbol{R}\right) \right\} \cdot \mathrm{N}_4(\boldsymbol{\beta} \mid \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_\beta).\end{aligned}$$

The product of the probability density functions above leads to the following distribution:

$$\boldsymbol{\beta} \mid \text{rest} \sim \text{N}_4(\boldsymbol{W}\boldsymbol{V}, \boldsymbol{W})$$

where

$$\boldsymbol{W} = \left\{ \boldsymbol{\Sigma}_\beta^{-1} + \sum_{i=1}^{N} \boldsymbol{X}_1^{(i)\top} \Big(\boldsymbol{G}^{-1} + \boldsymbol{\Gamma}_{s_{i,1}}^{-1} \Big) \boldsymbol{X}_1^{(i)} + \sum_{k=2}^{n_i} \Big(\boldsymbol{X}_k^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{X}_{k-1}^{(i)} \Big)^\top \boldsymbol{R}^{-1} \Big(\boldsymbol{X}_k^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{X}_{k-1}^{(i)} \Big) \right\}^{-1}$$

$$\begin{aligned}\boldsymbol{V} = \boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\beta}_0 + \sum_{i=1}^{N} \boldsymbol{X}_1^{(i)\top} \Big(\boldsymbol{G}^{-1} + \boldsymbol{\Gamma}_{s_{i,1}}^{-1} \Big) \Big(\boldsymbol{y}_1^{(i)} - \gamma^{(i)} - \boldsymbol{D}_{\omega,1}^{(i)} \boldsymbol{\omega} \Big) \\ + \sum_{k=2}^{n_i} (\boldsymbol{X}_k^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{X}_{k-1}^{(i)})^\top \boldsymbol{R}^{-1} \Big[\boldsymbol{y}_k^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{y}_{k-1}^{(i)} - (\boldsymbol{D}_{\alpha,k}^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{D}_{\alpha,k-1}^{(i)}) \text{vec}(\boldsymbol{\alpha}_*^{(i)}) \\ - (\boldsymbol{D}_{\omega,k}^{(i)} - \mathbf{A}_{s_{i,k}} \boldsymbol{D}_{\omega,k-1}^{(i)}) \boldsymbol{\omega} - (\mathbf{I} - \mathbf{A}_{s_{i,k}}) \boldsymbol{\gamma}^{(i)} \Big]\end{aligned}$$

Full Conditional Distribution for vec($\tilde{\boldsymbol{\alpha}}_*$)

$$\pi(\text{vec}(\tilde{\boldsymbol{\alpha}}_*) \mid \text{rest}) \propto \text{N}_{16}(\text{vec}(\tilde{\boldsymbol{\alpha}}_*) \mid \text{vec}(\tilde{\boldsymbol{\alpha}}_0), \boldsymbol{\Sigma}_\alpha) \cdot \prod_{i=1}^{N} \text{N}_{16}(\text{vec}(\boldsymbol{\alpha}_*^{(i)}) \mid \text{vec}(\tilde{\boldsymbol{\alpha}}_*), \boldsymbol{\Upsilon}_\alpha)$$

The product of the probability density functions above leads to the following distribution:

$$\text{vec}(\tilde{\boldsymbol{\alpha}}_*) \mid \text{rest} \sim \text{N}_{20}(\boldsymbol{W}\boldsymbol{V}, \boldsymbol{W})$$

where

$$\begin{aligned}\boldsymbol{W} &= \left(\boldsymbol{\Sigma}_\alpha^{-1} + N \cdot \boldsymbol{\Upsilon}_\alpha^{-1} \right)^{-1} \\ \boldsymbol{V} &= \boldsymbol{\Sigma}_\alpha^{-1} \text{vec}(\tilde{\boldsymbol{\alpha}}_0) + \boldsymbol{\Upsilon}_\alpha^{-1} \cdot \sum_{i=1}^{N} \text{vec}(\boldsymbol{\alpha}_*^{(i)})\end{aligned}$$

Full Conditional Distribution for $\boldsymbol{\Upsilon}_\alpha$

$$\pi(\boldsymbol{\Upsilon}_\alpha \mid \text{rest}) \propto \left\{ \prod_{i=1}^{N} \text{N}_{16}\big(\text{vec}(\boldsymbol{\alpha}_*^{(i)}) \mid \text{vec}(\tilde{\boldsymbol{\alpha}}_*), \boldsymbol{\Upsilon}_\alpha\big) \right\} \cdot \text{InvWish}(\boldsymbol{\Upsilon}_\alpha \mid \boldsymbol{\Psi}_\alpha, \nu_\alpha)$$

The product of the probability density functions above leads to the following distribution:

$$\boldsymbol{\Upsilon}_\alpha \mid \text{rest} \sim \text{InvWish}\left(\boldsymbol{\Psi}_\alpha + \sum_{i=1}^{N} \big[\text{vec}(\boldsymbol{\alpha}_*^{(i)}) - \text{vec}(\tilde{\boldsymbol{\alpha}}_*) \big] \big[\text{vec}(\boldsymbol{\alpha}_*^{(i)}) - \text{vec}(\tilde{\boldsymbol{\alpha}}_*) \big]^\top, \nu_\alpha + N \right)$$

Full Conditional Distribution for $\boldsymbol{G}$

$$\pi(\boldsymbol{G} \mid \text{rest}) \propto \left\{ \prod_{i=1}^{N} \text{N}_4(\boldsymbol{\gamma}^{(i)} \mid \boldsymbol{y}_1^{(i)} - \boldsymbol{D}_{\omega,1}^{(i)} \cdot \boldsymbol{\omega} - \boldsymbol{X}_1^{(i)} \cdot \boldsymbol{\beta},\ \boldsymbol{G}) \right\} \cdot \text{InvWish}(\boldsymbol{G} \mid \boldsymbol{\Psi}_G, \nu_G)$$

The product of the probability density functions above leads to the following distribution:

$$\boldsymbol{G} \mid \text{rest} \sim \text{InvWish}\left(\boldsymbol{\Psi}_G + \sum_{i=1}^{N} \left[\boldsymbol{\gamma}^{(i)} - \boldsymbol{y}_1^{(i)} + \boldsymbol{D}_{\omega,1}^{(i)} \cdot \boldsymbol{\omega} + \boldsymbol{X}_1^{(i)} \cdot \boldsymbol{\beta}\right]\left[\boldsymbol{\gamma}^{(i)} - \boldsymbol{y}_1^{(i)} + \boldsymbol{D}_{\omega,1}^{(i)} \cdot \boldsymbol{\omega} + \boldsymbol{X}_1^{(i)} \cdot \boldsymbol{\beta}\right]^{\top}, \nu_G + N\right)$$

"Approximate-Gibbs" Metropolis-Hastings Update for $\boldsymbol{R}$

Because of the autoregressive nature of our model, the conditional distribution of the initial time point has error covariance $\boldsymbol{\Gamma}_{s_{i,1}}$ instead of $\boldsymbol{R}$ [which the remaining time points use for the error covariance]. As a result, deriving a Gibbs update for $\boldsymbol{R}$ is infeasible; however, because $\boldsymbol{\Gamma}_{s_{i,1}}$ is a function of $\boldsymbol{R}$, we can derive an "approximate Gibbs" update for $\boldsymbol{R}$ which is in theory a Metropolis-Hastings update, but the proposal distribution emulates that of a full conditional distribution for $\boldsymbol{R}$. In particular, the full conditional distribution for $\boldsymbol{R}$ is proportional to the following

$$\begin{aligned}
\pi(\boldsymbol{R} \mid \text{rest}) \propto \pi(\boldsymbol{R} \mid \boldsymbol{\Psi}_R, \nu_R) \cdot \prod_{i=1}^{N} f\left(\boldsymbol{y}_1^{(i)} \mid \boldsymbol{b}_1^{(i)} = s_{i,1}, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_{s_{i,1}}, \boldsymbol{R}\right) \\
\times \prod_{k=2}^{n_i} f\left(\boldsymbol{y}_k^{(i)} \mid \boldsymbol{y}_{k-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{k}, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_{s_{i,k}}, \boldsymbol{R}\right) \\
\propto \text{InvWish}(\boldsymbol{R} \mid \boldsymbol{\Psi}_R, \nu_R) \cdot \prod_{i=1}^{N} \text{N}_4(\boldsymbol{y}_1^{(i)} \mid \boldsymbol{\gamma}^{(i)} + \boldsymbol{D}_{\omega,1}^{(i)} \cdot \boldsymbol{\omega} + \boldsymbol{X}_1^{(i)} \cdot \boldsymbol{\beta},\ \boldsymbol{\Gamma}_{s_{i,1}}) \\
\times \prod_{k=2}^{n_i} \text{N}_4\left(\boldsymbol{y}_k^{(i)} \mid \boldsymbol{\nu}_k^{(i)} + \mathbf{A}_{s_{i,k}} \cdot (\boldsymbol{y}_{k-1}^{(i)} - \boldsymbol{\nu}_{k-1}^{(i)}), \boldsymbol{R}\right).
\end{aligned}$$

Now, define $\boldsymbol{B}_i$ to be a matrix such that $\boldsymbol{B}_i \boldsymbol{\Gamma}_{s_{i,1}} \boldsymbol{B}_i^{\top} = \boldsymbol{R}$. Then, $\boldsymbol{B}_i = \boldsymbol{R}^{1/2} \boldsymbol{\Gamma}_{s_{i,1}}^{-1/2}$. As a result, we know the following:

$$\boldsymbol{B}_i \cdot \boldsymbol{y}_1^{(i)} \sim \text{N}_4\left(\boldsymbol{B}_i \cdot \left[\boldsymbol{\gamma}^{(i)} + \boldsymbol{D}_{\omega,1}^{(i)} \cdot \boldsymbol{\omega} + \boldsymbol{X}_1^{(i)} \cdot \boldsymbol{\beta}\right], \boldsymbol{R}\right)$$

