Supporting Information for Predicting Treatment Efficacy via Quantitative MRI: A Bayesian Joint Model

by

Jincao Wu[1]*, Timothy D. Johnson[1],
Craig J. Galbán[2], Thomas L. Chenevert[2], Charles R. Meyer[2],
Alnawaz Rehemtulla[2], Daniel A. Hamstra[3], Brain D. Ross[2]

[1]Department of Biostatistics, University of Michigan
[2]Department of Radiology, University of Michigan
[3]Department of Radiation Oncology, University of Michigan

*corresponding author, email: jincaowu@umich.edu; Tel: 734-272-9631;
address: 1415 Washington Heights, Ann Arbor, MI 48109–2029

Throughout this supplementary material, we will rely on the notation set out in the main manuscript.

1 Model and Algorithm Details

1.1 Joint Model:

Let $C_{i,j} = \{\{\boldsymbol{\mu}_{i'1,j}\}_{i' \in N_{i,j}}, \{\widetilde{\boldsymbol{\mu}}_{i'2,j}\}_{i' \in N_{i,j}}, \{\boldsymbol{\nu}_{i'2,j}\}_{i' \in N^e_{i,j} \cap \mathcal{S}_j}\}$. We use notation $\mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j})$ to denote the jth covariate vector which emphasizes its dependence on the parameters $\boldsymbol{\mu}_{2,j}$ and $\widetilde{\boldsymbol{\mu}}_{2,j}$. The joint distribution of all the data and parameters is

$$
\begin{aligned}
&\prod_{j=1}^{M}\prod_{i=1}^{n_j} \pi(\mathbf{Y}_{i,j} \mid \boldsymbol{\mu}_{i,j}, \Sigma_j)\pi(\Sigma_j)\pi(\boldsymbol{\mu}_{i,j} \mid \{\boldsymbol{\mu}_{i'j}\}_{i' \in N_{i,j}}, \Psi_j)\pi(\Psi_j) \qquad (1)\\
&\times \prod_{j=1}^{M}\prod_{i=1}^{n^e_j} \pi(\mathbf{W}_{i,j} \mid \boldsymbol{\nu}_{i,j}, \Delta_j)\pi(\Delta_j)\pi(\boldsymbol{\nu}_{i,j} \mid \{\boldsymbol{\nu}_{i,j}\}_{i' \in N^e_{i,j}}, \Omega_j)\pi(\Omega_j)\\
&\times \prod_{j=1}^{M}\prod_{i=1}^{n_j} \pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j)\\
&\times \prod_{j=1}^{M} \pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K})\pi(\boldsymbol{\Omega}_{2K}),
\end{aligned}
$$

where

$$
\begin{aligned}
\left[\mathbf{Y}_{i,j} \mid \boldsymbol{\mu}_{i,j}, \Sigma_j\right] &\sim \mathrm{N}(\boldsymbol{\mu}_{i,j}, \Sigma_j), \qquad [\mathbf{W}_{i,j} \mid \boldsymbol{\nu}_{i,j}, \Delta_j] \sim \mathrm{N}(\boldsymbol{\nu}_{i,j}, \Delta_j), \qquad (2)\\
[\Sigma_j] &\sim \mathrm{W}^{-1}(\mathrm{I}_4, 5), \qquad [\Delta_j] \sim \mathrm{W}^{-1}(\mathrm{I}_4, 5), \qquad (3)\\
\left[\boldsymbol{\mu}_{i,j} \mid \{\boldsymbol{\mu}_{i'j}\}_{i' \in N_{i,j}}, \Psi_j\right] &\sim \mathrm{N}\left[\boldsymbol{\mu}^*_{i,j}, |N_{i,j}|^{-1}\Psi_j\right], \qquad \boldsymbol{\mu}^*_{i,j} = |N_{i,j}|^{-1}\sum_{i' \in N_{i,j}} \boldsymbol{\mu}_{i',j}, \qquad (4)\\
\left[\boldsymbol{\nu}_{i,j} \mid \{\boldsymbol{\nu}_{i,j}\}_{i' \in N^e_{i,j}}, \Omega_j\right] &\sim \mathrm{N}\left(\boldsymbol{\nu}^*_{i,j}, |N^e_{i,j}|^{-1}\Omega_j\right], \qquad \boldsymbol{\nu}^*_{i,j} = |N^e_{i,j}|^{-1}\sum_{i' \in N^e_{i,j}} \boldsymbol{\nu}_{i',j}, \qquad (5)\\
[\Psi_j] &\sim \mathrm{W}^{-1}(\mathrm{I}_4, 5), \qquad [\Omega_j] \sim \mathrm{W}^{-1}(\mathrm{I}_4, 5), \qquad (6)
\end{aligned}
$$

and

$$\left[\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j\right] \tag{7}$$
$$\sim \mathrm{N}\left(\widetilde{\boldsymbol{\mu}}_{i2,j} + \Delta_{21,j}\left(\Delta_{11,j}\right)^{-1}\left(\mathbf{Y}_{i1,j} - \boldsymbol{\mu}_{i1,j}\right),\ \Delta_{22,j} - \Delta_{21,j}(\Delta_{11,j})^{-1}\Delta_{12,j}\right),$$

$$\left[\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j\right] \tag{8}$$
$$\sim \mathrm{N}\left(\widetilde{\boldsymbol{\mu}}^*_{i2,j} + \Omega_{21,j}(\Omega_{11,j})^{-1}(\boldsymbol{\mu}_{i1,j} - \boldsymbol{\mu}^*_{i1,j}), |N^e_{i,j}|^{-1}(\Omega_{22,j} - \Omega_{21,j}(\Omega_{11,j})^{-1}\Omega_{12,j})\right),$$

$$\left[d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K}\right] \sim \mathrm{N}(\eta_{jK}, 1), \tag{9}$$

where

$$\begin{pmatrix} \boldsymbol{\mu}^*_{i1,j} \\ \widetilde{\boldsymbol{\mu}}^*_{i2,j} \end{pmatrix} = |N^e_{i,j}|^{-1}\left[\sum_{i' \in N^e_{i,j} \cap \mathcal{S}_j} \begin{pmatrix} \boldsymbol{\nu}_{i'1,j} \\ \boldsymbol{\nu}_{i'2,j} \end{pmatrix} + \sum_{i' \in N^e_{i,j} \cap \mathcal{H}_j} \begin{pmatrix} \boldsymbol{\mu}_{i'1,j} \\ \widetilde{\boldsymbol{\mu}}_{i'2,j} \end{pmatrix}\right] \tag{10}$$

and η_{jK} is defined in (41).

1.2 Stage I:

For stage I parameters, some of the full conditional distributions have nice distributional forms from which we can directly sample, others require a Metropolis-Hastings update. The algorithm we use to draw stage I parameters is thus an hybrid Metropolis-within-Gibbs algorithm.

1.2.1 Updating Σ_j

By conjugacy, it is straightforward to derive the full conditional for the covariance matrix Σ_j for subject j. A priori, $\Sigma_j \sim \mathrm{W}^{-1}(\mathbf{I}_4, 5)$. Combining this prior with the data distribution

(first distribution in (2)), we obtain

$$\begin{aligned}
&\pi\left(\Sigma_j \mid \left\{\boldsymbol{\mu}_{i,j}\right\}_{i=1}^{n_j}, \left\{\mathbf{Y}_{i,j}\right\}_{i=1}^{n_j}\right) \\
&\quad\propto \prod_{i=1}^{n_j} \pi(\mathbf{Y}_{i,j} \mid \boldsymbol{\mu}_{i,j}, \Sigma_j)\pi(\Sigma_j) \\
&\quad\propto |\Sigma_j|^{-n_j/2} \exp\left\{-0.5 \sum_{i=1}^{n_j} (\mathbf{Y}_{i,j} - \boldsymbol{\mu}_{i,j})^{\mathrm{T}} \Sigma_j^{-1} (\mathbf{Y}_{i,j} - \boldsymbol{\mu}_{i,j})\right\} \times \\
&\qquad |\Sigma_j|^{-\frac{5+4+1}{2}} \exp\left\{-0.5 \operatorname{tr}(\Sigma_j^{-1})\right\} \\
&\quad= |\Sigma_j|^{-(n_j+10)/2} \exp\left\{-0.5 \operatorname{tr}\left(\Sigma_j{}^{-1} \left[\sum_{i=1}^{n_j} (\mathbf{Y}_{i,j} - \boldsymbol{\mu}_{i,j})(\mathbf{Y}_{i,j} - \boldsymbol{\mu}_{i,j})^{\mathrm{T}} + \mathrm{I}_4\right]\right)\right\},
\end{aligned}$$

which is the kernel of an inverse Wishart distribution. Let

$$\mathrm{S}_{1,j} = \sum_{i=1}^{n_j} \left(\mathbf{Y}_{i,j} - \boldsymbol{\mu}_{i,j}\right) \left(\mathbf{Y}_{i,j} - \boldsymbol{\mu}_{i,j}\right)^{\mathrm{T}}.$$

Then,

$$\left[\Sigma_j \mid \left\{\boldsymbol{\mu}_{i,j}\right\}_{i=1}^{n_j}, \left\{\mathbf{Y}_{i,j}\right\}_{i=1}^{n_j}\right] \sim \mathrm{W}^{-1}\left(\mathrm{S}_{1,j} + \mathrm{I}_4, n_j + 5\right). \tag{11}$$

1.2.2 Updating Ψ_j

Given the prior $\Psi_j \sim \mathrm{W}^{-1}(\mathrm{I}_4, 5)$ we can easily derive its full conditional:

$$\begin{aligned}
&\pi\left(\Psi_j \mid \left\{\boldsymbol{\mu}_{i,j}\right\}_{i=1}^{n_j}\right) \\
&\quad\propto \prod_{i=1}^{n_j} \pi(\boldsymbol{\mu}_{i,j} \mid \Psi_j, \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}})\pi(\Psi_j) = \pi\left(\left\{\boldsymbol{\mu}_{i,j}\right\}_{i=1}^{n_j} \mid \Psi_j\right)\pi(\Psi_j) \\
&\quad\propto |\Psi_j|^{-n_j/2} \exp\left\{-0.25 \sum_{i \sim i'} (\boldsymbol{\mu}_{i,j} - \boldsymbol{\mu}_{i',j})^{\mathrm{T}} \Psi_j{}^{-1} (\boldsymbol{\mu}_{i,j} - \boldsymbol{\mu}_{i',j})\right\} \times \\
&\qquad |\Psi_j|^{-10/2} \exp\left\{-0.5 \operatorname{tr}(\Psi_j^{-1})\right\} \\
&\quad= |\Psi_j|^{-(n_j+10)/2} \exp\left\{-0.5 \operatorname{tr}\left(\Psi_j^{-1} \left[0.5 \sum_{i \sim i'} (\boldsymbol{\mu}_{i,j} - \boldsymbol{\mu}_{i',j})(\boldsymbol{\mu}_{i,j} - \boldsymbol{\mu}_{i',j})^{\mathrm{T}} + \mathrm{I}_4\right]\right)\right\},
\end{aligned}$$

where we have generalized results from Higdon et al. (1997) to obtain

$$\pi\left(\left\{\boldsymbol{\mu}_{i,j}\right\}_{i=1}^{n_j} \mid \Psi_j\right) \propto |\Psi_j|^{-n_j/2} \exp\left\{-0.25 \sum_{i \sim i'} (\boldsymbol{\mu}_{i,j} - \boldsymbol{\mu}_{i',j})^{\mathrm{T}} \Psi_j{}^{-1} (\boldsymbol{\mu}_{i,j} - \boldsymbol{\mu}_{i',j})\right\}.$$

Let $\mathrm{S}_{2,j} = 0.5 \sum_{i \sim i'} \left(\boldsymbol{\mu}_{i,j} - \boldsymbol{\mu}_{i',j}\right) \left(\boldsymbol{\mu}_{i,j} - \boldsymbol{\mu}_{i',j}\right)^{\mathrm{T}}$. Then,

$$\left[\Psi_j \mid \left\{\boldsymbol{\mu}_{i,j}\right\}_{i=1}^{n_j}\right] \sim \mathrm{W}^{-1}\left(\mathrm{S}_{2,j} + \mathrm{I}_4, n_j + 5\right). \tag{12}$$

