Predicting Treatment Efficacy via Quantitative MRI: A Bayesian Joint Model

Jincao Wu*
University of Michigan
Department of Biostatistics
School of Public Health

Timothy D. Johnson
University of Michigan
Department of Biostatistics
School of Public Health

The prognosis for patients with high-grade gliomas is poor, with a median survival of one year. Treatment efficacy assessment is typically unavailable until 5–6 months post diagnosis. Investigators hypothesize that quantitative MRI (qMRI) can assess treatment efficacy three weeks after therapy starts, thereby allowing salvage treatments to begin earlier. The purpose of this work is to build a predictive model of treatment efficacy using qMRI data and to assess its performance. The outcome is one-year survival status. We propose a joint, two-stage Bayesian model. In stage I, we smooth the image data with a multivariate spatio-temporal pairwise difference prior. We propose four summary statistics that are functionals of posterior parameters from the first stage model. In stage II, these statistics enter a generalized non-linear model (GNLM) as predictors of survival status. We use the probit link and a multivariate adaptive regression spline basis. Gibbs sampling and reversible jump Markov chain monte carlo are applied iteratively between the two stages to estimate the posterior distribution. Through both simulation studies and model performance comparisons we find that we are able to attain higher overall correct classification rates by accounting for the spatio-temporal correlation in the images and by allowing for a more complex and flexible decision boundary provided by the GNLM.

KEY WORDS: Bayesian analysis; image analysis; multivariate adaptive regression splines; multivariate pairwise difference prior; quantitative MRI; spatio-temporal model.

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

1 Introduction

Our work is motivated by a need to appropriately analyze data collected from quantitative magnetic resonance imaging (qMRI) studies, and to determine whether qMRI can be used as an early predictor of treatment efficacy as measured by survival for patients with malignant gliomas. The data come from a pilot study of 53 high-grade glioma patients (Hamstra et al. (2005)). The prognosis for patients with the high-grade gliomas is poor. The mortality rate, at the time of data collection, is high with a median survival of one year after diagnosis (Laws et al. (2003)). Treatment is a combined approach of surgery (if possible), radiation therapy followed by chemotherapy. Assessment of treatment efficacy is based on radiological response approximately 8–10 weeks post therapy, or approximately five to six months after diagnosis (Moffat et al. (2005) and Hamstra et al. (2008)). Radiological response is determined by the change in tumor size from baseline as measured on anatomical MR images. For those with progressive disease, salvage therapy is given. However, it is typically too late for the salvage therapy to have any effect in prolonging survival (Moffat et al. (2005)). If treatment efficacy can be assessed earlier, salvage therapies can begin earlier or therapy can be modified.

In the pilot study, two different qMRI studies (diffusion and perfusion) and standard anatomical MRI studies were conducted at each of two time points: baseline (one week before therapy) and three weeks after therapy begins. All four quantitative images were registered to the pre-treatment anatomical MRI via a mutual information algorithm (Meyer et al. (1997), i.e., an affine translation and rotation). Full imaging data was available on 47 of the 53 patients, therefore we analyze the data from these 47 patients. Tumors were identified on contrast-enhanced T1-weighted MR images at both time points and segmented (outlined) by a radiologist. We use the intersection of the segmented tumors as the region of interest (Hamstra et al. (2005), Moffat et al. (2005) and Hamstra et al. (2008)). Using

the intersection of the segmented tumors, as opposed to the union, avoids the potential comparison of tumor in one image with healthy tissue or edema in the other image that may occur in the symmetric difference of the segmented tumors due to small changes in tumor volume, swelling of tissue caused by therapy, and errors in segmentation.

The apparent diffusion coefficient (ADC) is a measure of the magnitude of Brownian motion of water molecules in the extracellular space of tissue (Hamstra et al. (2005), Moffat et al. (2005) and Hamstra et al. (2008)). Diffusion in biological systems is a complex phenomenon, influenced directly by tissue microstructure. Its measurement can provide information about the organization of this structure in normal and diseased tissue (Basser and Jones (2002)). As tumor cells lyse, the ratio of extracellular to intracellular fluid increases thus causing a temporary increase in ADC (Moffat et al. (2005) and Moffat et al. (2006)). Perfusion is a measure of tissue-specific blood flow and blood volume and reflects the delivery of essential nutrients to tissue (Galbán et al. (2009)). It is hypothesized that effective therapy will disrupt tumor blood supply by damaging tumor neovascularity, resulting in decreased tumor perfusion. Furthermore, recent studies have suggested that qMRI can be used for early prediction of therapeutic efficacy. Early changes detected in mean tumor ADC values were first found to be correlated with treatment response in rodent tumor models (Ross et al. (1994), Zhao et al. (1996) and Chinnaiyan et al. (2000)). Previous studies investigating perfusion MRI for tumor diagnosis and response monitoring, relied on the whole-tumor mean value as the summary statistic of the perfusion maps for quantification of hemodynamic parameters, with varying success (Young et al. (2007) and Law et al. (2007)). The functional diffusion map (fDM), a voxel-by-voxel approach, was recently reported as an early, quantitative biomarker for clinical brain tumor treatment outcome (Hamstra et al. (2005), Moffat et al. (2005) and Hamstra et al. (2008)). Galbán et al. (2009) have also shown that the functional perfusion map (fPM) based on perfusion MRI (obtained in the same way

as fDM) is predictive of overall survival. However, both the fDM and fPM treat voxels as independent observations thus ignoring spatial structure in the images. Treating the data as independent observations may lead to incorrect variance estimates and invalid inference. Our work is motivated by all of these studies and aims to build a statistically robust and predictive model for treatment efficacy based on both the ADC and rCBF (relative cerebral blood flow, a measure of perfusion, Galbán et al. (2009)) images. An axial slice of a registered ADC image, a rCBF and a T1-weighted, contrast enhanced MR image are shown in Figure 1.

We propose a joint, two-stage Bayesian predictive model. In the first stage, we smooth the images (two images at each of two time points) using a multivariate pairwise difference prior (mPWDP) that models the spatio-temporal correlation in the images. The pairwise difference prior (PWDP) was first introduced by Besag (1993). It is a member of the class of pairwise interaction Markov random field models and captures general and local characteristics of the image. A priori, it assumes that the mean values of neighboring voxels are positively correlated. We extend the PWDP to the multivariate setting. We then propose four summary statistics that are functionals of the parameters in stage I. The statistics enter the second stage model as predictors of one-year survival status. The second stage model is a generalized nonlinear model (GNLM) proposed by Holmes and Denison (2003). The GNLM uses a probit link, for computational efficiency, and a Bayesian multivariate adaptive regression spline (BMARS) basis. The MARS model was introduced by Friedman (1991). The BMARS basis allows the predictors to enter the GNLM model nonlinearly; thus allowing for a very flexible decision boundary. The two models are fitted jointly and the model is validated via cross-validated prediction. Algorithmically, the models are joined by iterating between the two stages in a generalized Markov chain Monte Carlo simulation (Metropolis-within-Gibbs updates in stage I and an hybrid reversible jump Markov chain Monte Carlo

(RJMCMC, MCMC) and Gibbs updates of hyperparameters in stage II).

Compared to current methods, our joint model has several new features and improvements. In the first stage, our model: 1) accounts for spatio-temporal correlation in the images, as well as the correlation between the ADC and rCBF images; 2) increases the signal to noise ratio by smoothing the images; and 3) reduces the data dimension via subject level summary statistics. In the second stage, our model allows for a more flexible classification boundary than that allowed by the standard linear systematic component of a GLM. The joint model we proposed propagates the sampling error from stage I into stage II model. We adopt the Bayesian paradigm for estimating and predicting outcomes. Furthermore, model uncertainty is captured by model averaging.