Therefore, we get the following

$$\begin{aligned}
\pi(\boldsymbol{R} \mid \text{rest}) \approx \text{InvWish}(\boldsymbol{R} \mid \boldsymbol{\Psi}_R, \nu_R) \cdot \prod_{i=1}^{N} \text{N}_4\left(\boldsymbol{B}_i \cdot \boldsymbol{y}_1^{(i)} \mid \boldsymbol{B}_i \cdot \left[\boldsymbol{\gamma}^{(i)} + \boldsymbol{D}_{\omega,1}^{(i)} \cdot \boldsymbol{\omega} + \boldsymbol{X}_1^{(i)} \cdot \boldsymbol{\beta}\right], \boldsymbol{R}\right) \\
\times \prod_{k=2}^{n_i} \text{N}_4\left(\boldsymbol{y}_k^{(i)} \mid \boldsymbol{\nu}_k^{(i)} + \mathbf{A}_{s_{i,k}} \cdot (\boldsymbol{y}_{k-1}^{(i)} - \boldsymbol{\nu}_{k-1}^{(i)}), \boldsymbol{R}\right).
\end{aligned}$$

The approximate full conditional distribution for $\boldsymbol{R}$ above simplifies to an Inverse-Wishart distribution dependent on the value of $\boldsymbol{B}_i = \boldsymbol{R}^{1/2} \boldsymbol{\Gamma}_{s_{i,1}}^{-1/2}$. Therefore, given the current value of $\boldsymbol{B}_i$ (i.e., $\boldsymbol{R}$), we can write the proposal distribution for some new $\boldsymbol{R}^*$ as

$$q(\boldsymbol{R}^* \mid \boldsymbol{R}, \text{rest}) \sim \text{InvWish}(\boldsymbol{\Psi}_q, \nu_q)$$

where

$$\boldsymbol{\Psi}_q = \boldsymbol{\Psi}_R + \sum_{i=1}^{N} (\boldsymbol{R}^{1/2}\boldsymbol{\Gamma}_{s_{i,1}}^{-1/2})(\boldsymbol{y}_1^{(i)} - \boldsymbol{\nu}_1^{(i)})(\boldsymbol{y}_1^{(i)} - \boldsymbol{\nu}_1^{(i)})^\top (\boldsymbol{R}^{1/2}\boldsymbol{\Gamma}_{s_{i,1}}^{-1/2})^\top$$
$$+ \sum_{k=2}^{n_i} \left[\boldsymbol{y}_k^{(i)} - \boldsymbol{\nu}_k^{(i)} - \mathbf{A}_{s_{i,k}} \cdot (\boldsymbol{y}_{k-1}^{(i)} - \boldsymbol{\nu}_{k-1}^{(i)})\right] \left[\boldsymbol{y}_k^{(i)} - \boldsymbol{\nu}_k^{(i)} - \mathbf{A}_{s_{i,k}} \cdot (\boldsymbol{y}_{k-1}^{(i)} - \boldsymbol{\nu}_{k-1}^{(i)})\right]^\top$$
$$\nu_q = \nu_R + \sum_{i=1}^{N} n_i.$$

where $\boldsymbol{\nu}_1^{(i)} = \boldsymbol{\gamma}^{(i)} + \boldsymbol{D}_{\omega,1}^{(i)} \cdot \boldsymbol{\omega} + \boldsymbol{X}_1^{(i)} \cdot \boldsymbol{\beta}$.

3 Exact Prior Specifications

$\boxed{\mathbf{A}_1, \ldots, \mathbf{A}_5}$

$$\mathbf{A}_j = \begin{pmatrix} a_{1,j} & 0 & 0 & 0 \\ 0 & a_{2,j} & 0 & 0 \\ 0 & 0 & a_{3,j} & 0 \\ 0 & 0 & 0 & a_{4,j} \end{pmatrix},$$

and

$$(\text{logit}(a_{1,j}), \text{logit}(a_{2,j}), \text{logit}(a_{3,j}), \text{logit}(a_{4,j})) \overset{iid}{\sim} \text{N}_4(\mathbf{0}, \boldsymbol{I}),$$

for $j \in \{1, 2, 3, 4, 5\}$.

$\boxed{\boldsymbol{R}}$

$$\boldsymbol{R} \sim \text{InvWish}(\boldsymbol{\Psi}_R, \nu_R)$$

where

$$\boldsymbol{\Psi}_R = \nu_R \cdot \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{3}{2} & 0 & 0 \\ 0 & 0 & \frac{3}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

and $\nu_R = 2 \cdot \sum_{i=1}^N n_i$.

$\boxed{\boldsymbol{\zeta}}$

$$\boldsymbol{\zeta} \sim \text{N}_{24}(\boldsymbol{\mu}_\zeta, \boldsymbol{\Sigma}_\zeta)$$

where $\boldsymbol{\mu}_\zeta$ = (-7.2405, 2.5, -6.2152, 1, -2.6473, -1, -6.1475, -1, -9.4459, -1, -7.2404, 2.5, -7.2151, 1, -7.1778, 2.5, -5.2151, 0, -9.4459, -1, -7.2404, 2.5, -5.2151, 0), and $\boldsymbol{\Sigma}_\zeta = \boldsymbol{I}_{24}$.

$\boxed{\boldsymbol{\pi}}$

$$\text{logit}(\boldsymbol{\pi}) \sim \text{N}_4(\mathbf{0}, 100 \cdot \boldsymbol{I})$$

$\boxed{\boldsymbol{\omega}}$

$$\boldsymbol{\omega} \sim \mathrm{N}_{84}(\boldsymbol{\omega}_0,\ \boldsymbol{\Sigma}_\omega)$$

where

$$\begin{aligned}\boldsymbol{\omega}_0 = \frac{3}{2} \cdot (&-1, 1, 1, -1, -1, 1, 1, -1, 1, 1, -1, -1, 1, -1, 1, 1, -1, -1, -1, -1, 1, -1, 1, -1, 1,\\ &-1, -1, -1, -1, -1, 1, 1, -1, -1, -1, -1, -1, 1, 1, 1, -1, 1, -1, -1, -1, 1, -1, 1,\\ &-1, 1, -1, -1, -1, 1, 1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 1, 1, 1, -1, -1,\\ &-1, 1, -1, 1, -1, -1, -1, -1, 1, -1, -1, -1, -1, -1),\end{aligned}$$

$\boldsymbol{\Sigma}_\omega = \frac{1}{16} \cdot \boldsymbol{I}_{84}$, and $[\boldsymbol{\Sigma}_\omega]_{6,6} = [\boldsymbol{\Sigma}_\omega]_{42,42} = [\boldsymbol{\Sigma}_\omega]_{49,49} = \frac{1}{625}$.

$\boxed{\boldsymbol{\beta}}$

$$\boldsymbol{\beta} \sim \mathrm{N}_4(\boldsymbol{\beta}_0,\ \boldsymbol{\Sigma}_\beta)$$

$\boldsymbol{\beta}_0 = \left(\frac{1}{4}, -2, 2, -\frac{1}{4}\right)$ and

$$\boldsymbol{\Sigma}_\beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 16 & 0 & 0 \\ 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 16 \end{pmatrix}$$

$\boxed{\boldsymbol{\Upsilon}_\alpha}$

$$\boldsymbol{\Upsilon}_\alpha \sim \mathrm{InvWish}(\boldsymbol{\Psi}_\alpha,\ \nu_\alpha)$$

where

$$\boldsymbol{\Psi}_\alpha = (\nu_\alpha - 16 - 1) \cdot \mathrm{diag}\left(\frac{1}{4}, \frac{1}{4}, 4, 4, \frac{9}{4}, \frac{9}{4}, 25, 25, \frac{9}{4}, \frac{9}{4}, 25, 25, \frac{1}{4}, \frac{1}{4}, 4, 4\right)$$

and $\nu_\alpha = 80$.

$\boxed{\mathbf{vec}(\tilde{\boldsymbol{\alpha}}_*)}$

$$\mathrm{vec}(\tilde{\boldsymbol{\alpha}}_*) \sim \mathrm{N}_{16}(\mathrm{vec}(\tilde{\boldsymbol{\alpha}}_0),\ \boldsymbol{\Sigma}_\alpha)$$

where

$$\begin{aligned}\mathrm{vec}(\tilde{\boldsymbol{\alpha}}_0) &= (-1, 1, 0, 0, 7, -7, 0, 0, -7, 7, 0, 0, 1, -1, 0, 0),\\ \boldsymbol{\Sigma}_\alpha &= \mathrm{diag}\left(\frac{1}{4}, \frac{1}{4}, 4, 4, 4, 4, 64, 64, 4, 4, 64, 64, \frac{1}{4}, \frac{1}{4}, 4, 4\right)\end{aligned}$$

$\boxed{\boldsymbol{G}}$

$$\boldsymbol{G} \sim \mathrm{InvWish}(\boldsymbol{\Psi}_G,\ \nu_G)$$

where $\boldsymbol{\Psi}_G = (\nu_G - 4 - 1) \cdot \mathrm{diag}(8, 32, 32, 8)$ and $\nu_G = 8$.

4 Data Imputation

Since we have missingness in the data, data imputation for the response vector is necessary. The following are the distributions from which we sample the missing observations from.