1.2.3 Updating $\boldsymbol{\mu}_{i,j}$

Recall that the summary statistic vector, $\mathbf{X}_j$, depends on $\boldsymbol{\Omega}_{1j}$ only through $\{\boldsymbol{\mu}_{i2,j}\}_{i=1}^{n_j}$ and $\{\widetilde{\boldsymbol{\mu}}_{i2,j}\}_{i=1}^{n_j}$. The full conditional of $\boldsymbol{\mu}_{i,j}$ is

$$\begin{aligned}
&\pi(\boldsymbol{\mu}_{i,j} \mid \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}}, \mathbf{Y}_{i,j}, \Sigma_j, \Psi_j, d_j, \widetilde{\boldsymbol{\mu}}_{i2,j}, \boldsymbol{\Omega}_{2K}) \qquad (13)\\
&\quad \propto \pi(\mathbf{Y}_{i,j} \mid \boldsymbol{\mu}_{i,j}, \Sigma_j)\pi(\boldsymbol{\mu}_{i,j} \mid \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}}, \Psi_j)\pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K})\\
&\qquad \times \pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)\pi(\left\{\widetilde{\boldsymbol{\mu}}_{i2,j}\right\}_{i=1}^{n_j} \mid \left\{\boldsymbol{\mu}_{i1,j}\right\}_{i=1}^{n_j}, \{\boldsymbol{\nu}_{i2,j}\}_{i \in \mathcal{S}_j}, \Omega_j)\\
&\quad \propto \pi\left(\boldsymbol{\mu}_{i,j} \mid \Sigma_j, \Psi_j, \mathbf{Y}_{i,j}, \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}}\right) \pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K}) \qquad (14)\\
&\qquad \times \pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)\pi(\left\{\widetilde{\boldsymbol{\mu}}_{i2,j}\right\}_{i=1}^{n_j} \mid \left\{\boldsymbol{\mu}_{i1,j}\right\}_{i=1}^{n_j}, \{\boldsymbol{\nu}_{i2,j}\}_{i \in \mathcal{S}_j}, \Omega_j),
\end{aligned}$$

which does not have a nice distributional form from which we can easily sample and so we resort to a Metropolis-Hastings update. Note that the first term in (14) does have a nice distribution form:

$$\begin{aligned}
\pi\left(\boldsymbol{\mu}_{i,j} \mid \Sigma_j, \Psi_j, \mathbf{Y}_{i,j}, \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}}\right) &\propto \pi(\mathbf{Y}_{i,j} \mid \boldsymbol{\mu}_{i,j}, \Sigma_j)\pi(\boldsymbol{\mu}_{i,j} \mid \Psi_j, \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}})\\
&\propto \exp\left\{-0.5(\mathbf{Y}_{i,j} - \boldsymbol{\mu}_{i,j})^{\mathrm{T}} {\Sigma_j}^{-1} (\mathbf{Y}_{i,j} - \boldsymbol{\mu}_{i,j})\right\} \times\\
&\quad \exp\left\{-0.5|N_{i,j}|(\boldsymbol{\mu}_{i,j} - \boldsymbol{\mu}^*_{i,j})^{\mathrm{T}} \Psi_j^{-1} (\boldsymbol{\mu}_{i,j} - \boldsymbol{\mu}^*_{i,j})\right\}\\
&= \exp\left\{-0.5\left[\boldsymbol{\mu}_{i,j}^{\mathrm{T}}(\Sigma_j^{-1} + |N_{i,j}|\Psi_j^{-1})\boldsymbol{\mu}_{i,j}\right.\right.\\
&\qquad \left.\left. -2\boldsymbol{\mu}_{i,j}^{\mathrm{T}}\left(\Sigma_j^{-1}\mathbf{Y}_{i,j} + |N_{i,j}|\Psi_j^{-1}\boldsymbol{\mu}^*_{i,j}\right)\right]\right\},
\end{aligned}$$

which is the kernel of a normal distribution. Thus

$$\begin{aligned}
&\left[\boldsymbol{\mu}_{i,j} \mid \Sigma_j, \Psi_j, \mathbf{Y}_{i,j}, \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}}\right] \sim\\
&\quad \mathrm{N}\left[\left(|N_{i,j}|{\Psi_j}^{-1} + {\Sigma_j}^{-1}\right)^{-1}\left(|N_{i,j}|{\Psi_j}^{-1}\boldsymbol{\mu}^*_{i,j} + {\Sigma_j}^{-1}\mathbf{Y}_{i,j}\right), \left(|N_{i,j}|{\Psi_j}^{-1} + {\Sigma_j}^{-1}\right)^{-1}\right]. \qquad (15)
\end{aligned}$$

Therefore, we propose a new value of $\boldsymbol{\mu}_{i,j}^{prop}$ from (15), which simplifies the acceptance probability and results in a high acceptance rate. The acceptance probability is given by

$$\alpha_{\mu} = \min\left\{1, \mathcal{R}\right\}, \tag{16}$$

where the superscript *prop* represents a new proposed sample and *current* represents the current sample and

$$\begin{aligned}\mathcal{R} \;=\; & \frac{\pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{\{-i\}2,j}, \boldsymbol{\mu}_{i2,j}^{prop}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K})\pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}^{prop}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)}{\pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{\{-i\}2,j}, \boldsymbol{\mu}_{i2,j}^{current}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K})\pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}^{current}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)} \\ & \times \frac{\pi\left(\left\{\widetilde{\boldsymbol{\mu}}_{i2,j}\right\}_{i=1}^{n_j} \mid \boldsymbol{\mu}_{i1,j}^{prop}, \left\{\boldsymbol{\mu}_{k1,j}\right\}_{k=1}^{n_j}, \left\{\boldsymbol{\nu}_{i2,j}\right\}_{i\in\mathcal{S}_j}\Omega_j\right)}{\pi\left(\left\{\widetilde{\boldsymbol{\mu}}_{i2,j}\right\}_{i=1}^{n_j} \mid \boldsymbol{\mu}_{i1,j}^{current}, \left\{\boldsymbol{\mu}_{k1,j}\right\}_{k=1}^{n_j}, \left\{\boldsymbol{\nu}_{i2,j}\right\}_{i\in\mathcal{S}_j}\Omega_j\right)}.\end{aligned}$$

1.2.4 Updating $\boldsymbol{\nu}_{i,j}$

For healthy tissue voxels, analogous to (15), the full conditional of $\boldsymbol{\nu}_{i,j}$ for voxel $i \in N_{i',j}^e \cap \mathcal{H}_j$, is:

$$\begin{aligned}&\left[\boldsymbol{\nu}_{i,j} \mid \Omega_j, \Delta_j, \mathbf{W}_{i,j}, \left\{\boldsymbol{\nu}_{i',j}\right\}_{i'\in N_{i,j}^e}\right] \sim \\ &\qquad \mathrm{N}\left[\left(|N_{i,j}^e|\Omega_j^{-1} + \Delta_j{}^{-1}\right)^{-1}\left(|N_{i,j}^e|\Omega_j^{-1}\boldsymbol{\nu}_{i,j}^{*} + \Delta_j^{-1}\mathbf{W}_{i,j}\right), \left(|N_{i,j}^e|\Omega_j^{-1} + \Delta_j^{-1}\right)^{-1}\right]. \end{aligned} \tag{17}$$

While for voxel $i \in N_{i',j}^e \cap \mathcal{S}_j$, we have that the full conditional of $\boldsymbol{\nu}_{i,j}$ is

$$\pi(\boldsymbol{\nu}_{i,j} \mid \left\{\boldsymbol{\nu}_{i',j}\right\}_{i'\in N_{i,j}^e}, \mathbf{W}_{i,j}, \Delta_j, \Omega_j, \left\{\widetilde{\boldsymbol{\mu}}_{i2,j}\right\}_{i'\in N_i}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \left\{\boldsymbol{\mu}_{i'1,j}\right\}_{i'\in N_i}, \boldsymbol{\mu}_{i1,j}) \tag{18}$$

$$\begin{aligned}&\propto \pi(\mathbf{W}_{i,j} \mid \boldsymbol{\nu}_{i,j}, \Delta_j)\pi(\boldsymbol{\nu}_{i,j} \mid \left\{\boldsymbol{\nu}_{i',j}\right\}_{i'\in N_{i,j}^e}, \Omega_j)\pi\left(\left\{\widetilde{\boldsymbol{\mu}}_{i2,j}\right\}_{i=1}^{n_j} \mid \left\{\boldsymbol{\mu}_{i1,j}\right\}_{i=1}^{n_j}, \left\{\boldsymbol{\nu}_{i2,j}\right\}_{i\in\mathcal{S}_j}, \Omega_j\right) \\ &\propto \pi\left(\boldsymbol{\nu}_{i,j} \mid \Delta_j, \Omega_j, \mathbf{W}_{i,j}, \left\{\boldsymbol{\nu}_{i',j}\right\}_{i'\in N_{i,j}^e}\right)\pi\left(\left\{\widetilde{\boldsymbol{\mu}}_{i2,j}\right\}_{i=1}^{n_j} \mid \left\{\boldsymbol{\mu}_{i1,j}\right\}_{i=1}^{n_j}, \left\{\boldsymbol{\nu}_{i2,j}\right\}_{i\in\mathcal{S}_j}, \Omega_j\right), \end{aligned} \tag{19}$$

which does not have a nice distributional form. However, the first term in (19) has distribution (17). Thus, we propose a new value $\boldsymbol{\nu}_{i2,j}^{prop}, i \in N_{i',j}^e \cap \mathcal{S}_j$ from (17) and accept it with probability:

$$\alpha_{\nu} = \min\left\{1, \frac{\pi\left(\left\{\widetilde{\boldsymbol{\mu}}_{i2,j}\right\}_{i=1}^{n_j} \mid \left\{\boldsymbol{\mu}_{i1,j}\right\}_{i=1}^{n_j}, \boldsymbol{\nu}_{i2,j}^{prop}, \left\{\boldsymbol{\nu}_{k2,j}\right\}_{k\in\mathcal{S}_j}, \Omega_j\right)}{\pi\left(\left\{\widetilde{\boldsymbol{\mu}}_{i2,j}\right\}_{i=1}^{n_j} \mid \left\{\boldsymbol{\mu}_{i1,j}\right\}_{i=1}^{n_j}, \boldsymbol{\nu}_{i2,j}^{current}, \left\{\boldsymbol{\nu}_{k2,j}\right\}_{k\in\mathcal{S}_j}, \Omega_j\right)}\right\}. \tag{20}$$

1.2.5 Updating Δ_j

The full conditional distribution of the covariance matrix Δ_j for subject j is

$$\begin{aligned}
&\pi(\Delta_j \mid \{\mathbf{W}_{i,j}\}_{i=1}^{n_j^e}, \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n_j^e}, \{\widetilde{\mathbf{Y}}_{i2,j}\}_{i=1}^{n_j}, \{\mathbf{Y}_{i1,j}\}_{i=1}^{n_j}, \{\boldsymbol{\mu}_{i1,j}\}_{i=1}^{n_j}, \{\widetilde{\boldsymbol{\mu}}_{i2,j}\}_{i=1}^{n_j}) \qquad (21)\\
&\quad \propto \prod_{i=1}^{n_j^e} \pi(\mathbf{W}_{i,j} \mid \boldsymbol{\nu}_{i,j}, \Delta_j)\pi(\Delta_j) \prod_{i=1}^{n_j} \pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)\\
&\quad \propto \pi(\Delta_j \mid \{\mathbf{W}_{i,j}\}_{i=1}^{n_j^e}, \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n_j^e}) \prod_{i=1}^{n_j} \pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j), \qquad (22)
\end{aligned}$$

which does not have a nice distributional form from which to draw. However, the first term in (22) has an inverse Wishart distribution:

$$\left[\Delta_j \mid \{\mathbf{W}_{i,j}\}_{i=1}^{n_j^e}, \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n_j^e}\right] \;\sim\; \mathrm{W}^{-1}\left(\mathrm{S}_{3,j} + \mathrm{I}_4, n_j^e + 5\right), \qquad (23)$$

where $S_{3,j} = \sum_{i=1}^{n_j^e} (\mathbf{W}_{i,j} - \boldsymbol{\nu}_{i,j}) (\mathbf{W}_{i,j} - \boldsymbol{\nu}_{i,j})^{\mathrm{T}}$ for all $i \in \mathcal{H} \cup \mathcal{S}$. Therefore, we propose a new value of Δ_j^{prop} from (23) and accept this value with probability

$$\alpha_\Delta = \min\left\{1, \prod_{i=1}^{n_j} \frac{\pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j^{prop})}{\pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j^{current})}\right\}. \qquad (24)$$