This manuscript is organized as follows. In Section 2, we first outline, at a high level, our joint model, then specify the two stages of the model and propose our model evaluation strategy that we have adopted. The pilot study data are then analyzed in Section 3. We show our model outperforms simpler models in Section 4. The paper concludes with a discussion, summarizing the strengths and limitations of our approach. Due to manuscript length restriction, results from simulation studies and sensitivity analyses can be found in a supplementary web-based materials (SWBM) document available at `http://www.bepress.com/umichbiostat/paper83` or by contacting the authors. Detailed mathematical derivations of the posterior distributions, algorithm details and pseudo code are also provided in the SWBM.

2 Bayesian Joint Model

To begin, we briefly describe the joint model. Let $\mathcal{Y}$ denote the set of all images for all subjects and let $\mathbf{Z}$ denote the 1-year survival status (1-dead,0-alive) vector. Let $\boldsymbol{\Omega} = \boldsymbol{\Omega}_1 \cup \boldsymbol{\Omega}_2$ denote the set of all model parameters where $\boldsymbol{\Omega}_1$ is the set of stage I model parameters and

predictive values and $\mathbf{\Omega}_2$ is the set of stage II model parameters. We further note that the set (over all subjects) of all summary statistics, $\mathcal{X}$, calculated in stage I is a functional vector of $\mathbf{\Omega}_1$ and that $\mathbf{Z}$ depends on $\mathbf{\Omega}_1$ only through $\mathcal{X} = F(\mathbf{\Omega}_1)$. The posterior distribution can be factored as follows:

$$\pi(\mathbf{\Omega} \mid \mathcal{Y}, \mathbf{Z}) \propto \pi(\mathcal{Y} \mid \mathbf{\Omega}_1)\pi(\mathbf{\Omega}_1)\pi(\mathbf{Z} \mid \mathbf{\Omega}_1, \mathbf{\Omega}_2)\pi(\mathbf{\Omega}_2). \tag{1}$$

We will use $\pi(\mathbf{Z} \mid \mathbf{\Omega}_1, \mathbf{\Omega}_2)$, $\pi(\mathbf{Z} \mid \mathbf{\Omega})$ and $\pi(\mathbf{Z} \mid \mathcal{X}, \mathbf{\Omega}_2)$ interchangeably depending on the context. We draw from the posterior (1) via Markov chain Monte Carlo (MCMC) simulation by iteratively drawing between the full conditional distribution of $\mathbf{\Omega}_1$:

$$\pi(\mathbf{\Omega}_1|\mathcal{Y}, \mathbf{Z}, \mathbf{\Omega}_2) \propto \pi(\mathcal{Y} \mid \mathbf{\Omega}_1)\pi(\mathbf{\Omega}_1)\pi(\mathbf{Z} \mid \mathbf{\Omega}_1, \mathbf{\Omega}_2) \tag{2}$$

and the full conditional distribution of $\mathbf{\Omega}_2$:

$$\pi(\mathbf{\Omega}_2|\mathbf{Z}, \mathbf{\Omega}_1) \propto \pi(\mathbf{Z} \mid \mathbf{\Omega}_1, \mathbf{\Omega}_2)\pi(\mathbf{\Omega}_2). \tag{3}$$

The full conditionals in (2) and (3) are easily derived from (1) and repeated use of Bayes' theorem.

The remainder of this section is broken up into four subsections. In Subsection 2.1 we define the mPWDP model, in Subsection 2.2 we define the GNLM and then in Subsection 2.3 we give an overview of how we sample from the joint posterior distribution specified in (1). The last subsection, 2.4, we describe how we evaluate our model.

2.1 Stage I

In this subsection, patient subscripts are suppressed to reduce notational burden. Tumor voxels (short for volume element—a cube) are indexed by $i = 1, 2, \ldots, n$, where the tumor size n (i.e. the size of the tumor ROI defined in the introduction section) ranges from 770 to 20380 voxels with a mean of 6143 and standard deviation of 4721. Two voxels, i and i',

that share a common face are called neighbors, denoted by $i \sim i'$. Let $N_i = \{i' : i' \sim i\}$ denote the set of neighbors of voxel i with $|N_i|$ denoting the number of neighbors. Let Y_{ith} represent the image intensity at voxel i, time $t = 1, 2$ (baseline and week 3, respectively) and image type h ($h = 1$—diffusion, $h = 2$—perfusion). The vector of image intensities at voxel i is $\mathbf{Y}_i = (Y_{i11}, Y_{i12}, Y_{i21}, Y_{i22})^{\mathrm{T}}$. We split $\mathbf{Y}_i$ into two sub-vectors by time: $\mathbf{Y}_{it} = (Y_{it1}, Y_{it2})^{\mathrm{T}}$. Furthermore, let $\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \ldots, \mathbf{Y}_n^{\mathrm{T}})^{\mathrm{T}}$. Each Y_{ith} is measured with error with mean μ_{ith}. Let $\boldsymbol{\mu}_i = (\mu_{i11}, \mu_{i12}, \mu_{i21}, \mu_{i22})^{\mathrm{T}}$ with corresponding subvectors $\boldsymbol{\mu}_{it} = (\mu_{it1}, \mu_{it2})^{\mathrm{T}}$. Let $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^{\mathrm{T}}, \ldots, \boldsymbol{\mu}_n^{\mathrm{T}})^{\mathrm{T}}$. Note that the components in $\mathbf{Y}_i$ are correlated with covariance Σ.

We extend Besag's (1993) PWDP model to the multivariate setting. First,

$$[\mathbf{Y} \mid \boldsymbol{\mu}, \Sigma^*] \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma^*),$$

where $\Sigma^* = \mathrm{diag}(\Sigma)$—a block diagonal matrix with Σ along the main diagonal. The prior distribution of the mean vector $\boldsymbol{\mu}$ is

$$\pi(\boldsymbol{\mu} \mid \Psi) \propto \exp\left\{-0.5 \sum_{i \sim i'} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i'})^{\mathrm{T}} \Psi^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i'})\right\}.$$

Spatial correlation in the image is modeled through the diagonal elements of the 4×4 covariance matrix Ψ. The off diagonal elements of Ψ account for temporal correlation within an image type, correlation between image types at a particular time and correlation over time and across image types. The covariance matrix Σ accounts for residual covariances.

A priori, Σ and Ψ are assigned inverse Wishart distributions: $\mathrm{W}^{-1}(\mathrm{I}_4, 5)$. The scale matrix I_4 is the 4×4 identity matrix and the degrees of freedom is 5. The degrees of freedom can be regarded as the a priori sample size. Given the large n, this results in a rather weak prior.

Predicting tumor response under the "null": Ideally we would compare the observed tumor response to its counterfactual: tumor response given no treatment. Given that this is impossible, our summary statistics will be based on comparing the observed tumor

response to the predicted tumor response in the contralateral hemisphere of the brain under the assumption that the change in ADC/rCBF values in healthy tissue in the contralateral brain and those of tumor in the contralateral brain, if they could be observed, are similar. In the contralateral brain the healthy tissue receives a low dose of radiation and little damage from chemotherapy due to the blood-brain barrier which blocks large chemotherapy molecules. Thus, the healthy tissue in the contralateral brain is protected from therapy and diffusion and perfusion are stable over the short time period between imaging sessions. We define a healthy tissue region of interest (ROI) in the contralateral brain. The healthy tissue ROI is obtained by reflecting the tumor ROI, approximately about the midline of the brain, to the contralateral hemisphere. We then ensure, visually, that the healthy tissue ROI lies within the gray matter of the brain (some white matter is fine). If the healthy tissue ROI intersects the ventricles, meninges or skull, we manually shift the ROI, to avoid this overlap (details can be found in the SWBM). We now describe how we predict tumor response in the healthy tissue ROI, which we refer to as the *null response*.