$\boxed{k = 1}$

$$\pi(\boldsymbol{y}_1^{(i)} \mid \text{rest}) \propto \pi\Big(\boldsymbol{\gamma}^{(i)} \mid \boldsymbol{y}_1^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{G}\Big) \cdot f\Big(\boldsymbol{y}_1^{(i)} \mid \boldsymbol{b}_1^{(i)} = s_{i,1}, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_{s_{i,1}}, \boldsymbol{R}\Big)$$
$$\times f\Big(\boldsymbol{y}_2^{(i)} \mid \boldsymbol{y}_1^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{2}, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_{s_{i,2}}, \boldsymbol{R}\Big).$$

The product of the probability density functions above leads to the following distribution

$$\boldsymbol{y}_1^{(i)} \mid \text{rest} \sim \text{N}_4(\boldsymbol{W}_{i,1}\boldsymbol{V}_{i,1}, \boldsymbol{W}_{i,1})$$

where

$$\boldsymbol{W}_{i,1} = \Big(\boldsymbol{G}^{-1} + \boldsymbol{\Gamma}_{s_{i,1}}^{-1} + \mathbf{A}_{s_{i,2}}^{\top}\boldsymbol{R}^{-1}\mathbf{A}_{s_{i,2}}\Big)^{-1}$$
$$\boldsymbol{V}_{i,1} = \Big(\boldsymbol{G}^{-1} + \boldsymbol{\Gamma}_{s_{i,1}}^{-1}\Big)\Big(\boldsymbol{\gamma}^{(i)} + \boldsymbol{D}_{\omega,1}^{(i)} \cdot \boldsymbol{\omega} + \boldsymbol{X}_1^{(i)} \cdot \boldsymbol{\beta}\Big)$$
$$+ \mathbf{A}_{s_{i,2}}^{\top}\boldsymbol{R}^{-1}\Big(\boldsymbol{y}_2^{(i)} - \boldsymbol{D}_{\alpha,2}^{(i)}\text{vec}(\boldsymbol{\alpha}_*^{(i)}) - (\boldsymbol{D}_{\omega,2}^{(i)} - \mathbf{A}_{s_{i,2}}\boldsymbol{D}_{\omega,1}^{(i)})\boldsymbol{\omega} - (\boldsymbol{X}_2^{(i)} - \mathbf{A}_{s_{i,2}}\boldsymbol{X}_1^{(i)})\boldsymbol{\beta} - (\mathbf{I} - \mathbf{A}_{s_{i,2}})\boldsymbol{\gamma}^{(i)}\Big).$$

$\boxed{k \in \{2, 3, \ldots, n_i - 1\}}$

$$\pi(\boldsymbol{y}_k^{(i)} \mid \text{rest}) \propto f\Big(\boldsymbol{y}_k^{(i)} \mid \boldsymbol{y}_{k-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{k}, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_{s_{i,k}}, \boldsymbol{R}\Big)$$
$$\times f\Big(\boldsymbol{y}_{k+1}^{(i)} \mid \boldsymbol{y}_k^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{k+1}, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_{s_{i,k+1}}, \boldsymbol{R}\Big).$$

The product of the probability density functions above leads to the following distribution

$$\boldsymbol{y}_k^{(i)} \mid \text{rest} \sim \text{N}_4(\boldsymbol{W}_{i,k}\boldsymbol{V}_{i,k}, \boldsymbol{W}_{i,k})$$

where

$$\boldsymbol{W}_{i,k} = \Big(\boldsymbol{R}^{-1} + \mathbf{A}_{s_{i,k+1}}^{\top}\boldsymbol{R}^{-1}\mathbf{A}_{s_{i,k+1}}\Big)^{-1}$$
$$\boldsymbol{V}_{i,k} = \boldsymbol{R}^{-1} \cdot \Big[\mathbf{A}_{s_{i,k}}\boldsymbol{y}_{k-1}^{(i)} + (\boldsymbol{D}_{\alpha,k}^{(i)} - \mathbf{A}_{s_{i,k}}\boldsymbol{D}_{\alpha,k-1}^{(i)})\text{vec}(\boldsymbol{\alpha}_*^{(i)}) + (\boldsymbol{D}_{\omega,k}^{(i)} - \mathbf{A}_{s_{i,k}}\boldsymbol{D}_{\omega,k-1}^{(i)})\boldsymbol{\omega}$$
$$+ (\boldsymbol{X}_k^{(i)} - \mathbf{A}_{s_{i,k}}\boldsymbol{X}_{k-1}^{(i)})\boldsymbol{\beta} + (\mathbf{I} - \mathbf{A}_{s_{i,k}})\boldsymbol{\gamma}^{(i)}\Big]$$
$$+ \mathbf{A}_{s_{i,k+1}}^{\top}\boldsymbol{R}^{-1} \cdot \Big[\boldsymbol{y}_{k+1}^{(i)} - (\boldsymbol{D}_{\alpha,k+1}^{(i)} - \mathbf{A}_{s_{i,k+1}}\boldsymbol{D}_{\alpha,k}^{(i)})\text{vec}(\boldsymbol{\alpha}_*^{(i)}) - (\boldsymbol{D}_{\omega,k+1}^{(i)} - \mathbf{A}_{s_{i,k+1}}\boldsymbol{D}_{\omega,k}^{(i)})\boldsymbol{\omega}$$
$$- (\boldsymbol{X}_{k+1}^{(i)} - \mathbf{A}_{s_{i,k+1}}\boldsymbol{X}_k^{(i)})\boldsymbol{\beta} - (\mathbf{I} - \mathbf{A}_{s_{i,k+1}})\boldsymbol{\gamma}^{(i)}\Big].$$

$\boxed{k = n_i}$

$$\pi(\boldsymbol{y}_{n_i}^{(i)} \mid \text{rest}) \propto f\Big(\boldsymbol{y}_{n_i}^{(i)} \mid \boldsymbol{y}_{n_i-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{n_i}, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_{s_{i,n_i}}, \boldsymbol{R}\Big).$$

Thus, we trivially get the conditional response distribution for $\boldsymbol{y}_{n_i}^{(i)}$:

$$\boldsymbol{y}_{n_i}^{(i)} \mid \text{rest} \sim \text{N}_4\Big(\boldsymbol{\nu}_{n_i}^{(i)} + \mathbf{A}(\boldsymbol{y}_{n_i-1}^{(i)} - \boldsymbol{\nu}_{n_i-1}^{(i)}), \boldsymbol{R}\Big)$$

5 Our Metropolis-Hastings (MH) State-Sampler

Algorithm 1: MH state-sampling for subject $i \in \{1, 2, \ldots, N\}$

Input: Current $\{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{n_i}$ and model parameters

Data: $\boldsymbol{Y}^{(i)}$ for a given $i \in \{1, 2, \ldots, N\}$

p = randomly select from $\{2, 3, \ldots, 50\}$ with equal probability

if $p \geq n_i$ **then**

- Use Algorithm 2

else if $p = 2$ **then**

- Use Algorithm 3

else

- $k = 1$
- **while** $k \leq n_i - 2$ **do**
  - Initialize $s^*_{i,j} = s_{i,j}, \forall j \in \{1, 2, \ldots, n_i\}$
  - $k_{\max} = \min\{k + p - 1, n_i\}$
  - `/* Propose state sequence` $\{\boldsymbol{b}_j^{(i)} = s^*_{i,j}\}_{j=k}^{k_{\max}}$ `for time points` $k, k+1, \ldots, k_{\max}$ `*/`
  - **for** $t \in \{k, k+1, \ldots, k_{max} - 2\}$ **do**
    - `like_vals`: vector with length equal to number of distinct possible states (i.e., size of state-space)
    - **for** $m \in \{1, 2, \ldots, 5\}$ **do**
      - **if** $t = 1$ **then**
        - `like_vals`$[m] = \boldsymbol{\pi}_m \cdot f\Big(\boldsymbol{y}_1^{(i)} \mid \boldsymbol{b}_1^{(i)} = m, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_m, \boldsymbol{R}\Big)$
      - **else**
        - `like_vals`$[m] =$ $\mathbf{P}_{s^*_{i,t-1},m} \cdot f\Big(\boldsymbol{y}_t^{(i)} \mid \boldsymbol{y}_{t-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s^*_{i,j}\}_{j=1}^{t-1}, \boldsymbol{b}_t^{(i)} = m, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_m, \boldsymbol{R}\Big)$
    - $s^*_{i,t}$ = sample from $\{1, 2, \ldots, 5\}$ proportional to `like_vals`
  - **if** $k_{max} < n_i$ **then**
    - Randomly select $\{s^*_{i,k_{\max}-1}, s^*_{i,k_{\max}}\} \in \mathcal{B}^{(2)}_{s^*_{i,k_{\max}-2}, s^*_{i,k_{\max}+1}}$ with equal probability
  - **else**
    - Randomly select $\{s^*_{i,k_{\max}-1}, s^*_{i,k_{\max}}\} \in \mathcal{B}^{(2)}_{s^*_{i,k_{\max}-2},\cdot}$ with equal probability
  - a = MH acceptance ratio between the $\{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{n_i}$ and $\{\boldsymbol{b}_j^{(i)} = s^*_{i,j}\}_{j=1}^{n_i}$
  - u = sample from uniform$[0, 1]$
  - **if** $u < a$ **then**
    - $s_{i,j} = s^*_{i,j}, \forall j \in \{1, 2, \ldots, n_i\}$ `// proposal accepted`
  - $k = k + p - 2$
- **return** $\{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{n_i}$

Algorithm 2: MH state-sampling for subject $i \in \{1, 2, \ldots, N\}$ and $p \geq n_i$

Input: Current $\{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{n_i}$ and model parameters

Data: $\boldsymbol{Y}^{(i)}$ for a given $i \in \{1, 2, \ldots, N\}$

/* Proposing a full state sequence $\{\boldsymbol{b}_j^{(i)} = s^*_{i,j}\}_{j=1}^{n_i}$ */

for $k \in \{1, 2, \ldots, n_i\}$ **do**
- `like_vals`: vector with length equal to number of distinct possible states (i.e., size of state-space)
- **for** $m \in \{1, 2, \ldots, 5\}$ **do**
  - **if** $k = 1$ **then**
    - `like_vals`$[m] = \boldsymbol{\pi}_m \cdot f\left(\boldsymbol{y}_1^{(i)} \mid \boldsymbol{b}_1^{(i)} = m, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_m, \boldsymbol{R}\right)$
  - **else**
    - `like_vals`$[m] = \mathbf{P}_{s^*_{i,k-1}, m} \cdot f\left(\boldsymbol{y}_k^{(i)} \mid \boldsymbol{y}_{k-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s^*_{i,j}\}_{j=1}^{k-1}, \boldsymbol{b}_k^{(i)} = m, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\omega}, \boldsymbol{\beta}, \mathbf{A}_m, \boldsymbol{R}\right)$
- $s^*_{i,k}$ = sample from $\{1, 2, \ldots, 5\}$ proportional to `like_vals`

a = MH acceptance ratio between the $\{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{n_i}$ and $\{\boldsymbol{b}_j^{(i)} = s^*_{i,j}\}_{j=1}^{n_i}$

u = sample from uniform$[0, 1]$

if $u < a$ **then**
- $s_{i,j} = s^*_{i,j}, \forall j \in \{1, 2, \ldots, n_i\}$ // proposal accepted

return $\{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{n_i}$

Algorithm 3: MH state-sampling for subject $i \in \{1, 2, \ldots, N\}$ and $p = 2$