1.2.6 Updating Ω_j

From (1), we have

$$\begin{aligned}
&\pi(\Omega_j \mid \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n_j^e}, \{\boldsymbol{\mu}_{i1,j}\}_{i=1}^{n_j}, \{\widetilde{\boldsymbol{\mu}}_{i2,j}\}_{i=1}^{n_j}, \{\boldsymbol{\nu}_{i'2}\}_{i' \in N_i^e \cap \mathcal{S}_j}) \qquad (25)\\
&\quad \propto \prod_{i=1}^{n_j^e} \pi(\boldsymbol{\nu}_{i,j} \mid \{\boldsymbol{\nu}_{i,j}\}_{i' \in N_{i,j}^e}, \Omega_j)\pi(\Omega_j)\pi\left(\left\{\widetilde{\boldsymbol{\mu}}_{i2,j}\right\}_{i=1}^{n_j} \mid \left\{\boldsymbol{\mu}_{i1,j}\right\}_{i=1}^{n_j}, \{\boldsymbol{\nu}_{i2,j}\}_{i \in \mathcal{S}_j}\, \Omega_j\right)\\
&\quad \propto \pi(\Omega_j \mid \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n_j^e})\pi\left(\left\{\widetilde{\boldsymbol{\mu}}_{i2,j}\right\}_{i=1}^{n_j} \mid \left\{\boldsymbol{\mu}_{i1,j}\right\}_{i=1}^{n_j}, \{\boldsymbol{\nu}_{i2,j}\}_{i \in \mathcal{S}_j}\, \Omega_j\right). \qquad (26)
\end{aligned}$$

The first term in (26) has an inverse Wishart distribution:

$$\left[\Omega_j \mid \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n^e}\right] \sim \mathrm{W}^{-1}\left(\mathrm{S}_{4,j} + \mathrm{I}_4, n_j^e + 5\right), \qquad (27)$$

Thus we propose a new value Ω_j^{prop} from (27) and accept this value as a draw from the full conditional with probability

$$\alpha_\Omega = \min\left\{1, \frac{\pi\left(\left\{\widetilde{\boldsymbol{\mu}}_{i2,j}\right\}_{i=1}^{n_j} \mid \left\{\boldsymbol{\mu}_{i1,j}\right\}_{i=1}^{n_j}, \left\{\boldsymbol{\nu}_{i2,j}\right\}_{i\in\mathcal{S}_j}, \Omega_j^{prop}\right)}{\pi\left(\left\{\widetilde{\boldsymbol{\mu}}_{i2,j}\right\}_{i=1}^{n_j} \mid \left\{\boldsymbol{\mu}_{i1,j}\right\}_{i=1}^{n_j}, \left\{\boldsymbol{\nu}_{i2,j}\right\}_{i\in\mathcal{S}_j}, \Omega_j^{current}\right)}\right\} \tag{28}$$

1.2.7 Updating $\widetilde{\mathbf{Y}}_{i2,j}$ and $\widetilde{\boldsymbol{\mu}}_{i2,j}$

Now we derive the conditional predictive distribution of $\widetilde{\mathbf{Y}}_{i2,j}$ and the posterior distribution of $\widetilde{\boldsymbol{\mu}}_{i2,j}$. Since under the null, we define the joint distribution of $\mathbf{Y}_{i1,j}$ and $\widetilde{\mathbf{Y}}_{i2,j}$ by

$$\left[\begin{pmatrix}\mathbf{Y}_{i1,j}\\ \widetilde{\mathbf{Y}}_{i2,j}\end{pmatrix} \mid \begin{pmatrix}\boldsymbol{\mu}_{i1,j}\\ \widetilde{\boldsymbol{\mu}}_{i2,j}\end{pmatrix}, \begin{pmatrix}\Delta_{11,j} & \Delta_{12,j}\\ \Delta_{21,j} & \Delta_{22,j}\end{pmatrix}\right] \sim \mathrm{N}\left[\begin{pmatrix}\boldsymbol{\mu}_{i1,j}\\ \widetilde{\boldsymbol{\mu}}_{i2,j}\end{pmatrix}, \begin{pmatrix}\Delta_{11,j} & \Delta_{12,j}\\ \Delta_{21,j} & \Delta_{22,j}\end{pmatrix}\right], \tag{29}$$

where $\mathbf{Y}_{i1,j} = (Y_{i11,j}, Y_{i12,j})^{\mathrm{T}}$, representing the baseline diffusion and perfusion intensities at voxel i, while $\widetilde{\mathbf{Y}}_{i2,j} = \left(\widetilde{Y}_{i21,j}, \widetilde{Y}_{i22,j}\right)^{\mathrm{T}}$ is the predicted null response at time point 2. Let

$$\begin{pmatrix}\boldsymbol{\mu}_{i1,j}^*\\ \widetilde{\boldsymbol{\mu}}_{i2,j}^*\end{pmatrix} = |N_{i,j}^e|^{-1}\left[\sum_{i'\in N_{i,j}^e\cap\mathcal{S}_j}\begin{pmatrix}\boldsymbol{\nu}_{i'1,j}\\ \boldsymbol{\nu}_{i'2,j}\end{pmatrix} + \sum_{i'\in N_{i,j}^e\cap\mathcal{H}_j}\begin{pmatrix}\boldsymbol{\mu}_{i'1,j}\\ \widetilde{\boldsymbol{\mu}}_{i'2,j}\end{pmatrix}\right]. \tag{30}$$

The prior for the mean vector in (29) is

$$\left[\begin{pmatrix}\boldsymbol{\mu}_{i1,j}\\ \widetilde{\boldsymbol{\mu}}_{i2,j}\end{pmatrix} \mid \begin{pmatrix}\boldsymbol{\mu}_{i1,j}^*\\ \widetilde{\boldsymbol{\mu}}_{i2,j}^*\end{pmatrix}, \begin{pmatrix}\Omega_{11,j} & \Omega_{12,j}\\ \Omega_{21,j} & \Omega_{22,j}\end{pmatrix}\right] \sim \mathrm{N}\left[\begin{pmatrix}\boldsymbol{\mu}_{i1,j}^*\\ \widetilde{\boldsymbol{\mu}}_{i2,j}^*\end{pmatrix}, |N_{i,j}^e|^{-1}\begin{pmatrix}\Omega_{11,j} & \Omega_{12,j}\\ \Omega_{21,j} & \Omega_{22,j}\end{pmatrix}\right]. \tag{31}$$

The conditional distribution of $\widetilde{\mathbf{Y}}_{i2,j}$ given $\mathbf{Y}_{i1,j}$ and model parameters has a nice distributional form from which we can directly sample. It is a normal distribution (Rao (1973), Chapter 8):

$$\begin{aligned}&\left[\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j\right] \\ &\quad\sim \mathrm{N}\left(\widetilde{\boldsymbol{\mu}}_{i2,j} + \Delta_{21,j}\left(\Delta_{11,j}\right)^{-1}\left(\mathbf{Y}_{i1,j} - \boldsymbol{\mu}_{i1,j}\right), \Delta_{22,j} - \Delta_{21,j}(\Delta_{11,j})^{-1}\Delta_{12,j}\right).\end{aligned} \tag{32}$$

The posterior distribution of $\widetilde{\boldsymbol{\mu}}_{i2,j}$ is

$$\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \widetilde{\mathbf{Y}}_{i2,j}, \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, C_{i,j}, \Delta_j, \Omega_j, d_j, \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K}) \tag{33}$$

$$\propto \pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j)\pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K})$$

$$\propto \pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \widetilde{\mathbf{Y}}_{i2,j}, \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, C_{i,j}, \Delta_j, \Omega_j)\pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K}). \tag{34}$$

We propose a new value $\widetilde{\boldsymbol{\mu}}_{i2,j}^{prop}$ from the first term in (34) which has a normal distribution:

$$\left[\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \widetilde{\mathbf{Y}}_{i2,j}, \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, C_{i,j}, \Delta_j, \Omega_j\right] \tag{35}$$

$$\sim \mathrm{N}\left(\boldsymbol{\theta}_{i2,j} + \Lambda_{i_{21,j}}\left(\Lambda_{i_{11,j}}\right)^{-1}\left(\boldsymbol{\mu}_{i1,j} - \boldsymbol{\theta}_{i1,j}\right),\ \Lambda_{i_{22,j}} - \Lambda_{i_{21,j}}\left(\Lambda_{i_{11,j}}\right)^{-1}\Lambda_{i_{12,j}}\right),$$

where

$$\Lambda_{i,j} = \begin{pmatrix} \Lambda_{i11,j} & \Lambda_{i12,j} \\ \Lambda_{i21,j} & \Lambda_{i22,j} \end{pmatrix} = \left[|N_{i,j}^e| \begin{pmatrix} \Omega_{11,j} & \Omega_{12,j} \\ \Omega_{21,j} & \Omega_{22,j} \end{pmatrix}^{-1} + \begin{pmatrix} \Delta_{11,j} & \Delta_{12,j} \\ \Delta_{21,j} & \Delta_{22,j} \end{pmatrix}^{-1}\right]^{-1}$$

and

$$\boldsymbol{\theta}_{i,j} = \begin{pmatrix} \boldsymbol{\theta}_{i1,j} \\ \boldsymbol{\theta}_{i2,j} \end{pmatrix} = \Lambda_{i,j}\left[|N_{i,j}^e| \begin{pmatrix} \Omega_{11,j} & \Omega_{12,j} \\ \Omega_{21,j} & \Omega_{22,j} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\mu}_{i1,j}^* \\ \widetilde{\boldsymbol{\mu}}_{i2,j}^* \end{pmatrix} + \begin{pmatrix} \Delta_{11,j} & \Delta_{12,j} \\ \Delta_{21,j} & \Delta_{22,j} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}_{i1,j} \\ \widetilde{\mathbf{Y}}_{i2,j} \end{pmatrix}\right].$$

We then accept this proposed value with probability

$$\alpha_{\widetilde{\mu}} = \min\left\{1, \frac{\pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{\{-i\}2,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}^{prop}), \boldsymbol{\Omega}_{2K})}{\pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{\{-i\}2,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}^{current}), \boldsymbol{\Omega}_{2K})}\right\}. \tag{36}$$

1.3 Checking Covariance Structures

Next we check whether the covariance structures are similar (see the discussion in the main manuscript under the heading "Predicting tumor response under the 'null' "). Note that all the calculations in this part are for each subject j. We suppress the subject subscript j to simplify notation.

To investigate whether Σ_{11} and Δ_{11} are similar as well as Ψ_{11} and Ω_{11} as these describe the baseline residual covariances and spatial covariances, we compare the posterior expected

values of these leading sub-matrices after fitting our model to the data. Assume

$$\Sigma_{11}^{(t)} = \begin{pmatrix} \sigma_{11}^{(t)} & \sigma_{12}^{(t)} \\ \sigma_{21}^{(t)} & \sigma_{22}^{(t)} \end{pmatrix}, \quad \Delta_{11}^{(t)} = \begin{pmatrix} \delta_{11}^{(t)} & \delta_{12}^{(t)} \\ \delta_{21}^{(t)} & \delta_{22}^{(t)} \end{pmatrix},$$
$$\Psi_{11}^{(t)} = \begin{pmatrix} \psi_{11}^{(t)} & \psi_{12}^{(t)} \\ \psi_{21}^{(t)} & \psi_{22}^{(t)} \end{pmatrix}, \quad \Omega_{11}^{(t)} = \begin{pmatrix} \omega_{11}^{(t)} & \omega_{12}^{(t)} \\ \omega_{21}^{(t)} & \omega_{22}^{(t)} \end{pmatrix},$$

where $\sigma_{12}^{(t)} = \sigma_{21}^{(t)}$, $\delta_{12}^{(t)} = \delta_{21}^{(t)}$,$\omega_{12}^{(t)} = \omega_{21}^{(t)}$, $\psi_{12}^{(t)} = \psi_{21}^{(t)}$, and (t) indicates t^{th} posterior draw.

