Joint Bayesian Image and Prediction Modeling

by

Jincao Wu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2011

Doctoral Committee:

Research Associate Professor Timothy D. Johnson, Chair
Associate Professor Thomas Braun
Associate Professor Bin Nan
Professor Brian Ross

To Jun and my parents

ACKNOWLEDGEMENTS

I would like to gratefully and sincerely thank my advisor, Dr. Timothy D. Johnson for his patient guidance, understanding, and most importantly, for continuously supporting me with his profound insight, and constant encouragement. My appreciation also goes to my committee members: Dr. Bin Nan and Dr. Thomas Braun and Dr. Brian Ross for being on my dissertation committee and for the valuable advice on my dissertation. Without their help, the dissertation would not be successfully finished. I would also like to thank Dr. Brain Ross for kindly providing the data for my first two chapters.

I am grateful to all the knowledgeable faculty members who have advised me during my six-years of study at the University of Michigan, and to my fellow students in the Department of Biostatistics, from whom I have received many helpful suggestions. I have had very pleasant experiences working with colleagues at the Cancer Center and the Center for Molecular Imaging and have received their help and support. I thank them for the collegiality they have shown me.

Last but not the least, I would also like to express my gratitude to my parents and my husband Jun. Thanks to my parents for all they have done for me, for their strong support and endless love in my life. Thanks to my husband, Jun, for leading me to this exciting area and for always supporting me and helping me during all these years.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER I

Introduction

1.1 Background

Brain imaging allows scientists and doctors to view and monitor areas of the brain. In the past decade, brain imaging has played a prominent role in advancing our understanding of brain diseases and improving treatment management for patients. In MRI's infancy, information was largely anatomical, in the sense that relatively large structures would be observed. As MRI technology evolves, the imaging techniques are becoming more accurate with a finer resolution and more reliable with reproducible results. And imaging can move from a process of picture-taking, where reports are made on the basis of unusually bright, dark, small or large objects, to a process of quantitative measurements where a whole range of quantities can be tested to see whether they lie in a normal range, and whether they have changed from the time of a previous examination. The benefits of quantification are that fundamental research about biological changes in disease, and their response to potential treatments, can proceed in a more satisfactory way. We can also substantially reduce the problems of bias, reproducibility and interpretation.

In this dissertation, we focus on studying quantitative Magnetic Resonance Imaging (qMRI). Magnetic Resonance Imaging (MRI) is a type of non-invasive visu-

alization tool of the inside of living organisms, compared to other neuroimaging technologies (e.g. CAT, PET) that all require exposure to radiation, and are considered ethically problematic in children, particularly healthy children (Mason (2006)). Quantitative MRI measures the quantitative changes caused by the disease. It is increasingly used by scientists to study brain diseases and to measure treatment efficacy. In this dissertation, we studied three types of MRI: diffusion MRI, perfusion MRI and T1-weighted MRI. Diffusion and perfusion MRIs are two of the most commonly used quantitative MRI techniques and are evolving rapidly, with applications ranging from diagnosis of disease to the study of microvascular changes associated with functional cerebral activation (Galbán et al. (2009), Basser and Jones (2002) and Rao et al. (2007)). Diffusion MRI is based on the measurement of Brownian motion of water molecules. This technique can characterize water diffusion properties at each pixel/voxel of an image. Perfusion MRI provides a measurement of the parameters of cerebral micro-vascularisation. The main quantitative parameters measured are blood volume and blood flow. Both of these MRI techniques measure quantities that have direct physiological relevance. T1-weighted MRI produces an image with greater signal intensity from fat-containing tissue and can provide good gray matter/white matter contrast. It is mostly used to demonstrate pathological or other physiological alterations in the brain. In our study, we focus on the quantitative intensity measurements of this kind of MRI.

1.2 Bayesian Joint Model

There is a large body of research on medical image analysis, in particular functional MRI (fMRI). The analysis of qMRI is still an open research area. In this dissertation, we proposed three models for predicting specific clinical outcomes by

jointly modeling qMRI data and outcomes using Bayesian statistical methods.

Bayesian methods have proved to be particularly useful in analyzing high-dimensional data: (1). By constructing a suitable prior, Bayesian models assign a natural penalty against complex models and in favor of simple models. This is called Occam's razor (Denison et al. (2002)). Therefore, Bayesian methods are especially suitable for problems of dimension reduction or variable selection. (2). Instead of modeling the huge covariance structure of the images directly in the frequentist framework, spatial correlations in the images can be naturally accounted for by constructing an appropriate prior (e.g. the pairwise difference prior proposed by Besag (1993)). (3). It is straightforward to obtain inference for the parameters of interest, especially when the variances for those parameters are very difficult to derive or have no closed form. (4). Bayesian models will yield many final models instead of only one single model usually obtained in the frequentist framework. By using Bayesian model averaging, we account for model uncertainty. (5). Treating all the parameters in the system as random variables greatly clarifies the methods of analysis. It forms the basis on which our joint models can be clearly specified.

To formulate the general form of our joint models, we begin by specifying notation. In one study, let $\mathcal{Y}$ denote all the image data and let $\mathbf{Z}$ denote a patient's survival time or clinical outcome such as disease status. All of our joint models are divided into two stages with stage I focusing on modeling the image data and stage II aiming to build connections between the image data and the clinical/survival outcome; in other words, using image data to predict the outcome. We use the notation $\mathbf{\Omega}_1$ to denote the set of model parameters in stage I and $\mathbf{\Omega}_2$ in stage II. Let $\mathbf{\Omega} = \mathbf{\Omega}_1 \cup \mathbf{\Omega}_2$ denote the set of all model parameters. Then the model in stage I is formulated as: $\pi(\mathcal{Y} \mid \mathbf{\Omega}_1)$ with prior $\pi(\mathbf{\Omega}_1)$ and the model in stage II as: $\pi(\mathbf{Z} \mid \mathbf{\Omega}_1, \mathbf{\Omega}_2)$ with prior

$\pi(\mathbf{\Omega}_2)$. Then the posterior distribution can be factored as:

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

Based on Bayes' theorem, we can obtain posterior draws for $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ iteratively from the full conditional distributions:

$$\begin{aligned} \pi(\mathbf{\Omega}_1|\mathcal{Y}, \mathbf{Z}, \mathbf{\Omega}_2) &\propto \pi(\mathcal{Y} \mid \mathbf{\Omega}_1)\pi(\mathbf{\Omega}_1)\pi(\mathbf{Z} \mid \mathbf{\Omega}_1, \mathbf{\Omega}_2), \\ \pi(\mathbf{\Omega}_2|\mathbf{Z}, \mathbf{\Omega}_1) &\propto \pi(\mathbf{Z} \mid \mathbf{\Omega}_1, \mathbf{\Omega}_2)\pi(\mathbf{\Omega}_2). \end{aligned}$$

1.3 Statistical Challenges in Imaging Analysis

Brain imaging has many characteristics of typical problems facing modern statistics. Here we discuss three problems with image analyses that we have encountered in this dissertation.

Large Volume: One key issue in the prediction models with images as the predictors is how to appropriately reduce the image dimension and select the most relevant information from the images. The image data usually has a huge volume. A single slice from an MRI scan usually consists of 256×256 pixels (short for "picture elements"), and MRI images of size 512×512 are not uncommon. A whole-brain quantitative MRI scan usually has 20 to 36 such slices. Hence, there are millions of voxels, which is much larger than the number of subjects in the study. It is obviously a "large p and small n" problem. Here we review three main ways of dimension reduction and feature extraction in image analysis.

1.3.1 Region of Interest (ROI) Analysis

This type of analysis focuses on a particular part or parts of the brain, such as visible tumor or large volumes of normal-appearing white matter, where intensities are to be measured. In ROI-based comparisons of intensity changes across groups,

anatomical regions are first defined on each individuals unwarped images. These may be defined, for example, by manually circumscribing the boundary of the structure on the anatomical images, and upon these ROIs are superimposed the summary of intensity changes of the images at each voxel.

In a ROI analysis, we may need multiple MR images of different types, for example conventional structural MRI (showing tumor or lesions) and diffusion MRI (with the intensity values to be measured and tested). The appearance of tumors or lesions may be different on the two types of MRIs. The ROIs should ideally be defined on the conventional MR images, then transferred to the diffusion MRI. If the ROIs were defined directly on the diffusion MRI, the image intensity would influence where the ROI boundary was placed. Thus any conclusions about the intensity values in the region would be biased since those values are used to define which pixels/voxels would be included in the region.

To transfer ROIs between images in this way requires that the various images are all spatially registered. However, spatial registration will bring in other sources of bias as subject movement may be larger than the spatial resolution in the image.

1.3.2 Histogram Analysis

A solution to the problem of ROI placement, and possible bias arising from this process, is to test the whole brain. This is particularly appropriate for diseases where the biological effects are diffuse and widespread. The histogram is a frequency distribution showing the number of voxels with a particular range of MRI parameter values. Histogram analysis avoids any bias or pre-judgement about which parts of the brain may be affected by disease. It also avoids the need to place ROIs on the images.

Histograms do have the disadvantage that localization information is lost, and if

the disease only affects part of the brain, sensitivity will be reduced by pooling data from the entire brain. Moreover, histogram analysis ignores the inherent correlation between the pixels/voxels, which may significantly bias results.

To extract information from the images, histogram features (i.e. summary parameters) are calculated to reduce the information in the histogram down to a few parameters that are intended to contain important information in the histogram. The features can then be used to look for differences between groups of subjects.

1.3.3 Voxel-Based Analysis

The voxel-based approach compares signal changes across groups on a voxel by voxel basis. This is done after "warping" each individual brain onto a similar volume, shape, and spatial orientation, and then superimposing these brains on top of one another. Corresponding pixels/voxels in the images from each person are then assumed to represent the same region of tissue across the brains of individual subjects. A statistical model is built at each voxel for all individuals in the study. Every voxel therefore is associated with a population of signal changes for each diagnostic group in the study. These populations of signal change are then compared with one another using any number of standard parametric or nonparametric statistics, one test for each pixel/voxel in the image.

The greatest limitation of the voxel-based approach is the assumption that the anatomy and corresponding localization of function is the same across individuals, across ages, and across diagnostic groups, an assumption that anatomical imaging studies of normal and patient groups have been demonstrated to be false and is easily violated if misregistration of images is present. The ROI-based approach, in contrast, is not particularly vulnerable to this kind of artifact because it defines boundaries of the ROI based on the relevant anatomy and anatomical landmarks of each individual

in the study.

In every study, there is no correct or definite answer to the choice of feature extraction approach. However, one approach may be more suitable than an other under certain data settings and study designs. In Chapters II and III, we study malignant gliomas, which are a kind of solid tumor. Changes in MRI intensity in the tumor region are highly heterogeneous. Thus, the approach that we used to extract information from the images is a ROI based method combined with histogram analysis . While in Chapter IV, we focus on white matter changes which are highly heterogenous and differ in size and location. Histogram analysis is employed in this case.

Spatial Correlation: Another feature in image analysis is the intrinsic local correlation (the spatial structure) among the image units (pixel for 2D image and voxel for 3D image). Pixels/voxels located near each other can be expected to behave in a similar fashion because of communication leading to connections between different parts of the brain and also the temporal correlation due to the repeated imaging of the brain over time. Data are sometimes collected as images of the brain over the time course of an experiment, resulting in a large amount of available information on the timing and location of neuronal activity, the analysis of which is made more complicated by the presence of both spatial and temporal correlations. In Chapter II and III of our dissertation, we employ a pairwise difference prior model (Besag (1993)) to account for the spatial and temporal correlations in and between the images. In Chapter IV, we apply the discrete wavelet transformation to de-correlate the image data.

Noise: Furthermore, there is also a high level of noise in the images coming from sources as variable as the equipment used to perform the scans, the movement of

the subject within the scanner, and the effects of respiration and heartbeat. There are two ways of dealing with this noise – one is to try to remove the source of the noise; the other is to model it statistically. Both approaches are essential and should be pursued in brain imaging analysis. In our dissertation, we use two different statistical methods to reduce the noise in the images. In Chapter II and III, the pairwise difference prior model is utilized to smooth the images. The Bayesian Lasso is employed to de-noise the images in Chapter IV.

1.4 Dissertation Outline

In this dissertation, we focus on the predictive applications of brain images to two important brain diseases: gliomas and Alzheimer's disease. We propose three joint models.

In Chapter II, we propose a Bayesian joint classification model to predict treatment efficacy based on brain image data for patients diagnosed with malignant gliomas. 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 the project in Chapter II 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 sur-

vival status. We use the probit link and a multivariate adaptive regression spline basis. Metropolis-within-Gibbs algorithm and reversible jump MCMC 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 achieve higher 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.

In Chapter III, to assess therapy efficacy more efficiently, we extend the model in Chapter II. Instead of dichotomizing patients' survival time based on whether they survive beyond one year, we propose a Bayesian joint survival model for the patients' survival time with censoring. In stage I, we smooth the qMRI using a spatio-temporal multivariate pairwise difference prior and derive summary statistics. In stage II, we propose a Bayesian first hitting time (FHT) survival model for patients' survival time with censoring. We model patients' health statuses with a latent stochastic Wiener process. Patients' survival times are modeled as the FHT to an absorbing state (i.e. death). We link the summary statistics derived in stage I to the distribution parameters of the FHT via a Bayesian hierarchical model. Through both simulation and real data analyses, we find that our model can provide an early assessment of therapy efficacy that may aid in personalized therapy.

In Chapter IV, we propose a Bayesian joint classification model with wavelets to aid in the diagnosis of Alzheimer's disease. This study is motivated by the challenges of using MRI to diagnose Alzheimer's disease. In a MRI study of Alzheimers patients, white matter changes are highly heterogenous and differ in size and location making it difficult to use MRI as an accurate diagnostic tool. In our study, we propose to jointly model MRI data and polychotomous disease status (normal,

mild cognitive impairment or Alzheimers disease) using wavelets, which can mitigate these problems. In stage I, we apply a 3-D discrete wavelet transformation on the MRI data. A Bayesian Lasso is employed to denoise the wavelet transformed images. Summary statistics are then derived based on the images and included as covariates into the model in stage II. In stage II, we build a cumulative probit regression model to predict the polychotomous disease status. The selection of covariates is achieved by reversible jump MCMC.

CHAPTER II

Predicting Treatment Efficacy via Quantitative MRI: A Bayesian Joint Model

2.1 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 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 2.1.

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 mod-

els 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. We adopt the Bayesian paradigm for estimating and predicting outcomes. Furthermore, model uncertainty is captured by model averaging.

This chapter is organized as follows. In Section 2.2, 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 2.3. We show our model outperforms simpler models in Section 2.4. The paper concludes with a discussion, summarizing the strengths and limitations of our approach. We include the detailed mathematical derivations of the posterior distributions, algorithm details and pseudo code in the appendix in Section 2.6.

2.2 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 $\mathbf{\Omega} = \mathbf{\Omega}_1 \cup \mathbf{\Omega}_2$ denote the set of all model parameters where $\mathbf{\Omega}_1$ 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:

$$\pi(\mathbf{\Omega} \mid \mathcal{Y}, \mathbf{Z}) \propto \pi(\mathcal{Y} \mid \mathbf{\Omega}_1)\pi(\mathbf{\Omega}_1)\pi(\mathbf{Z} \mid \mathbf{\Omega}_1, \mathbf{\Omega}_2)\pi(\mathbf{\Omega}_2). \tag{2.1}$$

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 (2.1) via Markov chain Monte Carlo (MCMC) simulation by iteratively drawing between the full conditional distribution of $\mathbf{\Omega}_1$:

$$\pi(\mathbf{\Omega}_1|\mathcal{Y}, \mathbf{Z}, \mathbf{\Omega}_2) \propto \pi(\mathcal{Y} \mid \mathbf{\Omega}_1)\pi(\mathbf{\Omega}_1)\pi(\mathbf{Z} \mid \mathbf{\Omega}_1, \mathbf{\Omega}_2) \tag{2.2}$$

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

$$\pi(\mathbf{\Omega}_2|\mathbf{Z}, \mathbf{\Omega}_1) \propto \pi(\mathbf{Z} \mid \mathbf{\Omega}_1, \mathbf{\Omega}_2)\pi(\mathbf{\Omega}_2). \tag{2.3}$$

The full conditionals in (2.2) and (2.3) are easily derived from (2.1) and repeated use of Bayes' theorem.

The remainder of this section is broken up into four subsections. In Subsection 2.2.1 we define the mPWDP model, in Subsection 2.2.2 we define the GNLM and then in Subsection 2.2.3 we give an overview of how we sample from the joint posterior distribution specified in (2.1). The last subsection, 2.2.4, we describe how we evaluate our model.

2.2.1 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 sub-vectors $\boldsymbol{\mu}_{it} = (\mu_{it1}, \mu_{it2})^{\mathrm{T}}$. 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 (Besag (1993)) PWDP model to the multivariate setting. First,

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

where $\Sigma^* = \text{diag}(\Sigma)$—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 Section 2.6). 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 (2.4) and (2.5) below and Section 2.6). 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

$$\left[\begin{pmatrix} \mathbf{Y}_{i1} \\ \widetilde{\mathbf{Y}}_{i2} \end{pmatrix} \mid \begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix}, \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{pmatrix} \right] \sim \mathrm{N} \left[\begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix}, \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{pmatrix} \right] \tag{2.4}$$

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

$$\begin{pmatrix} \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix} = |N^e_i|^{-1} \left[\sum_{i' \in N^e_i \cap \mathcal{S}} \begin{pmatrix} \boldsymbol{\nu}_{i'1} \\ \boldsymbol{\nu}_{i'2} \end{pmatrix} + \sum_{i' \in N^e_i \cap \mathcal{H}} \begin{pmatrix} \boldsymbol{\mu}_{i'1} \\ \widetilde{\boldsymbol{\mu}}_{i'2} \end{pmatrix} \right]. \tag{2.5}$$

The prior for the mean vector in (2.4) is

$$\left[\begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix} \mid \begin{pmatrix} \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{i22} \end{pmatrix} \right] \sim \mathrm{N} \left[\begin{pmatrix} \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix}, |N^e_i|^{-1} \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \right] \tag{2.6}$$

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 (2.4) and Ω_{11} with Ψ_{11} in (2.6), 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 relative root mean squared 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 Section 2.6). 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 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 the posterior and 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 Section 2.6).

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 prediction 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 (Section 2.6) 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 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 $\widetilde{\mu}_{i22}$.

2.2.2 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, $\boldsymbol{\Omega}_2$, will now be subscripted by K, the number

of BMARS bases, as the number of basis 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 basis functions is:

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

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 larger amount of data 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 Section 2.6 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 $\mathbf{\Omega}_{2K} = \Theta_K \cup \{\boldsymbol{\beta}_K\}$.

We specify non-informative prior distributions for all parameters

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

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 Section 2.6.

2.2.3 Sampling from the joint posterior

Now we outline how we sample from the joint posterior given in (2.1). For more details and derivations, please consult Section 2.6.

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 Section 2.6. 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

$$\pi(Z_j = 1 \mid d_j) = 1 \quad \text{if} \quad d_j > 0, \quad \text{and} \quad = 0 \quad \text{if} \quad d_j \leq 0. \tag{2.7}$$

The full conditional distribution of d_j is straightforward to derive (Section 2.6 and Holmes and Denison (2003)) and is

$$[d_j \mid Z_j = z_j, \mathbf{X}_j, \boldsymbol{\Omega}_{2K}] \sim \begin{cases} \mathrm{N}(\eta_{jK}, 1) \text{ truncated at the left by 0 if } z_j = 1 \\ \mathrm{N}(\eta_{jK}, 1) \text{ truncated at the right by 0 if } z_j = 0. \end{cases} \tag{2.8}$$

We draw d_j, $j = 1, \ldots, M$ from (2.8). 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 $[v^{-1} \mid \boldsymbol{\beta}_K, K] \sim \text{Gamma}[0.001 + 0.5(K+1), 0.001 + 0.5\boldsymbol{\beta}_K^{\mathrm{T}}\boldsymbol{\beta}_K]$.

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 Section 2.6.

2.2.4 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 $\boldsymbol{\Omega}_2^{(t)}$ denote the value of $\boldsymbol{\Omega}_{2K}$ from the t^{th} draw from the posterior and that of $\boldsymbol{\Omega}_1$ by $\boldsymbol{\Omega}_1^{(t)}$. The cross-validated posterior predictive probability is estimated by MCMC output (see the Section 2.6) and is given by:

$$\begin{aligned}(2.9)\qquad &\pi\left(Z_j = 1 \mid \mathbf{Z}_{\{-j\}}, \mathcal{Y}\right) \\ &= \frac{\sum_{t=1}^{T} \pi\left(Z_j = 1 \mid \boldsymbol{\Omega}_{1j}^{(t)}, \boldsymbol{\Omega}_2^{(t)}\right) / \pi\left(Z_j = z_j \mid \boldsymbol{\Omega}_{1j}^{(t)}, \boldsymbol{\Omega}_2^{(t)}\right)}{\sum_{t=1}^{T} 1 / \pi\left(Z_j = z_j \mid \boldsymbol{\Omega}_{1j}^{(t)}, \boldsymbol{\Omega}_2^{(t)}\right)},\end{aligned}$$

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 is equal to one and zero 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 four. 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 2.2).

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)).

2.3 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 2.1, 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 2.2 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 2.2 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)).

The baseline prognostic factors age, surgery type, Karnofsky performance score, pathology grade and 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).

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.

2.4 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 2.1, 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 2.1, 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 2.1, 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 Section 2.6.

2.5 Discussion

In this chapter, 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 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 are easily extendable to more than 2 image types at more than 2 times points with more than four summary statistics.

Table 2.1: 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) fits 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 fits 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.