Input: Current $\{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{n_i}$ and model parameters

Data: $\boldsymbol{Y}^{(i)}$ for a given $i \in \{1, 2, \ldots, N\}$

for $k \in \{1, 2, \ldots, n_i\}$ **do**
- Initialize $s^*_{i,j} = s_{i,j}, \forall j \in \{1, 2, \ldots, n_i\}$
- /* Proposing $\{\boldsymbol{b}_k^{(i)} = s^*_{i,k}, \boldsymbol{b}_{k+1}^{(i)} = s^*_{i,k+1}\}$ */
- **if** $k = 1$ **then**
  - Randomly select $\{s^*_{i,1}, s^*_{i,2}\} \in \mathcal{B}^{(2)}_{\cdot, s^*_{i,3}}$ with equal probability
- **else if** $k = n_i - 1$ **then**
  - Randomly select $\{s^*_{i,n_i-1}, s^*_{i,n_i}\} \in \mathcal{B}^{(2)}_{s^*_{i,n_i-2}, \cdot}$ with equal probability
- **else**
  - Randomly select $\{s^*_{i,k+p-2}, s^*_{i,k+p-1}\} \in \mathcal{B}^{(2)}_{s^*_{i,k+p-3}, s^*_{i,k+p}}$ with equal probability
- a = MH acceptance ratio between the $\{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{n_i}$ and $\{\boldsymbol{b}_j^{(i)} = s^*_{i,j}\}_{j=1}^{n_i}$
- u = sample from uniform$[0, 1]$
- **if** $u < a$ **then**
  - $s_{i,j} = s^*_{i,j}, \forall j \in \{1, 2, \ldots, n_i\}$ // proposal accepted

return $\{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{n_i}$

6 Additional Information on Algorithms 1, 2, and 3

Although Algorithms 1, 2, and 3 provide pseudo-code for the logic of the state-sampling routine, the following provides further mathematical details for each of the algorithms. In particular, we define the proposal distributions for the cases of $p = 2$ and for $p \geq n_i$, for a given subject i.

In the case of Algorithm 3 where $p = 2$, the proposal distribution is precisely the same as *Approach (A)* defined in Supplementary Materials Section 7. In the case of Algorithm 2 where $p \geq n_i$, this more closely aligns with the proposal distribution provided in (6) in the manuscript. In particular, the proposal distribution for Algorithm 2 makes proposals over the entire state sequence for subject i, and is defined as

$$
\begin{aligned}
&q\Big(\{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{n_i} \mid \boldsymbol{Y}^{(i)}, \boldsymbol{\alpha}_*^{(i)}, \boldsymbol{\omega}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\beta}, \mathbf{A}, \boldsymbol{R}, \boldsymbol{\zeta}, \boldsymbol{\pi}\Big) \\
&:= \frac{\boldsymbol{\pi}_{s_{i,1}} \cdot f\Big(\boldsymbol{y}_1^{(i)} \mid \boldsymbol{b}_1^{(i)} = s_{i,1},\ \text{rest}\Big)}{\sum_{m=1}^{5} \boldsymbol{\pi}_m \cdot f\Big(\boldsymbol{y}_1^{(i)} \mid \boldsymbol{b}_1^{(i)} = m,\ \text{rest}\Big)} \times \prod_{t=2}^{n_i} \frac{\mathbf{P}_{s_{i,t-1}, s_{i,t}} \cdot f\Big(\boldsymbol{y}_t^{(i)} \mid \boldsymbol{y}_{t-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{t},\ \text{rest}\Big)}{\sum_{m=1}^{5} \mathbf{P}_{s_{i,t-1}, m} \cdot f\Big(\boldsymbol{y}_t^{(i)} \mid \boldsymbol{y}_{t-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{t-1}, \boldsymbol{b}_t^{(i)} = m,\ \text{rest}\Big)}
\end{aligned}
$$

Additionally, the acceptance ratio for the MH sampling in Algorithm 2 does not require computing the full joint posterior distribution. It can be shown (for $2 < p < n_i$) that the MH acceptance ratio between the proposed state sequence, $\{\boldsymbol{b}_j^{(i)} = s_{i,j}^*\}_{j=1}^{n_i}$, and the current state sequence, $\{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{n_i}$, simplifies to

$$
\begin{aligned}
MH_{\text{ratio}} = &\left\{ \prod_{t=k_{\max}-1}^{k_{\max}+1} \frac{\mathbf{P}_{s_{i,t-1}^*, s_{i,t}^*}}{\mathbf{P}_{s_{i,t-1}, s_{i,t}}} \right\} \cdot \left\{ \prod_{t=k_{\max}-1}^{n_i} \frac{f\Big(\boldsymbol{y}_t^{(i)} \mid \boldsymbol{y}_{t-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}^*\}_{j=1}^{t},\ \text{rest}\Big)}{f\Big(\boldsymbol{y}_t^{(i)} \mid \boldsymbol{y}_{t-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{t},\ \text{rest}\Big)} \right\} \\
&\times \left\{ \prod_{t=k}^{k_{\max}-2} \frac{\sum_{m=1}^{5} \mathbf{P}_{s_{i,t-1}^*, m} \cdot f\Big(\boldsymbol{y}_t^{(i)} \mid \boldsymbol{y}_{t-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}^*\}_{j=1}^{t-1}, \boldsymbol{b}_t^{(i)} = m,\ \text{rest}\Big)}{\sum_{m=1}^{5} \mathbf{P}_{s_{i,t-1}, m} \cdot f\Big(\boldsymbol{y}_t^{(i)} \mid \boldsymbol{y}_{t-1}^{(i)}, \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^{t-1}, \boldsymbol{b}_t^{(i)} = m,\ \text{rest}\Big)} \right\} \\
&\times \frac{|\mathcal{B}^{(2)}_{s_{i,k_{\max}-2}^*, s_{i,k_{\max}+1}^*}|}{|\mathcal{B}^{(2)}_{s_{i,k_{\max}-2}, s_{i,k_{\max}+1}}|},
\end{aligned}
$$

where each component above has already been computed from the derivation of the proposal distribution.

7 Alternative State-Sampling Routines

Approach (A)

$$q\Big(\{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=k}^{k+p-1} \mid \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j\notin\{k,\ldots,k+p-1\}},\ \text{rest}\Big) = \begin{cases} |\mathcal{B}^{(p)}_{\cdot,\, s_{i,k+p}}|^{-1} & ,\ k=1 \\ |\mathcal{B}^{(p)}_{s_{i,k-1},\, s_{i,k+p}}|^{-1} & ,\ k\in\{2,\ldots,n_i-p\} \\ |\mathcal{B}^{(p)}_{s_{i,k-1},\, \cdot}|^{-1} & ,\ k=n_i-p+1 \end{cases}$$

Approach (B)

$$q\Big(\{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=k}^{k+p-1} \mid \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j\notin\{k,\ldots,k+p-1\}},\ \text{rest}\Big)$$

$$= \begin{cases} \dfrac{\boldsymbol{\pi}_{s_{i,1}}\cdot f(\boldsymbol{y}_1^{(i)}|\boldsymbol{b}_1^{(i)}{=}s_{i,1},\,\text{rest})\prod_{t=2}^{n_i}\mathbf{P}_{s_{i,t-1},s_{i,t}}\cdot f\Big(\boldsymbol{y}_t^{(i)}|\boldsymbol{y}_{t-1}^{(i)},\{\boldsymbol{b}_j^{(i)}{=}s_{i,j}\}_{j=1}^{t},\,\text{rest}\Big)}{\sum_{\tilde{\boldsymbol{s}}\in\mathcal{B}^{(p)}_{\cdot,s_{i,k+p}}}\boldsymbol{\pi}_{\tilde{s}_1}\cdot f(\boldsymbol{y}_1^{(i)}|\boldsymbol{b}_1^{(i)}{=}\tilde{s}_1,\,\text{rest})\prod_{t=2}^{n_i}\mathbf{P}_{\tilde{s}_{t-1},\tilde{s}_t}\cdot f\Big(\boldsymbol{y}_t^{(i)}|\boldsymbol{y}_{t-1}^{(i)},\{\boldsymbol{b}_j^{(i)}{=}\tilde{s}_j\}_{j=1}^{t},\,\text{rest}\Big)} & ,\ k=1 \\[2ex] \dfrac{\prod_{t=k}^{n_i}\mathbf{P}_{s_{i,t-1},s_{i,t}}\cdot f\Big(\boldsymbol{y}_t^{(i)}|\boldsymbol{y}_{t-1}^{(i)},\{\boldsymbol{b}_j^{(i)}{=}s_{i,j}\}_{j=1}^{t},\,\text{rest}\Big)}{\sum_{\tilde{\boldsymbol{s}}\in\mathcal{B}^{(p)}_{s_{i,k-1},\,s_{i,k+p}}}\prod_{t=k}^{n_i}\mathbf{P}_{\tilde{s}_{t-1},\tilde{s}_t}\cdot f\Big(\boldsymbol{y}_t^{(i)}|\boldsymbol{y}_{t-1}^{(i)},\{\boldsymbol{b}_j^{(i)}{=}\tilde{s}_j\}_{j=1}^{t},\,\text{rest}\Big)} & ,\ k\in\{2,\ldots,n_i-p\} \\[2ex] \dfrac{\prod_{t=k}^{n_i}\mathbf{P}_{s_{i,t-1},s_{i,t}}\cdot f\Big(\boldsymbol{y}_t^{(i)}|\boldsymbol{y}_{t-1}^{(i)},\{\boldsymbol{b}_j^{(i)}{=}s_{i,j}\}_{j=1}^{t},\,\text{rest}\Big)}{\sum_{\tilde{\boldsymbol{s}}\in\mathcal{B}^{(p)}_{s_{i,k-1},\,\cdot}}\prod_{t=k}^{n_i}\mathbf{P}_{\tilde{s}_{t-1},\tilde{s}_t}\cdot f\Big(\boldsymbol{y}_t^{(i)}|\boldsymbol{y}_{t-1}^{(i)},\{\boldsymbol{b}_j^{(i)}{=}\tilde{s}_j\}_{j=1}^{t},\,\text{rest}\Big)} & ,\ k=n_i-p+1 \end{cases}$$