We computed the root mean squared relative difference between the three unique elements in the leading 2×2 sub-matrices, where the mean is computed over draws from the posterior. The relative root mean squared difference between the leading 2×2 sub-matrices of Δ and Σ (relative to Δ) is calculated as:

$$\text{rms}_1 = \sqrt{\frac{1}{3T} \sum_{t=1}^{T} \sum_{j \geq i}^{2} \sum_{i=1}^{2} \left(\frac{\sigma_{ij}^{(t)} - \delta_{ij}^{(t)}}{\delta_{ij}^{(t)}} \right)^2},$$

and the relative root mean squared difference between the leading 2×2 sub-matrices of Ω and Ψ (relative to Ω) is calculated as:

$$\text{rms}_2 = \sqrt{\frac{1}{3T} \sum_{t=1}^{T} \sum_{j \geq i}^{2} \sum_{i=1}^{2} \left(\frac{\psi_{ij}^{(t)} - \omega_{ij}^{(t)}}{\omega_{ij}^{(t)}} \right)^2}.$$

1.4 Summary Statistics

To compute the Kullback-Leibler divergence, we create two histograms with the posterior draws of μ_{i2h} and $\widetilde{\mu}_{i2h}$; one for diffusion, $h = 1$ and one for perfusion, $h = 2$. The bin width is $b = 3.5\sigma/n^{1/3}$ (Scott (1979)) where σ is the standard deviation of all draws of μ_{i2h} and $\widetilde{\mu}_{i2h}$ and n is number of tumor voxels. Let $(\mu_{2h}^{min}, \mu_{2h}^{max})$ denote the range of the histogram corresponding to image type h, where $\mu_{2h}^{min} = \min(\{\widetilde{\mu}_{i2h}\}_{i=1}^{n}, \{\mu_{i2h}\}_{i=1}^{n})$ and $\mu_{2h}^{max} = \max(\{\widetilde{\mu}_{i2h}\}_{i=1}^{n}, \{\mu_{i2h}\}_{i=1}^{n})$. The Kullback-Leibler divergence for image type h is approximated by $\sum_{\ell} P_{\ell h} \ln(P_{\ell h}/Q_{\ell h})$, where the summation is over all bins and $P_{\ell h}$ is the

proportion of the $\{\widetilde{\mu}_{i2h}\}_{i=1}^{n}$ that fall in bin ℓ and $Q_{\ell h}$ is the proportion of the $\{\mu_{i2h}\}_{i=1}^{n}$ that fall in bin ℓ. If $P_{\ell h} = 0$, we set $P_{\ell h}\ln(P_{\ell h}/Q_{\ell h})$ to zero and if $Q_{\ell h} = 0$, we set $Q_{\ell h} = 1.0e^{-5}$ so that the divergence is well-defined. Thus:

$$\begin{aligned} \text{dKLD} &= \sum_{\ell} P_{\ell 1}\ln(P_{\ell 1}/Q_{\ell 1}) && (37)\\ \text{pKLD} &= \sum_{\ell} P_{\ell 2}\ln(P_{\ell 2}/Q_{\ell 2}). && (38) \end{aligned}$$

The conditional diffusion and perfusion statistics are straightforward to calculate:

$$\begin{aligned} \text{cDS} &= n^{-1}\sum_{i=1}^{n} I\left[\mu_{i21} > q_{0.975}\left(\widetilde{\mu}_{i21}\right)\right] && (39)\\ \text{cPS} &= n^{-1}\sum_{i=1}^{n} I\left[\mu_{i22} < q_{0.025}\left(\widetilde{\mu}_{i22}\right)\right]. && (40) \end{aligned}$$

1.5 Stage II:

The GNLM-BMARS model with K bases functions is:

$$\begin{aligned} \pi(Z_j = 1 \mid \mathbf{X}_j, \mathbf{\Omega}_{2K}) &= g(\eta_{jK}), \qquad \eta_{jK} = \sum_{k=0}^{K}\beta_k B_k(\mathbf{X}_j),\\ B_k(\mathbf{X}_j) &= \begin{cases} 1, & k = 0,\\ \prod_{l=1}^{L_k}[s_{lk}(X_{jw_{lk}} - t_{lk})]_+, & k = 1,2,\ldots,K. \end{cases} \end{aligned} \tag{41}$$

1.5.1 Updating the latent vector d

Introduce a continuous latent variable, d_j, such that $[d_j \mid \mathbf{X}_j, \mathbf{\Omega}_{2K}] \sim \mathrm{N}\left(\eta_{jK}, 1\right)$ for each j. Let $\mathbf{d} = (d_1, \ldots, d_M)$. Define the conditional distribution of Z_j given d_j by

$$\pi(Z_j = 1 \mid d_j) = 1 \quad \text{if} \quad d_j > 0, \quad \text{and} \quad = 0 \quad \text{if} \quad d_j \leq 0. \tag{42}$$

Marginalizing (42) over d_j is equivalent to $\pi(Z_j = 1 \mid \mathbf{X}_j, \mathbf{\Omega}_{2K})$ in (41):

$$\pi(Z_j = 1 \mid \mathbf{X}_j, \mathbf{\Omega}_{2K}) = \int_{-\infty}^{\infty} \pi(Z_j = 1 \mid d_j)\pi(d_j \mid \mathbf{X}_j, \mathbf{\Omega}_{2K})\mathrm{d}d_j = \Phi(\eta_{jK}).$$

It is equally easy to show that

$$\pi(d_j \mid Z_j = 1, \mathbf{X}_j, \mathbf{\Omega}_{2K}) \quad = \quad \pi(d_j \mid \mathbf{X}_j, \mathbf{\Omega}_{2K}) I(d_j > 0) / \int_0^\infty \pi(d_j \mid \mathbf{X}_j, \mathbf{\Omega}_{2K}) \mathrm{d}d_j$$

and

$$\pi(d_j \mid Z_j = 0, \mathbf{X}_j, \mathbf{\Omega}_{2K}) \quad = \quad \pi(d_j \mid \mathbf{X}_j, \mathbf{\Omega}_{2K}) I(d_j \leq 0) / \int_{-\infty}^0 \pi(d_j \mid \mathbf{X}_j, \mathbf{\Omega}_{2K}) \mathrm{d}d_j$$

which are densities of truncated normal distributions. That is,

$$[d_j \mid Z_j = z_j, \mathbf{X}_j, \mathbf{\Omega}_{2K}] \sim \begin{cases} \mathrm{N}(\eta_{jK}, 1) \text{ truncated at the left by 0 if } z_j = 1 \\ \mathrm{N}(\eta_{jK}, 1) \text{ truncated at the right by 0 if } z_j = 0 \end{cases} . \tag{43}$$

1.5.2 Updating $\boldsymbol{\beta}_K$, ν and λ

A priori, $[\boldsymbol{\beta}_K \mid v, K] \sim \mathrm{N}(0, v\mathrm{I}_{K+1})$. By definition, $[d_j \mid \mathbf{X}_j, \mathbf{\Omega}_{2K}] \sim \mathrm{N}(\eta_{jK}, 1)$, independently, so that the distribution of the latent vector $[\mathbf{d} \mid \mathcal{X}, \mathbf{\Omega}_{2K}] \sim \mathrm{N}(\mathrm{B}_K \boldsymbol{\beta}_K, \mathrm{I}_M)$. Therefore,

$$\begin{aligned} \pi(\boldsymbol{\beta}_K | \mathbf{d}, v, \Theta_K, \mathcal{X}) &\propto \pi(\boldsymbol{\beta}_K \mid v, K) \pi(\mathbf{d} \mid \mathcal{X}, \mathbf{\Omega}_{2K}) \\ &\propto \exp\left\{-0.5 \left[v^{-1} {\boldsymbol{\beta}_K}^{\mathrm{T}} \boldsymbol{\beta}_K + (\mathbf{d} - \mathrm{B}_K \boldsymbol{\beta}_K)^{\mathrm{T}} (\mathbf{d} - \mathrm{B}_K \boldsymbol{\beta}_K)\right]\right\} \\ &\propto \exp\left\{-0.5 (\boldsymbol{\beta}_K - \mathbf{m}_K^*)^{\mathrm{T}} (V_K^*)^{-1} (\boldsymbol{\beta}_K - \mathbf{m}_K^*)\right\} \end{aligned} \tag{44}$$

where

$$V_K^* = [(v\mathrm{I}_{K+1})^{-1} + \mathrm{B}_K^{\mathrm{T}} \mathrm{B}_K]^{-1}, \tag{45}$$

$$\mathbf{m}_K^* = V_K^* \mathrm{B}_K^{\mathrm{T}} \mathbf{d}. \tag{46}$$

Thus,

$$[\boldsymbol{\beta}_K \mid \mathbf{d}, v, \Theta_K, \mathcal{X}] \sim \mathrm{N}(\mathbf{m}_K^*, V_K^*). \tag{47}$$

Equation (44) follows from the identity

$$\begin{aligned} &v^{-1} {\boldsymbol{\beta}_K}^{\mathrm{T}} \boldsymbol{\beta}_K + (\mathbf{d} - \mathrm{B}_K \boldsymbol{\beta}_K)^{\mathrm{T}} (\mathbf{d} - \mathrm{B}_K \boldsymbol{\beta}_K) = \\ &\quad (\boldsymbol{\beta}_K - \mathbf{m}_K^*)^{\mathrm{T}} (V_K^*)^{-1} (\boldsymbol{\beta}_K - \mathbf{m}_K^*) + \mathbf{d}^{\mathrm{T}} \mathbf{d} - (\mathbf{m}_K^*)^{\mathrm{T}} (V_K^*)^{-1} \mathbf{m}_K^*. \end{aligned} \tag{48}$$

Standard conjugacy results state that the full conditional distribution of v^{-1} is gamma:

$$[v^{-1} \mid \boldsymbol{\beta}_K,\, K] \sim \text{Gamma}[0.001 + 0.5(K+1),\, 0.001 + 0.5\boldsymbol{\beta}_K^{\mathrm{T}}\boldsymbol{\beta}_K], \tag{49}$$

and that

$$[\lambda \mid K] \sim \text{Gamma}(1 + K, 0.2 + 1). \tag{50}$$

1.5.3 RJMCMC moves

Now we derive the acceptance probabilities for the birth and death moves in the RJMCMC algorithm. The general form of the acceptance probability for the reversible jump algorithm is given in Green (1995). All parameters vectors in Θ_K change dimension as well as $\boldsymbol{\beta}_K$. At each iteration we randomly (with probability 0.5) choose to increase the number of BMARS bases by 1 (a birth move) or decrease it by 1 (a death move).

We begin by defining the acceptance probability of a birth move. The number of bases K, is allowed to increase by one to $K+1$. Thus, the dimension of the parameter space $\boldsymbol{\Omega}_{2K}$ changes by $2+3L_{K+1}$: $\boldsymbol{\beta}_K$ and $\mathbf{L}_K$ increase in dimension by 1 while $\mathbf{w}_K$, $\mathbf{s}_K$ and $\mathbf{t}_K$ increase in dimension by L_{K+1}. However, as we show below, $\boldsymbol{\beta}_K$ and $\boldsymbol{\beta}_{K+1}$ will be integrated out of the respective posterior distributions and thus we do not need to propose a new β_{K+1} in the birth step. If the birth proposal is accepted, a new vector $\boldsymbol{\beta}_{K+1}$ is drawn from it full conditional (47). Thus the dimension of the parameter space increases by $1 + 3L_{K+1}$ in the birth step. The RJMCMC algorithm relies on what Green (1995) calls dimension matching. We propose a random vector, say $\mathbf{U}$, of length $1 + 3L_{K+1}$ and append it to Θ_K. A bijective transformation, T, is then contrived between $\Theta_K \cup \mathbf{U}$ and Θ_{K+1}. The rate of acceptance crucially depends on this transformation and finding a good transformation can be the most difficult aspect of the RJMCMC algorithm. The transformation should be easy to compute, its Jacobian should be readily accessible and the acceptance rates of the moves should be high. The Jacobian of this transformation is multiplied into the acceptance ratio

of the proposal to account for the transformation. However, as will become evident in the next paragraph, T is the identity transformation and thus the Jacobian is 1. Integrating out $\boldsymbol{\beta}_K$ and $\boldsymbol{\beta}_{K+1}$ from the posterior distribution is key to achieving a high acceptance rate (Denison et al. (1998), Denison et al. (2002), Holmes and Denison (2003) and Mallick et al. (1999)).