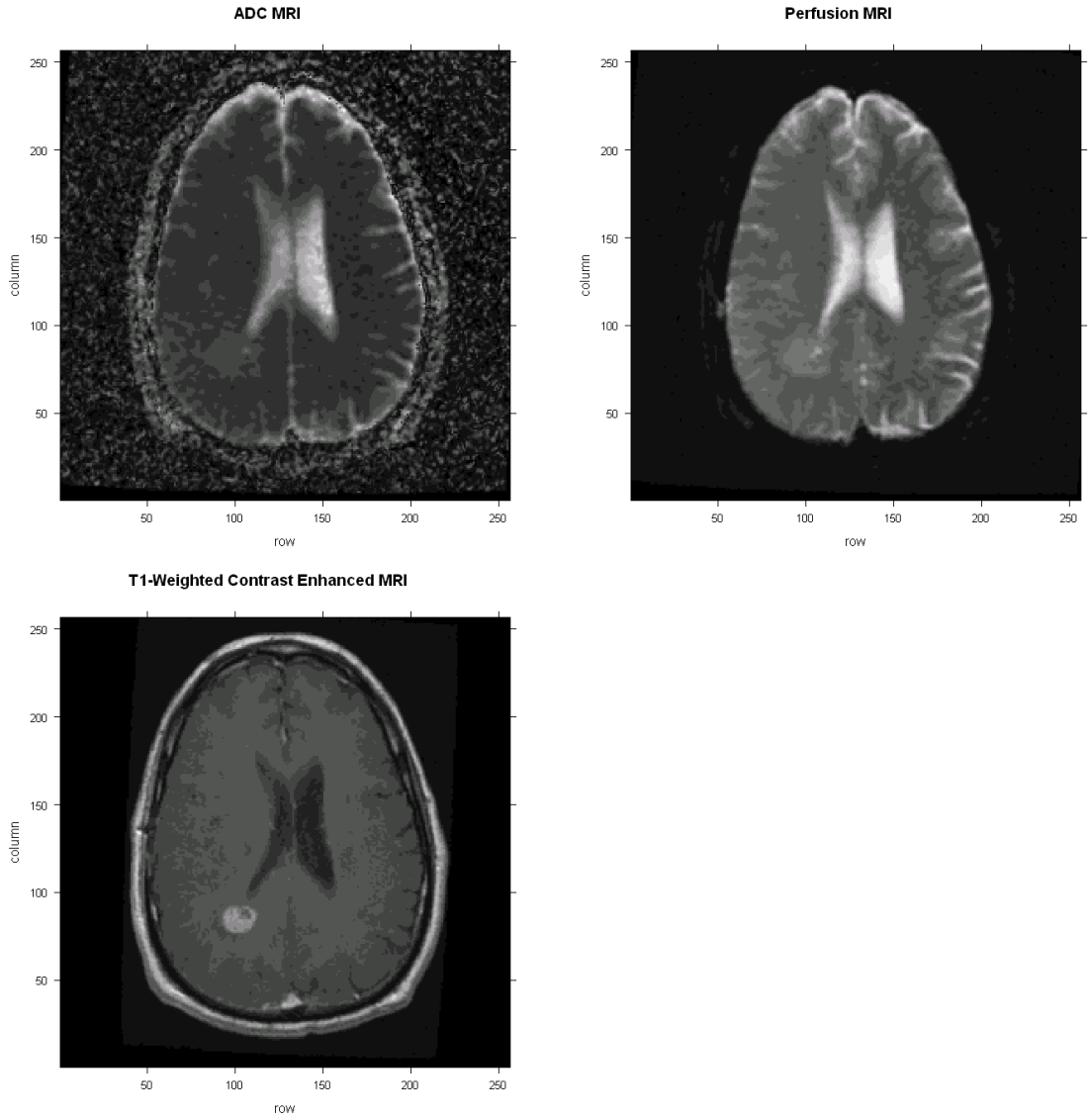

Figure 2.1: A single axial slice of pre-treatment MRI data. Upper left image: diffusion MRI; Upper right: perfusion MRI; Lower left: T1-weighted contrast enhanced MRI. The tumor is visible in all three images. It is located roughly at voxel (100,80) just below the left ventricle.

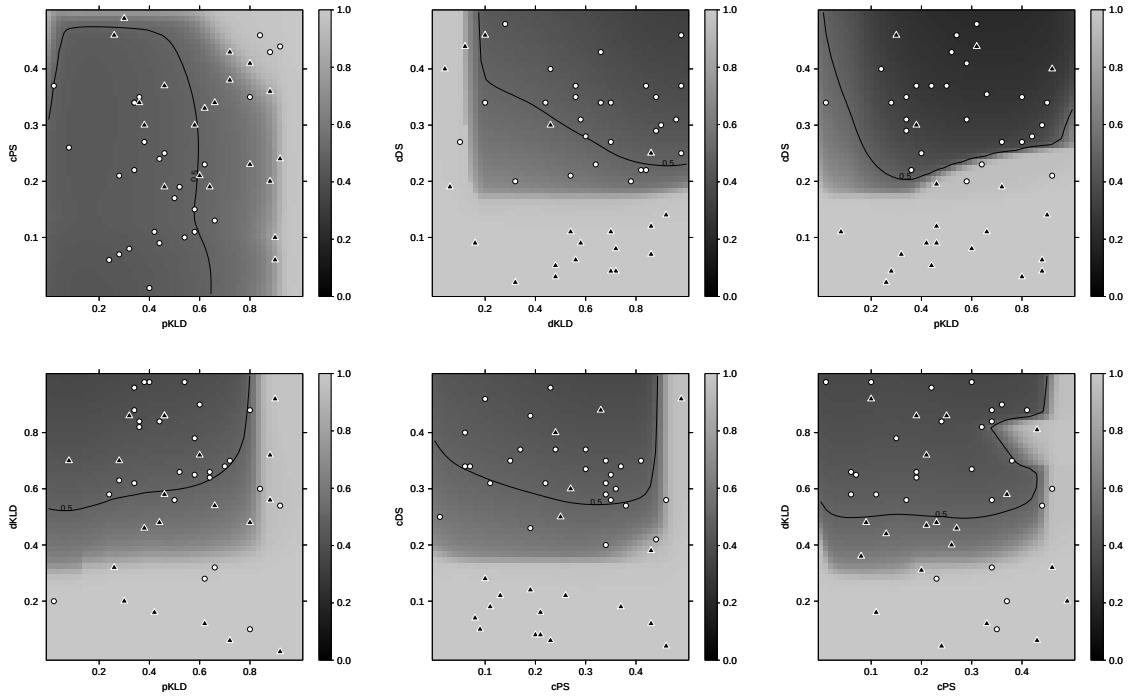

Figure 2.2: Marginal predictive probability maps from the glioma data. Each probability map is obtained by marginalizing over the two summary statistics not appearing in the x and y labels of the images. The darker gray indicates smaller probability of patients' death before one year (i.e. $\pi(Z_j = 1 \mid \mathbf{Z}_{(-j)}, \mathcal{Y}_{(-j)}, \mathcal{Y}_j)$). 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.

2.6 Appendix

Throughout this appendix, we will rely on the notation set out in the main part this chapter.

2.6.1 Model and Algorithm Details

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 denote the jth covariate vector using notation $\mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j})$ to emphasize 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

$$
\begin{aligned}
&\prod_{j=1}^{M}\prod_{i=1}^{n_j} \pi(\mathbf{Y}_{i,j} \mid \boldsymbol{\mu}_{i,j}, \Sigma_j)\pi(\Sigma_j)\pi(\boldsymbol{\mu}_{i,j} \mid \{\boldsymbol{\mu}_{i'j}\}_{i' \in N_{i,j}}, \Psi_j)\pi(\Psi_j) \\
&\times \prod_{j=1}^{M}\prod_{i=1}^{n^e_j} \pi(\mathbf{W}_{i,j} \mid \boldsymbol{\nu}_{i,j}, \Delta_j)\pi(\Delta_j)\pi(\boldsymbol{\nu}_{i,j} \mid \{\boldsymbol{\nu}_{i,j}\}_{i' \in N^e_{i,j}}, \Omega_j)\pi(\Omega_j) \\
&\times \prod_{j=1}^{M}\prod_{i=1}^{n_j} \pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j) \\
&\times \prod_{j=1}^{M} \pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K})\pi(\boldsymbol{\Omega}_{2K}),
\end{aligned} \tag{2.10}
$$

where

$$
\left[\mathbf{Y}_{i,j} \mid \boldsymbol{\mu}_{i,j}, \Sigma_j\right] \sim \mathrm{N}(\boldsymbol{\mu}_{i,j}, \Sigma_j), \qquad [\mathbf{W}_{i,j} \mid \boldsymbol{\nu}_{i,j}, \Delta_j] \sim \mathrm{N}(\boldsymbol{\nu}_{i,j}, \Delta_j), \tag{2.11}
$$

$$
[\Sigma_j] \sim \mathrm{W}^{-1}(\mathrm{I}_4, 5), \qquad [\Delta_j] \sim \mathrm{W}^{-1}(\mathrm{I}_4, 5), \tag{2.12}
$$

$$
\left[\boldsymbol{\mu}_{i,j} \mid \{\boldsymbol{\mu}_{i'j}\}_{i' \in N_{i,j}}, \Psi_j\right] \sim \mathrm{N}\left[\boldsymbol{\mu}^*_{i,j}, |N_{i,j}|^{-1}\Psi_j\right], \tag{2.13}
$$

$$
\left[\boldsymbol{\nu}_{i,j} \mid \{\boldsymbol{\nu}_{i,j}\}_{i' \in N^e_{i,j}}, \Omega_j\right] \sim \mathrm{N}\left(\boldsymbol{\nu}^*_{i,j}, |N^e_{i,j}|^{-1}\Omega_j\right], \tag{2.14}
$$

$$
[\Psi_j] \sim \mathrm{W}^{-1}(\mathrm{I}_4, 5), \qquad [\Omega_j] \sim \mathrm{W}^{-1}(\mathrm{I}_4, 5), \tag{2.15}
$$

and

$$\begin{aligned}
&\left[\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j\right] \\
&\quad \sim \mathrm{N}\left(\widetilde{\boldsymbol{\mu}}_{i2,j} + \Delta_{21,j}\left(\Delta_{11,j}\right)^{-1}\left(\mathbf{Y}_{i1,j} - \boldsymbol{\mu}_{i1,j}\right),\ \Delta_{22,j} - \Delta_{21,j}(\Delta_{11,j})^{-1}\Delta_{12,j}\right), \\
&\left[\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j\right] \\
&\quad \sim \mathrm{N}\left(\widetilde{\boldsymbol{\mu}}^*_{i2,j} + \Omega_{21,j}(\Omega_{11,j})^{-1}(\boldsymbol{\mu}_{i1,j} - \boldsymbol{\mu}^*_{i1,j}), |N^e_{i,j}|^{-1}(\Omega_{22,j} - \Omega_{21,j}(\Omega_{11,j})^{-1}\Omega_{12,j})\right), \\
&\left[d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K}\right] \sim \mathrm{N}(\eta_{jK}, 1),
\end{aligned}$$

where η_{jK} is defined in (2.48), and

$$\boldsymbol{\mu}^*_{i,j} = |N_{i,j}|^{-1} \sum_{i' \in N_{i,j}} \boldsymbol{\mu}_{i',j}\ ,\ \boldsymbol{\nu}^*_{i,j} = |N^e_{i,j}|^{-1} \sum_{i' \in N^e_{i,j}} \boldsymbol{\nu}_{i',j}, \tag{2.16}$$

$$\begin{pmatrix} \boldsymbol{\mu}^*_{i1,j} \\ \widetilde{\boldsymbol{\mu}}^*_{i2,j} \end{pmatrix} = |N^e_{i,j}|^{-1} \left[\sum_{i' \in N^e_{i,j} \cap \mathcal{S}_j} \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\nu}_{i'2,j} \end{pmatrix} + \sum_{i' \in N^e_{i,j} \cap \mathcal{H}_j} \begin{pmatrix} \boldsymbol{\mu}_{i'1,j} \\ \widetilde{\boldsymbol{\mu}}_{i'2,j} \end{pmatrix} \right]. \tag{2.17}$$

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.

Stage I – 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 (2.11)), 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,

$$\left[\Sigma_j \mid \left\{\boldsymbol{\mu}_{i,j}\right\}_{i=1}^{n_j}, \left\{\mathbf{Y}_{i,j}\right\}_{i=1}^{n_j}\right] \sim \mathrm{W}^{-1}\left(\mathrm{S}_{1,j} + \mathrm{I}_4, n_j + 5\right). \tag{2.18}$$

Stage I – 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,

$$\left[\Psi_j \mid \left\{\boldsymbol{\mu}_{i,j}\right\}_{i=1}^{n_j}\right] \sim \mathrm{W}^{-1}\left(\mathrm{S}_{2,j} + \mathrm{I}_4, n_j + 5\right). \tag{2.19}$$

Stage I – 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

$$\begin{aligned}
&\pi(\boldsymbol{\mu}_{i,j} \mid \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}}, \mathbf{Y}_{i,j}, \Sigma_j, \Psi_j, d_j, \widetilde{\boldsymbol{\mu}}_{i2,j}, \boldsymbol{\Omega}_{2K}) \qquad (2.20)\\
&\quad \propto \pi(\mathbf{Y}_{i,j} \mid \boldsymbol{\mu}_{i,j}, \Sigma_j)\pi(\boldsymbol{\mu}_{i,j} \mid \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}}, \Psi_j)\\
&\qquad \times \pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K})\pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)\\
&\qquad \times \pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j) \prod_{i' \in N_{i,j}} \pi(\widetilde{\boldsymbol{\mu}}_{i'2,j} \mid \boldsymbol{\mu}_{i'1,j}, C_{i',j}, \Omega_j)\\
&\quad \propto \pi\left(\boldsymbol{\mu}_{i,j} \mid \Sigma_j, \Psi_j, \mathbf{Y}_{i,j}, \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}}\right) \pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K}) \qquad (2.21)\\
&\qquad \times \pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j)\\
&\qquad \times \prod_{i' \in N_{i,j}} \pi(\widetilde{\boldsymbol{\mu}}_{i'2,j} \mid \boldsymbol{\mu}_{i'1,j}, C_{i',j}, \Omega_j),
\end{aligned}$$

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 (2.21) 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

$$\left[\boldsymbol{\mu}_{i,j} \mid \Sigma_j, \Psi_j, \mathbf{Y}_{i,j}, \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}}\right] \sim$$

$$\mathrm{N}\left[\mathcal{V}\left(|N_{i,j}|{\Psi_j}^{-1}\boldsymbol{\mu}^*_{i,j} + {\Sigma_j}^{-1}\mathbf{Y}_{i,j}\right), \mathcal{V}\right], \tag{2.22}$$

where $\mathcal{V} = \left(|N_{i,j}|{\Psi_j}^{-1} + {\Sigma_j}^{-1}\right)^{-1}$. Therefore, we propose a new value of $\boldsymbol{\mu}_{i,j}$, called it $\boldsymbol{\mu}^{prop}_{i,j}$ from (2.22), which simplifies the acceptance probability and results in a high acceptance rate. The acceptance probability is given by

$$\alpha_\mu = \min\{1, \mathcal{R}\}, \tag{2.23}$$

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

$$\begin{aligned}\mathcal{R} \;=\; & \frac{\pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{\{-i\}2,j}, \boldsymbol{\mu}^{prop}_{i2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K})\pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}^{prop}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)}{\pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{\{-i\}2,j}, \boldsymbol{\mu}^{current}_{i2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K})\pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}^{current}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)} \\ & \times \frac{\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}^{prop}_{i1,j}, C_{i,j}, \Omega_j)\prod_{i' \in N_{i,j}}\pi(\widetilde{\boldsymbol{\mu}}_{i'2,j} \mid \boldsymbol{\mu}_{i'1,j}, C^{prop}_{i',j}, \Omega_j)}{\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}^{current}_{i1,j}, C_{i,j}, \Omega_j)\prod_{i' \in N_{i,j}}\pi(\widetilde{\boldsymbol{\mu}}_{i'2,j} \mid \boldsymbol{\mu}_{i'1,j}, C^{current}_{i',j}, \Omega_j)},\end{aligned}$$

where $C^{prop}_{i',j} = \{\{\boldsymbol{\mu}^{prop}_{i1,j}, \boldsymbol{\mu}_{k1,j}\}_{k \in N_{i',j}}, \{\widetilde{\boldsymbol{\mu}}_{k2,j}\}_{k \in N_{i',j}}, \{\boldsymbol{\nu}_{k2,j}\}_{k \in N^e_{i',j} \cap \mathcal{S}_j}\}$.

Stage I – Updating $\boldsymbol{\nu}_{i,j}$:

For healthy tissue voxels, analogous to (2.22), the full conditional of $\boldsymbol{\nu}_{i,j}$ for a voxel $i \in N^e_{i',j} \cap \mathcal{H}_j$, is:

$$\left[\boldsymbol{\nu}_{i,j} \mid \Omega_j, \Delta_j, \mathbf{W}_{i,j}, \{\boldsymbol{\nu}_{i',j}\}_{i' \in N^e_{i,j}}\right] \sim$$

$$\mathrm{N}\left[\left(|N^e_{i,j}|\Omega_j^{-1} + {\Delta_j}^{-1}\right)^{-1}\left(|N^e_{i,j}|\Omega_j^{-1}\boldsymbol{\nu}^*_{i,j} + \Delta_j^{-1}\mathbf{W}_{i,j}\right), \left(|N^e_{i,j}|\Omega_j^{-1} + \Delta_j^{-1}\right)^{-1}\right]. \tag{2.24}$$

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

$$\begin{aligned}(2.25)\ \pi(\boldsymbol{\nu}_{i,j} \mid \{\boldsymbol{\nu}_{i',j}\}_{i' \in N^e_{i,j}}, \mathbf{W}_{i,j}, \Delta_j, \Omega_j, \{\widetilde{\boldsymbol{\mu}}_{i2,j}\}_{i' \in N_i}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \{\boldsymbol{\mu}_{i'1,j}\}_{i' \in N_i}, \boldsymbol{\mu}_{i1,j}) \\ \propto \pi(\mathbf{W}_{i,j} \mid \boldsymbol{\nu}_{i,j}, \Delta_j)\pi(\boldsymbol{\nu}_{i,j} \mid \{\boldsymbol{\nu}_{i',j}\}_{i' \in N^e_{i,j}}, \Omega_j) \\ \times \prod_{i' \in N_{i,j}} \pi(\widetilde{\boldsymbol{\mu}}_{i'2,j} \mid \boldsymbol{\mu}_{i'1,j}, C_{i',j}, \Omega_j) \\ (2.26)\qquad \propto \pi\left(\boldsymbol{\nu}_{i,j} \mid \Delta_j, \Omega_j, \mathbf{W}_{i,j}, \{\boldsymbol{\nu}_{i',j}\}_{i' \in N^e_{i,j}}\right) \prod_{i' \in N_{i,j}} \pi(\widetilde{\boldsymbol{\mu}}_{i'2,j} \mid \boldsymbol{\mu}_{i'1,j}, C_{i',j}, \Omega_j),\end{aligned}$$

which does not have a nice distributional form. However, the first term in (2.26) has distribution (2.24). Thus, we propose a new value $\boldsymbol{\nu}^{prop}_{i,j}, i \in N^e_{i',j} \cap \mathcal{S}_j$ from (2.24) and accept it with probability:

$$(2.27)\qquad \alpha_\nu = \min\left\{1, \prod_{i' \in N_{i,j}} \frac{\pi(\widetilde{\boldsymbol{\mu}}_{i'2,j} \mid \boldsymbol{\mu}_{i'1,j}, C^{prop}_{i',j}, \Omega_j)}{\pi(\widetilde{\boldsymbol{\mu}}_{i'2,j} \mid \boldsymbol{\mu}_{i'1,j}, C^{current}_{i',j}, \Omega_j)}\right\},$$

where $C^{prop}_{i',j} = \{\{\boldsymbol{\mu}_{k1,j}\}_{k \in N_{i',j}}, \{\widetilde{\boldsymbol{\mu}}_{k2,j}\}_{k \in N_{i',j}}, \{\boldsymbol{\nu}^{prop}_{i2,j}, \boldsymbol{\nu}_{k2,j}\}_{k \in N^e_{i',j} \cap \mathcal{S}_j}\}$.

Stage I – Updating Δ_j:

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

$$\begin{aligned}(2.28)\quad \pi(\Delta_j \mid \{\mathbf{W}_{i,j}\}_{i=1}^{n^e_j}, \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n^e_j}, \{\widetilde{\mathbf{Y}}_{i2,j}\}_{i=1}^{n_j}, \{\mathbf{Y}_{i1,j}\}_{i=1}^{n_j}, \{\boldsymbol{\mu}_{i1,j}\}_{i=1}^{n_j}, \{\widetilde{\boldsymbol{\mu}}_{i2,j}\}_{i=1}^{n_j}) \\ \propto \prod_{i=1}^{n^e_j} \pi(\mathbf{W}_{i,j} \mid \boldsymbol{\nu}_{i,j}, \Delta_j)\pi(\Delta_j) \prod_{i=1}^{n_j} \pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j) \\ (2.29)\qquad \propto \pi(\Delta_j \mid \{\mathbf{W}_{i,j}\}_{i=1}^{n^e_j}, \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n^e_j}) \prod_{i=1}^{n_j} \pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j),\end{aligned}$$

which does not have a nice distributional form from which to draw. However, the first term in (2.29) has an inverse Wishart distribution:

$$(2.30)\qquad \left[\Delta_j \mid \{\mathbf{W}_{i,j}\}_{i=1}^{n^e_j}, \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n^e_j}\right] \ \sim\ \mathrm{W}^{-1}\left(\mathrm{S}_{3,j} + \mathrm{I}_4, n^e_j + 5\right),$$

where $S_{3,j} = \sum_{i=1}^{n^e_j} (\mathbf{W}_{i,j} - \boldsymbol{\nu}_{i,j})(\mathbf{W}_{i,j} - \boldsymbol{\nu}_{i,j})^{\mathrm{T}}$ for all $i \in \mathcal{H} \cup \mathcal{S}$. Therefore, we propose a new value Δ^{prop}_j from (2.30) and accept this value with probability

$$(2.31)\qquad \alpha_\Delta = \min\left\{1, \prod_{i=1}^{n_j} \frac{\pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta^{prop}_j)}{\pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta^{current}_j)}\right\}.$$

Stage I – Updating Ω_j:

From (2.10), we have

$$\pi(\Omega_j \mid \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n_j^e}, \{\boldsymbol{\mu}_{i1,j}\}_{i=1}^{n_j}, \{\widetilde{\boldsymbol{\mu}}_{i2,j}\}_{i=1}^{n_j}, \{\boldsymbol{\nu}_{i'2}\}_{i' \in N_i^e \cap \mathcal{S}_j}) \tag{2.32}$$
$$\propto \prod_{i=1}^{n_j^e} \pi(\boldsymbol{\nu}_{i,j} \mid \{\boldsymbol{\nu}_{i,j}\}_{i' \in N_{i,j}^e}, \Omega_j) \pi(\Omega_j) \prod_{i=1}^{n_j} \pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j)$$
$$\propto \pi(\Omega_j \mid \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n_j^e}) \prod_{i=1}^{n_j} \pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j). \tag{2.33}$$

The first term in (2.33) has an inverse Wishart distribution:

$$\left[\Omega_j \mid \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n^e}\right] \sim \mathrm{W}^{-1}\left(\mathrm{S}_{4,j} + \mathrm{I}_4, n_j^e + 5\right), \tag{2.34}$$

Thus we propose a new value Ω_j^{prop} from (2.34) and accept this value as a draw from the full conditional with probability

$$\alpha_\Omega = \min\left\{1, \prod_{i=1}^{n_j} \frac{\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j^{prop})}{\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j^{current})}\right\}. \tag{2.35}$$

Stage I – 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

$$\left[\begin{pmatrix}\mathbf{Y}_{i1,j}\\ \widetilde{\mathbf{Y}}_{i2,j}\end{pmatrix} \mid \begin{pmatrix}\boldsymbol{\mu}_{i1,j}\\ \widetilde{\boldsymbol{\mu}}_{i2,j}\end{pmatrix}, \begin{pmatrix}\Delta_{11,j} & \Delta_{12,j}\\ \Delta_{21,j} & \Delta_{22,j}\end{pmatrix}\right] \sim \mathrm{N}\left[\begin{pmatrix}\boldsymbol{\mu}_{i1,j}\\ \widetilde{\boldsymbol{\mu}}_{i2,j}\end{pmatrix}, \begin{pmatrix}\Delta_{11,j} & \Delta_{12,j}\\ \Delta_{21,j} & \Delta_{22,j}\end{pmatrix}\right], \tag{2.36}$$

where $\mathbf{Y}_{i1,j} = (Y_{i11,j}, Y_{i12,j})^{\mathrm{T}}$ represents 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

$$\begin{pmatrix}\boldsymbol{\mu}_{i1,j}^*\\ \widetilde{\boldsymbol{\mu}}_{i2,j}^*\end{pmatrix} = |N_{i,j}^e|^{-1}\left[\sum_{i' \in N_{i,j}^e \cap \mathcal{S}_j} \begin{pmatrix}\mathbf{0}\\ \boldsymbol{\nu}_{i'2,j}\end{pmatrix} + \sum_{i' \in N_{i,j}^e \cap \mathcal{H}_j} \begin{pmatrix}\boldsymbol{\mu}_{i'1,j}\\ \widetilde{\boldsymbol{\mu}}_{i'2,j}\end{pmatrix}\right]. \tag{2.37}$$