8 Parameterization for Simulation Study

$$\boldsymbol{\beta} = \begin{pmatrix} 0.4322 \\ -0.7361 \\ 1.8589 \\ 0.0361 \end{pmatrix}, \quad \tilde{\boldsymbol{\alpha}}_* = \begin{pmatrix} -0.8429 & 6.6528 & -9.5501 & 0.7762 \\ 0.7443 & -6.9619 & 8.0447 & -0.7595 \\ 0.0919 & 0.0330 & 2.2708 & 0.2091 \\ -0.1177 & -1.3514 & -1.4781 & -0.0061 \end{pmatrix},$$

$$\begin{aligned} &\mathbf{A}_1 = \mathrm{diag}(0.1652, 0.9707, 0.8804, 0.8664), \quad \mathbf{A}_2 = \mathrm{diag}(0.1412, 0.0041, 0.0020, 0.8122), \\ &\mathbf{A}_3 = \mathrm{diag}(0.5263, 0.2666, 0.0068, 0.9035), \quad \mathbf{A}_4 = \mathrm{diag}(0.2314, 0.0009, 0.0013, 0.7314), \\ &\mathbf{A}_5 = \mathrm{diag}(0.1029, 0.3675, 0.1392, 0.2065), \end{aligned}$$

$$\boldsymbol{R} = \begin{pmatrix} 0.49852 & 0.00097 & 0.00236 & 0.00001 \\ 0.00097 & 5.61452 & 1.51487 & 0.00083 \\ 0.00236 & 1.51487 & 10.55849 & -0.00111 \\ 0.00001 & 0.00083 & -0.00111 & 0.49784 \end{pmatrix},$$

$$\boldsymbol{\zeta} = \begin{pmatrix} -2.9732, -2.1213, 1.9556, 0.6334, 1.0973, 0.1456, 0.3555, -1.0543, 0.7188, -0.1747, -1.9821, -0.4705 \\ 0.5475, 0.1003, -0.5696, 0.0869, -1.3116, -0.1631, -0.8327, 0.5962, 0.2341, -0.6008, 0.1264, -0.1105 \end{pmatrix}$$

$$\begin{aligned} \mathrm{diag}(\boldsymbol{\Upsilon}_\alpha) = (&0.1437, 0.1289, 0.9086, 0.9075, 33.012, 35.8889, 26.9045, 20.0647, \\ &54.0599, 42.9315, 43.4438, 26.2896, 0.1466, 0.1356, 1.0752, 1.1122), \end{aligned}$$

$$\begin{aligned} \boldsymbol{\omega} = (&-1.458, 1.307, 1.453, -1.046, -1.721, 0.95, 0.982, -1.914, 0.435, 0.742, -0.835, -1.516, \\ &1.666, -1.709, 1.482, 1.511, -1.133, -1.251, -0.76, -1.518, 1.497, -0.596, 1.915, -0.311, \\ &0.274, -1.436, -0.606, -0.688, -0.251, -0.941, 1.589, 0.816, -1.299, -1.368, -1.27, -1.292, \\ &-1.486, 1.051, 1.474, 1.063, -1.095, 1.22, -1.33, -1.562, -1.136, 1.88, -1.467, 0.703, \\ &-1.332, 0.928, -1.534, -0.733, -1.26, 1.616, 1.272, -1.489, -1.158, -1.494, -1.384, -0.874, \\ &-1.494, -0.341, -1.565, -1.229, -1.343, 0.574, 0.82, 0.924, -1.187, -1.351, -0.715, 0.018, \\ &-0.447, 1.918, -0.997, -1.204, -1.532, -1.09, 1.44, -1.604, -1.515, -1.589, -1.32, -1.341) \end{aligned}$$

9 Simple Simulation Study

The data for this simple simulation study are generated according to the description in Section 4.1 of the manuscript; however, the data generating mechanism here is simpler.

First, the state-space for this example contains only *three* states and the state sequences are assumed to follow a Markov process with transition probability matrix defined as

$$\mathbf{P} = \begin{pmatrix} 0.8808 & 0.1192 & 0 \\ 0 & 0.8808 & 0.1192 \\ 0.1543 & 0.1543 & 0.6914 \end{pmatrix}.$$

Mimicking Section 4.1 of the manuscript, for each subject $i \in \{1, 2, \ldots, 500\}$, we generate a state sequence of length $m_i + n_i$ (namely, $\boldsymbol{b}^{(i)}_{long}$) that always starts in state 1, but only the last n_i time points are used as the true state sequence (namely, $\boldsymbol{b}^{(i)}$). Then, for every i, let $m_i \sim \text{uniform}\{0, 1, \ldots, 50\}$ and $n_i \sim \text{Poisson}(\lambda = 100)$.

Next, we simplify the true data generating conditional response model from the simulation in Section 4 of the manuscript. Notably, the simulation here no longer contains random effects nor covariates, and the mean process is no longer autoregressive. The only similarity to the simulation study in the manuscript is the state-dependence structure. Thus, the data generating mechanism for the conditional response is defined as:

$$\begin{aligned} \boldsymbol{y}^{(i)}_1 \mid \boldsymbol{b}^{(i)}_1 = s_{i,1}, \boldsymbol{\alpha}, \boldsymbol{R} \quad & \sim \text{N}_4\Big(g(\boldsymbol{\alpha}, \boldsymbol{b}^{(i)}_1),\ \boldsymbol{R}\Big) \\ \boldsymbol{y}^{(i)}_k \mid \{\boldsymbol{b}^{(i)}_j = s_{i,j}\}_{j=1}^k, \boldsymbol{\alpha}, \boldsymbol{R} \quad & \sim \text{N}_4\left(g(\boldsymbol{\alpha}, \boldsymbol{b}^{(i)}_1) + \Big[\sum_{j=2}^k \mathbf{1}\{\boldsymbol{b}^{(i)}_j = 2\}\Big]\boldsymbol{\alpha}_{\cdot,2} + \Big[\sum_{j=2}^k \mathbf{1}\{\boldsymbol{b}^{(i)}_j = 3\}\Big]\boldsymbol{\alpha}_{\cdot,3},\ \boldsymbol{R}\right), \end{aligned}$$

for $k \in \{2, 3, \ldots, n_i\}$ and $s_{i,j} \in \{1, 2, 3\}$ for all i and for all $j \in \{1, \ldots, n_i\}$. Define $g(\boldsymbol{\alpha}, \boldsymbol{b}^{(i)}_1) := \boldsymbol{\alpha}_{\cdot,1} + t^{(i)}_2 \boldsymbol{\alpha}_{\cdot,2} + t^{(i)}_3 \boldsymbol{\alpha}_{\cdot,3}$ and $t^{(i)}_l := \sum_{j=1}^{m_i+1} \mathbf{1}\{\boldsymbol{b}^{(i)}_{long,\, j} = l\}$ for $l \in \{2, 3\}$. The true parameter values are then given by

$$\boldsymbol{\alpha} = \begin{pmatrix} 50 & -5 & 5 \\ 100 & 10 & -10 \\ 100 & -10 & 10 \\ 50 & 5 & -5 \end{pmatrix}, \qquad \boldsymbol{R} = \text{diag}(4, 4, 4, 4).$$

The purpose of this simulation study is to illustrate the inferential consequences of *not* accounting for pre-ICU-admission physiological changes, and verifying that our approach (as described in Section 3.2.4 of the manuscript) works for handling these pre-ICU-admission physiological changes. We fit two models (*Model A* and *Model B*). For both, we assume the latent state process is defined by a Markov process, and we will learn the discrete, initial state probability distribution $\boldsymbol{\pi}$ and the 3×3 transition probability matrix

P. These two models vary are in their definitions of the conditional response:

Model A

$$\boldsymbol{y}_k^{(i)} \mid \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=1}^k, \boldsymbol{\alpha}, \boldsymbol{R} \sim \mathrm{N}_4\left(\boldsymbol{\alpha}_{\cdot,1} + \Big[\sum_{j=1}^k \mathbf{1}\{\boldsymbol{b}_j^{(i)} = 2\}\Big]\boldsymbol{\alpha}_{\cdot,2} + \Big[\sum_{j=1}^k \mathbf{1}\{\boldsymbol{b}_j^{(i)} = 3\}\Big]\boldsymbol{\alpha}_{\cdot,3},\ \boldsymbol{R}\right),$$

for $k \in \{1, 2, \dots, n_i\}$.

Model B

$$\boldsymbol{y}_1^{(i)} \mid \boldsymbol{\gamma}^{(i)}, \boldsymbol{R} \sim \mathrm{N}_4(\boldsymbol{\gamma}^{(i)},\ \boldsymbol{R})$$

$$\boldsymbol{y}_k^{(i)} \mid \{\boldsymbol{b}_j^{(i)} = s_{i,j}\}_{j=2}^k, \boldsymbol{\alpha}_*, \boldsymbol{\gamma}^{(i)}, \boldsymbol{R} \sim \mathrm{N}_4\left(\boldsymbol{\gamma}^{(i)} + \Big[\sum_{j=2}^k \mathbf{1}\{\boldsymbol{b}_j^{(i)} = 2\}\Big]\boldsymbol{\alpha}^*_{\cdot,1} + \Big[\sum_{j=2}^k \mathbf{1}\{\boldsymbol{b}_j^{(i)} = 3\}\Big]\boldsymbol{\alpha}^*_{\cdot,2},\ \boldsymbol{R}\right),$$

for $k \in \{2, \dots, n_i\}$ where $\boldsymbol{\gamma}^{(i)} \mid \boldsymbol{y}_1^{(i)}, \boldsymbol{G} \sim \mathrm{N}_4(\boldsymbol{y}_1^{(i)},\ \boldsymbol{G})$, $\boldsymbol{G}$ is a 4×4 covariance matrix, and $\boldsymbol{\alpha}^*$ is a 4×2 matrix of slope coefficients for the expected change in the four-dimensional response as a result of being in states 2 or 3, respectively for each column.