Suppose there are K bases in the BMARS model. We first describe how we draw the augmentation vector $\mathbf{U}$. Each basis can consist of either a main effect or an interaction. We first draw an interaction level, $L_{K+1} \in \{1, 2\}$ for the $K+1$ basis with

$$\pi(L_{K+1} = 1) = \pi(L_{K+1} = 2) = 1/2. \tag{51}$$

Next, we draw L_{K+1} elements, $\{w_{1,K+1}, \ldots, w_{L_{K+1},K+1}\}$, from the set $\{1, 2, 3, 4\}$ without replacement. These are the covariate elements from the vectors $\mathbf{X}_j$, $j = 1, \ldots, M$. Each subset of size L_{K+1} from $\{1, 2, 3, 4\}$ is drawn with equal probability $\binom{4}{L_{K+1}}^{-1}$. Thus,

$$\pi(w_{1,K+1} = w \mid L_{K+1} = 1) \quad = \quad 1/4 \quad \text{for} \quad w = 1, 2, 3, 4. \tag{52}$$

$$\pi[(w_{1,K+1}, w_{2,K+1}) = (w, w') \mid L_{K+1} = 2] \quad = \quad 1/6 \tag{53}$$

$$\text{for} \quad (w, w') \quad = (1,2), (1,3), (1,4), (2,3), (2,4), (3,4).$$

Next, we draw knot locations: draw $t_{l,K+1}$, at random, from $\{X_{1w_{l,K+1}}, \ldots, X_{Mw_{l,K+1}}\}$ for $l = 1, \ldots, L_{K+1}$. That is,

$$\pi(t_{l,K+1} = X_{jw_{l,K+1}} \mid w_{l,K+1}) = 1/M \quad \text{for} \quad l = 1, \ldots, L_{K+1}. \tag{54}$$

Finally, we draw $s_{l,K+1} \in \{-1, 1\}$ with the following probabilities:

$$\pi(s_{l,K+1} = -1) = \pi(s_{l,K+1} = 1) = 1/2 \quad \text{for} \quad l = 1, \ldots, L_{K+1}. \tag{55}$$

Set $\mathbf{U} = (L_{K+1}, w_{1,K+1}, \ldots, w_{L_{K+1},K+1}, t_{1,K+1}, \ldots, t_{L_{K+1},K+1}, s_{1,K+1}, \ldots, s_{L_{K+1},K+1})$. Let $q(\mathbf{U})$

denote the proposal probability of the set of parameters $\mathbf{U}$. Then

$$q(\mathbf{U}) = \pi(L_{K+1})\pi[(w_{1,K+1},\ldots,w_{L_{K+1},K+1}) \mid L_{K+1}] \prod_{l=1}^{L_{K+1}} \pi(t_{l,K+1} \mid w_{l,K+1})\pi(s_{l,K+1}).$$

The acceptance probability of the birth of a new BMARS basis can now be written as

$$\alpha = \min\left\{1, \frac{\pi(\mathbf{d} \mid \mathcal{X}, \boldsymbol{\Omega}_{2,K+1})\pi(\Theta_{K+1} \mid \lambda)\pi(\boldsymbol{\beta}_{K+1} \mid v, K+1)\pi(v^{-1})\pi(\lambda)\pi_{\text{death}}}{\pi(\mathbf{d} \mid \mathcal{X}, \boldsymbol{\Omega}_{2K})\pi(\Theta_K \mid \lambda)\pi(\boldsymbol{\beta}_K \mid v, K)\pi(v^{-1})\pi(\lambda)q(\mathbf{U})\pi_{\text{birth}}}\right\}. \tag{56}$$

Now $\pi_{\text{death}} = 0.5$ is the probability of a proposing a death and $\pi_{\text{birth}} = 0.5$ is the probability of a proposing a birth. Furthermore, it is easy to show that

$$\frac{\pi(\Theta_{K+1} \mid \lambda)\pi_{\text{death}}}{\pi(\Theta_K \mid \lambda)q(\mathbf{U})\pi_{\text{birth}}} = \lambda/(K+1),$$

so that the acceptance probability reduces to

$$\alpha = \min\left\{1, \frac{\pi(\mathbf{d} \mid \mathcal{X}, \boldsymbol{\Omega}_{2,K+1})\pi(\boldsymbol{\beta}_{K+1} \mid v, K+1)\lambda}{\pi(\mathbf{d} \mid \mathcal{X}, \boldsymbol{\Omega}_{2K})\pi(\boldsymbol{\beta}_K \mid v, K)(K+1)}\right\}. \tag{57}$$

Also, it is straightforward to show that

$$\begin{aligned}
\pi(\mathbf{d} \mid \mathcal{X}, \boldsymbol{\Omega}_{2K})\pi(\boldsymbol{\beta}_K \mid v, K) &\propto (2\pi\, v)^{-(K+1)/2} \exp\left\{-0.5\left[v^{-1}\boldsymbol{\beta}_K^{\mathrm{T}}\boldsymbol{\beta}_K + \right.\right. \\
&\qquad \left.\left.(\mathbf{d} - \mathrm{B}_K\boldsymbol{\beta}_K)^{\mathrm{T}}(\mathbf{d} - \mathrm{B}_K\boldsymbol{\beta}_K)\right]\right\} \\
&= (2\pi\, v)^{-(K+1)/2} \exp\left\{-0.5(\boldsymbol{\beta}_K - \mathbf{m}_K^*)^{\mathrm{T}}(V_K^*)^{-1}(\boldsymbol{\beta}_K - \mathbf{m}_K^*)\right\} \times \\
&\quad \exp\left\{-0.5\left[2\mathbf{d}^{\mathrm{T}}\mathbf{d} - (\mathbf{m}_K^*)^{\mathrm{T}}(V_K^*)^{-1}\mathbf{m}_K^*\right]\right\},
\end{aligned} \tag{58}$$

where the equality follows from (48). Now integrating out $\boldsymbol{\beta}_{K+1}$ and $\boldsymbol{\beta}_K$ from their respective joint full conditionals (58) the acceptance probability simplifies to:

$$\begin{aligned}
\alpha_{\text{birth}} &= \min\left\{1, \frac{\pi(\mathbf{d} \mid \mathcal{X}, \boldsymbol{\Omega}_{2,K+1}, v)\lambda}{\pi(\mathbf{d} \mid \mathcal{X}, \boldsymbol{\Omega}_{2K}, v)(K+1)}\right\} \\
&= \min\left\{1, \frac{|V_{K+1}^*|^{1/2}\exp(a_K - a_{K+1})\lambda}{v^{1/2}|V_K^*|^{1/2}(K+1)}\right\}
\end{aligned} \tag{59}$$

where $a_K = (\mathbf{d}^{\mathrm{T}}\mathbf{d} - \mathbf{m}_K^{*\mathrm{T}}(V_K^*)^{-1}\mathbf{m}_K^*)/2$.

For a death move suppose there are K BMARS basis excluding the intercept term. We randomly draw one of the current K BMARS bases to delete each with probability $1/K$. The acceptance probability of this death step is

$$\alpha_{\text{death}} = \min\left\{1, \frac{v^{1/2}|V_{K-1}^*|^{1/2}K}{|V_K^*|^{1/2}\exp(a_{K-1} - a_K)\lambda}\right\}. \tag{60}$$

1.5.4 Updating Knot Locations

The final step in the algorithm for stage II is to propose a move of a knot location. To move a knot we first draw a basis at random each with probability $1/K$. Suppose the chosen basis has index k. Given this basis we draw a factor, ℓ, from the set $\{1, \ldots, L_k\}$ with equal probability (if there is a single factor ($L_k = 1$), $\ell = 1$ with probability 1, if there are two factors ($L_k = 2$), $\ell = 1$ with probability 0.5). Propose to move knot t_{lk} from its current position by sampling a new position from $\{X_{1w_{lk}}, \ldots, X_{Mw_{lk}}\}$ each with probability $1/M$ (note that there is probability of $1/M$ that the knot will not move from its current position. The current knot location is t_{lk}. Call the proposed position t_{lk}^{prop}. Update column k of B_K and call the propose matrix $\mathrm{B}_K^{\text{prop}}$. Compute the proposed vector $\mathbf{m}_K^{*,\text{prop}}$ and matrix $V_K^{*,\text{prop}}$ from (46) and (45) using $\mathrm{B}_K^{\text{prop}}$. Compute $a_K^{\text{prop}} = (\mathbf{d}^{\mathrm{T}}\mathbf{d} - \mathbf{m}_K^{*,\text{prop}^{\mathrm{T}}}(V_K^{*,\text{prop}})^{-1}\mathbf{m}_K^{*,\text{prop}})/2$. The acceptance probability of this move is

$$\alpha_{\text{move}} = \min\left\{1, \frac{|V_K^{*,\text{prop}}|^{1/2}\exp(a_K - a_K^{\text{prop}})}{|V_K^*|^{1/2}}\right\}. \tag{61}$$

2 Cross-validated Prediction

Cross-validated prediction was introduced by Gelfand et al. (1992). Following the notation in the parent manuscript, the predictive probability that $Z_j = 1$ given $\mathbf{Z}_{\{-j\}}$ and $\mathcal{Y}$ is

$$\pi(Z_j = 1 \mid \mathbf{Z}_{\{-j\}}, \mathcal{Y}) = \int\int \pi(Z_j = 1 \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2)\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z}_{\{-\mathbf{j}\}} = \mathbf{z}_{\{-\mathbf{j}\}}, \mathcal{Y})d\mathbf{\Omega}_1 d\mathbf{\Omega}_2. \tag{62}$$

Note that after marginalizing over the hyperprior parameters v and λ, we have

$$\begin{aligned}
\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z} = \mathbf{z}, \mathcal{Y}) &\propto \prod_{j=1}^{M} \pi(Z_j = z_j, \mathcal{Y}_j \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2)\pi(\mathbf{\Omega}_{1j})\pi(\mathbf{\Omega}_2) \\
\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z}_{\{-\mathbf{j}\}} = \mathbf{z}_{\{-\mathbf{j}\}}, \mathcal{Y}) &\propto \prod_{i=1; i\neq j}^{M} \pi(Z_i = z_i, \mathcal{Y}_i \mid \mathbf{\Omega}_{1i}, \mathbf{\Omega}_2)\pi(\mathbf{\Omega}_{1i})\pi(\mathbf{\Omega}_2) \times \\
&\qquad \pi(\mathcal{Y}_j \mid \mathbf{\Omega}_{1j})\pi(\mathbf{\Omega}_{1j})
\end{aligned}$$

and

$$\frac{\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z} = \mathbf{z}, \mathcal{Y})}{\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z}_{\{-\mathbf{j}\}} = \mathbf{z}_{\{-\mathbf{j}\}}, \mathcal{Y})} = \frac{\pi(Z_j = z_j, \mathcal{Y}_j \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2)}{\pi(\mathcal{Y}_j \mid \mathbf{\Omega}_{1j})} = \pi(Z_j = z_j \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2). \tag{63}$$

Now rewrite (62) using (63):

$$\begin{aligned}
&\pi(Z_j = 1 \mid \mathbf{Z}_{\{-j\}}, \mathcal{Y}) \\
&= \frac{\int\int \pi(Z_j = 1 \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2) \left[\frac{\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z}_{\{-\mathbf{j}\}} = \mathbf{z}_{\{-\mathbf{j}\}}, \mathcal{Y})}{\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z} = \mathbf{z}, \mathcal{Y})}\right] \pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z} = \mathbf{z}, \mathcal{Y}) d\mathbf{\Omega}_1 d\mathbf{\Omega}_2}{\int\int \left[\frac{\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z}_{\{-\mathbf{j}\}} = \mathbf{z}_{\{-\mathbf{j}\}}, \mathcal{Y})}{\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z} = \mathbf{z}, \mathcal{Y})}\right] \pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z} = \mathbf{z}, \mathcal{Y}) d\mathbf{\Omega}_1, \mathbf{\Omega}_2} \qquad (64) \\
&= \frac{\int\int \pi(Z_j = 1 \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2) \left[1/\pi(Z_j = z_j \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2)\right] \pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z} = \mathbf{z}, \mathcal{Y}) d\mathbf{\Omega}_1 d\mathbf{\Omega}_2}{\int\int \left[1/\pi(Z_j = z_j \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2)\right] \pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z} = \mathbf{z}, \mathcal{Y}) d\mathbf{\Omega}_1 d\mathbf{\Omega}_2} \qquad (65) \\
&\approx \frac{\frac{1}{T}\sum_{t=1}^{T} \pi\left(Z_j = 1 \mid \mathbf{\Omega}_{1j}^{(t)}, \mathbf{\Omega}_2^{(t)}\right) / \pi\left(Z_j = z_j \mid \mathbf{\Omega}_{1j}^{(t)}, \mathbf{\Omega}_2^{(t)}\right)}{\frac{1}{T}\sum_{t=1}^{T} 1/\pi\left(Z_j = z_j \mid \mathbf{\Omega}_{1j}^{(t)}, \mathbf{\Omega}_2^{(t)}\right)}. \qquad (66)
\end{aligned}$$

We note that in (64) the denominator equals 1 and that (65) only depends on the posterior distribution of the parameters given the entire data. Furthermore, (66) is computed from the MCMC draws from the posterior distribution of the parameters given the full data. Thus, we only need to run the algorithm once, on the entire data set, and estimate the cross-validated predictive probability for each subject j using (66).