The prior for the mean vector in (2.36) is

$$\left[\begin{pmatrix}\boldsymbol{\mu}_{i1,j}\\ \widetilde{\boldsymbol{\mu}}_{i2,j}\end{pmatrix} \mid \begin{pmatrix}\boldsymbol{\mu}^*_{i1,j}\\ \widetilde{\boldsymbol{\mu}}^*_{i2,j}\end{pmatrix}, \begin{pmatrix}\Omega_{11,j} & \Omega_{12,j}\\ \Omega_{21,j} & \Omega_{22,j}\end{pmatrix}\right] \tag{2.38}$$
$$\sim \mathrm{N}\left[\begin{pmatrix}\boldsymbol{\mu}^*_{i1,j}\\ \widetilde{\boldsymbol{\mu}}^*_{i2,j}\end{pmatrix}, |N^e_{i,j}|^{-1}\begin{pmatrix}\Omega_{11,j} & \Omega_{12,j}\\ \Omega_{21,j} & \Omega_{22,j}\end{pmatrix}\right].$$

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):

$$\left[\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j\right] \sim \tag{2.39}$$
$$\mathrm{N}\left(\widetilde{\boldsymbol{\mu}}_{i2,j} + \Delta_{21,j}\left(\Delta_{11,j}\right)^{-1}\left(\mathbf{Y}_{i1,j} - \boldsymbol{\mu}_{i1,j}\right),\ \Delta_{22,j} - \Delta_{21,j}(\Delta_{11,j})^{-1}\Delta_{12,j}\right).$$

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

$$\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \widetilde{\mathbf{Y}}_{i2,j}, \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, C_{i,j}, \Delta_j, \Omega_j, d_j, \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K}) \tag{2.40}$$
$$\propto \pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j)$$
$$\times\pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K})$$
$$\propto \pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \widetilde{\mathbf{Y}}_{i2,j}, \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, C_{i,j}, \Delta_j, \Omega_j) \tag{2.41}$$
$$\times\pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_{2K}).$$

We propose a new value $\widetilde{\boldsymbol{\mu}}^{prop}_{i2,j}$ from the first term in (2.41) which has a normal distribution:

$$\left[\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \widetilde{\mathbf{Y}}_{i2,j}, \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, C_{i,j}, \Delta_j, \Omega_j\right] \sim \tag{2.42}$$
$$\mathrm{N}\left(\boldsymbol{\theta}_{i2,j} + \Lambda_{i_{21,j}}\left(\Lambda_{i_{11,j}}\right)^{-1}\left(\boldsymbol{\mu}_{i1,j} - \boldsymbol{\theta}_{i1,j}\right),\ \Lambda_{i_{22,j}} - \Lambda_{i_{21,j}}\left(\Lambda_{i_{11,j}}\right)^{-1}\Lambda_{i_{12,j}}\right)$$

where

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

and

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

We then accept this proposed value with probability

$$\alpha_{\widetilde{\mu}} = \min \left\{ 1, \frac{\pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{\{-i\}2,j}, \widetilde{\boldsymbol{\mu}}^{prop}_{i2,j}), \boldsymbol{\Omega}_{2K})}{\pi(d_j \mid \mathbf{X}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{\{-i\}2,j}, \widetilde{\boldsymbol{\mu}}^{current}_{i2,j}), \boldsymbol{\Omega}_{2K})} \right\}. \tag{2.43}$$

Checking Covariance Structures:

Next we check whether the covariance structures are similar (see section 2.2.1). Note that all 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

$$\Sigma^{(t)}_{11} = \begin{pmatrix} \sigma^{(t)}_{11} & \sigma^{(t)}_{12} \\ \sigma^{(t)}_{21} & \sigma^{(t)}_{22} \end{pmatrix}, \quad \Delta^{(t)}_{11} = \begin{pmatrix} \delta^{(t)}_{11} & \delta^{(t)}_{12} \\ \delta^{(t)}_{21} & \delta^{(t)}_{22} \end{pmatrix},$$
$$\Psi^{(t)}_{11} = \begin{pmatrix} \psi^{(t)}_{11} & \psi^{(t)}_{12} \\ \psi^{(t)}_{21} & \psi^{(t)}_{22} \end{pmatrix}, \quad \Omega^{(t)}_{11} = \begin{pmatrix} \omega^{(t)}_{11} & \omega^{(t)}_{12} \\ \omega^{(t)}_{21} & \omega^{(t)}_{22} \end{pmatrix},$$

where $\sigma^{(t)}_{12} = \sigma^{(t)}_{21}$, $\delta^{(t)}_{12} = \delta^{(t)}_{21}$,$\omega^{(t)}_{12} = \omega^{(t)}_{21}$, $\psi^{(t)}_{12} = \psi^{(t)}_{21}$, and (t) indicates the 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:

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

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 used 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:

$$\text{dKLD} = \sum_\ell P_{\ell 1}\ln(P_{\ell 1}/Q_{\ell 1}) \tag{2.44}$$

$$\text{pKLD} = \sum_\ell P_{\ell 2}\ln(P_{\ell 2}/Q_{\ell 2}). \tag{2.45}$$

The conditional diffusion and perfusion statistics are straightforward to calculate:

$$\text{cDS} = n^{-1}\sum_{i=1}^{n} I\left[\mu_{i21} > q_{0.975}\left(\widetilde{\mu}_{i21}\right)\right] \tag{2.46}$$

$$\text{cPS} = n^{-1}\sum_{i=1}^{n} I\left[\mu_{i22} < q_{0.025}\left(\widetilde{\mu}_{i22}\right)\right]. \tag{2.47}$$

Stage II:

The GNLM-BMARS model with K bases functions is:

$$\pi(Z_j = 1 \mid \mathbf{X}_j, \mathbf{\Omega}_{2K}) = g(\eta_{jK}), \qquad \eta_{jK} = \sum_{k=0}^{K} \beta_k B_k(\mathbf{X}_j), \tag{2.48}$$

$$B_k(\mathbf{X}_j) = \begin{cases} 1, & k = 0, \\ \prod_{l=1}^{L_k} [s_{lk}(X_{jw_{lk}} - t_{lk})]_+, & k = 1, 2, \ldots, K. \end{cases}$$

Updating the latent vector d

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

$$\pi(Z_j = 1 \mid d_j) = 1 \quad \text{if} \quad d_j > 0, \quad \text{and} \quad = 0 \quad \text{if} \quad d_j \leq 0. \tag{2.49}$$

Marginalizing (2.49) over d_j is equivalent to $\pi(Z_j = 1 \mid \mathbf{X}_j, \mathbf{\Omega}_{2K})$ in (2.48):

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

It is equally easy to show that

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

and

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

which are densities of truncated normal distributions. That is,

$$[d_j \mid Z_j = z_j, \mathbf{X}_j, \boldsymbol{\Omega}_{2K}] \sim \begin{cases} \mathrm{N}(\eta_{jK}, 1) \text{ truncated at the left by 0 if } z_j = 1 \\ \mathrm{N}(\eta_{jK}, 1) \text{ truncated at the right by 0 if } z_j = 0 \end{cases}. \tag{2.50}$$

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

$$\begin{aligned} \pi(\boldsymbol{\beta}_K|\mathbf{d}, v, \Theta_K, \mathcal{X}) &\propto \pi(\boldsymbol{\beta}_K \mid v, K)\pi(\mathbf{d} \mid \mathcal{X}, \boldsymbol{\Omega}_{2K}) \\ &\propto \exp\left\{-0.5\left[v^{-1}{\boldsymbol{\beta}_K}^{\mathrm{T}}\boldsymbol{\beta}_K + (\mathbf{d} - \mathrm{B}_K\boldsymbol{\beta}_K)^{\mathrm{T}}(\mathbf{d} - \mathrm{B}_K\boldsymbol{\beta}_K)\right]\right\} \\ &\propto \exp\left\{-0.5(\boldsymbol{\beta}_K - \mathbf{m}_K^*)^{\mathrm{T}}(V_K^*)^{-1}(\boldsymbol{\beta}_K - \mathbf{m}_K^*)\right\} \end{aligned} \tag{2.51}$$

where

$$V_K^* = [(v\mathrm{I}_{K+1})^{-1} + \mathrm{B}_K^{\mathrm{T}}\mathrm{B}_K]^{-1}, \tag{2.52}$$

$$\mathbf{m}_K^* = V_K^*\mathrm{B}_K^{\mathrm{T}}\mathbf{d}. \tag{2.53}$$

Thus,

$$[\boldsymbol{\beta}_K \mid \mathbf{d}, v, \Theta_K, \mathcal{X}] \sim \mathrm{N}(\mathbf{m}_K^*, V_K^*). \tag{2.54}$$

Equation (2.51) follows from the identity

$$\begin{aligned} &v^{-1}{\boldsymbol{\beta}_K}^{\mathrm{T}}\boldsymbol{\beta}_K + (\mathbf{d} - \mathrm{B}_K\boldsymbol{\beta}_K)^{\mathrm{T}}(\mathbf{d} - \mathrm{B}_K\boldsymbol{\beta}_K) = \\ &\qquad (\boldsymbol{\beta}_K - \mathbf{m}_K^*)^{\mathrm{T}}(V_K^*)^{-1}(\boldsymbol{\beta}_K - \mathbf{m}_K^*) + \mathbf{d}^{\mathrm{T}}\mathbf{d} - (\mathbf{m}_K^*)^{\mathrm{T}}(V_K^*)^{-1}\mathbf{m}_K^*. \end{aligned} \tag{2.55}$$

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

$$[v^{-1} \mid \boldsymbol{\beta}_K,\, K] \sim \mathrm{Gamma}[0.001 + 0.5(K+1),\, 0.001 + 0.5\boldsymbol{\beta}_K^{\mathrm{T}}\boldsymbol{\beta}_K], \tag{2.56}$$

and that

$$[\lambda \mid K] \sim \text{Gamma}(1 + K, 0.2 + 1). \tag{2.57}$$

Stage II –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 (2.54). 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

$$\pi(L_{K+1} = 1) = \pi(L_{K+1} = 2) = 1/2. \tag{2.58}$$

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,

$$\pi(w_{1,K+1} = w \mid L_{K+1} = 1) = 1/4 \quad \text{for} \quad w = 1, 2, 3, 4. \tag{2.59}$$

$$\pi[(w_{1,K+1}, w_{2,K+1}) = (w, w') \mid L_{K+1} = 2] = 1/6 \tag{2.60}$$
$$\text{for} (w, w') = (1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4).$$

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,

$$\pi(t_{l,K+1} = X_{jw_{l,K+1}} \mid w_{l,K+1}) = 1/M \quad \text{for} \quad l = 1, \ldots, L_{K+1}. \tag{2.61}$$

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

$$\pi(s_{l,K+1} = -1) = \pi(s_{l,K+1} = 1) = 1/2 \quad \text{for} \quad l = 1, \ldots, L_{K+1}. \tag{2.62}$$

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

(2.63)
$$\alpha = \min\left\{1, \frac{\pi(\mathbf{d} \mid \mathcal{X}, \mathbf{\Omega}_{2,K+1})\pi(\Theta_{K+1} \mid \lambda)\pi(\boldsymbol{\beta}_{K+1} \mid v, K+1)\pi(v^{-1})\pi(\lambda)\pi_{\text{death}}}{\pi(\mathbf{d} \mid \mathcal{X}, \mathbf{\Omega}_{2K})\pi(\Theta_K \mid \lambda)\pi(\boldsymbol{\beta}_K \mid v, K)\pi(v^{-1})\pi(\lambda)q(\mathbf{U})\pi_{\text{birth}}}\right\}.$$

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

$$\alpha = \min\left\{1, \frac{\pi(\mathbf{d} \mid \mathcal{X}, \mathbf{\Omega}_{2,K+1})\pi(\boldsymbol{\beta}_{K+1} \mid v, K+1)\lambda}{\pi(\mathbf{d} \mid \mathcal{X}, \mathbf{\Omega}_{2K})\pi(\boldsymbol{\beta}_K \mid v, K)(K+1)}\right\}. \tag{2.64}$$

Also, it is straightforward to show that

$$\begin{aligned}
&\pi(\mathbf{d} \mid \mathcal{X}, \mathbf{\Omega}_{2K})\pi(\boldsymbol{\beta}_K \mid v, K) \\
&\quad \propto (v)^{-(K+1)/2} \exp\left\{-0.5\left[v^{-1}\boldsymbol{\beta}_K^{\mathrm{T}}\boldsymbol{\beta}_K + (\mathbf{d} - \mathrm{B}_K\boldsymbol{\beta}_K)^{\mathrm{T}}(\mathbf{d} - \mathrm{B}_K\boldsymbol{\beta}_K)\right]\right\} \\
&\quad = (v)^{-(K+1)/2} \exp\left\{-0.5(\boldsymbol{\beta}_K - \mathbf{m}_K^*)^{\mathrm{T}}(V_K^*)^{-1}(\boldsymbol{\beta}_K - \mathbf{m}_K^*)\right\} \times \\
&\qquad \exp\left\{-0.5\left[2\mathbf{d}^{\mathrm{T}}\mathbf{d} - (\mathbf{m}_K^*)^{\mathrm{T}}(V_K^*)^{-1}\mathbf{m}_K^*\right]\right\},
\end{aligned} \tag{2.65}$$

where the equality follows from (2.55). Now integrating out $\boldsymbol{\beta}_{K+1}$ and $\boldsymbol{\beta}_K$ from their respective joint full conditionals (2.65) the acceptance probability simplifies to:

$$\begin{aligned}
\alpha_{\text{birth}} &= \min\left\{1, \frac{\pi(\mathbf{d} \mid \mathcal{X}, \mathbf{\Omega}_{2,K+1}, v)\lambda}{\pi(\mathbf{d} \mid \mathcal{X}, \mathbf{\Omega}_{2K}, v)(K+1)}\right\} \\
&= \min\left\{1, \frac{|V_{K+1}^*|^{1/2}\exp(a_K - a_{K+1})\lambda}{v^{1/2}|V_K^*|^{1/2}(K+1)}\right\}
\end{aligned} \tag{2.66}$$

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

$$\alpha_{\text{death}} = \min\left\{1, \frac{v^{1/2}|V_{K-1}^*|^{1/2}K}{|V_K^*|^{1/2}\exp(a_{K-1} - a_K)\lambda}\right\}. \tag{2.67}$$

Stage II –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 proposed matrix $\mathrm{B}_K^{\text{prop}}$. Compute the proposed vector $\mathbf{m}_K^{*,\text{prop}}$ and matrix $V_K^{*,\text{prop}}$ from (2.53) and (2.52) 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

$$\alpha_{\text{move}} = \min\left\{1, \frac{|V_K^{*,\text{prop}}|^{1/2}\exp(a_K - a_K^{\text{prop}})}{|V_K^*|^{1/2}}\right\}. \tag{2.68}$$

2.6.2 Cross-validated Prediction:

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

$$\pi(Z_j = 1 \mid \mathbf{Z}_{\{-j\}}, \mathcal{Y}) = \int\int \pi(Z_j = 1 \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2)\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z}_{\{-\mathbf{j}\}} = \mathbf{z}_{\{-\mathbf{j}\}}, \mathcal{Y})d\mathbf{\Omega}_1 d\mathbf{\Omega}_2. \tag{2.69}$$

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

$$\begin{aligned}
&\frac{\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z} = \mathbf{z}, \mathcal{Y})}{\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z}_{\{-\mathbf{j}\}} = \mathbf{z}_{\{-\mathbf{j}\}}, \mathcal{Y})} \\
&\quad = \frac{\pi(Z_j = z_j, \mathcal{Y}_j \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2)}{\pi(\mathcal{Y}_j \mid \mathbf{\Omega}_{1j})} = \pi(Z_j = z_j \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2).
\end{aligned} \tag{2.70}$$

Now rewrite (2.69) using (2.70):

$$\begin{aligned}
&\pi(Z_j = 1 \mid \mathbf{Z}_{\{-j\}}, \mathcal{Y}) \\
&\quad = \frac{\int\int \pi(Z_j = 1 \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2) \left[\frac{\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 | \mathbf{Z}_{\{-\mathbf{j}\}} = \mathbf{z}_{\{-\mathbf{j}\}}, \mathcal{Y})}{\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 | \mathbf{Z} = \mathbf{z}, \mathcal{Y})}\right] \pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z} = \mathbf{z}, \mathcal{Y}) d\mathbf{\Omega}_1 d\mathbf{\Omega}_2}{\int\int \left[\frac{\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 | \mathbf{Z}_{\{-\mathbf{j}\}} = \mathbf{z}_{\{-\mathbf{j}\}}, \mathcal{Y})}{\pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 | \mathbf{Z} = \mathbf{z}, \mathcal{Y})}\right] \pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z} = \mathbf{z}, \mathcal{Y}) d\mathbf{\Omega}_1, \mathbf{\Omega}_2} \\
&\quad = \frac{\int\int \pi(Z_j = 1 \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2) \left[1/\pi(Z_j = z_j \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2)\right] \pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z} = \mathbf{z}, \mathcal{Y}) d\mathbf{\Omega}_1 d\mathbf{\Omega}_2}{\int\int \left[1/\pi(Z_j = z_j \mid \mathbf{\Omega}_{1j}, \mathbf{\Omega}_2)\right] \pi(\mathbf{\Omega}_1, \mathbf{\Omega}_2 \mid \mathbf{Z} = \mathbf{z}, \mathcal{Y}) d\mathbf{\Omega}_1 d\mathbf{\Omega}_2}
\end{aligned}$$

We note that $\pi(Z_j = 1 \mid \mathbf{Z}_{\{-j\}}, \mathcal{Y})$ only depends on the posterior distribution of the parameters given the entire data after some derivations. We can calculate $\pi(Z_j = 1 \mid \mathbf{Z}_{\{-j\}}, \mathcal{Y})$ using MCMC draws from the posterior distribution of the parameters

given the full data:

$$\pi(Z_j = 1 \mid \mathbf{Z}_{\{-j\}}, \mathcal{Y}) \approx \frac{\frac{1}{T}\sum_{t=1}^{T} \pi\left(Z_j = 1 \mid \boldsymbol{\Omega}_{1j}^{(t)}, \boldsymbol{\Omega}_2^{(t)}\right) / \pi\left(Z_j = z_j \mid \boldsymbol{\Omega}_{1j}^{(t)}, \boldsymbol{\Omega}_2^{(t)}\right)}{\frac{1}{T}\sum_{t=1}^{T} 1/\pi\left(Z_j = z_j \mid \boldsymbol{\Omega}_{1j}^{(t)}, \boldsymbol{\Omega}_2^{(t)}\right)}. \tag{2.71}$$

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 (2.71).

2.6.3 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 (2.54) ($\boldsymbol{\beta} = \beta_0$ 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 part of this chapter).

1. For tumor ROI:

   (a) For each voxel $i = 1, \ldots, n$, propose $\boldsymbol{\mu}_i^{prop}$ from (2.22).

   Accept $\boldsymbol{\mu}_i^{prop}$ with the probability (2.23).

   (b) Draw $[\Sigma \mid \{\boldsymbol{\mu}_i\}_{i=1}^n, \{\mathbf{Y}_i\}_{i=1}^n]$ from (2.18).

   (c) Draw $[\Psi \mid \{\boldsymbol{\mu}_i\}_{i=1}^n]$ from (2.19).

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 (2.24).

   (b) For each $i \in N_{i'}^e \cap \mathcal{S}$, propose $\boldsymbol{\nu}_i^{prop}$ from (2.24).

   Accept $\boldsymbol{\nu}_i^{prop}$ with probability (2.27).

   (c) Propose Ω^{prop} from (2.34).

   Accept Ω^{prop} with probability (2.35).

   (d) Propose $\Delta^{prop}$from (2.30).

   Accept Δ with probability (2.31).

3. Predict tumor response under null:

   (a) Draw $\left[\widetilde{\mathbf{Y}}_{i2} \mid \cdot\right]$, $i = 1, \ldots, n$, from (2.39).

   (b) For $i = 1, \ldots, n$, propose $\widetilde{\boldsymbol{\mu}}_{i2}^{prop}$ from (2.42).

   Accept $\widetilde{\boldsymbol{\mu}}_{i2}^{prop}$ with probability (2.43).

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

   (a) Calculate dKLD using equation (2.44) and pKLD using equation (2.45).

(b) Calculate cDS using equation (2.46) and cPS using equation (2.47).

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} (2.68).

      i. Draw latent variables $[d_j \mid Z_j = z_j, \mathbf{X}_j, \boldsymbol{\Omega}_{2K}], j = 1, \ldots, M$, from (2.50).

      ii. Draw $[v^{-1} \mid \boldsymbol{\beta}_K, K]$ from (2.56).

      iii. Draw $[\boldsymbol{\beta}_K \mid \mathbf{d}, v, \Theta_K, \mathcal{X}]$ from (2.54).

      iv. Draw $[\lambda \mid K]$ from (2.57).

      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 (2.58).

      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 (2.59).

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 (2.60).

iii. Draw the knot point(s) $t_{l,K+1}$, $l = 1, \ldots, L_{K+1}$, see (2.61).

iv. Draw $s_{l,K+1}$, $l = 1, \ldots, L_{K+1}$, see (2.62).

v. Accept the birth with probability α_{birth} (2.66).

(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} (2.67).

3. Draw latent variables $[d_j \mid Z_j = z_j, \mathbf{X}_j, \boldsymbol{\Omega}_{2K}], j = 1, \ldots, M$, from (2.50).

4. Draw $[v^{-1} \mid \boldsymbol{\beta}_K, K]$ from (2.56).

5. Draw $[\boldsymbol{\beta}_K \mid \mathbf{d}, v, \Theta_K, \mathcal{X}]$ from (2.54).

end oversample

End Iterate

2.6.4 Image Processing:

All the MR images for each subject are spatially co-registered by using the pre-treatment anatomical image as the reference data set. This step allows all images from 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 2.3). 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.

2.6.5 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 (2.36) and (2.38). Let

$$\begin{pmatrix} \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix} = |N^e_i \cap \mathcal{H}|^{-1} \sum_{i' \in N^e_i \cap \mathcal{H}} \begin{pmatrix} \boldsymbol{\mu}_{i'1} \\ \widetilde{\boldsymbol{\mu}}_{i'2} \end{pmatrix}. \tag{2.72}$$

Then

$$\begin{aligned} \pi\left[\widetilde{\mathbf{Y}}_{i2}, \widetilde{\boldsymbol{\mu}}_{i2} \mid \mathbf{Y}_{i1}, \boldsymbol{\mu}_{i1}, \cdot\right] \;\propto\; & \exp\left[-0.5 \begin{pmatrix} \mathbf{Y}_{i1} - \boldsymbol{\mu}_{i1} \\ \widetilde{\mathbf{Y}}_{i2} - \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix}^{\mathrm{T}} P \begin{pmatrix} \mathbf{Y}_{i1} - \boldsymbol{\mu}_{i1} \\ \widetilde{\mathbf{Y}}_{i2} - \widetilde{\boldsymbol{\mu}}_{i2} \end{pmatrix}\right] \times \\ & \exp\left[-0.5 \begin{pmatrix} \boldsymbol{\mu}_{i1} - \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} - \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix}^{\mathrm{T}} Q \begin{pmatrix} \boldsymbol{\mu}_{i1} - \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}_{i2} - \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix}\right], \end{aligned} \tag{2.73}$$

for all $i \in \mathcal{H}$. Now it is obvious that the density (2.73) 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 (2.73), further condition on the $\boldsymbol{\nu}_i$, $i \in \mathcal{S}$ so that (2.72) becomes

$$\begin{pmatrix} \boldsymbol{\mu}^*_{i1} \\ \widetilde{\boldsymbol{\mu}}^*_{i2} \end{pmatrix} = |N^e_i|^{-1} \left[\sum_{i' \in N^e_i \cap \mathcal{S}} \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\nu}_{i'2} \end{pmatrix} + \sum_{i' \in N^e_i \cap \mathcal{H}} \begin{pmatrix} \boldsymbol{\mu}_{i'1} \\ \widetilde{\boldsymbol{\mu}}_{i'2} \end{pmatrix} \right]. \tag{2.74}$$

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 (2.73). 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 (2.73) 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.