Model A serves as a naive strategy that does *not* account for pre-ICU-admission physiological changes because it assumes that if the initial state is not state 1, then the subject has *only* been in either state 2 or state 3 for one time instance. Model B, however, follows from the strategy discussed in Section 3.2.4 of the manuscript. Notice that from the data generating mechanism, we have no way of learning the true intercept term $\boldsymbol{\alpha}_{\cdot,1}$ without knowing $t_2^{(i)}$ and $t_3^{(i)}$, for all i. In lieu of this, Model B has no intercept term to learn; instead we learn $\boldsymbol{\alpha}_{\cdot,2}$ and $\boldsymbol{\alpha}_{\cdot,3}$ by estimating $\boldsymbol{\alpha}^*_{\cdot,1}$ and $\boldsymbol{\alpha}^*_{\cdot,2}$, respectively.

We fit Models A and B to 25 distinct datasets by running a Metropolis-within-Gibbs MCMC sampling routine for 10,000 steps, discarding the first 5,000 steps for burnin. Figure 1 presents box plots of the posterior means (from fitting Models A and B) for the slope coefficients corresponding to the effect of state 2 and state 3 on the mean of the response. We see from Figure 1 that our proposed approach (Model B) correctly centers the posterior means around the true parameter value, while Model A exhibits biased and more variable estimation of these slopes. We can attribute these estimation results from Model A to the fact that it does not properly account for the state changes that occur before the first observation.

	Q1	Median	Mean	Q3
Model A	0.6427	0.7092	0.7082	0.7753
Model B	1.0000	1.0000	0.9957	1.0000

Table 1: Summary of how Model A and Model B perform with respect to identifying the latent state sequences across all subjects, for one of the 25 simulated datasets.

Next, we can compare how Model A and Model B perform with respect to accurately identifying the true underlying state sequence for each subject ("accuracy" defined as the proportion of time instances where the posterior modal state correctly corresponds to the true state). Table 1 provides a summary of each models' performance. We again see that our proposed approach (Model B) outperforms Model A, this time with respect to learning the underlying latent state sequences for each subject in the data.

Trace plots and the remaining box plots for the other model parameters can be found by either emailing the author or visiting `https://ebkendall.github.io/research.html`.

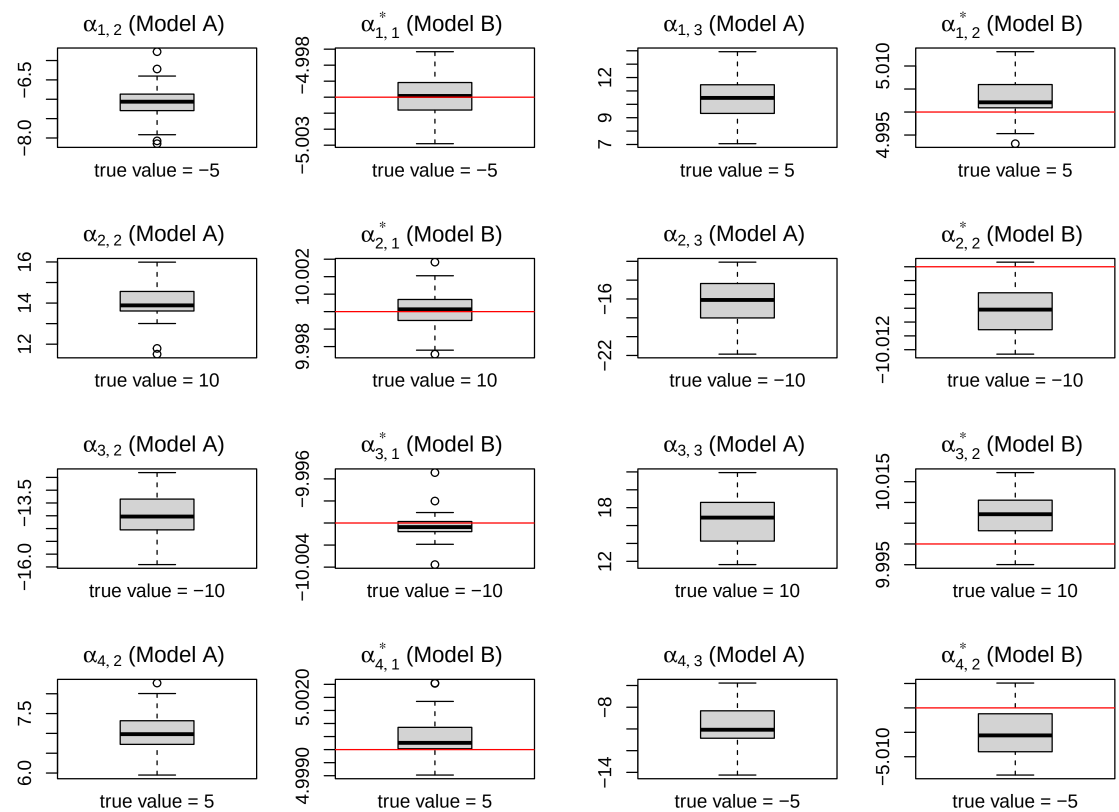

Figure 1: Box plots of the posterior means across 25 simulated datasets. These four columns specifically compare the estimated slope coefficients corresponding to state 2 (columns 1 and 2) and state 3 (columns 3 and 4) using Model A and Model B to fit the data. The *true* parameter value is provided on the x-axis, as well as marked by a red horizontal line, in each plot.

10 Estimated Medication Effects

Medication	Heart Rate Effect			MAP Effect		
	Exp.	Cont.	Disc.	Exp.	Cont.	Disc.
ANGIOTENSIN		–	–	↑	1.627*	–
EPHEDRINE	↑	–	1.912*	↑	–	0.801*
EPINEPHRINE	↑	0.517*	1.505*	↑	0.754*	1.92*
NOREPINEPHRINE	↑	0.971*	–	↑	1.234*	–
PHENYLEPHRINE	↓	−1.1*	−0.302*	↑	1.003*	0.891*
VASOPRESSIN		–	–	↑	1.968*	–
DOBUTAMINE	↑	1.691*	–	↓	−0.817*	–
DOPAMINE	↑	1.504*	–	↑	1.276*	–
MILRINONE	↑	1.182*	–	↓	−1.281*	–
ATROPINE	↑	–	0.879*	↑	–	1.436*
CALCIUM CHLORIDE		–	–	↑	0.914*	0.014
CALCIUM GLUCONATE		–	–	↑	–	0.56*
ISOPROTERENOL	↑	1.507*	–	↓	−1.496*	–
SODIUM BICARBONATE		–	–	↑	1.15*	–
ALBUMIN	↓	−1.612*	–	↑	1.522*	–
DILTIAZEM	↓	−1.464*	−0.69*	↓	−1.227*	−1.313*
ADENOSINE	↓	–	−0.946*		–	–
AMIODARONE	↓	−1.925*	−1.119*	↓	−1.403*	−1.461*
AMLODIPINE		–	–	↓	–	−0.821*
ATENOLOL	↓	–	−1.502*	↓	–	−1.503*
BISOPROLOL	↓	–	−1.186*	↓	–	−1.332*
CAPTOPRIL		–	–	↓	–	−1.509*
CARVEDILOL	↓	–	−0.752*	↓	–	−0.794*
CLEVIDIPINE		–	–	↓	−1.098*	–
CLONIDINE		–	–	↓	–	−1.06*
DIGOXIN	↓	–	−0.626*		–	–
ESMOLOL	↓	−1.722*	−1.268*	↓	−1.354*	−1.571*
ISOSORBIDE DINITRATE		–	–	↓	–	−1.621*
ISOSORBIDE MONONITRATE		–	–	↓	–	−1.493*
LABETALOL	↓	−1.531*	−0.78*	↓	−1.528*	−1.405*
LISINOPRIL		–	–	↓	–	−0.387*
LOSARTAN		–	–	↓	–	−1.401*
METOPROLOL	↓	–	−0.354*	↓	–	−0.455*
METOPROLOL SUCCINATE	↓	–	−1.489*	↓	–	−1.238*
METOPROLOL TARTRATE	↓	–	−1.102*	↓	–	−1.045*
NICARDIPINE		–	–	↓	−1.43*	–
NIFEDIPINE		–	–	↓	–	−1.18*
NIMODIPINE		–	–	↓	–	−1.497*
NITROGLYCERIN	↑	0.959*	1.502*	↓	−1.339*	−1.584*
NITROPRUSSIDE	↑	1.397*	–	↓	−1.564*	–
SILDENAFIL		–	–	↓	–	−1.319*
DEXMEDETOMIDINE	↓	−0.9*	–	↓	−1.328*	–
KETAMINE	↑	0.751*	0.321*		–	–
PROPOFOL		–	–	↓	−1.082*	−1.043*

Table 2: Posterior median medication effects from the real data analysis. The * indicates that the empirical 95% credible intervals exclude zero. The expected effect of the medication on heart rate and MAP is provided as a directional arrow where ↑ indicates an upward or positive effect, while ↓ indicates a downward or negative effect.

11 Trace Plots and Box Plots (Simulation)

For clarity and conciseness, `pdf` documents of the trace plots for the 100 simulations and the box plots of the 100 posterior medians can be found by either emailing the author or visiting `https://ebkendall.github.io/research.html`.

12 Trace Plots (Real Data Analysis)

For clarity and conciseness, `pdf` documents of the trace plots for the real data analysis can be found by either emailing the author or visiting `https://ebkendall.github.io/research.html`.

13 Chart Plots (Test Set)

13.1 Subject A

Figure 2 demonstrates the course for a 63-year-old male who underwent liver transplant for alcoholic cirrhosis. He had been admitted for more than 30 days prior to transplant. Intraoperatively, the case proceeded well. He required 8 units of cryoprecipitate, 16 units of fresh frozen plasma (FFP), 4 units of platelets, and 11 units of RBCs during the procedure. Postoperatively, he continued to require transfusion of blood products as well as vasopressor support with two agents. Due to this, he was taken back to the operating room within 12 hours where a large amount of clot was evacuated and areas of bleeding were controlled.

The model predicted a bleeding event as can be seen in Figure 2 which was confirmed in this patient and ultimately required return to the operating room for resolution.