First we build a mPWDP model for the healthy tissue data in the healthy tissue ROI. The model is identical to that described above with the following notational changes. For healthy tissue, in the healthy tissue ROI, let $\mathbf{W}_i$ denote the image intensities for voxel i with mean vector $\boldsymbol{\nu}_i$. The covariance of the $\mathbf{W}_i$ will be denoted Δ and the covariance of the mean vector $\boldsymbol{\nu}_i$ will be denoted Ω. The number of voxels in the healthy tissue ROI is also n. Denote the set of voxels in the healthy ROI by $\mathcal{H}$. We extend the healthy tissue ROI by a one-voxel thick shell and denote the set of voxels in this shell by $\mathcal{S}$. Without this extension, $\widetilde{\mathbf{Y}}_{i2}$ and $\widetilde{\boldsymbol{\mu}}_{i2}$ are not identifiable (see equations (4) and (5) below and Section 5 in the SWBM). Let n^s denote the number of voxels in the shell and let $n^e = n + n^s$ be the number in the extended ROI. Let $N_i^e = \{i' : i' \sim i\}$ denote the set of neighbors of voxel i in the extended ROI and $|N_i^e|$ denote the number in this set.

Now to predict tumor null response translate the tumor baseline values $\mathbf{Y}_{i1}$ to the healthy tissue ROI, using the same reflection and shift that created the healthy tissue ROI. We partition the 4×4 covariance matrices into 4, 2×2 matrices. The mPWDP for prediction is

$$\left[\begin{pmatrix} \mathbf{Y}_{i1} \\ \widetilde{\mathbf{Y}}_{i2} \end{pmatrix} \mid \begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix}, \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{pmatrix} \right] \sim \mathrm{N} \left[\begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix}, \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{pmatrix} \right] \tag{4}$$

where $\widetilde{\mathbf{Y}}_{i2} = \left(\widetilde{Y}_{i21}, \widetilde{Y}_{i22} \right)^{\mathrm{T}}$ is the predicted null response at time point 2 and $\widetilde{\boldsymbol{\mu}}_{i2}$ is its mean. The $\boldsymbol{\mu}_{i1}$ are obtained from the posterior distribution of the tumor mPWDP model and the covariances from the healthy tissue mPWDP models. Let

$$\begin{pmatrix} \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix} = |N^e_i|^{-1} \left[\sum_{i' \in N^e_i \cap \mathcal{S}} \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\nu}_{i'2} \end{pmatrix} + \sum_{i' \in N^e_i \cap \mathcal{H}} \begin{pmatrix} \boldsymbol{\mu}_{i'1} \\ \widetilde{\boldsymbol{\mu}}_{i'2} \end{pmatrix} \right]. \tag{5}$$

The prior for the mean vector in (4) is

$$\left[\begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix} \mid \begin{pmatrix} \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{i22} \end{pmatrix} \right] \sim \mathrm{N} \left[\begin{pmatrix} \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix}, |N^e_i|^{-1} \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \right] \tag{6}$$

where the covariances are obtained from the posterior of the healthy tissue mPWDP model. The covariances taken from the posterior of the healthy tissue mPWDP model describe the spatio-temporal relationship between the baseline tumor ADC/rCBF values and the predicted values under our assumption that tumor changes would be similar to healthy tissue changes in the environment of the contralateral hemisphere. We need to ensure that 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. If they are much different, the inequality in the baseline covariances may result in biased predictions. One may be tempted to replace Δ_{11} with Σ_{11} in (4) and Ω_{11} with Ψ_{11} in (6), however, there is no guarantee that the resulting covariance matrices would be positive definite. After fitting our model to the data we investigated whether these assumptions hold by comparing the posterior expected values of these leading sub-matrices. To compare them, 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 (see details in the SWBM, section 1.1). The relative root mean squared difference between the leading 2×2 sub-matrices of Δ and Σ (relative to Δ) is 0.038 (sd = 0.029) and that between the leading sub-matrices of Ω and Ψ (relative to Ω) is 0.039 (sd = 0.018)—both small relative differences—hence we feel that this assumption is justified in our model.

Now we can explicitly define the stage I parameter set $\mathbf{\Omega}_1$ and, at the same time, add a subject specific index, j. Gather all parameters and predictive values into a set of parameters for subject j: $\mathbf{\Omega}_{1j} = \{\{\boldsymbol{\mu}_{i,j}\}_{i=1}^{n_j}, \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n_j^e}, \Sigma_j, \Psi_j, \Omega_j, \Delta_j, \{\widetilde{\mathbf{Y}}_{i2,j}\}_{i=1}^{n_j}, \{\widetilde{\boldsymbol{\mu}}_{i2_j}\}_{i=1}^{n_j}\}$. Then $\mathbf{\Omega}_1 = \cup_j \mathbf{\Omega}_{1j}$.

Summary Statistics: The summary statistics are based on comparing the observed tumor response with the predicted tumor response under the null. Previous work suggests that the mean change in tumor ADC values is not predictive of treatment efficacy in humans (Chenevert et al. (2000) and Moffat et al. (2005)). Empirically, however, the baseline tumor ADC (rCBF) histogram and the week 3 tumor ADC (rCBF) histogram are notably different. This gave us the idea to investigate whether the Kullback-Leibler divergence (Kullback and Leibler (1951)) between posterior/predictive draws of μ_{i2h} and $\widetilde{\mu}_{i2h}$, $h = 1, 2$, respectively, would be good predictors of treatment efficacy. Specifically, we draw μ_{i2h} from its full conditional posterior for all i in the tumor ROI and create a histogram and draw $\widetilde{\mu}_{i2h}$ from its full conditional posterior for all i, create a histogram and then compute the Kullback-Leibler divergence between these two histograms (See details in SWBM, Section 1.2).

Hamstra et al. (2005), Moffat et al. (2005) and Hamstra et al. (2008) have demonstrated that fDM, a statistical approach for segmenting tumors into regions of response and non-response, based on a defined upper threshold of ADC change following therapy, is a

good biomarker for predicting early tumor response to therapy (this threshold is basically an upper confidence limit of the regression slope of the week 3 tumor ADC values regressed on the baseline ADC values). The fDM approach is based on the rationale that early ADC changes due to therapy are heterogeneous within the tumor. Parts of the tumor respond to therapy and show an increase in ADC, while other regions show no change or even a decrease in ADC. However, successful therapy should result in tumor cells lysing with a corresponding increase in ADC, thus the rationale for defining an upper threshold. Furthermore, a successful treatment should result in a decrease in rCBF, as discussed in the introduction. However, again, tumor response is heterogeneous and the mean change is minimal, whereas changes in the tails of the distribution are more pronounced. Inspired by fDM, we sought statistics that summarize the proportion of extreme expected values, $\boldsymbol{\mu}_{i2}$, in the tumor response relative to the conditional distribution (SWBM, formula (29)) of means of predicted null tumor voxel responses. We propose two additional summary statistics: the conditional diffusion statistic (cDS) and the conditional perfusion statistic (cPS). The first, cDS, is defined as the proportion of tumor voxels that have a mean response that is greater than the 0.975 quantile of the conditional distribution (SWBM, formula (29)) of the same voxel under the null assumption: $\text{cDS} = n^{-1} \sum_{i=1}^{n} I\left[\mu_{i21} > q_{0.975}\left(\widetilde{\mu}_{i21}\right)\right]$, where $I[\cdot]$ is the indicator function and $q_{0.975}\left(\widetilde{\mu}_{i21}\right)$ is the 0.975 quantile of the conditional posterior distribution of $\widetilde{\mu}_{i21}$. The summary measure cPS is similarly defined: $\text{cPS} = n^{-1} \sum_{i=1}^{n} I\left[\mu_{i22} < q_{0.025}\left(\widetilde{\mu}_{i22}\right)\right]$, where $q_{0.025}\left(\widetilde{\mu}_{i22}\right)$ is the 0.025 quantile of the conditional posterior distribution of $\widetilde{\mu}_{i22}$.