2.6.6 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

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 $\varepsilon_{i1} \sim N(0, 0.03)$, $\varepsilon_{i2} \sim$ 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 2.2). When $\theta = 0.05$ and $\sigma^2 = 0.01$ the simulated data follow the null response. From Table 2.2 (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 $[Y_{i2} \mid Y_{i1}] \sim$ N$(Y_{i1} + \eta_1, \sigma_1^2)$ and $[W_{i2} \mid W_{i1}] \sim$ N$(W_{i1} + \eta_2, \sigma_2^2)$, independently. We predict tumor response, $[\widetilde{Y}_{i2} \mid Y_{i1}] \sim$ 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 2.2. 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 2.3. 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 part of this chapter.

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 2.4 and 2.5 we graph the marginal posterior densities of the four summary statistics for two subjects. In Figure 2.4 we show them for the subject with the smallest tumor and in Figure 2.5 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.

Table 2.2: 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 2.3: Sensitivity analysis of different hyperprior distributions for λ in the proposed model. CCR_{CV} denotes the leave-one-out cross-validated classification rate. Mean and variance are calculated for different prior distributions.

Prior of λ	Prior mean	Prior variance	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

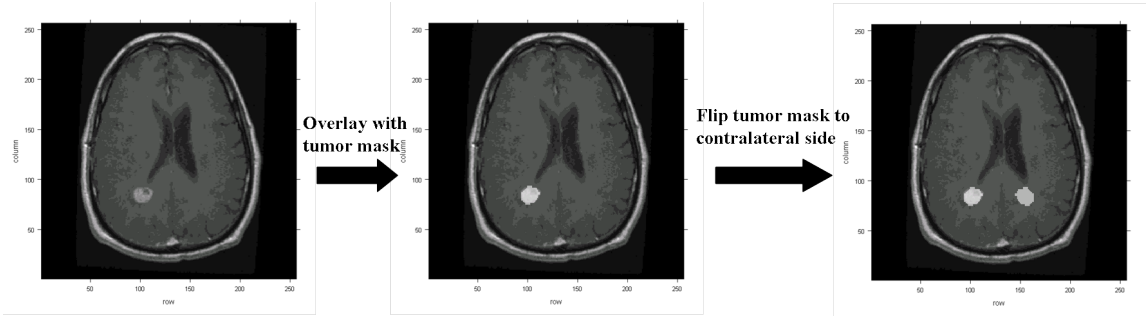

Figure 2.3: Obtaining the healthy tissue ROI. Left image: original T1-weighted contrast enhanced MRI; Middle image: overlay the original MRI with tumor mask to obtain the tumor region of interest; Right image: mirror the tumor mask to the contralateral hemisphere of the brain to get the healthy tissue region of interest.

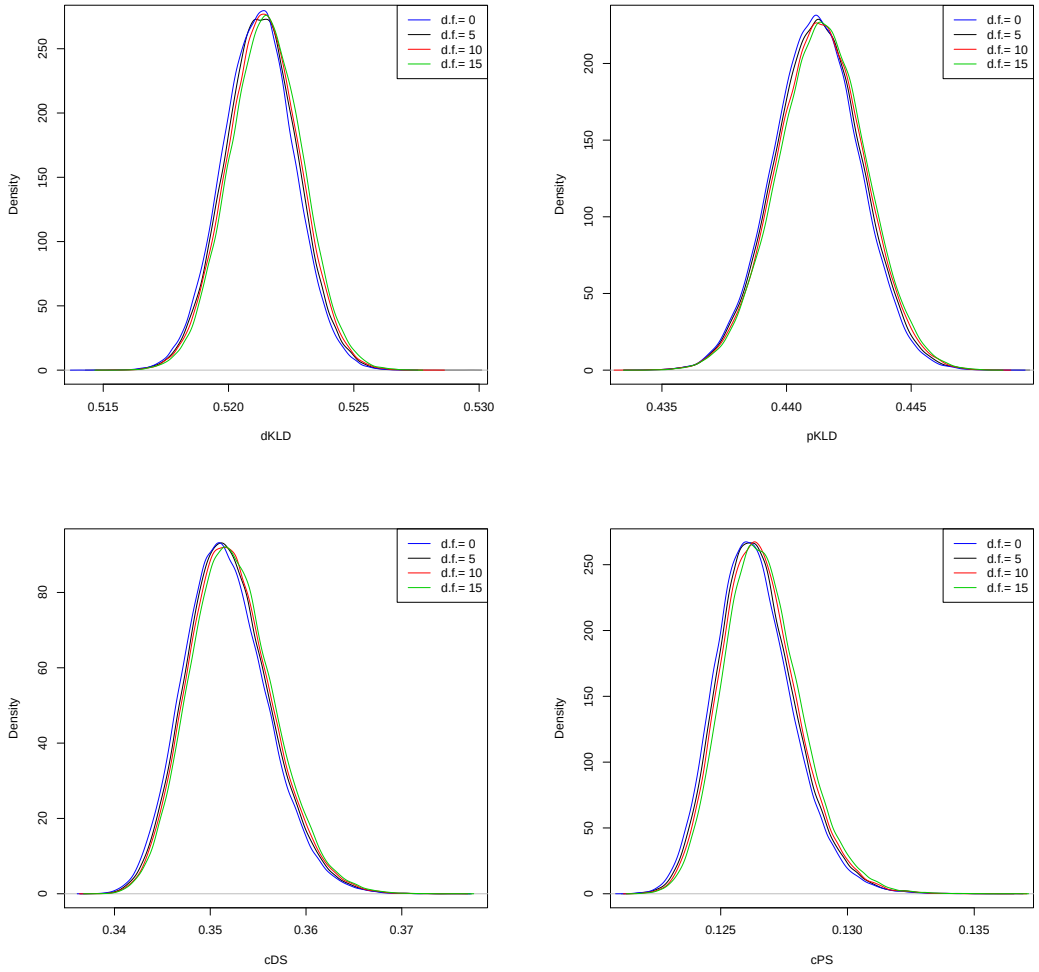

Figure 2.4: Sensitivity of the summary statistics to the prior number of degrees of freedom for the covariance matrices Σ, Ψ, Ω and Δ for the subject with the smallest tumor. The statistics are robust to changes in the degrees of freedom due to the large number of voxels in the tumors.

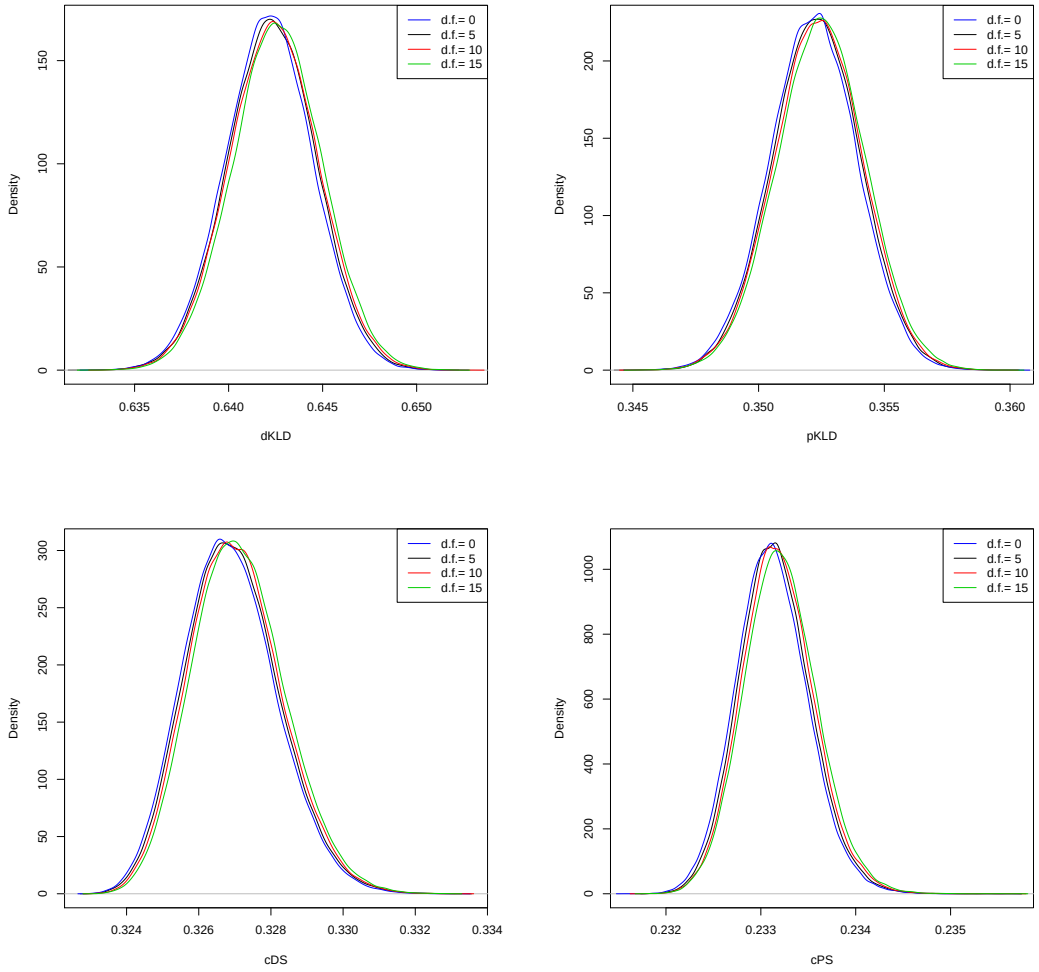

Figure 2.5: Sensitivity of the summary statistics to the prior number of degrees of freedom for the covariance matrices Σ, Ψ, Ω and Δ for a randomly selected subject. The statistics are robust to changes in the degrees of freedom due to the large number of voxels in the tumors.

CHAPTER III

A Bayesian Joint Survival Model for Assessing Treatment Efficacy via Quantitative MRI

3.1 Introduction

In Chapter II, in order to determine whether qMRI data can be used as an early predictor for treatment efficacy for patients with malignant gliomas, we dichotomize patients survival time and build a Bayesian joint classification model. However, this may lead to inefficient estimation due to a loss of information. A survival model is required to more efficiently assess the early predictive role of qMRI for predicting treatment efficacy. With this goal in mind, we propose a Bayesian joint survival model in this chapter. In the first stage, the qMRI data are smoothed using the multivariate pairwise difference prior (mPWDP) as detailed in Chapter II. We then derive summary statistics as described in Chapter II. In the second stage, we propose a Bayesian first hitting time survival model for patients' survival time with censoring. Patients' health status is modeled by a stochastic process and the survival time by the first passage time of the process sample path to a threshold. Parameters of the process and threshold are related to summary statistics by fitting a Bayesian hierarchical model with multivariate adaptive regression spline (MARS) basis functions (Friedman (1991)).

The most well-known and widely used model in survival analysis is the Cox pro-

portional hazards model. It is a semi-parametric model with an unspecified baseline hazard function comprising the nonparametric component and the regression coefficients comprising the parametric component. However, the simplicity of the Cox model imposes unrealistic assumptions on the data. Most significantly, the model assumes the proportionality of the hazard functions, and in many contexts, this is unacceptable. For example, in a study of patients' survival after surgery, age is a more important predictor of risk immediately after surgery, than some time after the surgery (after initial recovery). In our study, the summary statistics derived based on smoothed images in the first stage reflect the treatment effects on tumors only 3 weeks after the treatment starts. It is unreasonable to assume a proportional hazard assumption for patients with similar treatment effects, since the risk of death can change as treatment progresses. In other words, the hazard function will depend on whether the treatment is effective, but also for how long the treatment lasts. Furthermore, there may be other risk factors or competing risks that we do not consider in this study, which may also affect the hazard function.

In the second stage of our study, we propose a Bayesian first hitting time survival model, with summary statistics derived from the mPWDP model included as covariates. Previous research in several fields dealing with time-to-event data has considered models in which an event occurs when the sample path of a stochastic process first satisfies a specified condition. The time until this occurrence is the first hitting time. Lee and Whitmore (2006) proposed to model patients' health status as a latent stochastic process. Death occurs if the stochastic process reaches a certain threshold (i.e. hits level zero), so that the survival time of each patient can be modeled as the first hitting time (FHT) of the latent stochastic process. There are several advantages of the FHT model: (1) it is practical as the mechanism of

disease progression and death can be well described by the stochastic process; (2) it is flexible as different types of subject level covariates can be can accommodated in the model; (3) it is simple and easy to understand; (4) and most importantly, the FHT model does not require the validity of the proportional hazard assumption, which is a necessary assumption for the Cox proportional hazards model.

In the second stage of our joint survival model, we choose a Wiener diffusion process to model the latent health status process. One important reason to choose this kind of process is that the bidirectional movement of the Wiener process can capture the fluctuations of health status. The stochastic properties of the Wiener process makes it a suitable model for many physical processes that exhibit random variation over time. Moreover, there are many applications of the Wiener process to model survival and time-to-event data. Whitmore (1975) modeled hospital stay for mental illness using this process, Doksum and Hóyland (1992) applied this model to equipment degradation. Lu and Meeker (1993) considered longitudinal degradation and censored survival data in modeling equipment failure. Lee et al. (2004) assessed lung cancer risk in railroad workers with this model.

The parameters of the Wiener process are connected to the covariates via a Bayesian hierarchical model using multivariate adaptive regression splines (MARS) as basis functions. MARS basis functions allow us to build a more flexible and complex regression curve and can better explain the inherent dispersion of the data (Denison et al. (1998) and Holmes and Denison (2003)). In the Bayesian hierarchical model, hyper-prior distributions are placed on the parameters of Wiener process.

The current research builds on previous work by 1) accounting for spatio-temporal correlation in the images via the multivariate pairwise difference prior; 2) extending the first hitting time model into the Bayesian framework and introducing a Bayesian

hierarchical model; 3) incorporating the multivariate adaptive regression spline basis into the regression structure; 4) proposing a Bayesian joint survival model which allows us to draw inference directly on glioma patients' survival data with censoring. Our proposed survival model in the second stage is particularly suited to the prediction of the survival functions of censored data because it is easy to obtain the predictive estimates for the survival function from the posterior draws. The Bayesian paradigm allows a stochastic search over the model space which can improve the predictions made from models derived deterministically. This is commonly done by sampling from the posterior distributions of the model given the data.

This chapter is organized as follows. In Section 3.2, we introduce the notation and specify the joint model. In Section 3.3, we discuss posterior simulation. In Section 3.4, we check our proposed model through model diagnostic methods. In Section 3.5, we present results of simulation studies. Results from the motivating study are presented in Section 3.6. We conclude by summarizing the strengths and limitations of our approach, and discuss future work in Section 3.7. We include detailed mathematical derivations of the posterior distributions, algorithm details and pseudo code in the Appendix in Section 3.8.

3.2 Model

3.2.1 Bayesian FHT Regression Model

In our Bayesian FHT regression model, the fluctuations in patients' health status are described by a stochastic process. We denote this stochastic process in time as $\{X(t), t \geq 0\}$ with an absorbing set $\mathcal{B}$ in the state space of the process. The first hitting time is the random variable $\mathcal{T}$ defined as: $\mathcal{T} = \{t : X(t) \in \mathcal{B}\}$. In other words, the first hitting time is the time until the stochastic process first enters or hits set $\mathcal{B}$. With censoring survival data, the stochastic process $X(t)$ is latent

(unobservable) and the absorbing state is level 0: the death of the patient.

We use a Wiener diffusion process to model the latent health status process. The essential structure of the FHT model of a Wiener process is illustrated in Figure 3.1. The Wiener process is well defined by three subject-level parameters: the initial process value $X(0)$, the mean parameter η and the variance parameter σ^2. These three parameters describe the trajectory of the latent health status sample path. The parameter $X(0)$ defines the initial health status with larger values of $X(0)$ indicating a better initial health of the subject with respect to the disease. The mean parameter η denotes the rate at which a subject's health status decreases to the threshold (i.e. zero). The parameter σ^2 represents the inherent variability of the process. In our model, we set the threshold to zero (where death occurs) (Figure 3.1) and the latent health status has an arbitrary unit. Therefore, we can rescale the trajectory based on the value of σ^2. Without loss of generality, we fix the interval of the health status scale by setting σ^2 to one. We assume that these parameters depend on covariates that vary across individuals and are connected to the covariates through suitable regression link functions.

The FHT, $\mathcal{T}$, for a Wiener process has an inverse Gaussian distribution (Medhi (1994)) given by:

$$\pi\left(\mathcal{T}|X(0),\eta,\sigma\right)=\left(2\pi\sigma^2\mathcal{T}^3\right)^{-1/2}X(0)\exp\left\{-\left(2\sigma^2\mathcal{T}\right)^{-1}\left(X(0)-\eta\mathcal{T}\right)^2\right\} \tag{3.1}$$

where $\eta > 0$ and $X(0) > 0$.

We denote the observed time for the j^{th} patient as T_j and the failure indicator as δ_j with a value of one indicating that the observed T_j is a FHT. In other words, we have $\mathcal{T}_j = T_j$ if $\delta_j = 1$ and $\mathcal{T}_j > T_j$ if $\delta_j = 0$ for $j = 1, \ldots, M$. The covariate vector for the j^{th} patient is denoted by $\boldsymbol{Z}_j$. The parameter η_j and $X_j(0)$ of the Wiener

process are linked to relevant covariates by:

$$\log \eta_j = \mathbf{B}_1 \left(\boldsymbol{Z}_j \right) \boldsymbol{\alpha} + \epsilon_j, \tag{3.2}$$

$$\log \left(X_j(0) \right) = \mathbf{B}_2 \left(\boldsymbol{Z}_j \right) \boldsymbol{\beta} + \epsilon'_j \tag{3.3}$$

where $\mathbf{B}_i \left(\cdot \right)$, $i = 1, 2$, is a row vector of basis functions (defined later in this section), $\epsilon_j \sim N \left(0, \omega^2 \right)$, $\epsilon'_j \sim N \left(0, \omega'^2 \right)$. The regression coefficient vectors $\boldsymbol{\beta} = \{\beta_0, \ldots, \beta_K\}$ and $\boldsymbol{\alpha} = \{\alpha_0, \ldots, \alpha_K\}$ where β_0 and α_0 are model intercepts. By introducing random errors into formula (3.2) and (3.3), we transform the fixed effects models into random effects models.

In the Bayesian framework, the random effects model is equivalent to a Bayesian hierarchical model. We assign priors to the coefficient parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and to the precisions and rewrite formula (3.2) and (3.3) as:

$$\begin{aligned} \left[\eta_j | \mathbf{B}_1 \left(\boldsymbol{Z}_j \right), \boldsymbol{\alpha}, \omega \right] &\sim \text{log-Norm} \left(\mathbf{B}_1 \left(\boldsymbol{Z}_j \right) \boldsymbol{\alpha}, \omega^2 \right), \\ \left[\boldsymbol{\alpha} | \kappa \right] &\sim \text{N} \left(0, \kappa \boldsymbol{I}_{K+1} \right), \end{aligned} \tag{3.4}$$

$$\begin{aligned} \left[X_j(0) | \mathbf{B}_2 \left(\boldsymbol{Z}_j \right), \boldsymbol{\beta}, \omega' \right] &\sim \text{log-Norm} \left(\mathbf{B}_2 \left(\boldsymbol{Z}_j \right) \boldsymbol{\beta}, \omega'^2 \right), \\ \left[\boldsymbol{\beta} | \lambda \right] &\sim \text{N} \left(0, \lambda \boldsymbol{I}_{K+1} \right), \end{aligned} \tag{3.5}$$

where log-Norm represents the log normal distribution. We assign diffuse prior distributions, Gamma $(0.001, 0.001)$, to the precision parameters: ω^{-2}, κ^{-1}, ω'^{-2} and λ^{-1}.

We use the multivariate adaptive regression splines (MARS, Friedman (1991), Denison et al. (1998) and Holmes and Denison (2003)) as the basis functions, defined by:

$$B_k(\mathbf{Z}_j) = \begin{cases} 1, & k = 0, \\ \prod_{l=1}^{L_k} \left[s_{lk} (Z_{j w_{lk}} - t_{lk}) \right]_+, & k = 1, 2, \ldots, K. \end{cases}$$

where the function $[\,\cdot\,]_+ = \max(0\ ,\ \cdot\ )$. The parameter K denotes the number of basis functions in the model. L_k is the degree of interaction in the basis $B_k(\,\cdot\,)$ and the maximum number of interactions in any basis function is 2. The w_{lk} are the predictor indices corresponding to the knots t_{lk}. All w_{lk} are constrained to be distinct so that each predictor only appears once in each interaction term to maintain the linear nature of the basis functions. The t_{lk} is restricted to the set of predictor values $\{Z_{1w_{lk}}, \ldots, Z_{Mw_{lk}}\}$. We use $\mathbf{B}_i,\ i = 1, 2$ to denote the matrix of BMARS basis functions with (j,k)th element $B_{i,k}(\mathbf{Z}_j)$. Let $\mathbf{B}_i(\boldsymbol{Z}_j)$ denote the jth row vector of the matrix $\mathbf{B}_i,\ i = 1, 2$, in distributions (3.4) and (3.5).

We place priors on the MARS parameters. We assume that the number of basis functions, K, is unknown and we assign it a Poisson prior with parameter τ, where τ follows Gamma$(1, 0.2)$. The impact of this informative prior on the results of our model is later assessed in the sensitivity analysis (Section 3.8). For the other parameters in the MARS basis, we specify non-informative priors – discrete uniform priors over the possible values of parameters (i.e. over $\{-1, 1\}$ for s_{lk}, over $\{1, 2\}$ for L_k, over $\{Z_{1w_{lk}}, \ldots, Z_{Mw_{lk}}\}$ for t_{lk} and over all predictor indices for w_{lk}).

3.2.2 Multivariate Pairwise Difference Prior Model and the Derivation of Summary Statistics

We now describe how we derive the covariate vector $\boldsymbol{Z}_j$ for each patient j from the mPWDP model based on the image data. We begin by introducing some notation in the mPWDP model. In the image data, the smallest unit is called a voxel. For each patient j, tumor voxels are indexed by $i = 1, 2, \ldots, n_j$, where n_j ranges from 770 to 20380 with a mean of 6143 and standard error of 4721. We define two voxels to be neighbors if they share a common face, denoted by $i \sim i'$. The set of neighbors for voxel i is then denoted by $N_i = \{i' : i' \sim i\}$, so for patient j, this set is denoted

by $N_{i,j}$. The number of neighbors is then defined by $|N_{i,j}|$. The image intensity at voxel i for patient j, is denoted by $Y_{ith,j}$ with t indexing time (1 at baseline and 2 week three into therapy) and h indexing image type (1 for diffusion MRI and 2 for perfusion MRI). Therefore, for each voxel i, we can write $\mathbf{Y}_{i,j} = (\mathbf{Y}_{i1,j}, \mathbf{Y}_{i2,j})^{\mathrm{T}} = (Y_{i11,j}, Y_{i12,j}, Y_{i21,j}, Y_{i22,j})^{\mathrm{T}}$ with the subvector $\mathbf{Y}_{it,j} = (Y_{it1,j}, Y_{it2,j})^{\mathrm{T}}$ for $t = 1, 2$.