13.2 Subject B

Figure 3 demonstrates the course for a medically complex 65-year-old woman transferred from her local hospital for management of an enterocutaneous fistula. During her stay at the local hospital, she underwent surgery to manage a small bowel obstruction which required resection of portions of her small bowel and colon and the creation of a colostomy. Unfortunately, her course there was complicated by sepsis and multiorgan failure including respiratory failure requiring mechanical ventilation, renal failure requiring continuous dialysis, liver failure, and delirium. During the ICU course outlined in Figure 3 following transfer to Mayo Clinic, she required vasopressor support for hypotension but did not develop a lactic acidosis. No bleeding was identified. She was transfused during the timeline

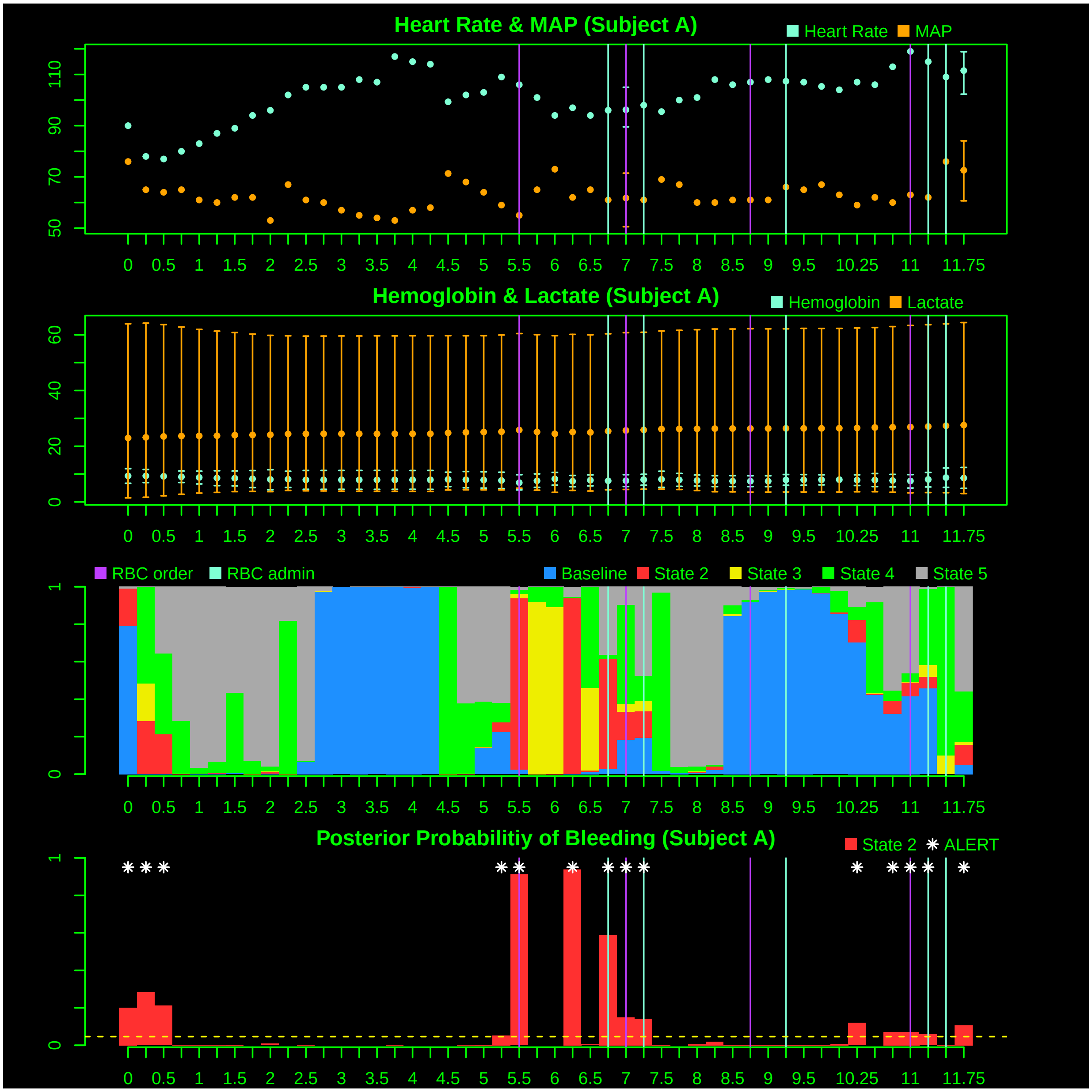

Figure 2: The top two panels correspond to the longitudinal vital measurements. The points with the error bars correspond to missing values; the error bars are empirical 95% credible intervals for the imputed response values. The third panel depicts the discrete posterior probability distributions of the latent states, at each time point. The bottom panel is the posterior probability of state 2 at each time point, and the yellow dashed line represents the threshold $\hat{c} = 0.0465$ determined from Section 4.2. The white stars indicate that the posterior probability of state 2 exceeds the threshold. The purple and turquoise vertical lines represent RBC transfusion order and administration times, respectively.

in Figure 3, but for a subacute anemia related to phlebotomy, critical illness, liver and renal failure rather than hemorrhage.

Although Figure 3 suggested a likely bleeding event occurring during the time depicted, the patient did not suffer a bleeding event. This may represent over-detection of hemorrhage by the model related to heart rate and MAP trends.

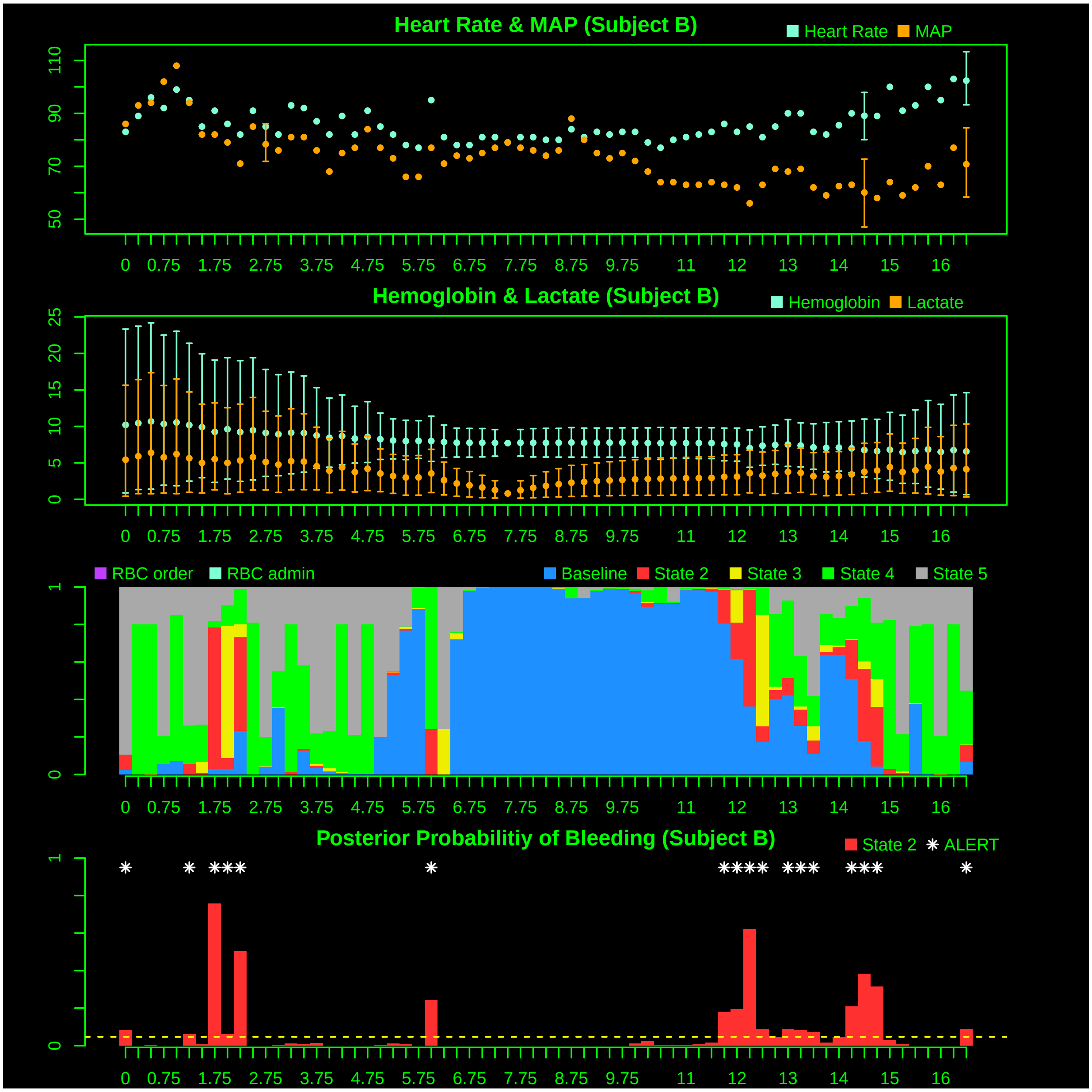

Figure 3: The top two panels correspond to the longitudinal vital measurements. The points with the error bars correspond to missing values; the error bars are empirical 95% credible intervals for the imputed response values. The third panel depicts the discrete posterior probability distributions of the latent states, at each time point. The bottom panel is the posterior probability of state 2 at each time point, and the yellow dashed line represents the threshold $\hat{c} = 0.0465$ determined from Section 4.2. The white stars indicate that the posterior probability of state 2 exceeds the threshold. The purple and turquoise vertical lines represent RBC transfusion order and administration times, respectively.