2.2 Stage II

For stage II, we borrow the generalized non-linear model with a Bayesian MARS basis (GNLM-BMARS) proposed by Holmes and Denison (2003) to predict patients' one-year survival status. For patient j, let $\mathbf{X}_j = (X_{j1}, \ldots, X_{j4})^{\mathbf{T}}$ denote the vector of the summary

statistics obtained in stage I. Hence $\mathcal{X} = \cup_j \{\mathbf{X}_j\}_j$. Let Z_j index the survival status of patient j, with $Z_j = 1$, representing the death of patient j within one year, and $Z_j = 0$ otherwise, for $j = 1, \ldots, M$. Set $\mathbf{Z} = (Z_1, \ldots, Z_M)$. The set of all GNLM-BMARS parameters, $\mathbf{\Omega}_2$, will now be subscripted by K, the number of BMARS bases, as the number of bases is treated as a parameter to be estimated and the number of parameters in $\mathbf{\Omega}_{2K}$ depends on K. All parameters in $\mathbf{\Omega}_{2K}$ will be defined shortly. The GNLM-BMARS model with K bases functions is:

$$\begin{aligned} \pi(Z_j = 1 \mid \mathbf{X}_j, \mathbf{\Omega}_{2K}) &= g(\eta_{jK}), \qquad \eta_{jK} = \sum_{k=0}^{K} \beta_k B_k(\mathbf{X}_j), \\ B_k(\mathbf{X}_j) &= \begin{cases} 1, & k = 0, \\ \prod_{l=1}^{L_k} [s_{lk}(X_{jw_{lk}} - t_{lk})]_+, & k = 1, 2, \ldots, K. \end{cases} \end{aligned} \tag{7}$$

The link function g could be the cumulative distribution function (CDF) from any of the commonly used distributions for modeling binary data such as the logistic, normal or extreme value distributions. Due to the flexibility in the decision boundary afforded by the BMARS basis, we argue that the choice of link function is not crucial. Thus, for computational efficiency and simplicity, we use the probit link function, $g(\cdot) = \Phi(\cdot)$, where Φ is the standard normal CDF. The function $[\,\cdot\,]_+ = \max(0\ ,\ \cdot\ )$. K is the number of basis functions in the model. L_k is the degree of interaction in basis function $B_k(\,\cdot\,)$. For our application, we set the highest order of interaction to 2. Thus, only main effects and two-way interactions are allowed to enter the model. Estimating higher order interactions with any certainty would require a large amount of data (sample size of our study is $M = 47$) due to the curse of dimensionality (Denison et al. (2002)). The variable s_{lk} is a sign indicator, taking values in $\{-1, 1\}$, t_{lk} is the location of the spline knot associated with the covariate indexed by $w_{lk} \in \{1, 2, 3, 4\}$. Further t_{lk} is restricted to the set of covariate values $\{X_{1w_{lk}}, \ldots, X_{Mw_{lk}}\}$ and all w_{lk} are distinct for each k (that is, each basis function is at most linear in any one

variable). Consult Holmes and Denison (2003) and Denison et al. (2002), Chapter 4 or the SWBM for further details. Let $\boldsymbol{\beta}_K = (\beta_0, \ldots, \beta_K)^{\mathrm{T}}$ where β_0 is the model intercept. Also, let $\mathbf{L}_K = \{L_1, \ldots, L_K\}$, $\mathbf{s}_K = \{s_{11}, \ldots, s_{L_K K}\}$, $\mathbf{w}_K = \{w_{11}, \ldots, w_{L_K K}\}, \mathbf{t}_K = \{t_{11}, \ldots, t_{L_K K}\}$ and $\Theta_K = \{K, \mathbf{s}_K, \mathbf{w}_K, \mathbf{t}_K, \mathbf{L}_K\}$. Then $\boldsymbol{\Omega}_{2K} = \Theta_K \cup \{\boldsymbol{\beta}_K\}$.

We specify non-informative prior distributions for all parameters

$$\begin{aligned}
\pi(L_k = 1) = \pi(L_k = 2) &= 1/2 \\
\pi\left(w_{1k} = w \mid L_K = 1\right) &= 1/4,\ w = 1, 2, 3, 4 \\
\pi\left[(w_{1k}, w_{2k}) = (w, w') \mid L_K = 2\right] &= 1/6,\ (w, w') = (1,2), (1,3), (1,4), (2,3), (2,4), (3,4) \\
\pi(t_{lk} = X_{jw_{lk}} \mid w_{lk}) &= 1/M,\ j = 1, \ldots, M \\
\pi(s_{lk} = -1) = \pi(s_{lk} = 1) &= 1/2 \\
[\boldsymbol{\beta}_K \mid v, K] &\sim \mathrm{N}(0, v\mathrm{I}_{K+1}), \quad \left[v^{-1}\right] \sim \mathrm{Gamma}(0.001, 0.001),
\end{aligned}$$

with one exception: $[K \mid \lambda] \sim \mathrm{Poisson}(\lambda)$ and $[\lambda] \sim \mathrm{Gamma}(1, 0.2)$. We assess the impact of this prior on classification results in the SWBM.

2.3 Sampling from the joint posterior

Now we outline how we sample from the joint posterior given in (1). For more details and derivations, please consult the SWBM.

We begin with the sampling of stage I parameters. We sample parameters $\boldsymbol{\mu}_{i,j}$, $\boldsymbol{\nu}_{i,j}$, $\widetilde{\boldsymbol{\mu}}_{i2,j}$, Σ_j, Ψ_j, Δ_j and Ω_j for $i = 1, \ldots, n_j$ and $j = 1, \ldots, M$ from their full conditional distributions via a hybrid Metropolis-within-Gibbs algorithm. Both, Ψ_j and Σ_j are drawn directly from their full conditionals (inverse Wishart distributions). The remaining parameters are drawn from their full conditionals via Metropolis-Hastings updates (Hastings (1970)). Full details are provided in the SWBM, Section 1. We note here that all parameters in a stand-alone MPWDP model can be updated by a Gibbs algorithm. However, due to the joint nature of

our full model, all stage 1 parameters other than Ψ_j and Σ_j are linked to stage II through the summary statistics and thus require Metropolis-within-Gibbs updates as the full conditionals no longer have a nice distributional form.