We assume that all the $\mathbf{Y}_{i,j}$ are conditionally independent given the means and covariance matrix: $\left[\mathbf{Y}_{i,j} \mid \boldsymbol{\mu}_{i,j}, \Sigma_j\right] \sim \mathrm{N}\left(\boldsymbol{\mu}_{i,j}, \Sigma_j\right)$, with $\boldsymbol{\mu}_{i,j} = \left(\boldsymbol{\mu}_{i1,j}, \boldsymbol{\mu}_{i2,j}\right)^{\mathrm{T}} = (\mu_{i11,j}, \mu_{i12,j}, \mu_{i21,j}, \mu_{i22,j})^{\mathrm{T}}$ and $\boldsymbol{\mu}_{it,j} = (\mu_{it1,j}, \mu_{it2,j})^{\mathrm{T}}$. We employ a multivariate pairwise difference prior (mPWDP) on $\boldsymbol{\mu}_{i,j}$: $\left[\boldsymbol{\mu}_{i,j} | \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}}, \Psi_j\right] \sim \mathrm{N}\left(\boldsymbol{\mu}^*_{i,j}, |N_{i,j}|^{-1}\Psi_j\right)$, where $\boldsymbol{\mu}^*_{i,j} = |N_{i,j}|^{-1} \sum_{i' \in N_{i,j}} \boldsymbol{\mu}_{i',j}$. We assign conjugate priors to Σ_j and Ψ_j: the inverse Wishart distribution: $\mathrm{W}^{-1}(\mathbf{I}_4, 5)$ with degrees of freedom of 5 and a 4×4 identity scale matrix $\mathbf{I}_4$.

To evaluate treatment efficacy, we would like to compare the observed tumor response to its counterfactual: tumor response given no treatment in an ideal situation. However, all the patients have received the treatment in our study. Therefore, our summary statistics are derived based on comparing the observed tumor response under treatment to the predicted tumor response in the contralateral hemisphere of the brain under the "null" scenario. We assume that under the "null" scenario, the changes in diffusion/perfusion of the tumor, if they could be observed, and those of healthy tissue in the contralateral brain are similar. The healthy tissue in the contralateral hemisphere of the brain receives relatively low radiation as the treatment focuses on destroying diseased tissue and sparing healthy tissue. The treatment should have minimal effect to that region. We define the healthy tissue region of interest (ROI) by reflecting the tumor ROI, approximately about the midline of the brain, to the contralateral hemisphere of the brain (non-brain tissue structures are

excluded, details can be found in the Section 3.8 of Chapter II). Let $\tilde{\mu}_{i2h,j}$ denote the predicted tumor response under the null for voxel i at time point 2 for image type h ($h = 1, 2$) for patient j. Conditional on $\mu_{i1h,j}$, we estimate $\tilde{\mu}_{i2h,j}$ from the posterior predictive distribution of the healthy tissue obtained by fitting the mPWDP model separately for voxels in the healthy tissue ROI (details can be found in Section 3.8 of Chapter II). Then we derive summary statistics by comparing the predicted tumor null response ($\tilde{\mu}_{i2h,j}$) with the observed tumor response ($\mu_{i2h,j}$) for $h = 1, 2$.

We hypothesize that the distributional difference should be noticeable between the posterior draws of $\tilde{\mu}_{i2h,j}$ and $\mu_{i2h,j}$, if the treatment is effective. Previous work (Hamstra et al. (2005), Moffat et al. (2005) and Hamstra et al. (2008)) has found that tumor response is highly heterogeneous and the mean change in tumor ADC values is not as predictive of tumor response as other distributional changes. Therefore, we propose to use the Kullback-Leibler divergence (Kullback and Leibler (1951)) statistic to measure the distributional differences between $\tilde{\mu}_{i2h,j}$ and $\mu_{i2h,j}$ over all tumor voxels and denote it as dKLD_j for diffusion images and pKLD_j for perfusion images for patient j. Moreover, we also notice that the changes in the tails of the distributions are more pronounced. As successful therapy should result in tumor cells lysing with a corresponding increase in ADC, but a decrease in rCBF, as discussed in the introduction, we propose 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 ($\mu_{i21,j}$) that is greater than the 0.975 quantile of the conditional distribution of the tumor response under the "null": $\text{cDS}_j = n_j^{-1} \sum_{i=1}^{n_j} I\left(\mu_{i21,j} > q_{0.975}\left(\tilde{\mu}_{i21,j}\right)\right)$ for patient j (Section 3.8 of Chapter II). $I(\cdot)$ is the indicator function and $q_{0.975}\left(\tilde{\mu}_{i21,j}\right)$ is the 0.975 quantile of the conditional posterior distribution of $\tilde{\mu}_{i21,j}$. The summary measure cPS is similarly

defined: $\text{cPS}_j = n_j^{-1} \sum_{i=1}^{n_j} I\left(\mu_{i22,j} < q_{0.025}\left(\tilde{\mu}_{i22,j}\right)\right)$, where $q_{0.025}\left(\tilde{\mu}_{i22,j}\right)$ is the 0.025 quantile of the conditional posterior distribution of $\tilde{\mu}_{i22,j}$ for patient j. The four summary statistics are then included as covariates into the Bayesian FHT regression model.

3.2.3 Joint Modeling

Typically covariates in the FHT model will be assumed to be measured without error. In our case the covariates in the FHT model are functionals of the mPWDP model parameters. Let $\boldsymbol{Z}_j = \{Z_{j1}, Z_{j2}, Z_{j3}, Z_{j4}\}$, representing the four summary measures: dKLD_j, pKLD_j, cDS_j and cPS_j derived in the mPWDP model for patient j. Since these summary measures are not observed, taking a point estimate of the covariates from the mPWDP model and treating them as fixed and known in the Bayesian FHT regression model will result in biased variance estimation. Therefore, we specify the full joint posterior of all model parameters in a full Bayesian model.

We use notation $\boldsymbol{\Omega}_1$ to denote the set of all parameters in the mPWDP model and $\boldsymbol{\Omega}_2$ to denote the set of all parameters in the Bayesian FHT regression model. We denote $\boldsymbol{\Omega} = \boldsymbol{\Omega}_1 \cup \boldsymbol{\Omega}_2$. The set $\boldsymbol{\Omega}_1 = \cup_{j=1}^{M} \boldsymbol{\Omega}_{1j}$. The summary statistics $\boldsymbol{Z}_j$ for patient j is then a functional vector of $\boldsymbol{\Omega}_{1j}$: $\boldsymbol{Z}_j = F(\boldsymbol{\Omega}_{1j})$. Thus, in the Bayesian FHT model, $\pi(\boldsymbol{T}_j, \boldsymbol{\delta}_j \mid \boldsymbol{Z}_j, \boldsymbol{\Omega}_2) = \pi(\boldsymbol{T}_j, \boldsymbol{\delta}_j \mid \boldsymbol{\Omega}_{1j}, \boldsymbol{\Omega}_2)$.

Let $\mathcal{Y}$ denote the set of all images for all subjects, let $\boldsymbol{T}$ denote the vector of the observed survival times and let $\boldsymbol{\delta}$ denote the vector of censoring indicators. The posterior distribution can be factored as follows:

$$\pi(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2 \mid \mathcal{Y}, \boldsymbol{T}, \boldsymbol{\delta}) \propto \pi(\mathcal{Y} \mid \boldsymbol{\Omega}_1)\pi(\boldsymbol{\Omega}_1)\pi(\boldsymbol{T}, \boldsymbol{\delta} \mid \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)\pi(\boldsymbol{\Omega}_2). \tag{3.6}$$

In the mPWDP model, we obtain estimates for $\boldsymbol{\Omega}_1$ from the conditional posterior

distribution $[\boldsymbol{\Omega}_1|\mathcal{Y}, \boldsymbol{T}, \boldsymbol{\delta}]$:

$$\pi(\boldsymbol{\Omega}_1|\mathcal{Y}, \boldsymbol{T}, \boldsymbol{\delta}) \tag{3.7}$$
$$\propto \pi(\mathcal{Y} \mid \boldsymbol{\Omega}_1)\pi(\boldsymbol{\Omega}_1)\pi(\boldsymbol{T}, \boldsymbol{\delta} \mid \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) \propto \pi(\boldsymbol{\Omega}_1 \mid \mathcal{Y})\pi(\boldsymbol{T}, \boldsymbol{\delta} \mid \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2).$$

Similarly, we obtain estimates for $\boldsymbol{\Omega}_2$ from

$$\pi(\boldsymbol{\Omega}_2|\boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{\Omega}_1) \propto \pi(\boldsymbol{T}, \boldsymbol{\delta} \mid \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)\pi(\boldsymbol{\Omega}_2). \tag{3.8}$$

3.3 Implementation

3.3.1 Sampling from the Posterior

Below, we outline our sampling algorithm. Details and pseudo-code can be found in the Section 3.8.

As we fix $\sigma^2 = 1$, we rewrite the likelihood of the FHT for each individual as:

$$\pi\left(\mathcal{T}_j|X_j(0), \eta_j\right) = \left(2\pi\mathcal{T}_j^3\right)^{-1/2} X_j(0) \exp\left\{-\left(2\mathcal{T}_j\right)^{-1}\left(X_j(0) - \eta_j\mathcal{T}_j\right)^2\right\}. \tag{3.9}$$

If $\delta_j = 1$, $\mathcal{T}_j = T_j$. Otherwise, we impute the failure time from the truncated inverse Gaussian distribution with density (3.9) given $\mathcal{T}_j > T_j$.

Given $\boldsymbol{\mathcal{T}}$, we update MARS parameters, FHT parameters $\eta, X(0)$, regression coefficients $\boldsymbol{\beta}, \boldsymbol{\alpha}$ and precisions $\omega^{-2}, \omega'^{-2}, \kappa^{-2}, \lambda^{-2}$. Since the parameter η_j follows the log normal distribution with density:

$$\pi\left(\eta_j|\mathbf{B}_1(\boldsymbol{Z}_j)\boldsymbol{\alpha}, \omega\right) = \frac{1}{\eta_j\omega\sqrt{2\pi}} \exp\left\{-\frac{\left(\log\eta_j - \mathbf{B}_1(\boldsymbol{Z}_j)\boldsymbol{\alpha}\right)^2}{2\omega^2}\right\},$$

the posterior distribution of the parameter η_j is:

$$\begin{aligned}
[\eta_j|\cdot] &\propto \pi\left(\mathcal{T}_j|X_j(0), \eta_j\right)\pi\left(\eta_j|\mathbf{B}_1(\boldsymbol{Z}_j)\boldsymbol{\alpha}, \omega\right) \\
&\propto \exp\left\{-\log\eta_j - \frac{\eta_j^2\mathcal{T}_j^2 - 2X_j(0)\eta_j\mathcal{T}_j}{2\mathcal{T}_j} - \frac{\left(\log\eta_j\right)^2 - 2\mathbf{B}_1(\boldsymbol{Z}_j)\boldsymbol{\alpha}\ln\eta_j}{2\omega^2}\right\}.
\end{aligned} \tag{3.10}$$

The distribution (3.10) does not have a nice form from which we can directly sample and is not a log-concave function, therefore, we apply the adaptive rejection Metropolis sampling algorithm (ARMS, Gilks et al. (1995)) to simulate draws from the posterior.

Once we obtain draws of η_j for $j = 1, \ldots, M$, the model can be viewed as a linear regression (formula (3.2), (3.3)). Assume $\zeta_j = \log \eta_j$ and $\boldsymbol{\zeta} = \{\zeta_1, \ldots, \zeta_M\}$, then:

$$\pi(\boldsymbol{\zeta}|\boldsymbol{\alpha}, \mathbf{B}_1, \omega) = (2\pi\omega^2)^{-M/2} \exp\left\{-\frac{(\boldsymbol{\zeta} - \mathbf{B}_1\boldsymbol{\alpha})^{\mathrm{T}}(\boldsymbol{\zeta} - \mathbf{B}_1\boldsymbol{\alpha})}{2\omega^2}\right\}. \tag{3.11}$$

The posterior distribution of $[\boldsymbol{\alpha}|\boldsymbol{\zeta}, \kappa, \omega, \mathbf{B}_1]$ is $\mathrm{N}(\mathbf{m}^\star, \mathbf{V}^\star\omega^{\mathbf{2}})$ where $\mathbf{V}^\star = [\left(\frac{\kappa}{\omega^2}\mathbf{I}_{K+1}\right)^{-1} + \mathbf{B}_1^{\mathrm{T}}\mathbf{B}_1]^{-1}$ and $\mathbf{m}^\star = \mathbf{V}^\star\mathbf{B}_1^{\mathrm{T}}\boldsymbol{\zeta}$.

We assign a diffuse conjugate hyper-prior, Gamma$(0.001, 0.001)$, to the parameters κ^{-1} and ω^{-2}. Thus we update κ^{-1} from its full conditional: Gamma$(0.001 + (K+1)/2, 0.001 + \boldsymbol{\alpha}^{\mathbf{T}}\boldsymbol{\alpha}/2)$ and ω^{-2} from its full conditional: Gamma$(0.001 + M/2, 0.001 + \sum_{j=1}^{M}[\zeta_j - \mathbf{B}_1(\boldsymbol{Z}_j)\boldsymbol{\alpha}]^2/2)$. Moreover, we assume that the number of basis functions, K, is unknown a priori and assign a prior distribution: $[K|\tau] \sim$ Poisson(τ) and $[\tau] \sim$ Gamma$(1, 0.2)$. The posterior of τ is Gamma$(1 + K, 0.2 + 1)$.

For the MARS parameters in $\mathbf{B}_1$, we let $\mathbf{L} = \{L_1, \ldots, L_K\}$, $\mathbf{s} = \{s_{11}, \ldots, s_{L_K K}\}$, $\mathbf{w} = \{w_{11}, \ldots, w_{L_K K}\}, \mathbf{t} = \{t_{11}, \ldots, t_{L_K K}\}$ and $\Theta_1 = \{K, \mathbf{s}, \mathbf{w}, \mathbf{t}, \mathbf{L}\}$. The dimension of Θ_1 is allowed to vary at each iteration as K changes. Hence, the column dimension of $\mathbf{B}_1$ varies as does the dimension of $\boldsymbol{\alpha}$. 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 and are carried out via reversible jump MCMC (RJMCMC) (Green (1995)). Details of the RJMCMC algorithm and pseudo code for sampling from the posterior distribution of our joint model are given in Section 3.8.

Similarly, the parameter $X_j(0)$ follows a log normal distribution with mean and variance: $\mathbf{B}_2(\boldsymbol{Z}_j)\boldsymbol{\beta}$ and ω'^2 respectively. The posterior distribution of $X_j(0)$ for each subject is:

$$[X_j(0)|\cdot] \propto \exp\left\{-\frac{X_j(0)^2 - 2X_j(0)\eta_j\mathcal{T}_j}{2\mathcal{T}_j} - \frac{(\ln X_j(0))^2 - 2\mathbf{B}_2(\boldsymbol{Z}_j)\boldsymbol{\beta}\ln X_j(0)}{2\omega'^2}\right\}. \tag{3.12}$$

The distribution (3.12) is not a log-concave function, therefore, we apply the adaptive rejection Metropolis sampling algorithm (ARMS) to simulate draws from the full conditional.

We define the set of MARS parameters in $\mathbf{B}_2$ as Θ_2. The posterior draws of $\boldsymbol{\beta}$ and all the parameters in Θ_2 can be derived in the same way as the procedure of obtaining $\boldsymbol{\alpha}$ and those in Θ_1. Similarly, updating Θ_2 can be achieved using the same procedure as described for Θ_1 (see details in the Section 3.8).

The summary statistics are obtained by the posterior draws of $\tilde{\mu}_{i2h,j}$ and $\mu_{i2h,j}$ for $h = 1, 2$. We sample the parameters in the mPWDP model from their posterior full conditional distributions. However, the full conditional distributions for $\boldsymbol{\mu}_{i,j}$, $\widetilde{\boldsymbol{\mu}}_{i2,j}$ for $i = 1, \ldots, n_j$ and $j = 1, \ldots, M$ in the tumor ROI and the parameters in healthy tissue ROI do not have closed forms, hence we use the Metropolis-Hasting algorithm to sample from their full conditionals (details can be found in the Section 3.8).

3.3.2 Model Inference

We predict the survival functions for patients who are censored by obtaining predictive estimates for the survival function from the posterior draws. The survival function is denoted by $S(\mathcal{T})$ and is:

$$S(\mathcal{T}|X(0), \eta) = \Phi\left[\frac{(-\eta\mathcal{T} + X(0))}{\mathcal{T}^{1/2}}\right] - \exp\{2X(0)\eta\}\,\Phi\left[\frac{-\eta\mathcal{T} - X(0)}{\mathcal{T}^{1/2}}\right]. \tag{3.13}$$

By Bayesian model averaging, we approximate formula (3.13) using MCMC draws of $X(0)$ and η by $\sum_{d=1}^{D} S(\mathcal{T}|X^{(d)}(0), \eta^{(d)})/D$, where superscript (d) indicates the value at iteration d.

3.4 Model Evaluation

We evaluate our proposed joint model via a leave-one-out cross-validation approach. We denote $\boldsymbol{T} = \{T_1, \ldots, T_M\}$. We denote the remaining data by $\boldsymbol{T}_{\{-j\}}$ when the j^{th} observation is left out and we want to calculate $S(\mathcal{T}_j|\mathcal{Y}, \boldsymbol{T}_{\{-j\}}, \boldsymbol{\delta}_{\{-j\}})$, which is the predicted survival curve. By using the importance sampling method proposed by Gelfand et al. (1992) and Bayesian model averaging, we only need to estimate the posterior distribution of the full data once and then estimate $S(\mathcal{T}_j|\mathcal{Y}, \boldsymbol{T}_{\{-j\}}, \boldsymbol{\delta}_{\{-j\}})$. The cross-validated predicted survival function is given by:

$$\text{If } \delta_j = 1, \quad \widehat{\mathrm{S}(\mathcal{T}_j|\mathcal{Y}, \boldsymbol{T}_{\{-j\}}, \boldsymbol{\delta}_{\{-j\}})} = \frac{\sum_{d=1}^{D} S(\mathcal{T}_j|\boldsymbol{\Omega^{(d)}})/f(T_j, \delta_j, |\boldsymbol{\Omega^{(d)}})}{\sum_{d=1}^{D} 1/f(T_j, \delta_j, |\boldsymbol{\Omega^{(d)}})},$$

$$\text{If } \delta_j = 0, \quad \widehat{\mathrm{S}(\mathcal{T}_j|\mathcal{Y}, \boldsymbol{T}_{\{-j\}}, \boldsymbol{\delta}_{\{-j\}})} = \frac{\sum_{d=1}^{D} S(\mathcal{T}_j|\boldsymbol{\Omega^{(d)}})/\Pr(\mathcal{T}_j > T_j, \delta_j|\boldsymbol{\Omega^{(d)}})}{\sum_{d=1}^{D} 1/\Pr(\mathcal{T}_j > T_j, \delta_j|\boldsymbol{\Omega^{(d)}})}.$$

We further check the fit of our model to data by simulating values of a discrepancy measure from the posterior predictive distribution and compare these samples to the sample from the observed data. We adopt the model checking method proposed by Gelman et al. (1996). Assume we observe the failure time $\boldsymbol{\mathcal{T}}$ for all the subjects and H is our proposed model. To avoid confusion with the observed data, $\boldsymbol{\mathcal{T}}$, define $\boldsymbol{\mathcal{T}}^{\text{rep}}$ as the replicated data that is drawn from the posterior predictive distribution given model H: $f(\boldsymbol{\mathcal{T}}^{\text{rep}}|\boldsymbol{\mathcal{T}}, H) = \int f(\boldsymbol{\mathcal{T}}^{\text{rep}}|\boldsymbol{\Omega}, H)\pi(\boldsymbol{\Omega}|\boldsymbol{\mathcal{T}}, H)d\boldsymbol{\Omega}$.

Then we calculate the posterior predictive tail probability as:

$$p = \Pr\left[D(\boldsymbol{\mathcal{T}}^{\text{rep}}; \boldsymbol{\Omega}) \geq D(\boldsymbol{\mathcal{T}}; \boldsymbol{\Omega})|H, \boldsymbol{\mathcal{T}}\right]. \tag{3.14}$$

We adopt the χ^2 discrepancy as mentioned in Gelman, Meng and Stern (1996):

$$\chi^2(\boldsymbol{\mathcal{T}};\boldsymbol{\Omega}) = \sum_{j=1}^{M} \frac{(\mathcal{T}_j - \mathrm{E}(\mathcal{T}_j|\boldsymbol{\Omega}))^2}{\mathrm{Var}(\mathcal{T}_j|\boldsymbol{\Omega})}.$$

If all the $\boldsymbol{\mathcal{T}}$ are observed, then the calculation of p in (3.14) is straightforward. In our study, some of the subjects' survival times are censored and not observed. In this case, the definition of p must be slightly modified. We "impute" each right censored observation by sampling from the truncated posterior predictive distribution, where the truncation is taken to be larger than the observed censoring time T_j.

With the MC estimates $\boldsymbol{\Omega}^{(d)}$ for $d = 1, \ldots, D$, we can calculate $\chi^2(\boldsymbol{\mathcal{T}};\boldsymbol{\Omega}^{(d)})$ and $\chi^2(\boldsymbol{\mathcal{T}}^{\mathrm{rep}};\boldsymbol{\Omega}^{(d)})$. We can then draw a scatter plot to make a graphical assessment, and estimate p by the proportion of the D pairs for which $\chi^2(\boldsymbol{\mathcal{T}}^{\mathrm{rep}};\boldsymbol{\Omega}^{(d)})$ exceeds $\chi^2(\boldsymbol{\mathcal{T}};\boldsymbol{\Omega}^{(d)})$.

3.5 Simulation Studies

We describe several sets of simulation studies to evaluate the performance of our proposed Bayesian survival model. For each set, we generate $N = 1000$ simulations and compute the average relative bias ($\overline{\mathrm{rBias}}$) and average relative Mean Squared Error ($\overline{\mathrm{rMSE}}$) of the prediction of failure time for censored patients:

$$\begin{aligned}
\overline{\mathrm{rBias}}(\mathcal{T}) &= N^{-1} n_{cg}^{-1} \sum_{g=1}^{N} \sum_{j=1}^{n_{cg}} \left(\overline{\mathcal{T}}_{jg} - \mathcal{T}_{jg}^{True}\right) / \mathcal{T}_{jg}^{True}, \\
\overline{\mathrm{rMSE}}(\mathcal{T}) &= N^{-1} n_{cg}^{-1} \sum_{g=1}^{N} \sum_{j=1}^{n_{cg}} \left(\overline{\mathcal{T}}_{jg} - \mathcal{T}_{jg}^{True}\right)^2 / \left(\mathcal{T}_{jg}^{True}\right)^2
\end{aligned}$$

where $\overline{\mathcal{T}}_{jg}$ is the posterior mean and $\mathcal{T}_{jg}^{True}$ is the true failure time for jth censored patient of the gth simulation. We also calculate the 95% HPD (Highest Probability Density) interval of rBias. n_{cg} is the number of censoring subjects in each simulation g.