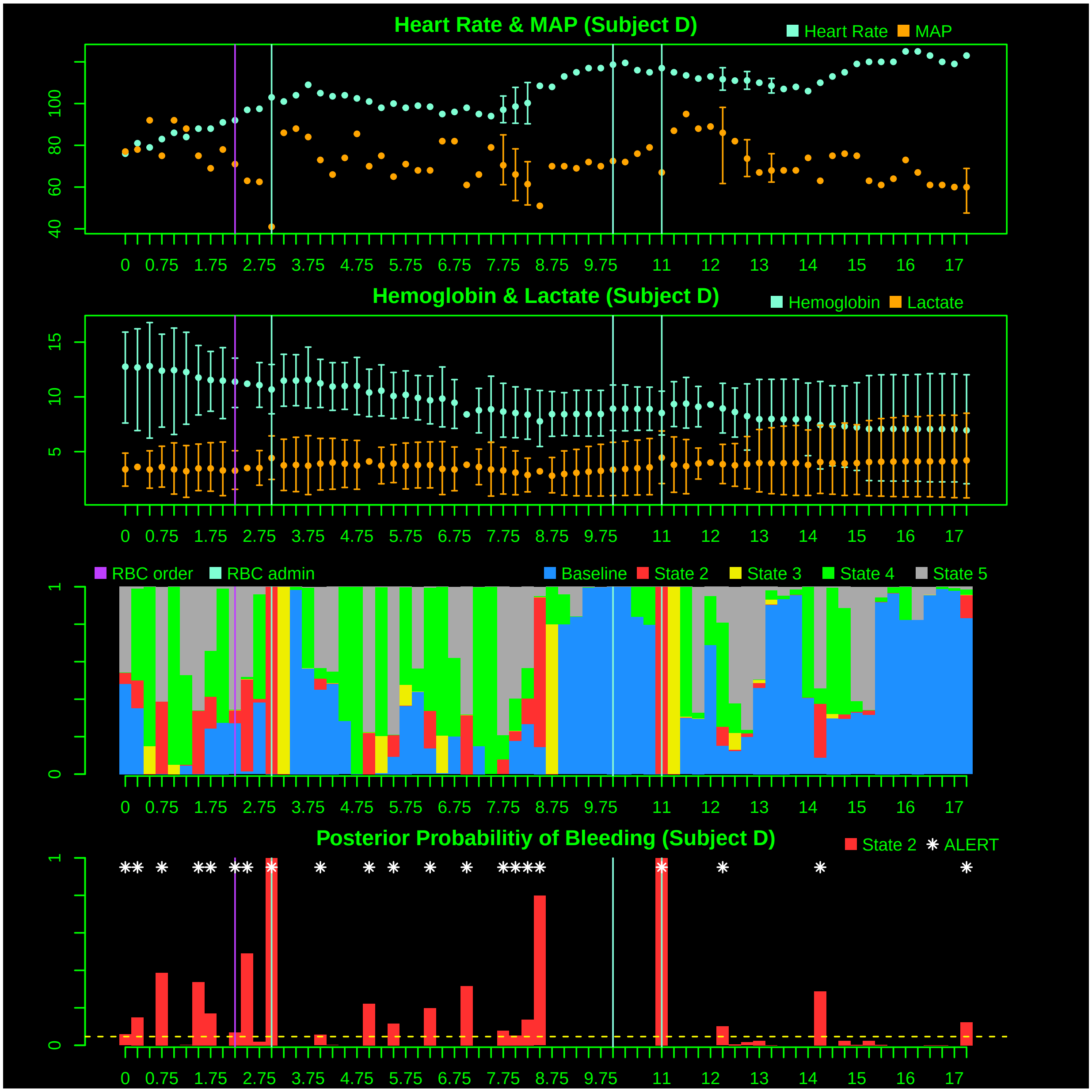

Figure 4: The top two panels correspond to the longitudinal vital measurements. The points with the error bars correspond to missing values; the error bars are empirical 95% credible intervals for the imputed response values. The third panel depicts the discrete posterior probability distributions of the latent states, at each time point. The bottom panel is the posterior probability of state 2 at each time point, and the yellow dashed line represents the threshold $\hat{c} = 0.0465$ determined from Section 4.2. The white stars indicate that the posterior probability of state 2 exceeds the threshold. The purple and turquoise vertical lines represent RBC transfusion order and administration times, respectively.

13.3 Subject D

Figure 4 demonstrates the course for a medically complex 76-year-old female who presented with right lower extremity non-healing wounds. She had previously undergone bilateral common femoral endarterectomies and retrograde iliac stents earlier in the year for critical limb ischemia. She developed thrombosis of the right common femoral artery and critical stenosis of the left common femoral artery. The wound on her right leg had enlarged in size and she was having severe rest pain. She went to surgery for revascularization via stenting or bypass. An endovascular approach was performed, however, before the intervention could be done, the patient became unstable from an intraoperative bleeding event. This was treated with a stent at the aortic bifurcation and the patient required a laparotomy due to elevated airway pressures, decreased urine output, and abdominal distention. The hematoma was evacuated and her abdomen was left open. During the case she received 14 units of packed red blood cells (PRBC), 3 units of platelets, 5 units of FFP, and 4 units of cryoprecipitate. On arrival to the ICU she was hemodynamically stable, but over the course of the day, she required additional blood products, had bright red drain output, required vasopressors, and had an increase in lactate. She returned to the operating room due to concern for limb ischemia and was found to have hemoperitoneum and ischemic bowel as well. She underwent bowel resection and thrombectomy. Bleeding sites were treated.

The model was able to detect that the patient was bleeding internally.

13.4 Subject E

Figure 5 relates to a 65-year-old man admitted with septic shock and found to have a small bowel perforation with fecal contamination of his peritoneum. He underwent an exploratory laparotomy, resection of the perforated bowel and temporary abdominal closure. He required significant vasopressor and inotropic support due to septic shock. Approximately one liter of blood loss was recorded for the operation. Following surgery, he was transferred to the ICU where he received two units of blood to correct anemia from blood loss during the operation. During his ICU stay, there was no concern for occult blood loss. He returned to the operating room multiple times during his hospital admission to re-evaluate the bowel and eventually underwent abdominal closure.

Figure 5 indicates several episodes of state 2 for which the model defines as a bleeding event. Although heart rate was rising, MAP was declining and hemoglobin was trending down, clinically, these time points were periods of post-operative resuscitation of a septic shock patient without occult or other hemorrhage.

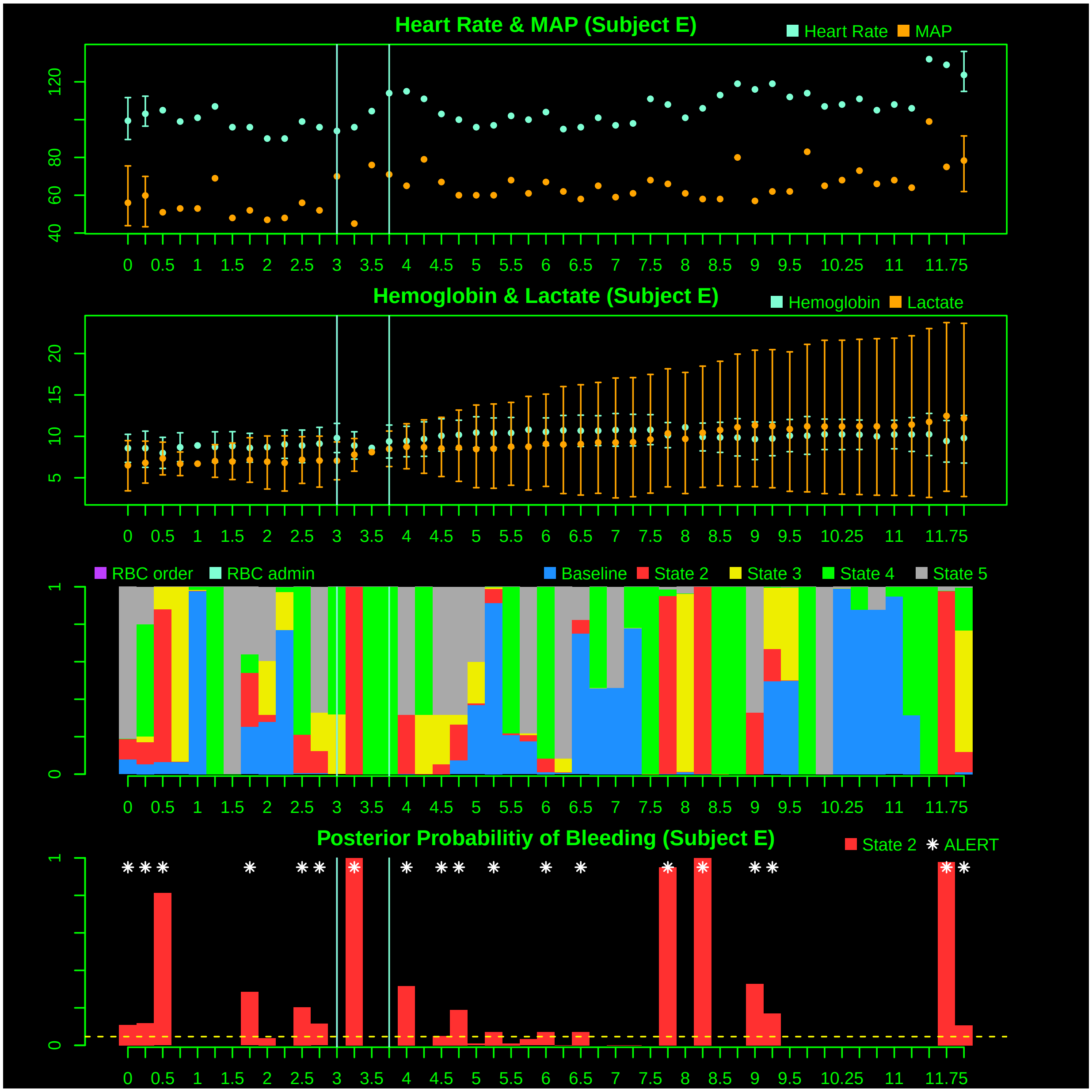

Figure 5: The top two panels correspond to the longitudinal vital measurements. The points with the error bars correspond to missing values; the error bars are empirical 95% credible intervals for the imputed response values. The third panel depicts the discrete posterior probability distributions of the latent states, at each time point. The bottom panel is the posterior probability of state 2 at each time point, and the yellow dashed line represents the threshold $\hat{c} = 0.0465$ determined from Section 4.2. The white stars indicate that the posterior probability of state 2 exceeds the threshold. The purple and turquoise vertical lines represent RBC transfusion order and administration times, respectively.