Now we outline our posterior sampling algorithm for stage II parameters. In Probit regression models, the posterior distribution can be simulated by a Metropolis-Hastings algorithm. However, to simplify computation, Albert and Chib (1993) derived a data augmentation scheme which relies on the latent variable model representation of a binary variable. This approach greatly simplifies sampling from the posterior distribution as the model is transformed from a Probit regression model into an equivalent linear model, thus the parameter vector $\boldsymbol{\beta}_K$ can be drawn from its full conditional as opposed to the Metropolis-Hastings algorithm.

We introduce a continuous latent vector $\mathbf{d} = (d_1, \ldots, d_M)^{\mathrm{T}}$. Define the conditional distribution of Z_j given d_j by

$$\pi(Z_j = 1 \mid d_j) = 1 \quad \text{if} \quad d_j > 0, \quad \text{and} \quad = 0 \quad \text{if} \quad d_j \leq 0. \tag{8}$$

The full conditional distribution of d_j is straightforward to derive (SWBM and Holmes and Denison (2003)) and is

$$[d_j \mid Z_j = z_j, \mathbf{X}_j, \boldsymbol{\Omega}_{2K}] \sim \begin{cases} \mathrm{N}(\eta_{jK}, 1) \text{ truncated at the left by 0 if } z_j = 1 \\ \mathrm{N}(\eta_{jK}, 1) \text{ truncated at the right by 0 if } z_j = 0. \end{cases} \tag{9}$$

We draw d_j, $j = 1, \ldots, M$ from (9). We then draw $\boldsymbol{\beta}_K$ from its full conditional distribution: $[\boldsymbol{\beta}_K \mid \mathbf{d}, v, \Theta_K, \mathcal{X}] \sim \mathrm{N}(\mathbf{m}_K^*, V_K^*)$, where $V_K^* = [(v\mathrm{I}_{K+1})^{-1} + \mathrm{B}_K^{\mathrm{T}}\mathrm{B}_K]^{-1}$ and $\mathbf{m}_K^* = V_K^*\mathrm{B}_K^{\mathrm{T}}\mathbf{d}$. Standard conjugacy results state that the full conditional distribution of v^{-1} is $[v^{-1} \mid \boldsymbol{\beta}_K, K] \sim \mathrm{Gamma}[0.001 + 0.5(K+1), 0.001 + 0.5\boldsymbol{\beta}_K^{\mathrm{T}}\boldsymbol{\beta}_K]$.

All parameters contained in Θ_K, are updated via the reversible jump MCMC algorithm (Green (1995)). Since K is random, the dimension of Θ_K varies as well as the column dimension of the matrix of BMARS bases, B_K, and the dimension of the vector $\boldsymbol{\beta}_K$. At each

iteration of the algorithm, we randomly (with equal probability) choose to add a new basis function (birth step) or to remove one of the existing basis functions (death step). Thus, covariates (summary statistics) and any two-way interactions enter the model via these birth and death steps. Details of the RJMCMC algorithm and pseudo code for sampling from the posterior distribution of our joint model are given in the SWBM.

2.4 Model Evaluation

The traditional way to evaluate classification models is by randomly partitioning the data into a training set for model building and a test set for model evaluation. However, due to the small sample size in our data set, we evaluate our proposed joint model via cross-validation. To implement cross-validation, a straightforward, but computationally expensive, approach is to run the algorithm multiple times with one observation left out each time. Instead, we adopt the importance sampling approach proposed by Gelfand et al. (1992) whereby one need only estimate the posterior distribution of the parameters given the full dataset and then by importance sampling compute the predictive probability that $Z_j = 1$ given $\mathbf{Z}_{\{-j\}}$ and $\mathcal{Y}$ for subject j where $\mathbf{Z}_{\{-j\}}$ denotes all observations except that of subject j. Let $\mathbf{\Omega}_2^{(t)}$ denote the value of $\mathbf{\Omega}_{2K}$ from the t^{th} draw from the posterior and that of $\mathbf{\Omega}_1$ by $\mathbf{\Omega}_1^{(t)}$. The cross-validated posterior predictive probability is estimated by MCMC output (see the SWBM) and is given by:

$$\pi\left(Z_j = 1 \mid \mathbf{Z}_{\{-j\}}, \mathcal{Y}\right) = \frac{\sum_{t=1}^{T} \pi\left(Z_j = 1 \mid \mathbf{\Omega}_{1j}^{(t)}, \mathbf{\Omega}_2^{(t)}\right) / \pi\left(Z_j = z_j \mid \mathbf{\Omega}_{1j}^{(t)}, \mathbf{\Omega}_2^{(t)}\right)}{\sum_{t=1}^{T} 1 / \pi\left(Z_j = z_j \mid \mathbf{\Omega}_{1j}^{(t)}, \mathbf{\Omega}_2^{(t)}\right)}, \tag{10}$$

where z_j is the observed value of Z_j. We assume that the losses incurred by a false negative and a false positive prediction are equal. Thus, if $\pi\left(Z_j = 1 \mid \mathbf{Z}_{\{-j\}}, \mathcal{Y}\right) \geq 0.5$, then the cross-validated prediction of $Z_j = 1$ and 0 otherwise.

Although not part of model evaluation, here is a good place to discuss the predictive

decision boundary. Theoretically, one could use the predictive distribution, $\pi(Z_{\text{new}} = 1 \mid \mathcal{Y}_{\text{new}}, \mathbf{Z}, \mathcal{Y})$, to define the decision boundary by varying $\mathcal{Y}_{\text{new}}$ over the space of all images. Obviously this is too daunting a task. Instead we will define the conditional predictive decision boundary in terms of the summary statistics. This decision boundary is a hypersurface in $\mathbb{R}^4$—the covariate space. It is defined as all solutions, $\mathbf{X}_{\text{new}}$, to the posterior predictive probability

$$1/2 = \pi(Z_{\text{new}} = 1 \mid \mathbf{Z}, \mathbf{X}_{\text{new}}) = \int \pi(Z_{\text{new}} = 1 \mid \mathbf{X}_{\text{new}}, \mathbf{\Omega}_{2K})\pi(\mathbf{\Omega}_{2K} \mid \mathbf{Z}, \mathbf{\Omega}_1)\mathrm{d}\mathbf{\Omega}_{2K}.$$

We are not able to visualize this decision boundary either as the dimension is 4. Therefore, to visualize the decision boundary, we will marginalize over pairs of covariates and plot the marginal predictive probability map as a function of the remaining pair of covariates by discretizing the marginal covariate space into a grid of values. The marginal decision boundary, then, is a curve in 2-dimensional space (see Figure 2).

We note here, that at each iteration the number of BMARS basis may change, thus implicit in the estimation of the cross-validated predictive probability and in building the marginal probability maps we average over all potential BMARS models. By doing so, we account for model uncertainty in our results along with the uncertainty in the model parameters, and thus inductively, the uncertainty in the covariates $\mathcal{X}$ (Raftery et al. (1996)).

3 Results

Stage I is computationally much more expensive than stage II due to the large number of voxels, n, in each patient's tumor. We run the algorithm (stage I and stage II combined) for 100,000 iterations. In each iteration, we over-sample (10:1) draws from the posterior of the stage II parameters. Stage I takes around 20 hours for all 47 patients, whereas stage II takes 5 minutes. The algorithm is programmed in C and implemented on a 3.0 GHz Mac

Xserve. The first 50,000 draws are discarded as burn-in. By visual inspection of the trace plots of the (fixed dimension) parameters, the burn-in is sufficient and the chain is sampling from the posterior (stationary) distribution.