In the first set of simulations, we generate a covariate vector $\boldsymbol{Z}_j = \{Z_{1j}, j = 1, \ldots, 300\}$ independently from $N(2, 0.1)$. The first 200 observations are used as training data, while the remaining 100 observations are used as testing data. We fix the regression coefficients $\beta_0 = 0.5$, $\beta_1 = 1$, $\alpha_0 = 1.5$ and $\alpha_1 = -4$. Assume $X_j(0)^{True} = \exp\{\beta_0 + \beta_1(Z_{1j} - t_1)_+\}$ and $\eta_j^{True} = \exp\{\alpha_0 + \alpha_1(Z_{1j} - t_1')_+\}$, where the two distinct knots t_1 and t_1' are uniformly selected from $\boldsymbol{Z}_1$ without replacement. We then generate patients' failure times by simulating $\mathcal{T}_j$ from an inverse Gaussian distribution with parameters η_j^{True} and $X_j(0)^{True}$. The censoring time C_j is simulated independently from different uniform distributions so that the we can control the censoring rate at 10%, 30%, 50% and 70%. The results are compared in Table 3.1. As the censoring rate increases, $\overline{\text{rBias}}$ and $\overline{\text{rMSE}}$ also increase. This is reasonable as there are fewer observations to build the model. In all cases, the 95% credible intervals of rBias cover the null value 0.

In the second set of simulations, we check the performance of our model if the true distribution of subject's survival time is mis-specified. First, we assume the true underlying stochastic process of patients' health status is not a Wiener process, but a Gamma process. A Gamma process has monotonic (nondecreasing) sample paths, hence it is widely used to model the cumulative hazard function. Singpurwalla (1995) and Park and Padgett (2005) consider the Gamma process as a model for degradation. If the latent health status of each subject follows a Gamma process, the failure time (FHT) will follow the inverse Gamma distribution. We simulate the covariate vector $\boldsymbol{Z}_1$ in the same way as described before. We fix the regression coefficients $\beta_0 = 2.25$, $\beta_1 = 1$, $\alpha_0 = 1.6$ and $\alpha_1 = -0.6$. Assume $X_j(0)^{True} = \exp\{\beta_0 + \beta_1(Z_{1j} - t_1)_+\}$ and $\eta_j^{True} = \exp\{\alpha_0 + \alpha_1(Z_{1j} - t_1')_+\}$. We then generate patients' failure times by simulating $\mathcal{T}_j$ from the inverse gamma distribution with

scale parameter η_j^{True} and shape parameter $X_j(0)^{True}$. We simulate the censoring time C_j from Unif$[0, 2.5]$, which controls the censoring rate at approximately 30%. The results are shown in the second row of Table 3.2. We can see that the $\overline{\text{rBias}}$ and $\overline{\text{rMSE}}$ are very close to those if the true underlying distribution is correctly specified as an inverse Gaussian distribution (the first row of Table 3.2).

We also consider two other commonly used distributions in survival analysis: the Weibull and the log-normal distributions. In one case, we fix the regression coefficients $\beta_0 = 1$, $\beta_1 = 0.5$, $\alpha_0 = 0.09$ and $\alpha_1 = -0.1$ and sample $\mathcal{T}_j$ from a Weibull distribution with shape parameter $X_j(0)^{True}$ and scale parameter η_j^{True}. C_j follows Unif$[0, 3]$. Similarly, in the other case, we fix the regression coefficients $\beta_0 = 0.05$, $\beta_1 = 0.1$, $\alpha_0 = -1$ and $\alpha_1 = -0.1$ and sample $\mathcal{T}_j$ from a log-normal distribution with mean $X_j(0)^{True}$ and variance η_j^{True}. C_j follows Unif$[0, 8]$. In both cases, the censoring rates are controlled at 30%. The results are shown in the third and fourth rows of Table 3.2. We can see that in both cases, the $\overline{\text{rBias}}$ and $\overline{\text{rMSE}}$ do not differ much from those by assuming an inverse Gaussian distribution for $\mathcal{T}$.

In Figure 3.2, we plot the predicted survival curve and its 95% credible interval from one simulation for a randomly selected patient by using Bayesian model averaging. The true survival curve for that patient is also plotted for comparison. From the figures, in all cases when the true underlying distributions are mis-specified, the predicted survival curves are very close to the true survival curves and the 95% credible intervals all cover the true survival curves. Therefore, our proposed model provides robust prediction even when the true underlying distribution of survival time is mis-specified.

In the last set of simulations, we construct two more complicated MARS bases for parameter $X_j(0)^{True}$ and η_j^{True} by introducing more covariates and an interac-

tion term. We simulate 3 sets of covariate variables: $\boldsymbol{Z}_1$ from $N(2, 0.1)$, $\boldsymbol{Z}_2$ from $N(-1, 0.05)$ and $\boldsymbol{Z}_3$ from $N(0.5, 0.01)$, and fix the regression coefficients $\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 1.2$, $\alpha_0 = 1.5$, $\alpha_1 = -4$ and $\alpha_2 = -1$. Assume

$$\begin{aligned} X_j(0)^{True} &= \exp\{\beta_0 + \beta_1(Z_{1j} - t_1)_+ + \beta_2(Z_{2j} - t_2)_+\} \\ \eta_j^{True} &= \exp\{\alpha_0 + \alpha_1(Z_{1j} - t_1')_+ + \alpha_2(Z_{2j} - t_2')_+(Z_{3j} - t_3')_+\}. \end{aligned}$$

We simulate $\mathcal{T}_j$ from an inverse Gaussian distribution and also control the censoring rate at 30% by sampling C_j from Unif$[0, 8]$. We refer to this model as Model 2. The simple model (with 30% censoring rate) described in the first set of simulations is denoted by Model 1. The $\overline{\text{rBias}}$ and $\overline{\text{rMSE}}$ are shown in Table 3.3. The results are not affected by introducing more covariates and an interaction term, even if we include an irrelevant variable (referred as noise, follows Unif$(0, 1)$. In Table 3.4, we calculate the percentage of iterations that each covariate is included in the model as either a main effect term or an interaction term. We can see that covariates $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are predictive of variable $X(0)$ and are selected more than 95% of the time by the model, while $\boldsymbol{Z}_3$ is only selected around 20% of time. Similar results can be observed with parameter η. The irrelevant noise is included into the model less than 20% of time. We conclude that our model can also correctly select the most predictive covariates.

3.6 Real Data Application

We iteratively sample from the posterior distributions between the two stages. Stage I is computationally much more expensive than stage II due to the large number of voxels per subject. We run the algorithm for 100,000 iterations and we over-sample draws (10:1) in stage II. It takes around 28 hours for all 47 patients. The algorithms are programmed in C and implemented on a 3.0 GHz Mac Xserve.

The first 50,000 draws are discarded as burn-in. The burn-in period is chosen by examining plots of all the parameters, which have converged by this point.

Among 47 patients, 32 died and 15 were censored. The censoring rate of the gliomas data is around 30%. By using the posterior draws of η and $X(0)$, we estimate the survival curve and its 95% credible interval for each patient (Figures 3.3 and Section 3.8). We also calculate the median survival time: $\mathcal{T}_{0.5} : S(\mathcal{T}_{0.5}) = 0.5$, as well as its corresponding 95% credible interval for each patient. Of the 32 patients that died, 28 patients have a 95% credible interval that covers the truth (Section 3.8). For the 15 patients who were censored, we predict their survival curves and the 95% credible intervals conditional on that they survive beyond the censoring time: $\mathcal{T}_j > T_j$ (Figures 3.3 and Section 3.8). Our model provides investigators with an estimated survival curve for each patient (observed or censored), which may better help investigators manage and personalize treatment.

In Table 3.6, we calculate the percentages of iterations that our four summary statistics are included in either hierarchical model for parameter η or $X(0)$. All four statistics are included more than 80% of time and the KLD statistics for both the diffusion and perfusion MRI are included more often than the conditional diffusion/perfusion statistics in both hierarchical models. Therefore all four summary statistics derived based on qMRI appear to affect both levels of the sample path (as set by its initial starting point) and its rate of change over time. Several baseline prognostic factors (e.g. age, surgery type, Karnofsky performance score, Pathology grade, tumor size) were also included in stage II as covariates. However, none of them appeared to be strong predictors of patients' survival time. They were included in the model around $20\% - 30\%$ of the time.

We further check the fit of our model to data using the methods described in

Section 3.4. For the realized discrepancy, Figure 3.4 shows a scatterplot of the realized discrepancy, $\chi^2(\boldsymbol{\mathcal{T}};\boldsymbol{\Omega})$ and the predictive discrepancy, $\chi^2(\boldsymbol{\mathcal{T}}^{\text{rep}};\boldsymbol{\Omega})$, in which each point corresponds to a different value of $(\boldsymbol{\mathcal{T}}^{\text{rep}};\boldsymbol{\Omega})$ drawn from the posterior distribution. We also calculate the tail-area probability of the realized discrepancy as the probability that the realized discrepancy exceeds the predictive discrepancy, which, in this case, equals 35.5%, the proportion of points above the 45° line in the figure (ideally, 50% is the best fit). Therefore, we conclude that our model fits the data well.

As argued in the Introduction, the Cox model is not a suitable model for this data set because the proportional hazard assumption is violated. Further proof is given in Table 3.5. The posterior means of the four summary statistics: dKLD, pKLD, cDS and cPS, derived in stage I are included into a Cox Model. The proportionality assumption is checked by incorporating a time dependent term: dKLD$\times \log(\mathcal{T})$, cDS$\times \log(\mathcal{T})$ and cPS$\times \log(\mathcal{T})$ (Chow and Liu (2004)). From the table, we can see that the coefficients of the dKLD$\times \log(\mathcal{T})$ and cDS$\times \log(\mathcal{T})$ are significant, which indicates that the proportional hazard assumption is not valid here.

3.7 Discussion

In this chapter, we proposed a Bayesian joint survival model to assess early treatment efficacy based on qMRI data from patients with malignant gliomas. In stage I, we smooth the qMRI data via a spatio-temporal multivariate pairwise difference prior model and derive summary statistics. In stage II, we propose a Bayesian FHT regression model for patients' survival time with censoring and link the summary statistics derived in stage I to the distribution parameters of the FHT via a Bayesian hierarchical model. By accounting for the spatio-temporal correlation in the images,

we increase the signal to noise ratio. We reduce the data dimension by deriving summary statistics based on the posterior distributions of the smoothed images. The Bayesian FHT model that we propose in the second stage is computationally easy to implement and provides a rich, plausible and flexible modeling structure for the data. The Bayesian paradigm allows us to explore the parameter space randomly while average over all the visited models via Bayesian model averaging which accounts for model uncertainty. Moreover, in the Bayesian framework, it is easy to obtain predictive estimates for the survival function based on the posterior distribution.

Lee and Whitmore (2006) and Lee et al. (2004) introduced the concept of operational time into the FHT model to distinguish from calendar time. Calendar time measures time in months and years, while operational time measures the cumulative exposure of a system to aggregate physical effects that cause its deterioration, such as the wear-out of a car which is related more to usage than the simple passage of time. And the aggregate effects may relate to multiple causes of death, such as different working exposures (Lee et al. (2004)). In our study, since the patients with gliomas have a high mortality rate and very short median survival time, their will be the main cause of the death. For the sake of model simplicity and interpretation, we assumed a linear monotonic relationship between operational time and calendar time.

Table 3.1: Average rMSE , rBias of predictions and the 95% HPD interval of bias for different censoring rates.

Censoring Rate	$\overline{\text{rMSE}} \pm$ SD (10^{-2})	$\overline{\text{rBias}} \pm$ SD (10^{-2})	95% HPD Interval of Bias
10%	0.40 ± 0.29	2.40 ± 5.90	$(-0.114, 0.102)$
30%	0.53 ± 0.44	3.91 ± 6.21	$(-0.101, 0.117)$
50%	0.89 ± 0.65	6.87 ± 6.52	$(-0.092, 0.144)$
70%	1.89 ± 1.41	11.8 ± 7.09	$(-0.087, 0.173)$

Table 3.2: Average rMSE , rBias of predictions and the 95% HPD interval of bias for different true distributions, censoring rate is controlled at 30%.

True Distribution	$\overline{\text{rMSE}} \pm \text{SD}\ (10^{-2})$	$\overline{\text{rBias}} \pm \text{SD}\ (10^{-2})$	95% HPD Interval of Bias
Inverse Gaussian	0.53 ± 0.44	3.91 ± 6.21	$(-0.101, 0.117)$
Inverse Gamma	0.56 ± 0.59	4.10 ± 6.31	$(-0.092, 0.115)$
log-Normal	0.56 ± 0.54	4.12 ± 6.32	$(-0.091, 0.121)$
Weibull	0.64 ± 0.61	5.02 ± 6.28	$(-0.109, 0.126)$

Table 3.3: Average rMSE , rBias of predictions and the 95% HPD interval of bias for different models, censoring rates are controlled at 30%.

Model	$\overline{\text{rMSE}} \pm \text{SD}\ (10^{-2})$	$\overline{\text{rBias}} \pm \text{SD}\ (10^{-2})$	95% HPD Interval of Bias
Model 1	0.39 ± 0.30	2.87 ± 5.62	$(-0.093, 0.098)$
Model 2	0.36 ± 0.32	2.93 ± 5.23	$(-0.090, 0.101)$
Model 2 with noise	0.41 ± 0.31	3.01 ± 5.70	$(-0.085, 0.103)$

Table 3.4: Percentage of iterations that the predictors are selected by different models, censoring rates are controlled at 30%.

Predictor	Model 2 (%)		Model 2 with noise (%)	
	X_0	η	X_0	η
Z_1	96.2	97.7	95.5	97.1
Z_2	98.7	96.5	97.9	96.2
Z_3	21.0	95.4	15.6	95.8
noise	–	–	13.7	18.4

Table 3.5: Posterior means of summary statistics: dKLD, pKLD, cDS and cPS are included into a Cox Model. The proportionality assumption is checked via incorporating the time dependent term: dKLD$\times \log(\mathcal{T})$, cDS$\times \log(\mathcal{T})$ and cPS$\times \log(\mathcal{T})$.

Parameter	Coefficients	SD	P-value
dKLD	-1.15	0.55	0.12
pKLD	0.82	0.29	0.005
cDS	-0.92	0.34	< 0.001
cPS	1.24	0.56	0.11
dKLD$\times \log(\mathcal{T})$	-1.63	1.30	0.009
cDS$\times \log(\mathcal{T})$	-0.37	0.28	< 0.001
cPS$\times \log(\mathcal{T})$	0.78	0.57	0.08

Table 3.6: Percentage of iterations that the predictors are included in the model.

Predictor	$X(0)$	η
dKLD	94.8	96.1
pKLD	92.3	91.7
cDS	87.4	88.2
cPS	84.9	82.6

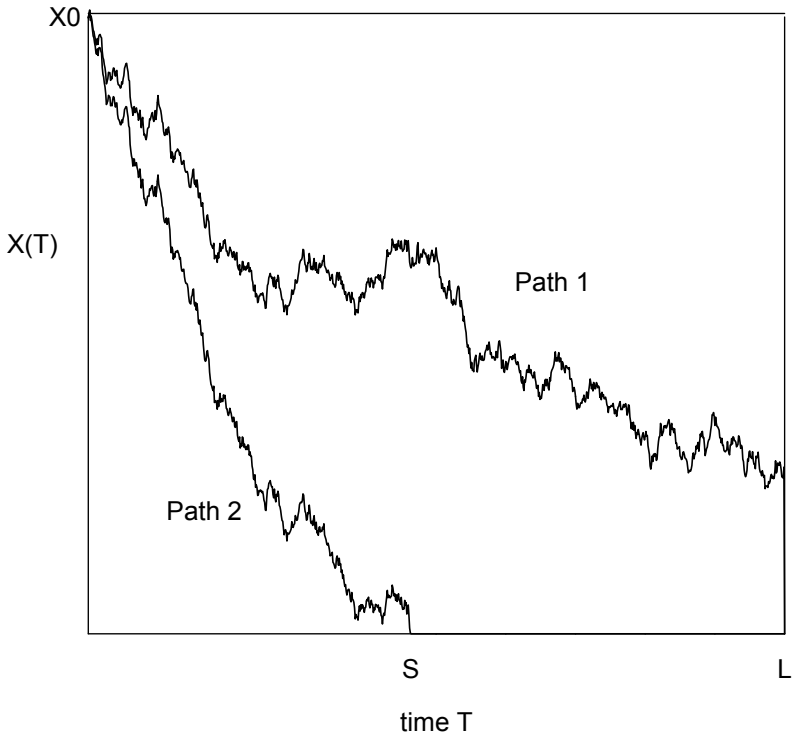

Figure 3.1: Two survival sample paths of health status starting from initial level X_0 until failure time S (path 2) or end of follow-up at time L (path 1).

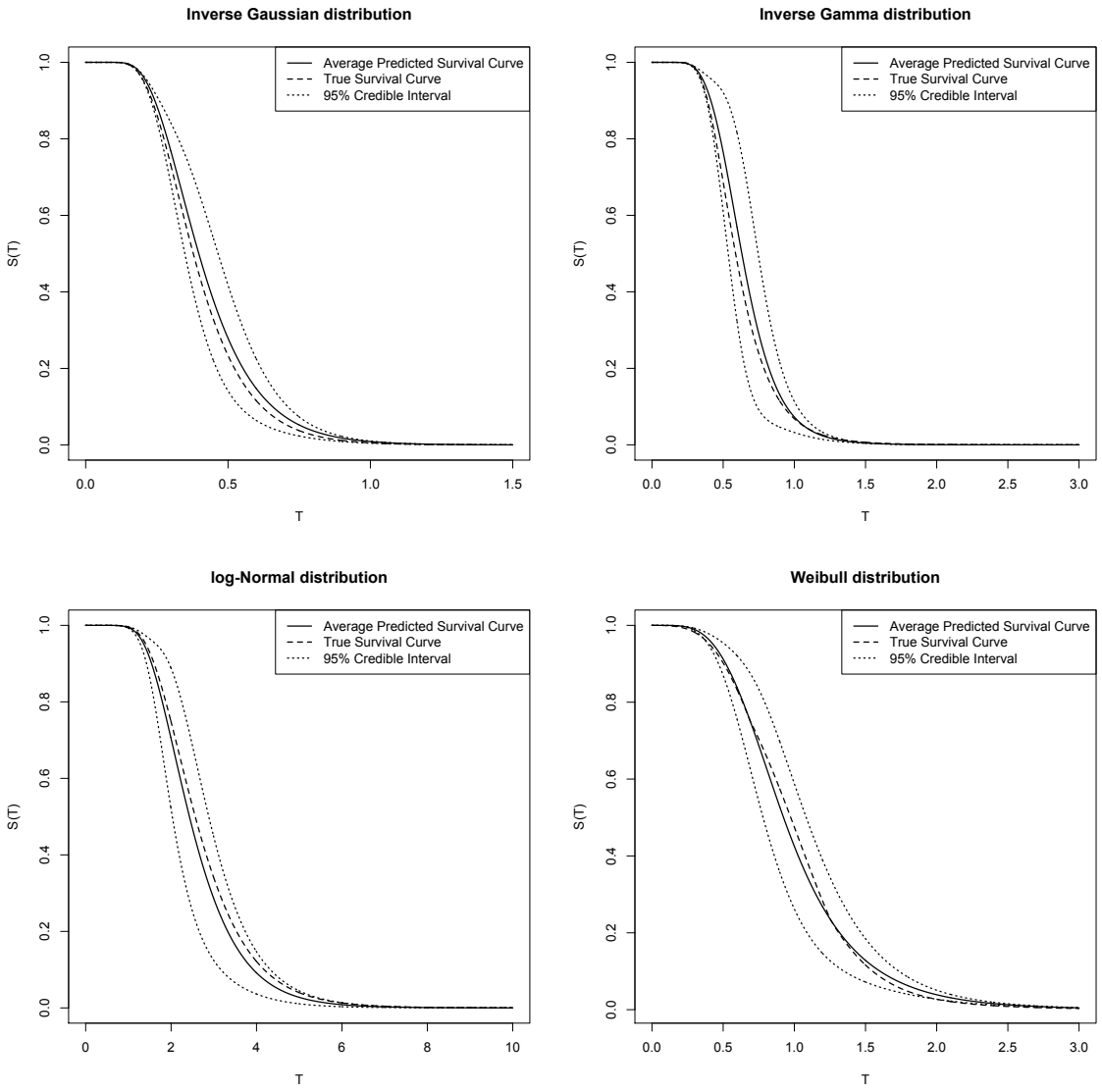

Figure 3.2: The survival curve for one patient from one simulation. The predicted survival curve and its corresponding 95% credible interval are estimated by Bayesian model averaging. The true survival curve is plotted for comparison.

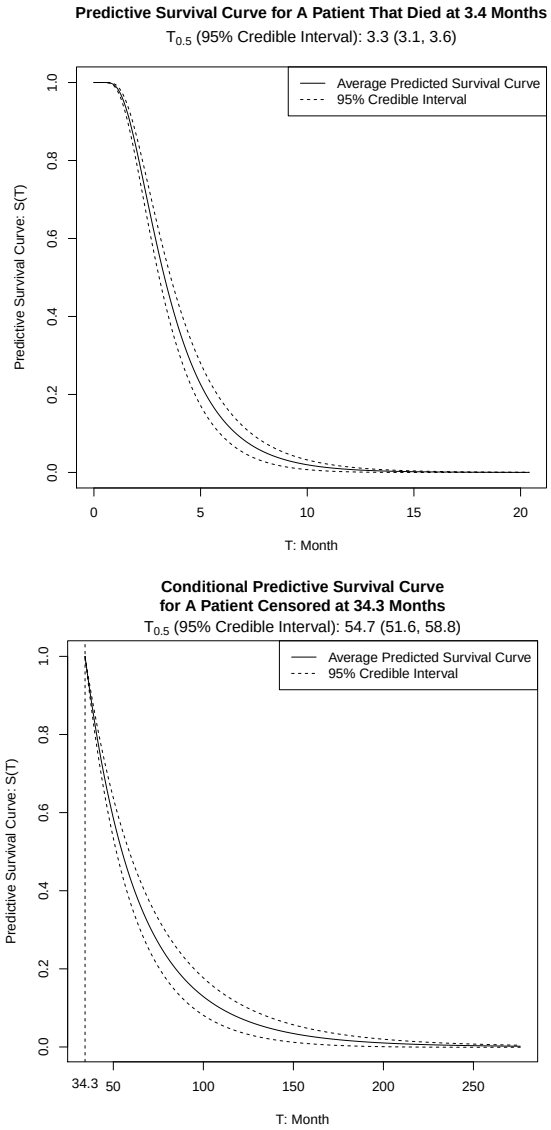

Figure 3.3: Upper figure: Predicted survival curve for one patient who failed at 3.4 months; Lower figure: Conditional predicted survival curve for one patient who was censored at 34.3 months, conditional on the patient survived after 34.4 months.

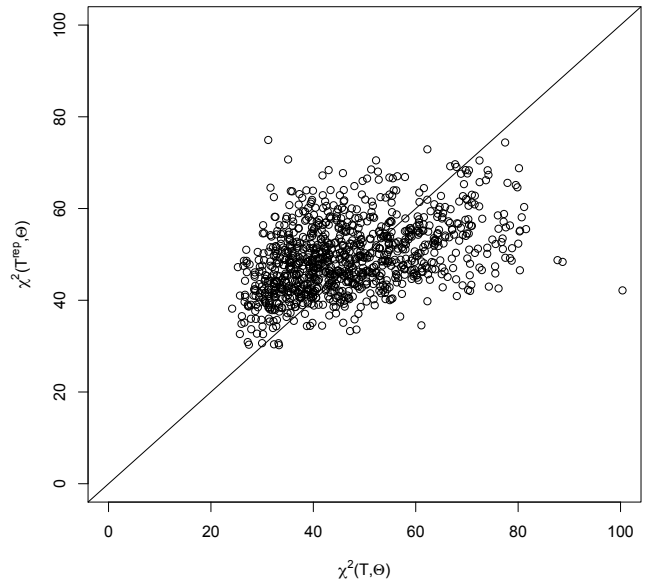

Figure 3.4: Scatterplot of predictive vs. realized χ^2 discrepancies for failure time $\boldsymbol{\mathcal{T}}$ given parameters $\boldsymbol{\Omega}$

3.8 Appendix

Throughout this appendix, we will rely on the notation set out in the main part of this chapter.