3 Pseudocode

Initialize parameters

Stage I:

For each subject

1. Set $\boldsymbol{\mu}_i = \mathbf{Y}_i$, $i = 1, \ldots, n$.
2. Set $\boldsymbol{\nu}_i = \mathbf{W}_i \, i = 1, \ldots, n^e$.
3. Set $\Sigma = \Psi = \Delta = \Omega = \mathrm{I}_4$.

End for each subject

Stage II:

1. Set $K = 0$ (intercept term only).
2. Set $v = 1$.
3. Set $\lambda = 5$.
4. Set $d_j = 1$ if $Z_j = 1$ an $d_j = -1$ if $Z_j = 0$, for $j = 1, \ldots, M$.
5. Draw $\boldsymbol{\beta}$ from distribution (47) ($\boldsymbol{\beta} = \beta_0$ when $K = 0$ and $\mathrm{B}_0 = B_0(\mathbf{X}_j) = (1, \ldots, 1)^{\mathrm{T}}$, a vector of ones of length M).

Iterate For $t = 1$ to $100,000$ discarding the first $50,000$ as burn-in.

Stage I:

Iterate over all subjects, $j = 1, \ldots, M$. (Each subject has her/his own set of parameters. The subject index j is suppressed to be consistent with the main manuscript).

1. For tumor ROI:

   (a) For each voxel $i = 1, \ldots, n$, propose $\boldsymbol{\mu}_i^{prop}$ from (15).

   Accept $\boldsymbol{\mu}_i^{prop}$ with the probability (16).

   (b) Draw $[\Sigma \mid \{\boldsymbol{\mu}_i\}_{i=1}^n, \{\mathbf{Y}_i\}_{i=1}^n]$ from (11).

(c) Draw $[\Psi \mid \{\boldsymbol{\mu}_i\}_{i=1}^n]$ from (12).

2. For healthy tissue ROI:

(a) Draw $\left[\boldsymbol{\nu}_i \mid \Omega, \Delta, \mathbf{W}_i, \{\boldsymbol{\nu}_{i'}\}_{i' \in N_i^e}\right]$, $i \in N_{i'}^e \cap \mathcal{H}$, from (17).

(b) For each $i \in N_{i'}^e \cap \mathcal{S}$, propose $\boldsymbol{\nu}_i^{prop}$ from (17).

Accept $\boldsymbol{\nu}_i^{prop}$ with probability (20).

(c) Propose Ω^{prop} from (27).

Accept Ω^{prop} with probability (28).

(d) Propose $\Delta^{prop}$from (23).

Accept Δ with probability (24).

3. Predict tumor response under null:

(a) Draw $\left[\widetilde{\mathbf{Y}}_{i2} \mid \cdot\right]$, $i = 1, \ldots, n$, from (32).

(b) For $i = 1, \ldots, n$, propose $\widetilde{\boldsymbol{\mu}}_{i2}^{prop}$ from (35).

Accept $\widetilde{\boldsymbol{\mu}}_{i2}^{prop}$ with probability (36).

4. Calculate the summary statistics for each subject j (covariate vector $\mathbf{X}_j$):

(a) Calculate dKLD using equation (37) and pKLD using equation (38).

(b) Calculate cDS using equation (39) and cPS using equation (40).

End iterate over subjects.

Stage II: Assume there are currently K basis functions.

Iterate 10 times (oversample) $q = 1$ to 10.

1. Attempt a **Move step** by altering a spline basis function if $K > 0$, else go to 2:

(a) Draw a BMARS basis, k, at random, with equal probability $1/K$, from the set of bases $\{1, \ldots, K\}$.

(b) Draw a factor, l, at random, with equal probability $1/L_k$, from the set of factors $\{1, \ldots, L_k\}$.

(c) Draw a knot location, t_{lk}, at random, with equal probability $1/M$, from $\{X_{1w_{lk}}, \ldots, X_{Mw_{lk}}\}$.

(d) If move (new knot location) accepted with probability α_{move} (61).

i. Draw latent variables $[d_j \mid Z_j = z_j, \mathbf{X}_j, \boldsymbol{\Omega}_{2K}], j = 1, \ldots, M$, from (43).

ii. Draw $[v^{-1} \mid \boldsymbol{\beta}_K, K]$ from (49).

iii. Draw $[\boldsymbol{\beta}_K \mid \mathbf{d}, v, \Theta_K, \mathcal{X}]$ from (47).

iv. Draw $[\lambda \mid K]$ from (50).

else, keep current knot location.

2. RJMCMC: Draw $U \sim$ Bernoulli(0.5) if $K > 0$ otherwise set $U = 0$.

(a) if $U = 0$ **Birth step.**

i. Draw L_{K+1} according to (51).

ii. If $L_{K+1} = 1$, draw $w_{1,K+1} \mid L_{K+1}$ from the set $\{1, 2, 3, 4\}$ with equal prob. 1/4, see (52).

else draw $(w_{1,K+1}, w_{2,K+1}) \mid L_{K+1}$, with equal prob., from the set $\{(1,2), (1,3), (1,4), (2,3), (2,4), (3,4)\}$, see (53).

iii. Draw the knot point(s) $t_{l,K+1}$, $l = 1, \ldots, L_{K+1}$, see (54).

iv. Draw $s_{l,K+1}$, $l = 1, \ldots, L_{K+1}$, see (55).

v. Accept the birth with probability α_{birth} (59).

(b) if $U = 1$ **Death step.**

i. Remove k^{th} basis from the model with probability $1/K$.

ii. Accept the death with probability α_{death} (60).

3. Draw latent variables $[d_j \mid Z_j = z_j, \mathbf{X}_j, \mathbf{\Omega}_{2K}], j = 1, \ldots, M$, from (43).

4. Draw $[v^{-1} \mid \boldsymbol{\beta}_K,\ K]$ from (49).

5. Draw $[\boldsymbol{\beta}_K \mid \mathbf{d}, v, \Theta_K, \mathcal{X}]$ from (47).

end oversample

End Iterate

4 Image Processing

All the MR images are spatially co-registered by using the pretreatment anatomical images as the reference data set. This step allows all images of a given patient to be viewed and analyzed from a fixed frame of reference. The co-registration was performed by using the "mutual information for automatic multi-modality image fusion" (MIAMI FUSE) program (Meyer et al. (1997)). After co-registration, tumors were manually segmented by a neuroradiologist. Only the intersection of the segmented tumors at the two time points were retained as our tumor ROI. To define the healthy tissue ROI, we reflected the tumor ROI to the contralateral hemisphere of the brain where the axis of reflection is determined on axial slices of the brain (Figure 1). The axial midline of the brain is not perfectly aligned with the vertical axis. Therefore, after reflection, we visually inspected whether the healthy tissue ROI intersected any non-brain tissue regions such as the ventricles, meninges or the skull. If it intersected any of these structures, we translated the ROI a small amount, but as large as necessary, to remove the intersection. Translations of 3 to 10 voxels was all that was required for our data set. Seventeen of the 47 subjects required healthy tissue ROI translations. A sensitivity analysis was conducted where the healthy tissue ROIs were translated by varying amounts (up to 20 voxels), while ensuring that the ROIs were completely within brain tissue.

The distribution of healthy tissue intensities were similar and the varying translations did not substantively affect the posterior distributions of the summary statistics.

5 Identifying $\widetilde{\mathbf{Y}}_{i2}$ and $\widetilde{\boldsymbol{\mu}}_{i2}$

There is an issue of identifiability when simultaneously predicting $\widetilde{\mathbf{Y}}_{i2}$ and estimating $\widetilde{\boldsymbol{\mu}}_{i2}$ (tumor response under the "null") in the healthy tissue ROI. To see this, note that

$$\begin{aligned}\pi\left[\widetilde{\mathbf{Y}}_{i2}, \widetilde{\boldsymbol{\mu}}_{i2} \mid \mathbf{Y}_{i1}, \boldsymbol{\mu}_{i1}, \cdot\right] &\propto \pi\left[\begin{pmatrix} \mathbf{Y}_{i1} \\ \widetilde{\mathbf{Y}}_{i2} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix} \mid \cdot\right] \\ &= \pi\left[\begin{pmatrix} \mathbf{Y}_{i1} \\ \widetilde{\mathbf{Y}}_{i2} \end{pmatrix} \mid \begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix}, \cdot\right] \pi\left[\begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix} \mid \cdot\right].\end{aligned}$$

Now,

$$\begin{aligned}\left[\begin{pmatrix} \mathbf{Y}_{i1} \\ \widetilde{\mathbf{Y}}_{i2} \end{pmatrix} \mid \begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix}, P^{-1}\right] &\sim \mathrm{N}\left[\begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix}, P^{-1}\right] \\ \left[\begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix} \mid \begin{pmatrix} \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix}, Q^{-1}\right] &\sim \mathrm{N}\left[\begin{pmatrix} \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix}, Q^{-1}\right],\end{aligned}$$

where P and Q are the inverses of the covariances in (29) and (31). Let

$$\begin{pmatrix} \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix} = |N_i^e \cap \mathcal{H}|^{-1} \sum_{i' \in N_i^e \cap \mathcal{H}} \begin{pmatrix} \boldsymbol{\mu}_{i'1} \\ \widetilde{\boldsymbol{\mu}}_{i'2} \end{pmatrix}. \tag{67}$$

Then

$$\begin{aligned}\pi\left[\widetilde{\mathbf{Y}}_{i2}, \widetilde{\boldsymbol{\mu}}_{i2} \mid \mathbf{Y}_{i1}, \boldsymbol{\mu}_{i1}, \cdot\right] \propto\ & \exp\left[-0.5 \begin{pmatrix} \mathbf{Y}_{i1} - \boldsymbol{\mu}_{i1} \\ \widetilde{\mathbf{Y}}_{i2} - \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix}^{\mathrm{T}} P \begin{pmatrix} \mathbf{Y}_{i1} - \boldsymbol{\mu}_{i1} \\ \widetilde{\mathbf{Y}}_{i2} - \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix}\right] \times \\ & \exp\left[-0.5 \begin{pmatrix} \boldsymbol{\mu}_{i1} - \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} - \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix}^{\mathrm{T}} Q \begin{pmatrix} \boldsymbol{\mu}_{i1} - \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} - \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix}\right],\end{aligned} \tag{68}$$

for all $i \in \mathcal{H}$. Now it is obvious that the density (68) is invariant when an arbitrary constant vector $\boldsymbol{\delta}$ is added to $\widetilde{\mathbf{Y}}_{i2}$ and $\widetilde{\boldsymbol{\mu}}_{i2}$ for all $i \in \mathcal{H}$. Hence, $\widetilde{\mathbf{Y}}_{i2}$ and $\widetilde{\boldsymbol{\mu}}_{i2}$ are not identifiable.