We calculate the cross-validated correct classification rate, CCR_{CV}—the proportion of correctly predicted survival statuses. The positive predictive value, PPV_{CV}—the probability of death within one year conditional on prediction of death within one year. And, the negative predictive value, NPV_{CV}—the probability of survival greater than one year given a prediction of survival greater than one year. The results are: $\text{CCR}_{\text{CV}} = 0.787$ (37/47); $\text{PPV}_{\text{CV}} = 0.813$ (13/16); and $\text{NPV}_{\text{CV}} = 0.774$ (24/31) (Table 1, row 1). Investigators are interested in therapy intervention or modification if the model accurately predicts death within one year. Therefore, the PPV_{CV} is of greater interest than the NPV_{CV}.

In Figure 2 we display the six bivariate marginal predictive probability maps. On each map is the marginal decision boundary separating the space of covariates into two regions based on whether $\pi(Z_{\text{new}} = 1 \mid \mathbf{Z}, \mathbf{X}_{\text{new}}) > 0.5$. Also shown in the figure are the posterior means of the covariates for all 47 subjects. The triangles represent those subjects who died before one year, and the circles represent those who lived greater than one year. The probabilities in the maps are $\pi(Z_{\text{new}} = 1 \mid \mathbf{Z}, \mathbf{X}_{\text{new}})$ marginalized over the six combinations of pairs of covariates. It is evident that the marginal decision boundaries are quite complex. From Figure 2 we see that, marginally, small values of dKLD and cDS are associated with poor survival and that large values of pKLD are also associated with poor survival. There are also substantial interactions between cPS and dKLD, between dKLD and cDS and between pKLD and dKLD. In general, the overall gross pattern of increases in the dKLD and cDS statistics are predictive of longer survival—consonant with what our colleagues hypothesized. However, the overall gross pattern of decreases in the pKLD and cPS statistics are predictive of shorter survival—dissonant with that hypothesized. One plausible explanation provided

by our colleagues is that a reduction in rCBF creates an hypoxic environment within the tumor and hypoxia is known to be protective against radiation damage. However, we caution that the exact mechanism is unknown and that it warrants further investigation (Galbán et al. (2009)).

Each of the four summary statistics were included in the joint model as either a main effect or as an interaction term a high percentage of the MCMC draws (dKLD, 95.9%; pKLD, 90.5%; cDS, 81.1% and cPS, 84.3%). This indicates their importance in predicting survival. Both dKLD and pKLD appear to be slightly stronger predictors than either cDS or cPS based on the amount of time spent in the model.

The baseline prognostic factors age, surgery type, Karnofsky performance score, Pathology grade, tumor size were also included in stage II as covariates. However, their inclusion did not increase the overall correct classification rate and each was included in the model less than 20% of the time (either as main effects or in an interaction term).

4 Model Assessment

Comparison with simpler models: Our first comparison is with two separate models (i.e. not modeled jointly). The image data are fitted with our mPWDP model. The posterior means of the summary statistics are treated as fixed, known values and used as covariates in our GNLM-BMARS model. Thus the only difference between this procedure and our joint model is that our joint model accounts for the uncertainty in the summary statistics. Using point estimates, such as the posterior means, of the summary statistics as covariates in stage II results in overly optimistic prediction errors (see, e.g. Little and Rubin (2002)). Ignoring the sampling variability in stage I, two additional patients are correctly classified (Table 1, row 2). The posterior means of the summary statistics for these patients are near the decision boundary and happened to fall on the correct side, while the maximum a posteriori

probability (MAP) estimate was on the other side. Accounting for the variability in these random statistics, therefore, is necessary for robust prediction.

Our second comparison is again with two separate models. We estimate the posterior means of the summary statistics from our mPWDP model. These point estimates are then treated as fixed, known covariates and put into a standard Probit regression model (Table 1, row 3). Both main effects and interaction terms are allowed in the Probit regression model. BIC is used for model selection. Correct prediction from our joint model is much higher even though uncertainty in the covariates is ignored, as well as model uncertainty, in the separate mPWDP + Probit regression model. The extra flexibility afforded by the BMARS basis has a large effect on prediction.

Our final comparison illustrates the benefits of the spatio-temporal modeling in stage I by comparing our results to those based on the observed images. Since the cDS and cPS statistics rely on the conditional distribution of tumor response under the null, it is not possible to derive these summary statistics on the observed images as we have no model to use to predict tumor null response. Thus, this comparison uses only dKLD and pKLD. We estimate dKLD and pKLD using the observed images by calculating the KL divergence between observed tumor response and observed healthy tissue response (in the contralateral hemisphere) at week 3 and plugging these statistics into our GNLM-BMARS model as fixed covariates. Cross-validation results are shown in the bottom half of Table 1, rows 4 and 5. Spatio-temporal modeling results in higher CCR_{CV}, PPV_{CV} and NPV_{CV}.

Our overall conclusion from these comparisons is that joint modeling of the spatio-temporal structure in the images and the complexity in the decision boundary afforded by the covariates entering the GNLM model non-linearly and interacting in a complex manner is warranted for this data set. The images have complex structure and there is a complex relationship between the image based summary statistics and one-year survival. Furthermore,

accounting for the uncertainty in the summary statistics and model averaging are necessary for robust prediction.

Results from simulation studies and sensitivity analyses can be found in the SWBM.

5 Discussion

In this manuscript, we propose a Bayesian joint model to predict early treatment efficacy based on qMRI data from patients with high-grade gliomas. In stage I, we model the spatio-temporal structure in the qMRI data via a mPWDP model and derive summary statistics as functionals of the parameters in the posterior. In stage II, a GNLM is used to classify each patient's one-year survival status with the summary statistics derived in stage I as random predictors. The final predictive power is evaluated by cross-validation. Compared to previous work, our proposed joint model integrates many of the ideas that have been previously discussed. First, we extend the idea of the PWDP model to a multivariate setting, and, in fact, use it in a full spatio-temporal setting. The mPWDP model accounts for spatio-temporal correlation in the images as well as the correlation between the diffusion and perfusion images. This results in an increase in the signal to noise ratio. Furthermore, data dimension reduction is realized by defining subject level summary statistics. Second, by utilizing the BMARS basis, we allow a flexible and complex classification model that can achieve high predictive power. Finally, our model accounts for the uncertainty in the stage II covariates and for the uncertainty in model selection, resulting in more robust predictions.

In our analysis, we dichotomize each patient's survival status at one year. However, there may not be any substantial difference between a patient who dies 11 months after diagnosis and a patient who dies at 13 months. Moreover, the censoring rate in the data is about 30% with a median follow-up of 23.1 months and all censored observations are greater than one year. Censoring may also play a role in the evaluation of tumor treatment efficacy and

dichotomizing survival may lead to inefficient estimation. We are currently building a joint imaging/survival model, where, in stage II, we model the censored survival times explicitly.

We note here that we propose four summary statistics that capture information about the early changes in ADC and rCBF due to treatment. Results show that they perform well in terms of good prediction. We do not claim that these summary statistics capture the most, or even the best, information. Information is always lost in data reduction. Much more research is needed to determine how much data reduction is tolerable. Reduction to four summary statistics does a good job, but perhaps five or six would be better. We did not attempt to use more than four summary statistics due to the limited sample size in the pilot study. With larger samples size, less data reduction may be beneficial.

Our results show that early changes in diffusion and perfusion appear to be valuable biomarkers for the early assessment of treatment efficacy. These result are promising, albeit preliminary. The ability to predict treatment response during therapy, as opposed to waiting to assess traditional radiologic response, has the potential to facilitate patient management and may allow second line or salvage therapies to begin earlier than current practice dictates. Lastly, our model and sampling algorithm is easily extendable to more than 2 image types at more than 2 times points with more than four summary statistics.