3.8.1 Model and Algorithm Details

In this section, we derive the posterior distribution for parameters in our proposed model. We begin by giving a detailed derivation of the Bayesian FHT regression model and then describe briefly how we sample parameters in the mPWDP model and derive the summary statistics.

Bayesian FHT Regression Model:

Since the parameter η_j follows the log normal distribution with density:

$$f\left(\eta_j|\mathbf{B}_1(\boldsymbol{Z}_j)\boldsymbol{\alpha},\omega\right)=\frac{1}{\eta_j\omega\sqrt{2\pi}}\exp\left\{-\frac{\left(\log\eta_j-\mathbf{B}_1(\boldsymbol{Z}_j)\boldsymbol{\alpha}\right)^2}{2\omega^2}\right\},$$

the posterior distribution of the parameter η_j for each subject is given by:

$$\begin{aligned}[\eta_j|\cdot\,] &\propto f\left(\mathcal{T}_j|X_j(0),\eta_j\right)f\left(\eta_j|\mathbf{B}_1(\boldsymbol{Z}_j)\boldsymbol{\alpha},\omega\right)\\ &\propto \exp\left\{-\log\eta_j-\frac{\eta_j^2\mathcal{T}_j^2-2X_j(0)\eta_j\mathcal{T}_j}{2\mathcal{T}_j}-\frac{\left(\log\eta_j\right)^2-2\mathbf{B}_1(\boldsymbol{Z}_j)\boldsymbol{\alpha}\ln\eta_j}{2\omega^2}\right\}. \qquad (3.15)\end{aligned}$$

The distribution (3.15) is not a log-concave function, therefore, we apply the adaptive rejection Metropolis sampling algorithm (ARMS, Gilks et al. (1995)) to simulate the posterior.

Once we obtained the draws of η_j for $j=1,2,\ldots,M$, we assume $\zeta_j=\log\eta_j$ and $\boldsymbol{\zeta}=\{\zeta_1,\ldots,\zeta_M\}$, therefore, we have:

$$p(\boldsymbol{\zeta}|\boldsymbol{\alpha},\mathbf{B}_1,\omega)=(2\pi\omega^2)^{-M/2}\exp\left\{-\frac{(\boldsymbol{\zeta}-\mathbf{B}_1\boldsymbol{\alpha})^{\mathrm{T}}(\boldsymbol{\zeta}-\mathbf{B}_1\boldsymbol{\alpha})}{2\omega^2}\right\}. \qquad (3.16)$$

A priori, $[\boldsymbol{\alpha}|\kappa]\sim\mathrm{N}\left(0,\kappa\boldsymbol{I}_{K+1}\right)$. We can write out the posterior distribution $\boldsymbol{\alpha}$ given $\boldsymbol{\zeta}$, ω, κ and the basis matrix, which is proportional to the product of formula (3.16)

and the prior distribution for $\boldsymbol{\alpha}$ given κ:

$$\begin{aligned} p(\boldsymbol{\alpha}|\boldsymbol{\zeta}, \kappa, \omega, \mathbf{B}_1) &\propto \pi(\boldsymbol{\alpha}|\kappa)\pi(\boldsymbol{\zeta}|\boldsymbol{\alpha}, \omega, \mathbf{B}_1) \\ &\propto \exp\left\{-\frac{(\boldsymbol{\alpha} - \mathbf{m}^\star)^{\mathrm{T}}(\mathbf{V}^\star)^{-\mathbf{1}}(\boldsymbol{\alpha} - \mathbf{m}^\star)}{2\omega^2}\right\} \end{aligned}$$

where $\mathbf{V}^\star = \left[\left(\frac{\kappa}{\omega^2}\boldsymbol{I}\right)^{-1} + \mathbf{B}_1^{\mathrm{T}}\mathbf{B}_1\right]^{-1}$ and $\mathbf{m}^\star = \mathbf{V}^\star\mathbf{B}_1^{\mathrm{T}}\boldsymbol{\zeta}$. Therefore, the full posterior for $\boldsymbol{\alpha}$ is:

$$[\boldsymbol{\alpha}|\boldsymbol{\zeta}, \omega, \kappa] \sim \mathrm{N}(\mathbf{m}^\star, \mathbf{V}^\star\omega^{\mathbf{2}}). \tag{3.17}$$

Standard conjugacy results state that the full conditional distributions of κ^{-1}, ω^{-2} and τ are Gamma distributions:

$$\left[\kappa^{-1}|\boldsymbol{\alpha}, K\right] \sim \mathrm{Gamma}\left(0.001 + \frac{K+1}{2}, 0.001 + \frac{\boldsymbol{\alpha}^{\mathbf{T}}\boldsymbol{\alpha}}{2}\right), \tag{3.18}$$

$$\begin{aligned} &\left[\omega^{-2}|\boldsymbol{\alpha}, \zeta_j, \mathbf{B}_1(\boldsymbol{Z}_j)\right] \\ &\qquad \sim \mathrm{Gamma}\left(0.001 + \frac{M}{2}, 0.001 + \frac{\sum_j \left[\zeta_j - \mathbf{B}_1(\boldsymbol{Z}_j)\boldsymbol{\alpha}\right]^2}{2}\right) \end{aligned} \tag{3.19}$$

and

$$[\tau \mid K] \sim \mathrm{Gamma}(1 + K, 0.2 + 1). \tag{3.20}$$

The acceptance probabilities for the birth and death moves in the RJMCMC algorithm can be derived as described in Chapter II. Here we briefly summarize the sampling formula and algorithm.

For a birth move suppose there are K bases in the BMARS model. We propose to draw an interaction level, $L_{K+1} \in \{1, 2\}$ for the $K+1$ basis with

$$\pi(L_{K+1} = 1) = \pi(L_{K+1} = 2) = 1/2. \tag{3.21}$$

We then draw L_{K+1} elements, $\{w_{1,K+1}, \ldots, w_{L_{K+1},K+1}\}$ with

$$\pi(w_{1,K+1} = w \mid L_{K+1} = 1) = 1/4 \quad \text{for} \quad w = 1, 2, 3, 4. \tag{3.22}$$

$$\pi[(w_{1,K+1}, w_{2,K+1}) = (w, w') \mid L_{K+1} = 2] = 1/6 \tag{3.23}$$
$$\text{for} (w, w') = (1,2), (1,3), (1,4), (2,3), (2,4), (3,4).$$

The knot point $t_{l,K+1}$ is draw at random from $\{Z_{1w_{l,K+1}}, \ldots, Z_{Mw_{l,K+1}}\}$ with

$$\pi(t_{l,K+1} = Z_{jw_{l,K+1}} \mid w_{l,K+1}) = 1/M \quad \text{for} \quad l = 1, \ldots, L_{K+1}, \tag{3.24}$$

and $s_{l,K+1}$ is drawn with equal probability from the set $\{-1, 1\}$:

$$\pi(s_{l,K+1} = -1) = \pi(s_{l,K+1} = 1) = 1/2 \quad \text{for} \quad l = 1, \ldots, L_{K+1}. \tag{3.25}$$

Then the acceptance rate of a birth step is:

$$\gamma_{\text{birth}} = \min\left\{1, \frac{|\mathbf{V}^*_{K+1}|^{1/2} \exp(a_K - a_{K+1})\tau}{(\kappa/\omega^2)^{1/2}|\mathbf{V}^*_K|^{1/2}(K+1)}\right\} \tag{3.26}$$

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

For a death move suppose there are K BMARS bases 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

$$\gamma_{\text{death}} = \min\left\{1, \frac{(\kappa/\omega^2)^{1/2}|\mathbf{V}^*_{K-1}|^{1/2} K}{|\mathbf{V}^*_K|^{1/2} \exp(a_{K-1} - a_K)\tau}\right\}. \tag{3.27}$$

For a move step, we choose a basis k with probability $1/K$ among the K existing bases, and then we propose to move the knot location in the main effect basis ($L_k = 1$) or in a term in an interaction basis ($L_k = 2$, the term is chosen with probability 0.5) (Holmes and Denison (2003), Denison et al. (2002)). The acceptance probability of this move is then

$$\gamma_{\text{move}} = \min\left\{1, \frac{|\mathbf{V}^{*,\text{prop}}_K|^{1/2} \exp(a_K - a^{\text{prop}}_K)}{|\mathbf{V}^*_K|^{1/2}}\right\} \tag{3.28}$$

where "prop" the proposed value of a_K.

Similarly, since the parameter $X_j(0)$ follows the log normal distribution, the posterior distribution of $X_j(0)$ for each subject is:

$$(3.29)\ [X_j(0)|\cdot] \propto \exp\left\{-\frac{X_j(0)^2 - 2X_j(0)\eta_j\mathcal{T}_j}{2\mathcal{T}_j} - \frac{(\ln X_j(0))^2 - 2\mathbf{B}_2(\boldsymbol{Z}_j)\boldsymbol{\beta}\ln X_j(0)}{2\omega'^2}\right\}.$$

The distribution (3.29) is also not a log-concave function, therefore, we apply the ARMS to draw from the full conditional.

The posterior draws of $\boldsymbol{\beta}$ and all the parameters in Θ_2 can be derived in the same way as the procedure for $\boldsymbol{\alpha}$ and those in Θ_1. Similarly, updating Θ_2 can be achieved using the same procedure as described for updating Θ_1.

mPWDP Model and Derivation of Summary Statistics:

To derive the posterior distribution for parameters in the mPWDP model, we need to write out the joint distribution of all the data and parameters in both stage I and stage II.

Joint Model:

Let $C_{i,j} = \{\{\boldsymbol{\mu}_{i'1,j}\}_{i'\in N_{i,j}}, \{\widetilde{\boldsymbol{\mu}}_{i'2,j}\}_{i'\in N_{i,j}}, \{\boldsymbol{\nu}_{i'2,j}\}_{i'\in N^e_{i,j}\cap S_j}\}$. We denote the jth covariate vector, which is a functional of the parameters $\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}$, by $\boldsymbol{Z}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j})$. The joint distribution of all the data and parameters is

$$
\begin{aligned}
(3.30) \qquad & \prod_{j=1}^{M}\prod_{i=1}^{n_j} \pi(\mathbf{Y}_{i,j} \mid \boldsymbol{\mu}_{i,j}, \Sigma_j)\pi(\Sigma_j)\pi(\boldsymbol{\mu}_{i,j} \mid \{\boldsymbol{\mu}_{i'j}\}_{i'\in N_{i,j}}, \Psi_j)\pi(\Psi_j) \\
& \times \prod_{j=1}^{M}\prod_{i=1}^{n_j^e} \pi(\mathbf{W}_{i,j} \mid \boldsymbol{\nu}_{i,j}, \Delta_j)\pi(\Delta_j)\pi(\boldsymbol{\nu}_{i,j} \mid \{\boldsymbol{\nu}_{i,j}\}_{i'\in N^e_{i,j}}, \Omega_j)\pi(\Omega_j) \\
& \times \prod_{j=1}^{M}\prod_{i=1}^{n_j} \pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j) \\
& \times \prod_{j=1}^{M} \pi(X_j(0) \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \Theta_2)\pi(\eta_j \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \Theta_1) \\
& \times \prod_{j=1}^{M} \pi(\mathcal{T}_j | X_j(0), \eta_j)\pi(\Theta_1)\pi(\Theta_2).
\end{aligned}
$$

We use the same notation as in Chapter II to define the parameters in the mPWDP model. In brief, let $\mathbf{W}_{i,j}$ denote the image intensities for voxel i with mean vector $\boldsymbol{\nu}_{i,j}$ in the healthy tissue ROI for patient j. The covariance of $\mathbf{W}_{i,j}$ will be denoted Δ_j and the covariance of the mean vector $\boldsymbol{\nu}_{i,j}$ will be denoted Ω_j. Denote the set of voxels in the healthy ROI by $\mathcal{H}_j$. The number of voxels in the healthy tissue ROI is also n_j. $\widetilde{\mathbf{Y}}_{i2,j} = \left(\widetilde{Y}_{i21,j}, \widetilde{Y}_{i22,j}\right)^{\mathrm{T}}$ is the predicted tumor response at time point 2 under the "null" and $\widetilde{\boldsymbol{\mu}}_{i2,j}$ is its mean. We extend the healthy tissue ROI by a one-voxel thick shell and denote the set of voxels in this shell by $\mathcal{S}_j$ to mitigate the identifiability issue that arises when drawing $\widetilde{\mathbf{Y}}_{i2,j}$ and $\widetilde{\boldsymbol{\mu}}_{i2,j}$ simultaneously. Let n_j^s denote the number of voxels in the shell and let $n_j^e = n_j + n_j^s$ be the number in the extended ROI. Let $N^e_{i,j} = \{i' : i' \sim i\}$ denote the set of neighbors of voxel i in the extended ROI and $|N^e_{i,j}|$ denote the number in this set for patient j. We build a separate mPWDP model for the image data in the healthy tissue ROI and predict tumor response under the null: $\widetilde{\boldsymbol{\mu}}_{i2,j}$ (Details can be found in the Chapter

II). Therefore, we have

$$[\mathbf{Y}_{i,j} \mid \boldsymbol{\mu}_{i,j}, \Sigma_j] \sim \mathrm{N}(\boldsymbol{\mu}_{i,j}, \Sigma_j), \qquad [\mathbf{W}_{i,j} \mid \boldsymbol{\nu}_{i,j}, \Delta_j] \sim \mathrm{N}(\boldsymbol{\nu}_{i,j}, \Delta_j), \tag{3.31}$$

$$[\Sigma_j] \sim \mathrm{W}^{-1}(\mathrm{I}_4, 5), \qquad [\Delta_j] \sim \mathrm{W}^{-1}(\mathrm{I}_4, 5), \tag{3.32}$$

$$\left[\boldsymbol{\mu}_{i,j} \mid \{\boldsymbol{\mu}_{i'j}\}_{i' \in N_{i,j}}, \Psi_j\right] \sim \mathrm{N}\left[\boldsymbol{\mu}^*_{i,j}, |N_{i,j}|^{-1}\Psi_j\right], \tag{3.33}$$

$$\left[\boldsymbol{\nu}_{i,j} \mid \{\boldsymbol{\nu}_{i,j}\}_{i' \in N^e_{i,j}}, \Omega_j\right] \sim \mathrm{N}\left(\boldsymbol{\nu}^*_{i,j}, |N^e_{i,j}|^{-1}\Omega_j\right], \tag{3.34}$$

$$[\Psi_j] \sim \mathrm{W}^{-1}(\mathrm{I}_4, 5), \qquad [\Omega_j] \sim \mathrm{W}^{-1}(\mathrm{I}_4, 5), \tag{3.35}$$

and

$$\begin{aligned}
&\left[\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j\right] \\
&\quad \sim \mathrm{N}\left(\widetilde{\boldsymbol{\mu}}_{i2,j} + \Delta_{21,j}\left(\Delta_{11,j}\right)^{-1}\left(\mathbf{Y}_{i1,j} - \boldsymbol{\mu}_{i1,j}\right), \Delta_{22,j} - \Delta_{21,j}(\Delta_{11,j})^{-1}\Delta_{12,j}\right), \\
&\left[\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j\right] \\
&\quad \sim \mathrm{N}\left(\widetilde{\boldsymbol{\mu}}^*_{i2,j} + \Omega_{21,j}(\Omega_{11,j})^{-1}(\boldsymbol{\mu}_{i1,j} - \boldsymbol{\mu}^*_{i1,j}), |N^e_{i,j}|^{-1}(\Omega_{22,j} - \Omega_{21,j}(\Omega_{11,j})^{-1}\Omega_{12,j})\right), \\
&\left[\eta_j \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_2\right] \sim \text{log-Norm}\left(\mathbf{B}_1(\boldsymbol{Z}_j)\boldsymbol{\alpha}_j, \omega_j^2\right), \\
&\left[X_j(0) \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_2\right] \sim \text{log-Norm}\left(\mathbf{B}_2(\boldsymbol{Z}_j)\boldsymbol{\beta}_j, \omega_j'^2\right),
\end{aligned}$$

where

$$\boldsymbol{\mu}^*_{i,j} = |N_{i,j}|^{-1} \sum_{i' \in N_{i,j}} \boldsymbol{\mu}_{i',j}\ ,\ \boldsymbol{\nu}^*_{i,j} = |N^e_{i,j}|^{-1} \sum_{i' \in N^e_{i,j}} \boldsymbol{\nu}_{i',j}, \tag{3.36}$$

and

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

Stage I – mPWDP Model:

In stage I, we fit the mPWDP model and derive summary statistics. The full conditional posterior sampling distributions for the parameters in the mPWDP model

are derived analogously to those in Chapter II. We employ an hybrid Metropolis-within-Gibbs algorithm to sample from the full conditionals. We use the Gibbs sampler to sample from those full conditional distributions with nice distributional forms, while we apply the Metropolis-Hastings algorithm to sample from full conditionals of the remaining parameters.

The full conditional for the covariance matrix Σ_j for subject j is:

$$\left[\Sigma_j \mid \left\{\boldsymbol{\mu}_{i,j}\right\}_{i=1}^{n_j}, \left\{\mathbf{Y}_{i,j}\right\}_{i=1}^{n_j}\right] \sim \mathrm{W}^{-1}\left(\mathrm{S}_{1,j} + \mathrm{I}_4, n_j + 5\right), \tag{3.37}$$

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

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

$$\left[\Psi_j \mid \left\{\boldsymbol{\mu}_{i,j}\right\}_{i=1}^{n_j}\right] \sim \mathrm{W}^{-1}\left(\mathrm{S}_{2,j} + \mathrm{I}_4, n_j + 5\right), \tag{3.38}$$

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

To update $\boldsymbol{\mu}_{i,j}$, recall that the summary statistics, $\boldsymbol{Z}_j$, depends on $\{\boldsymbol{\mu}_{i2,j}\}_{i=1}^{n_j}$ and $\{\widetilde{\boldsymbol{\mu}}_{i2,j}\}_{i=1}^{n_j}$. For voxel i, $\boldsymbol{\mu}_{i,j}$ has full conditional given by

$$\begin{aligned}
&\pi(\boldsymbol{\mu}_{i,j} \mid \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}}, \mathbf{Y}_{i,j}, \Sigma_j, \Psi_j, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Theta_1, \Theta_2, X_j(0), \eta_j) \\
&\quad \propto \pi(\mathbf{Y}_{i,j} \mid \boldsymbol{\mu}_{i,j}, \Sigma_j)\pi(\boldsymbol{\mu}_{i,j} \mid \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}}, \Psi_j) \\
&\qquad \times \pi(X_j(0) \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \Theta_2)\pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j) \\
&\qquad \times \pi(\eta_j \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \Theta_1)\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j) \\
&\qquad \times \prod_{i' \in N_{i,j}} \pi(\widetilde{\boldsymbol{\mu}}_{i'2,j} \mid \boldsymbol{\mu}_{i'1,j}, C_{i',j}, \Omega_j),
\end{aligned} \tag{3.39}$$

which does not have a nice distributional form from which we can easily sample and so we use a Metropolis-Hastings update. We propose a new value of $\boldsymbol{\mu}_{i,j}$, call it $\boldsymbol{\mu}_{i,j}^{prop}$ from:

$$\begin{aligned}
&\left[\boldsymbol{\mu}_{i,j} \mid \Sigma_j, \Psi_j, \mathbf{Y}_{i,j}, \{\boldsymbol{\mu}_{i',j}\}_{i' \in N_{i,j}}\right] \sim \\
&\qquad \mathrm{N}\left[\mathcal{V}\left(|N_{i,j}|{\Psi_j}^{-1}\boldsymbol{\mu}_{i,j}^{*} + {\Sigma_j}^{-1}\mathbf{Y}_{i,j}\right), \mathcal{V}\right].
\end{aligned} \tag{3.40}$$

where $\mathcal{V} = \left(|N_{i,j}|{\Psi_j}^{-1} + {\Sigma_j}^{-1}\right)^{-1}$. So the acceptance ratio can be simplified to achieve a high acceptance rate. The acceptance probability is given by

$$\alpha_\mu = \min\{1, \mathcal{R}\} \tag{3.41}$$

$$\begin{aligned}\text{where} \quad \mathcal{R} \;=\; & \frac{\pi(\eta_j \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{\{-i\}2,j}, \boldsymbol{\mu}_{i2,j}^{prop}, \widetilde{\boldsymbol{\mu}}_{2,j}), \Theta_1)\pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}^{prop}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)}{\pi(\eta_j \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{\{-i\}2,j}, \boldsymbol{\mu}_{i2,j}^{current}, \widetilde{\boldsymbol{\mu}}_{2,j}), \Theta_1)\pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}^{current}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)} \\ & \times \frac{\pi(X_j(0) \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{\{-i\}2,j}, \boldsymbol{\mu}_{i2,j}^{prop}, \widetilde{\boldsymbol{\mu}}_{2,j}), \Theta_2)}{\pi(X_j(0) \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{\{-i\}2,j}, \boldsymbol{\mu}_{i2,j}^{current}, \widetilde{\boldsymbol{\mu}}_{2,j}), \Theta_2)} \\ & \times \frac{\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}^{prop}, C_{i,j}, \Omega_j)\prod_{i' \in N_{i,j}} \pi(\widetilde{\boldsymbol{\mu}}_{i'2,j} \mid \boldsymbol{\mu}_{i'1,j}, C_{i',j}^{prop}, \Omega_j)}{\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}^{current}, C_{i,j}, \Omega_j)\prod_{i' \in N_{i,j}} \pi(\widetilde{\boldsymbol{\mu}}_{i'2,j} \mid \boldsymbol{\mu}_{i'1,j}, C_{i',j}^{current}, \Omega_j)},\end{aligned}$$

with $C_{i',j}^{prop} = \{\{\boldsymbol{\mu}_{i1,j}^{prop}, \boldsymbol{\mu}_{k1,j}\}_{k \in N_{i',j}}, \{\widetilde{\boldsymbol{\mu}}_{k2,j}\}_{k \in N_{i',j}}, \{\boldsymbol{\nu}_{k2,j}\}_{k \in N_{i',j}^e \cap \mathcal{S}_j}\}$ and superscript *prop* represents a new proposed sample and *current* represents the current sample.