To solve this identifiability problem, we expand the healthy tissue ROI by a one voxel thick shell and estimate the posterior distribution of the parameters for the healthy tissue expanded ROI. In (68), further condition on the $\boldsymbol{\nu}_i$, $i \in \mathcal{S}$ so that (67) becomes

$$\begin{pmatrix} \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix} = |N^e_i|^{-1} \left[\sum_{i' \in N^e_i \cap \mathcal{S}} \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\nu}_{i'2} \end{pmatrix} + \sum_{i' \in N^e_i \cap \mathcal{H}} \begin{pmatrix} \boldsymbol{\mu}_{i'1} \\ \widetilde{\boldsymbol{\mu}}_{i'2} \end{pmatrix} \right]. \tag{69}$$

We can no longer add a constant $\boldsymbol{\delta}$ to $\widetilde{\mathbf{Y}}_{i2}$ and $\widetilde{\boldsymbol{\mu}}_{i2}$ for all $i \in \mathcal{H}$ without changing the density (68). To see this, consider a voxel i in $\mathcal{H}$ such that $N^e_i \cap \mathcal{S}$ is non-empty. For this voxel, $\widetilde{\boldsymbol{\mu}}^*_{i2}$ depends on some $\boldsymbol{\nu}_i$, $i \in \mathcal{S}$, on which we have conditioned, therefore, the second exponential in (68) is no longer invariant to the addition of an arbitrary constant to all the $\widetilde{\boldsymbol{\mu}}_{i2}$, $i \in \mathcal{H}$. Furthermore, this lack of invariance propagates to all $\widetilde{\boldsymbol{\mu}}_{i2}$, $i \in \mathcal{H}$. In fact, in the PWDP or mPWDP model, the joint prior distribution of the means is not a proper distribution (Besag (1993)) and the means are not, a priori, identifiable. However, the posterior is a proper distribution (Besag (1993)) and the means are a posteriori identifiable. The difference here in the predictive setting is that some of the data, the $\widetilde{\mathbf{Y}}_{i2}$, are not observed.

6 Simulation Studies and Sensitivity Analyses

Simulation Studies: We perform a series of simulation studies to assess the mPWDP model performance. To simplify the simulations, we only consider the Kullback-Leibler divergence statistics. The cDS and cPS are completely dependent on the mPWDP model and are extremely complicated to generate, if at all possible (we do not see a way), in a proper simulation study. Under each simulation scenario, we generate $N = 1000$ simulated data sets and compute the average relative mean squared error (rMSE) and relative bias

(rBias) of the KLD statistics:

$$\begin{aligned} \text{rMSE} &= N^{-1}\sum_{i=1}^{N}\left[\left(\overline{\text{KLD}}_i - \text{KLD}_i^{true}\right)/\text{KLD}_i^{true}\right]^2 \\ \text{rBias} &= N^{-1}\sum_{i=1}^{N}\left(\overline{\text{KLD}}_i - \text{KLD}_i^{true}\right)/\text{KLD}_i^{true}, \end{aligned}$$

where $\overline{\text{KLD}}_i$ is the posterior mean and KLD_i^{true} is the true KLD from the ith simulation. We also calculate the percentage of time that the 95% HPD (Highest Probability Density) interval covers the truth.

Without loss of generality, we assume that there is only one image type. Rather than construct ROIs, we randomly select ROIs from the glioma data. Given a simulation scenario, we generate 1000 simulated image pairs—baseline and week 3 images (full brain images). For each simulated image pair, one tumor/healthy tissue ROI pair is selected with equal probability from the set of observed ROIs from the glioma data set. These ROIs are then placed within the brain template. To simulate a baseline image, we first assume that the baseline mean intensities of all tumor voxels, μ_{i1}, and healthy tissue voxels, ν_{i1}, are independently and identically distributed as $\text{N}(1, 3)$. We then generate week 3 mean images assuming different location shifts between the underlying distributions of all voxels in the healthy tissue and tumor ROIs: $\mu_{i2} = \mu_{i1} + \phi_i$, where $\phi_i \sim N(\theta, \sigma^2)$; $\nu_{i2} = \nu_{i1} + \varphi_i$, where $\varphi_i \sim N(0.05, \sigma^2)$. Next, spatial correlation is induced in the images by smoothing each image using an isotropic Gaussian kernel at three levels of smoothing: FWHM $= 3, 5$ and 7mm. FWHM is an acronym for *full width at half maximum*. For isotropic normally distributed data with a common variance σ^2 it is defined by FWHM $= 2\sqrt{2\ln 2}\sigma$. We consider these smoothed images as the truth. The voxel means of the smoothed images are distinguished from the voxel means of the unsmoothed images by the superscript *. The true KL divergence statistic is then calculated as the marginal distribution difference between tumor ROI voxels μ_{i2}^* and healthy ROI voxels ν_{i2}^*, $i = 1, \ldots, n$.

To obtain the final simulated images, we add random noise to the smoothed images: $W_{i1} = \nu^*_{i1} + \varepsilon_{i1}$, $W_{i2} = \nu^*_{i2} + \varepsilon_{i2}$, $Y_{i1} = \mu^*_{i1} + \epsilon_{i1}$ and $Y_{i2} = \mu^*_{i2} + \epsilon_{i2}$, where $\varepsilon_{i1} \sim N(0, 0.03)$, $\varepsilon_{i2} \sim \mathrm{N}(0, 0.04)$, $\epsilon_{i1} \sim N(0, 0.05)$ and $\epsilon_{i2} \sim N(0, 0.06)$. The values of all the parameters in the simulation study are determined based on the posterior parameter estimates given the glioma data. We then apply our mPWDP model on the simulated data sets with different combinations of location and scale shifts, θ, σ^2, as well as the three levels of smoothing (Table 1). When $\theta = 0.05$ and $\sigma^2 = 0.01$ the simulated data follow the null response. From Table 1 (mPWDP model), we can see that the relative MSE and bias are relatively small, and decrease as the location shift θ increases. Moreover, the 95% HPD interval coverage is close to the nominal 95% level.

Next, we compare the mPWDP model with a simpler model that ignores spatial correlation. This simpler model is $[Y_{i2} \mid Y_{i1}] \sim \mathrm{N}(Y_{i1} + \eta_1, \sigma_1^2)$ and $[W_{i2} \mid W_{i1}] \sim \mathrm{N}(W_{i1} + \eta_2, \sigma_2^2)$, independently. We predict tumor response, $[\widetilde{Y}_{i2} \mid Y_{i1}] \sim \mathrm{N}(Y_{i1} + \widehat{\eta}_2, \widehat{\sigma}_2^2)$, under the "null" by using the MLE estimates $\widehat{\eta}_2$ and $\widehat{\sigma}_2{}^2$ of η_2 and σ_2^2, respectively. The KL divergence is then estimated between the marginal distributions of the observed tumor ROI voxels and the predicted (under the null) tumor ROI voxels. The relative MSE and bias are tabulated in the last two columns of Table 1. Ignoring the spatial correlation in the data results in rMSE and rBias that are an order of magnitude larger than when spatial correlation is accounted for, demonstrating the importance of accounting for this correlation.

Sensitivity Analysis: In stage II, the only informative prior is that on K, the number of BMARS basis: $[K \mid \lambda] \sim \text{Poisson}(\lambda)$, $\lambda \sim \text{Gamma}(\alpha, \beta)$ where we set $\alpha = 1$, $\beta = 0.2$. Given the small sample size, 47 patients, we believe that a parsimonious model is in order. A prior on K that favors a large number of basis functions may result in over-fitting of the data and a potential decrease in predictive power (Denison et al. (2002), Chapter 2). Thus, we choose to place an informative prior on K with a small mean. We do note, however, that

marginalizing the joint distribution of $[K, \lambda]$ over λ results in a negative binomial distribution for K. Further, we estimate λ as well. Hierarchically modeling K and λ in this fashion removes some of the dependence of λ, and hence of K, on the prior and places more weight on the data.

We perform a sensitivity analysis on our choice of prior for λ, and hence, marginally, on K. We change the values of α and β as well as the distribution of λ to a uniform prior distribution on $[0, 10]$. Correct classification results are tabulated in Table 2. The overall classification rate is not very sensitive to these changes in the prior distribution of λ. For this sensitivity analysis, at most one extra subject is misclassified.

We also assess prediction sensitivity to the thresholds used to derive cDS and cPS. Recall the thresholds used are the 97.5th and the 2.5th percentile, respectively, of the conditional distribution of the means of the predicted tumor response at week 3 under the null. We change these thresholds to the 99.5th/0.5th and to the 95.0th/5.0th percentile. The CCR_{CV} is reduces to 0.74 for both sets of thresholds—two more subjects are misclassified. Nevertheless, this is still an acceptable classification rate and is higher than all other (simpler) models considered in the main manuscript.

In stage I, we assign inverse Wishart distributions with identity scale matrix and 5 degrees of freedom to the the covariance matrices in our model. We argue that these priors have little influence on the posterior due to the large number of tumor voxels. We now provide support in favor of our argument via a sensitivity analysis. We assess the change in the marginal posterior distributions of the four summary statistics as we vary the a priori degrees of freedom of the covariance matrices. We set the degrees of freedom to four values: 0, 5, 10 and 15. In Figures 2 and 3 we graph the marginal posterior densities of the four summary statistics for two subjects. In Figure 2 we show them for the subject with the smallest tumor and in Figure 3 we show them for a randomly selected subject. The marginal

posterior distributions of the four statistics are minimally affected, and this is true for all 47 patients.

References

Besag, J. (1993). Towards Bayesian image analysis. *Journal of Applied Statistics 20*, 107–119.

Denison, D. G. T., C. C. Holmes, B. K. Mallick, and A. F. M. Smith (2002). *Bayesian Methods for Nonlinear Classification and Regression*. Wiley.

Denison, D. G. T., B. K. Mallick, and A. F. M. Smith (1998). Bayesian MARS. *Statistics and Computing 8*, 337–346.

Gelfand, A. E., D. K. Dey, and H. Chang (1992). Model determination using predictive distributions with implementation via sampling-based methods. *Bayesian Statistics 4*, 147–167.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*, 711–732.

Higdon, D. M., J. E. Bowsher, V. E. Johnson, T. G. Turkington, D. R. Gilland, and R. J. Jaszczak (1997). Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data. *IEEE Transactions on Medical Imaging 16*, 516–526.

Holmes, C. C. and D. G. T. Denison (2003). Classification with Bayesian MARS. *Machine Learning 50*, 159–173.

Mallick, B. K., D. G. T. Denison, and A. F. M. Smith (1999). Bayesian survival analysis using a MARS model. *Biometrics 55*, 1071–1077.

Meyer, C. R., J. L. Boes, B. Kim, P. H. Bland, K. R. Zasadny, P. V. Kison, K. Koral,

K. A. Frey, and R. L. Wahl (1997). Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Medical Image Analysis 1*, 195–206.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications, 2nd Edition*. John Wiley & Sons.

Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika 66*, 605–610.

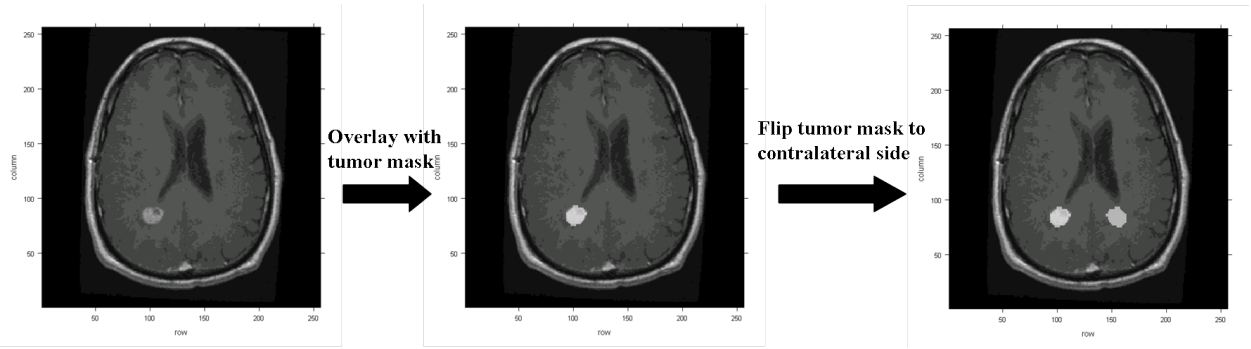

Figure 1: Obtaining the healthy tissue ROI. Left image: original T1-weighted contrast enhanced MRI; Middle image: overlay the original MRI with tumor mask to obtain the tumor region of interest; Right image: mirror the tumor mask to the contralateral hemisphere of the brain to get the healthy tissue region of interest.