Acknowledgements

The authors would like to thank Brian D. Ross, Department of Radiation Oncology, University of Michigan for providing us with the data. We also acknowledge Craig Galbán, Tom Chenevert, Chuch Meyer and Christina Tsien, who were all instrumental in the collection, reconstruction and registration of the imaging data. This work was funded by the U.S. NIH grant P01 CA085878.

References

Albert, J. H. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association 88*, 669–679.

Basser, P. J. and D. K. Jones (2002). Diffusion-tensor MRI: theory, experimental design and data analysis-a technical review. *NMR in Biomedicine 15*, 456–467.

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

Chenevert, T. L., L. D. Stegman, J. M. G. Taylor, P. L. Robertson, H. S. Greenberg, A. Rehemtulla, and B. D. Ross (2000). Diffusion magnetic resonance imaging: an early surrogate marker of therapeutic efficacy in brain tumors. *Journal of the National Cancer Institue 92*, 2029–2036.

Chinnaiyan, A. M., U. Prasad, S. Shankar, D. A. Hamstra, M. Shanaiah, T. L. Chenevert, B. D. Ross, and A. Rehemtulla (2000). Combined effect of tumor necrosis factor-related apoptosis-inducing ligand and ionizing radiation in breast cancer therapy. *Proceedings of the National Academy of Sciences 97*, 1754–1759.

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

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics 19*, 1–61.

Galbán, C. J., T. L. Chenevert, C. R. Meyer, C. Tsien, T. S. Lawrence, D. A. Hamstra, L. Junck, P. C. Sundgren, T. D. Johnson, D. J. Ross, et al. (2009). The parametric response map is an imaging biomarker for early cancer treatment outcome. *Nature medicine 15*, 572–576.

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.

Hamstra, D. A., T. L. Chenevert, B. A. Moffat, T. D. Johnson, C. R. Meyer, S. K. Mukherji, D. J. Quint, S. S. Gebarski, X. Fan, C. I. Tsien, T. S. Lawrence, L. Junck, A. Rehemtulla, and B. D. Ross (2005). Evaluation of the functional diffusion map as an early biomarker of time-to-progression and overall survival in high-grade glioma. *Proceedings of the National Academy of Sciences 102*, 16759–16764.

Hamstra, D. A., C. J. Galbn, C. R. Meyer, T. D. Johnson, P. C. Sundgren, C. Tsien, T. S. Lawrence, L. Junck, D. J. Ross, A. Rehemtulla, B. D. Ross, and T. L. Chenevert (2008). Functional diffusion map as an early imaging biomarker for high-grade glioma: correlation with conventional radiologic response and overall survival. *Journal of Clinical Oncology 26*, 3387–3394.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 97–109.

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

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics 22*, 79–86.

Law, M., R. Young, J. Babb, E. Pollack, and G. Johnson (2007). Histogram analysis versus region of interest analysis of dynamic susceptibility contrast perfusion MR imaging data in the grading of cerebral gliomas. *American Journal of Neuroradiology 28*, 761–766.

Laws, E. R., I. F. Parney, W. Huang, F. Anderson, A. M. Morris, A. Asher, K. O. Lillehei, M. Bernstein, H. Brem, A. Sloan, et al. (2003). Glioma Outcomes Investigators. Survival following surgery and prognostic factors for recently diagnosed malignant glioma: data from the Glioma Outcomes Project. *Journal of Neurosurgery 99*, 467–473.

Little, R. J. A. and D. B. Rubin (2002). *Statistical analysis with missing data*. Wiley-Interscience.

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.

Moffat, B. A., T. L. Chenevert, T. S. Lawrence, C. R. Meyer, T. D. Johnson, Q. Dong, C. Tsien, S. Mukherji, D. J. Quint, S. S. Gebarski, P. L. Robertson, L. R. Junck, A. Rehemtulla, and B. D. Ross (2005). Functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response. *Proceedings of the National Academy of Sciences 102*, 5524–5529.

Moffat, B. A., T. L. Chenevert, C. R. Meyer, P. E. McKeever, D. E. Hall, B. A. Hoff, T. D. Johnson, A. Rehemtulla, and B. D. Ross (2006). The functional diffusion map: an imaging biomarker for the early prediction of cancer treatment outcome. *Neoplasia 8*, 259–267.

Raftery, A. E., D. Madigan, and C. T. Volinsky (1996). Accounting for model uncertainty in survival analysis improves predictive performance. *Bayesian statistics 5*, 323–349.

Ross, B. D., T. L. Chenevert, B. Kim, and O. Ben-Yoseph (1994). Magnetic resonance imaging and spectroscopy: application to experimental neuro-oncology. *Quantitative Nuclear Magnetic Resonance in Medicine and Biology 1*, 89–106.

Young, R., J. Babb, M. Law, E. Pollack, and G. Johnson (2007). Comparison of region-of-interest analysis with three different histogram analysis methods in the determination of perfusion metrics in patients with brain gliomas. *Journal of Magnetic Resonance Imaging 26*, 1053–1063.

Zhao, M., J. G. Pipe, J. Bonnett, and J. L. Evelhoch (1996). Early detection of treatment response by diffusion-weighted 1 H-NMR spectroscopy in a murine tumour in vivo. *British Journal of Cancer 73*, 61–64.

Table 1: Model Comparisons. Upper part: comparisons based on all four statistics. First row: our proposed joint model. Second row: separate two-stage model (not joint). The second stage model (GNLM-BMARS) run conditional on the posterior expectations of the summary statistics from stage 1. Third row: separate two-stage model. The second stage model is a standard Probit regression model. Summary covariates are fixed at their posterior expectations from stage I. Bottom part: comparisons using only the two KLD statistics. Fourth row: our proposed model. Fifth row: Summary statistics computed on observed data. GNLM-BMARS runs conditional on "observed" dKLD and pKLD.

Model	[1]CCR_{cv}	[2]PPV_{cv}	[3]NPV_{cv}
Bayesian joint model	0.787 (37/47)	0.813 (13/16)	0.774 (24/31)
Separate models (two-stage model)	0.830 (39/47)	0.853 (15/18)	0.827 (24/29)
Separate models (stage I + Probit)	0.617 (29/47)	0.572 (11/20)	0.667 (18/27)
Bayesian joint model	0.723 (34/47)	0.733 (11/15)	0.719 (23/32)
Single model (stage II only)	0.638 (30/47)	0.600 (9/15)	0.656 (21/32)

[1] Correct cross-validated (CV) classification rate.
[2] Cross-validated positive predictive value.
[3] Cross-validated negative predictive value.