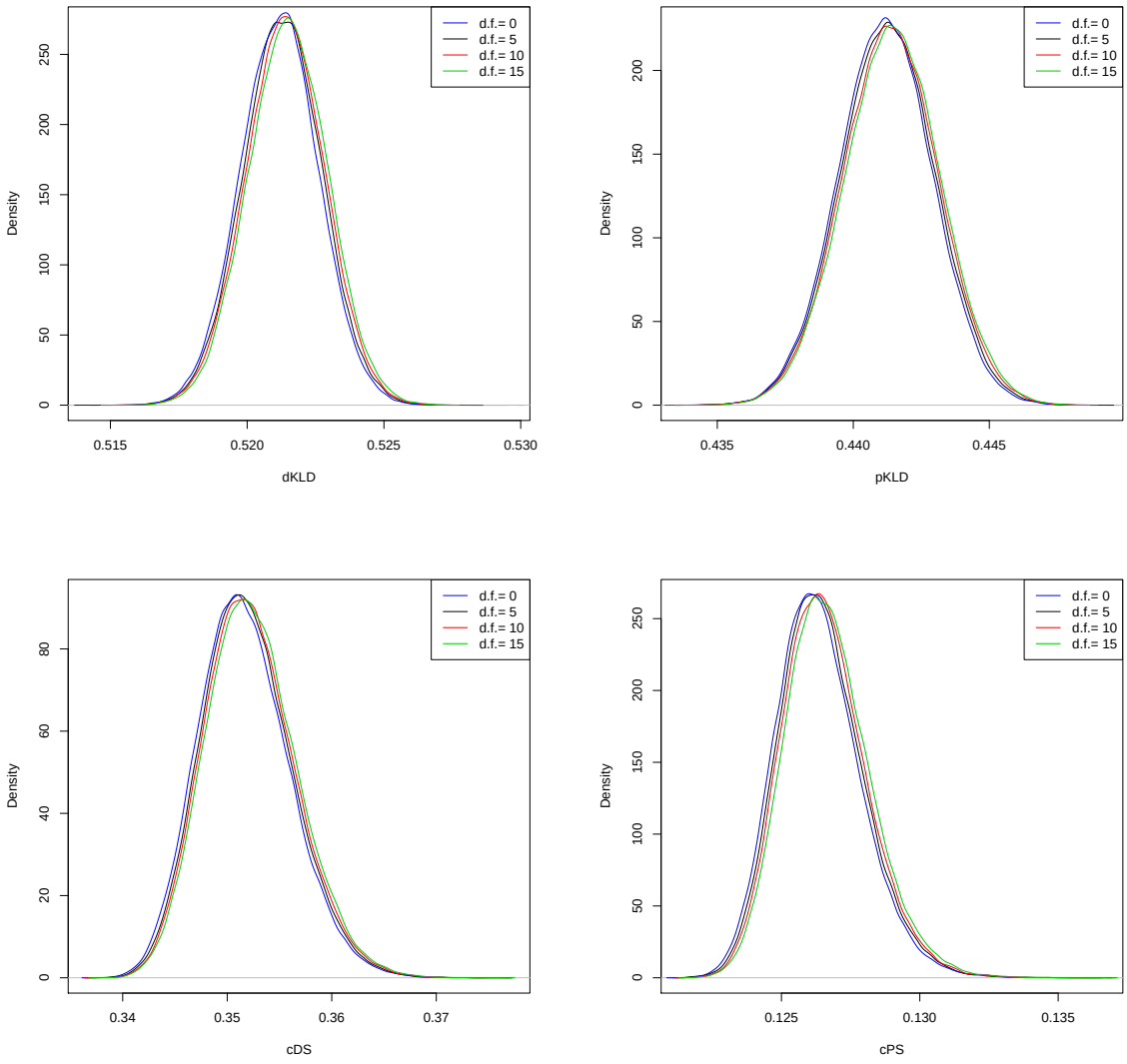

Figure 2: Sensitivity of the summary statistics to the prior number of degrees of freedom for the covariance matrices Σ, Ψ, Ω and Δ for the subject with the smallest tumor. The statistics are robust to changes in the degrees of freedom due to the large number of voxels in the tumors.

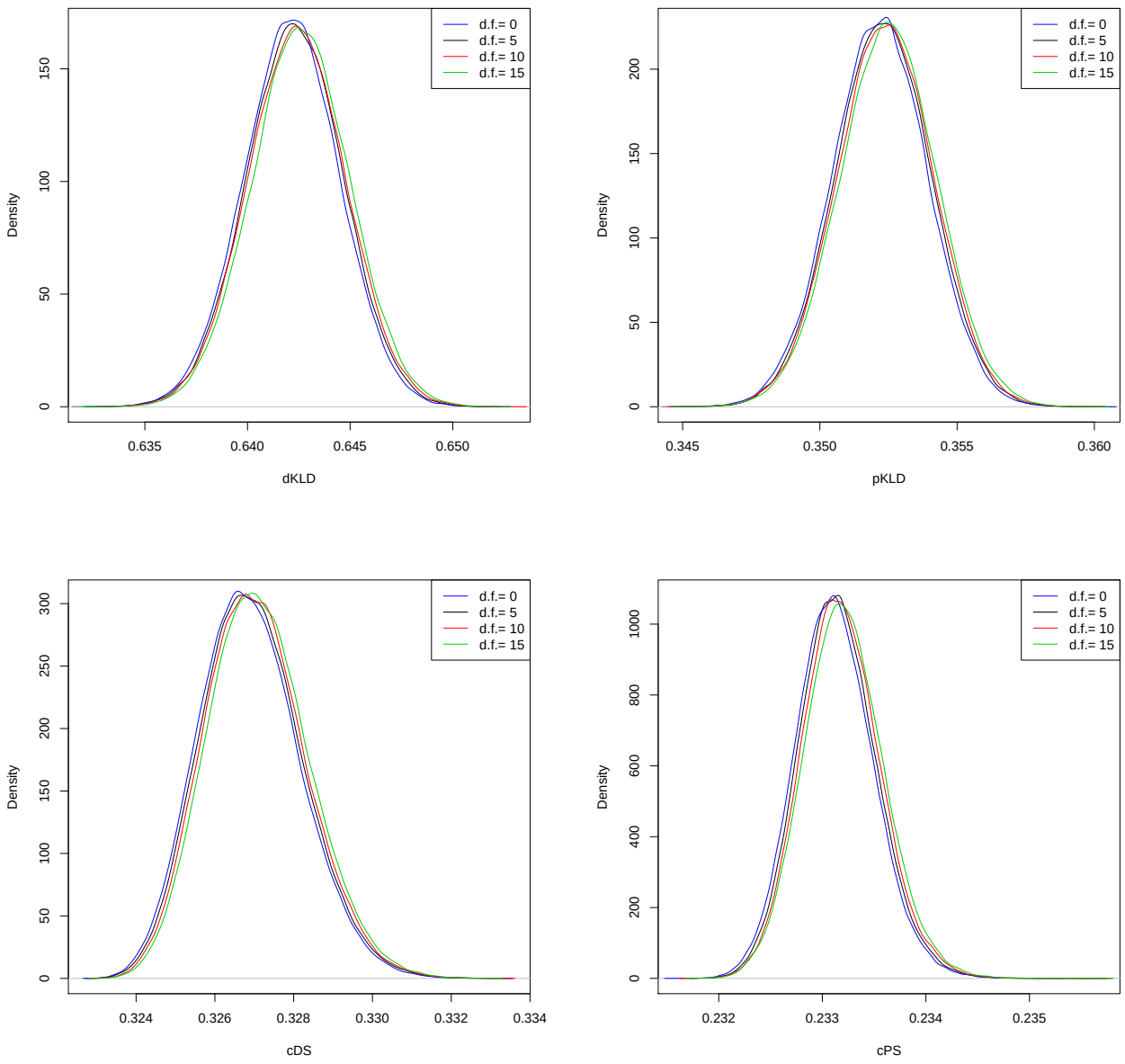

Figure 3: Sensitivity of the summary statistics to the prior number of degrees of freedom for the covariance matrices Σ, Ψ, Ω and Δ for a randomly selected subject. The statistics are robust to changes in the degrees of freedom due to the large number of voxels in the tumors.

Table 1: Simulation studies — rMSE and rBias of KLD in stage I of the mPWDP model vs. a spatial independence model.

FWHM[1]	θ	σ^2	mPWDP model			Independence model	
			rMSE (SD)[2]	rBias (SD)[3]	Coverage[4]	rMSE (SD)[5]	rBias (SD)[6]
3 mm	0.05	0.01	1.41(2.59)	4.60(9.01)	95.0	65.2(20.5)	74.5(19.2)
	0.05	0.05	1.48(2.65)	4.66(9.09)	95.1	66.1(20.7)	75.0(20.3)
	0.05	0.10	1.53(2.67)	4.73(9.15)	95.3	66.6(21.4)	75.7(20.5)
	0.10	0.01	1.27(2.43)	4.47(8.67)	95.2	56.1(18.9)	70.2(15.6)
	0.10	0.05	1.31(2.50)	4.51(8.83)	95.4	57.9(19.1)	70.7(16.3)
	0.10	0.10	1.33(2.52)	4.58(8.99)	94.3	59.2(19.5)	72.1(17.1)
	0.50	0.01	1.08(2.30)	−4.13(6.79)	95.1	37.8(28.3)	−36.4(36.6)
	0.50	0.05	1.15(2.31)	−4.17(6.98)	95.5	38.1(28.9)	−38.3(37.5)
	0.50	0.10	1.19(2.33)	−4.25(7.23)	95.3	39.3(29.5)	−38.1(38.3)
5 mm	0.05	0.01	1.67(2.60)	4.71(9.67)	95.3	70.3(21.2)	81.2(19.9)
	0.05	0.05	1.78(2.71)	4.81(9.77)	95.6	71.2(21.7)	81.9(20.6)
	0.05	0.10	1.75(2.67)	4.78(9.72)	95.4	72.0(22.4)	82.3(21.1)
	0.10	0.01	1.45(2.51)	4.63(9.27)	95.3	57.9(20.1)	75.6(16.0)
	0.10	0.05	1.59(2.47)	4.69(9.61)	95.7	59.3(20.5)	76.0(17.5)
	0.10	0.10	1.62(2.58)	4.75(9.93)	94.4	61.5(21.0)	77.1(18.0)
	0.50	0.01	1.13(2.32)	−4.21(7.65)	95.3	38.7(29.3)	−39.6(37.1)
	0.50	0.05	1.19(2.36)	−4.30(8.01)	95.5	39.5(31.5)	−40.5(38.0)
	0.50	0.10	1.24(2.41)	−4.36(8.36)	95.4	40.1(33.3)	−41.6(38.5)
7 mm	0.05	0.01	2.17(2.95)	5.10(10.3)	94.4	74.4(23.5)	85.3(20.8)
	0.05	0.05	2.25(3.01)	5.18(10.7)	94.5	74.9(24.0)	86.1(21.3)
	0.05	0.10	2.24(3.02)	5.19(10.9)	95.8	75.5(24.6)	86.5(21.6)
	0.10	0.01	1.99(2.80)	4.93(9.77)	94.6	61.7(21.0)	83.8(17.9)
	0.10	0.05	2.06(2.81)	5.05(9.82)	94.4	62.5(22.3)	84.1(18.4)
	0.10	0.10	2.13(2.87)	4.08(9.91)	94.7	63.8(23.1)	85.0(19.6)
	0.50	0.01	1.42(2.60)	−4.56(8.93)	95.7	44.7(31.3)	−48.5(41.8)
	0.50	0.05	1.51(2.71)	−4.67(9.09)	95.4	45.6(32.0)	−49.7(42.7)
	0.50	0.10	1.73(2.75)	−4.73(9.15)	94.3	47.8(33.1)	−50.6(44.0)

[1]full width at half-maximum. In the imaging literature, this a a common way to describe the variability of an isotropic gaussian. If the common variance is σ, then FWHM $= 2\sqrt{2\ln 2}\sigma$.

[2,3,5,6]$\times 10^{-2}$. That is, all numbers are to be multiplied by .01

[4]Percentage of time that the 95% HPD interval of the posterior draws of KLD covers the truth.

Table 2: Sensitivity analysis of different hyperprior distributions for λ in the proposed model. $\mathrm{CCR_{CV}}$ denotes the leave-one-out cross-validated classification rate.

Prior of λ	Prior mean	Prior variance	$\mathrm{CCR_{CV}}$
Gamma(0.6, 0.2)	3.0	15.0	0.766
Gamma(0.8, 0.2)	4.0	20.0	0.787
Gamma(1.0, 0.2)	5.0	25.0	0.787
Gamma(1.2, 0.2)	6.0	30.0	0.787
Gamma(1.4, 0.2)	7.0	35.0	0.787
Gamma(1.8, 0.2)	8.0	45.0	0.766
Gamma(0.5, 0.1)	5.0	50.0	0.787
Gamma(2, 0.2)	10.0	50.0	0.766
Gamma(2, 0.4)	5.0	12.5	0.787
U [0, 10]	5.0	8.3	0.766