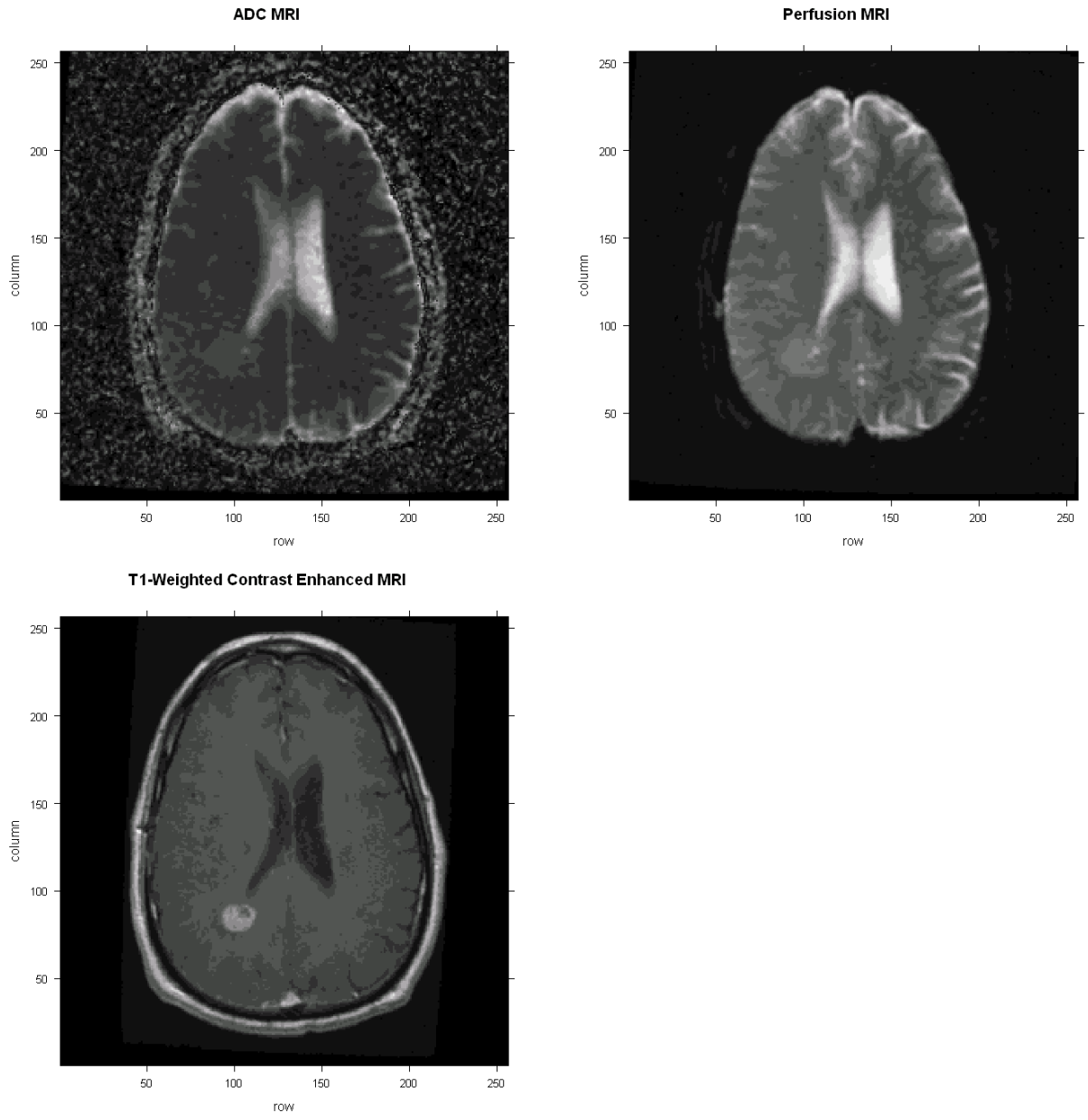

Figure 1: A single axial slice of pre-treatment MRI data. Upper left image: diffusion MRI; Upper right: perfusion MRI; Lower left: T1-weighted contrast enhanced MRI. The tumor is visible in all three images. It is located roughly at voxel (100,80) just below the left ventricle.

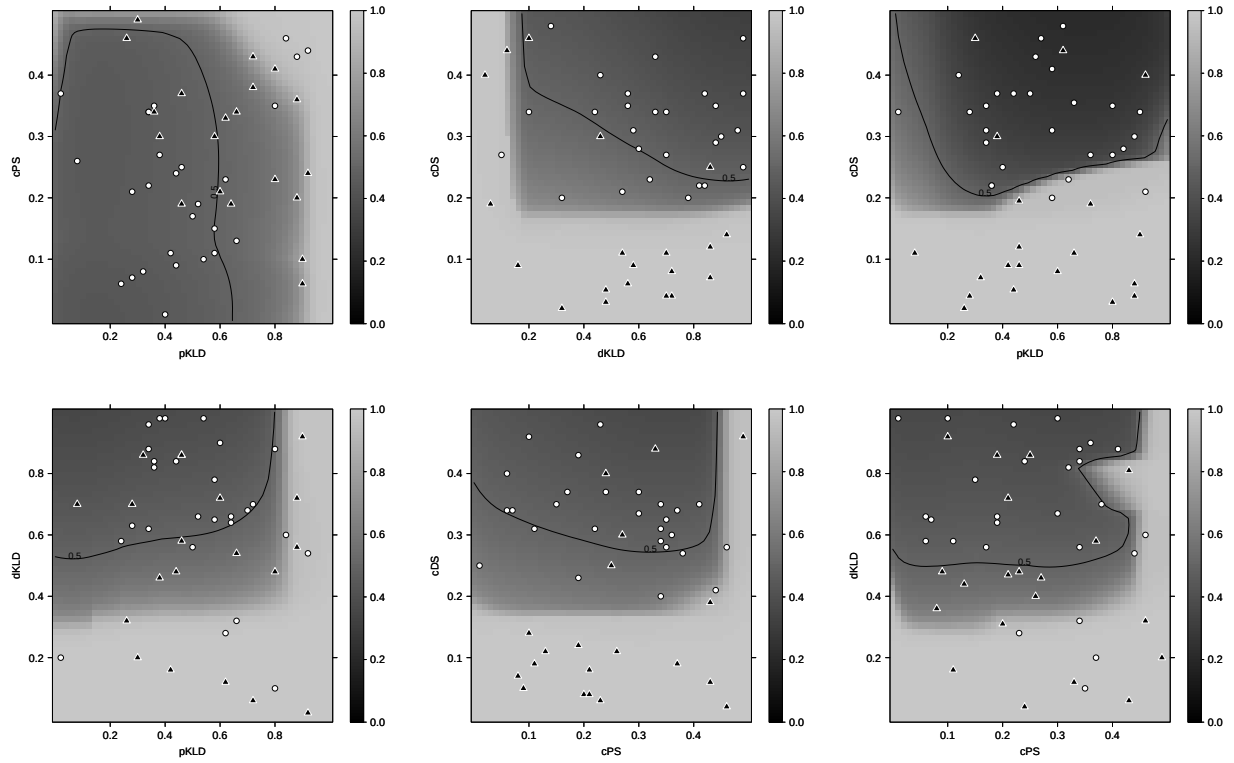

Figure 2: Marginal predictive probability maps from the glioma data. Each probability map is obtained by marginalizing over the two summary statistics not appearing in x and y labels of the images. The darker gray indicates smaller probability of patients' death before one year (i.e. $\pi(Z_j = 1 \mid \mathbf{Z}_{(-j)}, \mathcal{Y}_{(-j)}, \mathcal{Y}_j)$). The curved lines demarcate the marginal decision boundary (i.e. $\pi(Z_j = 1 \mid \mathbf{Z}_{(-j)}, \mathcal{Y}_{(-j)}, \mathcal{Y}_j) = 0.5$)—if $\pi(Z_j = 1 \mid \mathbf{Z}_{(-j)}, \mathcal{Y}_{(-j)}, \mathcal{Y}_j) > 0.5$, we predict death of that patient before one year. The symbols are located at the marginal posterior means of the statistics. The circles and triangles represent the true one year survival status for each patient. A circle indicates that the patient actually lived longer than one year. A triangle indicates that the patient died before one year.