For healthy tissue voxels, we then derive the full conditional posterior distribution of $\boldsymbol{\nu}_{i,j}$. Analogous to (3.40), the full conditional of $\boldsymbol{\nu}_{i,j}$ for voxel $i \in N_{i',j}^e \cap \mathcal{H}_j$, is

$$\begin{aligned}&\left[\boldsymbol{\nu}_{i,j} \mid \Omega_j, \Delta_j, \mathbf{W}_{i,j}, \{\boldsymbol{\nu}_{i',j}\}_{i' \in N_{i,j}^e}\right] \sim \\ &\mathrm{N}\left[\left(|N_{i,j}^e|\Omega_j^{-1} + {\Delta_j}^{-1}\right)^{-1}\left(|N_{i,j}^e|\Omega_j^{-1}\boldsymbol{\nu}_{i,j}^* + \Delta_j^{-1}\mathbf{W}_{i,j}\right), \left(|N_{i,j}^e|\Omega_j^{-1} + \Delta_j^{-1}\right)^{-1}\right].\end{aligned} \tag{3.42}$$

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

$$\begin{aligned}&\pi(\boldsymbol{\nu}_{i,j} \mid \{\boldsymbol{\nu}_{i',j}\}_{i' \in N_{i,j}^e}, \mathbf{W}_{i,j}, \Delta_j, \Omega_j, \{\widetilde{\boldsymbol{\mu}}_{i2,j}\}_{i' \in N_i}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \{\boldsymbol{\mu}_{i'1,j}\}_{i' \in N_i}, \boldsymbol{\mu}_{i1,j}) \\ &\quad \propto \pi(\mathbf{W}_{i,j} \mid \boldsymbol{\nu}_{i,j}, \Delta_j)\pi(\boldsymbol{\nu}_{i,j} \mid \{\boldsymbol{\nu}_{i',j}\}_{i' \in N_{i,j}^e}, \Omega_j) \prod_{i' \in N_{i,j}} \pi(\widetilde{\boldsymbol{\mu}}_{i'2,j} \mid \boldsymbol{\mu}_{i'1,j}, C_{i',j}, \Omega_j),\end{aligned}$$

which does not have a closed form. Thus, we propose a new value $\boldsymbol{\nu}_{i,j}^{prop}, i \in N_{i',j}^e \cap \mathcal{S}_j$ from (3.42) and accept with probability:

$$\alpha_\nu = \min\left\{1, \prod_{i' \in N_{i,j}} \frac{\pi(\widetilde{\boldsymbol{\mu}}_{i'2,j} \mid \boldsymbol{\mu}_{i'1,j}, C_{i',j}^{prop}, \Omega_j)}{\pi(\widetilde{\boldsymbol{\mu}}_{i'2,j} \mid \boldsymbol{\mu}_{i'1,j}, C_{i',j}^{current}, \Omega_j)}\right\} \tag{3.43}$$

where $C_{i',j}^{prop} = \{\{\boldsymbol{\mu}_{k1,j}\}_{k \in N_{i',j}}, \{\widetilde{\boldsymbol{\mu}}_{k2,j}\}_{k \in N_{i',j}}, \{\boldsymbol{\nu}_{i2,j}^{prop}, \boldsymbol{\nu}_{k2,j}\}_{k \in N_{i',j}^e \cap S_j}\}$.

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

$$\begin{aligned}(3.44)\quad &\pi(\Delta_j \mid \{\mathbf{W}_{i,j}\}_{i=1}^{n_j^e}, \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n_j^e}, \{\widetilde{\mathbf{Y}}_{i2,j}\}_{i=1}^{n_j}, \{\mathbf{Y}_{i1,j}\}_{i=1}^{n_j}, \{\boldsymbol{\mu}_{i1,j}\}_{i=1}^{n_j}, \{\widetilde{\boldsymbol{\mu}}_{i2,j}\}_{i=1}^{n_j}) \\ &\propto \prod_{i=1}^{n_j^e} \pi(\mathbf{W}_{i,j} \mid \boldsymbol{\nu}_{i,j}, \Delta_j)\pi(\Delta_j) \prod_{i=1}^{n_j} \pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j),\end{aligned}$$

which does not have a nice distributional form. To simulate from the posterior, we propose a new value Δ_j^{prop} from

$$(3.45)\qquad \left[\Delta_j \mid \{\mathbf{W}_{i,j}\}_{i=1}^{n_j^e}, \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n_j^e}\right] \;\sim\; \mathrm{W}^{-1}\left(\mathrm{S}_{3,j} + \mathrm{I}_4, n_j^e + 5\right),$$

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 $i \in \mathcal{H}_j \cup \mathcal{S}_j$. We then accept this value with probability

$$(3.46)\qquad \alpha_\Delta = \min\left\{1, \prod_{i=1}^{n_j} \frac{\pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j^{prop})}{\pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j^{current})}\right\}.$$

We then propose an update to Ω_j. From (3.30), we have

$$\begin{aligned}(3.47)\quad &\pi(\Omega_j \mid \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n_j^e}, \{\boldsymbol{\mu}_{i1,j}\}_{i=1}^{n_j}, \{\widetilde{\boldsymbol{\mu}}_{i2,j}\}_{i=1}^{n_j}, \{\boldsymbol{\nu}_{i'2}\}_{i' \in N_i^e \cap \mathcal{S}_j}) \\ &\propto \prod_{i=1}^{n_j^e} \pi(\boldsymbol{\nu}_{i,j} \mid \{\boldsymbol{\nu}_{i,j}\}_{i' \in N_{i,j}^e}, \Omega_j)\pi(\Omega_j) \prod_{i=1}^{n_j} \pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j).\end{aligned}$$

from which the posterior draws of Ω_j can not be directly obtained. We propose a new value Ω_j^{prop} from

$$(3.48)\qquad \left[\Omega_j \mid \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n^e}\right] \sim \mathrm{W}^{-1}\left(\mathrm{S}_{4,j} + \mathrm{I}_4, n_j^e + 5\right),$$

with probability

$$(3.49)\qquad \alpha_\Omega = \min\left\{1, \prod_{i=1}^{n_j} \frac{\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j^{prop})}{\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j^{current})}\right\}.$$

Similarly as illustrated in Section 3.8 of Chapter II, we can easily derive the conditional predictive distribution of $\widetilde{\mathbf{Y}}_{i2,j}$ and the posterior distribution of $\widetilde{\boldsymbol{\mu}}_{i2,j}$.

The conditional distribution of $\widetilde{\mathbf{Y}}_{i2,j}$ given $\mathbf{Y}_{i1,j}$ and model parameters is given by

$$(3.50)\left[\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j\right] \sim$$
$$\mathrm{N}\left(\widetilde{\boldsymbol{\mu}}_{i2,j} + \Delta_{21,j}\left(\Delta_{11,j}\right)^{-1}\left(\mathbf{Y}_{i1,j} - \boldsymbol{\mu}_{i1,j}\right),\ \Delta_{22,j} - \Delta_{21,j}(\Delta_{11,j})^{-1}\Delta_{12,j}\right).$$

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

$$(3.51)\qquad \pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \widetilde{\mathbf{Y}}_{i2,j}, \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, C_{i,j}, \Delta_j, \Omega_j, \boldsymbol{Z}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \boldsymbol{\Omega}_2, X_j(0), \eta_j)$$
$$\propto \pi(\widetilde{\mathbf{Y}}_{i2,j} \mid \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, \Delta_j)\pi(\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \boldsymbol{\mu}_{i1,j}, C_{i,j}, \Omega_j)$$
$$\times\pi(X_j(0) \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \Theta_2)\pi(\eta_j \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{2,j}), \Theta_1),$$

We propose a new value $\widetilde{\boldsymbol{\mu}}_{i2,j}^{prop}$ from

$$(3.52)\left[\widetilde{\boldsymbol{\mu}}_{i2,j} \mid \widetilde{\mathbf{Y}}_{i2,j}, \mathbf{Y}_{i1,j}, \boldsymbol{\mu}_{i1,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}, C_{i,j}, \Delta_j, \Omega_j\right] \sim$$
$$\mathrm{N}\left(\boldsymbol{\theta}_{i2,j} + \Lambda_{i_{21,j}}\left(\Lambda_{i_{11,j}}\right)^{-1}\left(\boldsymbol{\mu}_{i1,j} - \boldsymbol{\theta}_{i1,j}\right),\ \Lambda_{i_{22,j}} - \Lambda_{i_{21,j}}\left(\Lambda_{i_{11,j}}\right)^{-1}\Lambda_{i_{12,j}}\right)$$

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 draws with probability

$$(3.53)\qquad \alpha_{\widetilde{\mu}} = \min\left\{1, \mathrm{LR}_{\widetilde{\mu}}\right\}$$

where

$$\mathrm{LR}_{\widetilde{\mu}} = \frac{\pi(\eta_j \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{\{-i\}2,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}^{prop}), \Theta_1)\pi(X_j(0) \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{\{-i\}2,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}^{prop}), \Theta_2)}{\pi(\eta_j \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{\{-i\}2,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}^{current}), \Theta_1)\pi(X_j(0) \mid \boldsymbol{Z}_j(\boldsymbol{\mu}_{2,j}, \widetilde{\boldsymbol{\mu}}_{\{-i\}2,j}, \widetilde{\boldsymbol{\mu}}_{i2,j}^{current}), \Theta_2)}.$$

3.8.2 Pseudocode

Initialize parameters

Stage I:

For each subject

1. Set $\boldsymbol{\mu}_{i,j} = \mathbf{Y}_{i,j}$, $i = 1, \ldots, n_j$.
2. Set $\boldsymbol{\nu}_{i,j} = \mathbf{W}_{i,j}\, i = 1, \ldots, n_j^e$.
3. Set $\Sigma_j = \Psi_j = \Delta_j = \Omega_j = \mathrm{I}_4$.

End for each subject

Stage II:

For parameters in Θ_1:

1. Set $K = 0$ (intercept term only).
2. Set $\kappa = 1, \omega^2 = 1$.
3. Set $\tau = 5$.
4. Set $X_j(0) = 6$ and $\eta_j = 6$, for $j = 1, \ldots, M$.
5. Impute the death time for patients who are censored, by drawing $\mathcal{T}_j$ from the truncated inverse Gaussian distribution with $\mathcal{T}_j > T_j$, where T_j is the observed censoring time.
6. Draw $\boldsymbol{\alpha}$ from distribution (3.17) ($\boldsymbol{\alpha} = \alpha_0$ when $K = 0$ and $\mathbf{B}_1 = \mathbf{B}_1(\boldsymbol{Z}_j) = (1, \ldots, 1)^{\mathrm{T}}$, a vector of ones of length M).

7. Initialize the parameters in Θ_2, ω'^2, λ, and $\boldsymbol{\beta}$ the same way as described above.

Iterate For $d = 1$ 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 part of this chapter).

1. For tumor ROI:
   (a) For each voxel $i = 1, \ldots, n_j$, propose $\boldsymbol{\mu}_{i,j}^{prop}$ from (3.40).
   Accept $\boldsymbol{\mu}_{i,j}^{prop}$ with the probability (3.41).
   (b) Draw $\left[\Sigma_j \mid \{\boldsymbol{\mu}_{i,j}\}_{i=1}^{n_j}, \{\mathbf{Y}_{i,j}\}_{i=1}^{n_j}\right]$ from (3.37).
   (c) Draw $\left[\Psi_j \mid \{\boldsymbol{\mu}_{i,j}\}_{i=1}^{n_j}\right]$ from (3.38).
2. For healthy tissue ROI:
   (a) Draw $\left[\boldsymbol{\nu}_{i,j} \mid \Omega_j, \Delta_j, \mathbf{W}_{i,j}, \{\boldsymbol{\nu}_{i',j}\}_{i' \in N_{i,j}^e}\right]$, $i \in N_{i',j}^e \cap \mathcal{H}_j$, from (3.42).
   (b) For each $i \in N_{i',j}^e \cap \mathcal{S}_j$, propose $\boldsymbol{\nu}_{i,j}^{prop}$ from (3.42).
   Accept $\boldsymbol{\nu}_{i,j}^{prop}$ with probability (3.43).
   (c) Propose Ω_j^{prop} from (3.48).
   Accept Ω_j^{prop} with probability (3.49).
   (d) Propose $\Delta_j^{prop}$from (3.45).
   Accept Δ_j with probability (3.46).
3. Predict tumor response under null:
   (a) Draw $\left[\widetilde{\mathbf{Y}}_{i2,j} \mid \cdot\right]$, $i = 1, \ldots, n_j$, from (3.50).
   (b) For $i = 1, \ldots, n_j$, propose $\widetilde{\boldsymbol{\mu}}_{i2,j}^{prop}$ from (3.52).

Accept $\widetilde{\boldsymbol{\mu}}_{i2,j}^{prop}$ with probability (3.53).

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

(a) Calculate dKLD, pKLD, cDS and cPS.

End iterate over subjects.

Stage II:

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

I. Assume there are currently K basis functions in Θ_1.

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 $\{Z_{1w_{lk}}, \ldots, Z_{Mw_{lk}}\}$.

(d) If move (new knot location) accepted with probability γ_{move} (3.28).

i. Draw $[\eta_j \mid \cdot], j = 1, \ldots, M$, from (3.15) using ARMS.

ii. Draw $[\kappa^{-1} \mid \cdot]$ from (3.18).

iii. Draw $[\omega^{-2} \mid \cdot]$ from (3.20).

iv. Draw $[\boldsymbol{\alpha} \mid \cdot]$ from (3.17).

v. Draw $[\tau \mid K]$ from (3.20).

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 (3.21).

ii. If $L_{K+1} = 1$, draw $w_{1,K+1} \mid L_{K+1}$ from the set $\{1, 2, 3, 4\}$ with equal probability 1/4, see (3.22).

else draw $(w_{1,K+1}, w_{2,K+1}) \mid L_{K+1}$, with equal probability, from the set $\{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$, see (3.23).

iii. Draw the knot point(s) $t_{l,K+1}$, $l = 1, \ldots, L_{K+1}$, see (3.24).

iv. Draw $s_{l,K+1}$, $l = 1, \ldots, L_{K+1}$, see (3.25).

v. Accept the birth with probability γ_{birth} (3.26).

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

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

ii. Accept the death with probability γ_{death} (3.27).

3. Draw $[\eta_j \mid \cdot], j = 1, \ldots, M$, from (3.15) using ARMS.

4. Draw $[\kappa^{-1} \mid \cdot]$ from (3.18).

5. Draw $[\omega^{-2} \mid \cdot]$ from (3.20).

6. Draw $[\boldsymbol{\alpha} \mid \cdot]$ from (3.17).

II. Update the parameters in Θ_2 and $\boldsymbol{\beta}$ the same way as those in Θ_1 and $\boldsymbol{\alpha}$.

III. Impute the death time for patients who are censored, by drawing $\mathcal{T}_j$ from the truncated inverse Gaussian distribution with $\mathcal{T}_j > T_j$, where T_j is the observed censoring time.

end oversample

End Iterate

3.8.3 Sensitivity Analyses

In the Bayesian FHT regression model, we assign an informative prior to K, the number of BMARS basis: $[K \mid \tau] \sim \text{Poisson}(\tau)$, $\tau \sim \text{Gamma}(a, b)$ where we set $a = 1$, $b = 0.2$, when updating parameters in both Θ_1 and Θ_2. Given the small sample size, 47 patients in our study, we believe that a parsimonious model is warrated. 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 perform a sensitivity analysis on our choice of prior for τ, and hence, marginally, on K. We change the values of a and b as well as the distribution of τ to a uniform prior distribution on $[0, 10]$. Percentages of the HPD interval for the median survival time: $\mathcal{T}_{0.5}$ covering the observed $\mathcal{T}$ are shown in the last column of Table 3.7. The results from our model are not very sensitive to these changes in the prior distribution of τ.

To derive the summary statistics cDS/cPS, we used thresholds 97.5th and 2.5th percentile, respectively, of the conditional distribution of the means of the predicted tumor response 3 weeks post therapy initiation under the null. To assess the prediction sensitivity to those thresholds, we use the 99.5th/0.5th and the 95.0th/5.0th percentile as new thresholds to derive cDS/cPS. We found that at most one more subject has an HPD interval for $\mathcal{T}_{0.5}$ not covering the observed $\mathcal{T}$.

Table 3.7: Sensitivity analysis for different hyperprior distributions of τ in the proposed model. Percentages of the HPD interval of median survival time $\mathcal{T}_{0.5}$ covering the observed $\mathcal{T}$ are shown in the last column. Mean and variance are calculated for different prior distributions.

Prior of τ	Prior mean	Prior variance	$\mathcal{T}_{0.5}$ Coverage %
Gamma(0.6, 0.2)	3.0	15.0	84.4
Gamma(0.8, 0.2)	4.0	20.0	87.5
Gamma(1.0, 0.2)	5.0	25.0	87.5
Gamma(1.8, 0.2)	8.0	45.0	87.5
Gamma(0.5, 0.1)	5.0	50.0	87.5
Gamma(2, 0.2)	10.0	50.0	84.4
Gamma(2, 0.4)	5.0	12.5	87.5
U [0, 10]	5.0	8.3	84.4

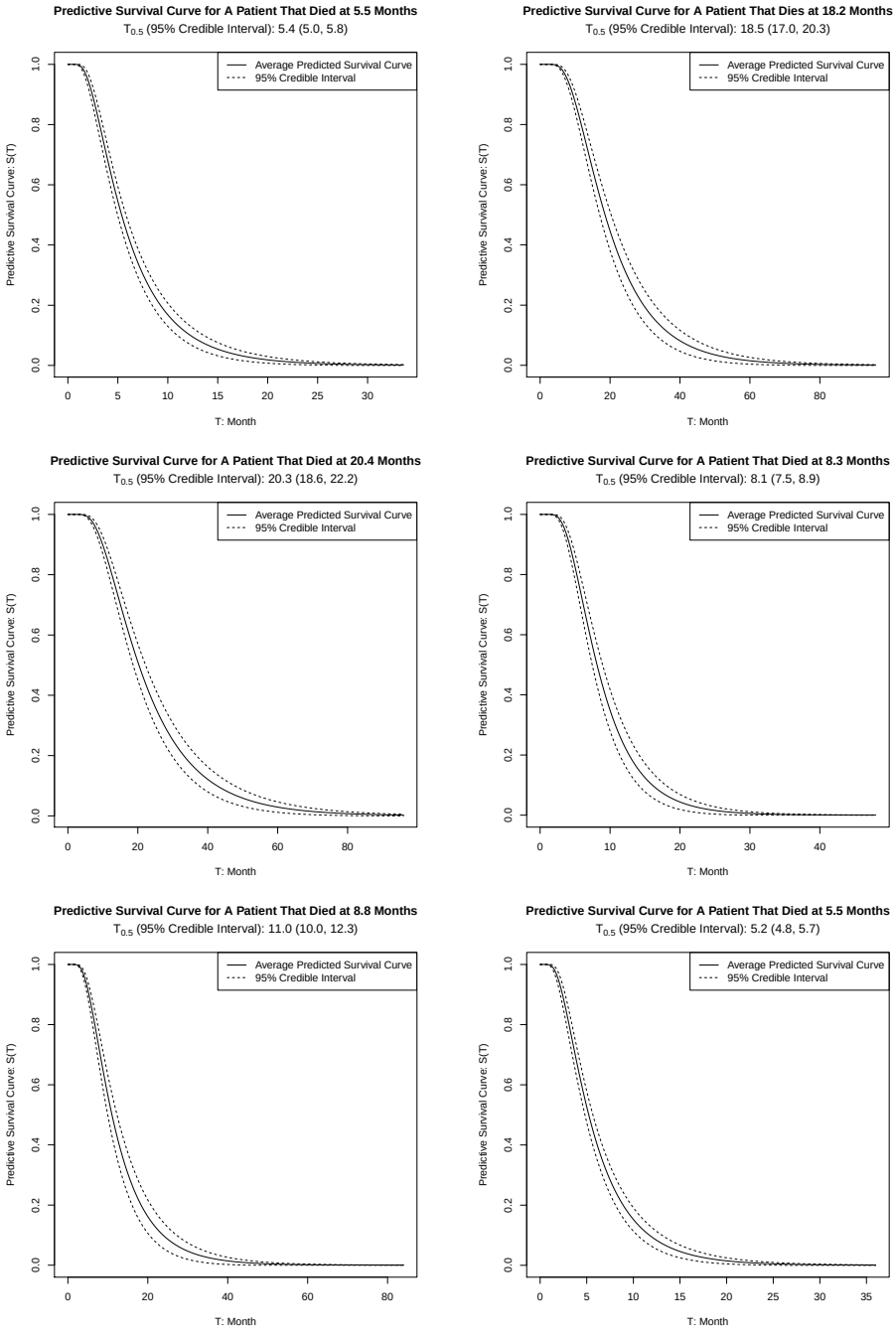

Figure 3.5: Predicted survival curves for patients who died

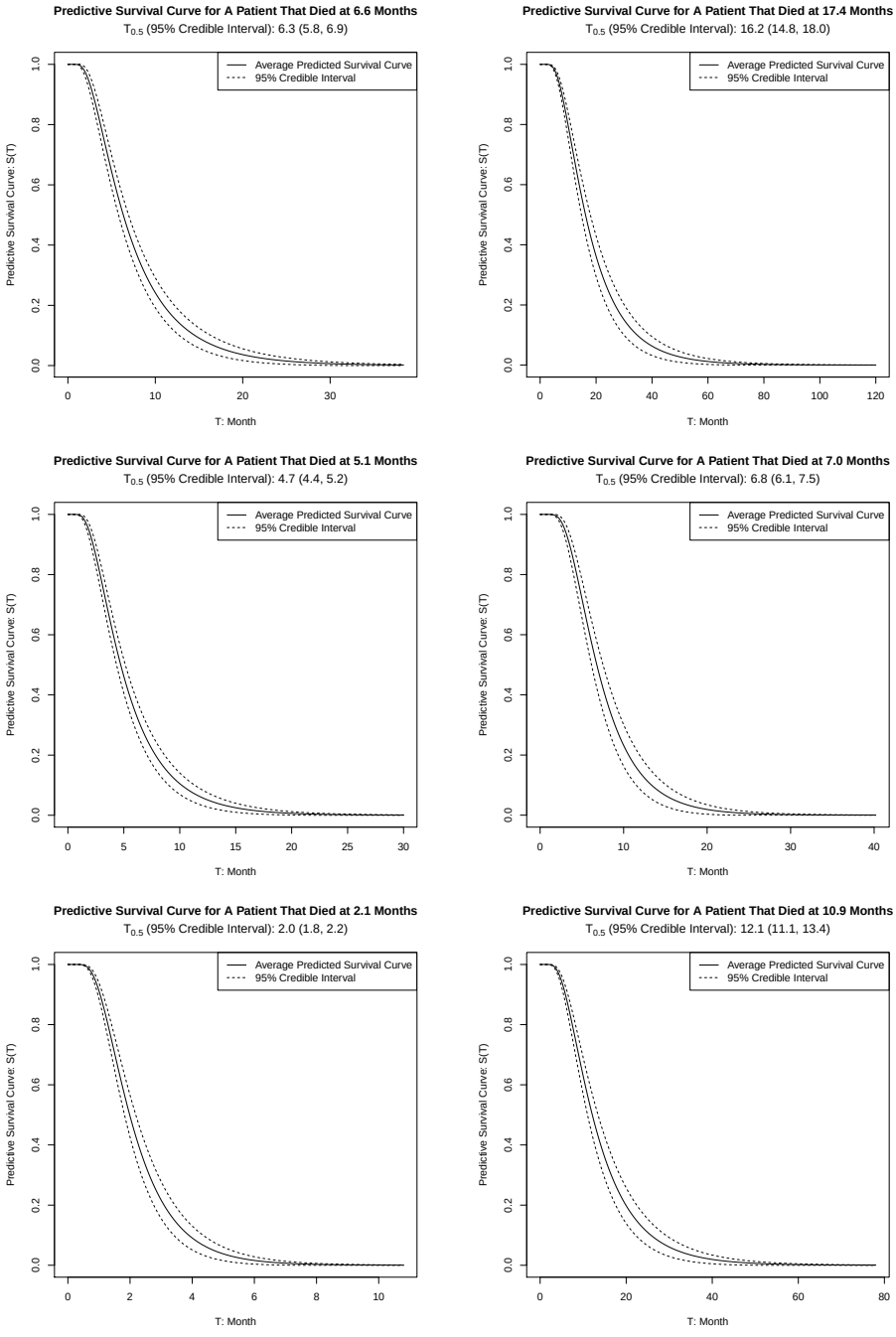

Figure 3.6: Predicted survival curves for patients who died

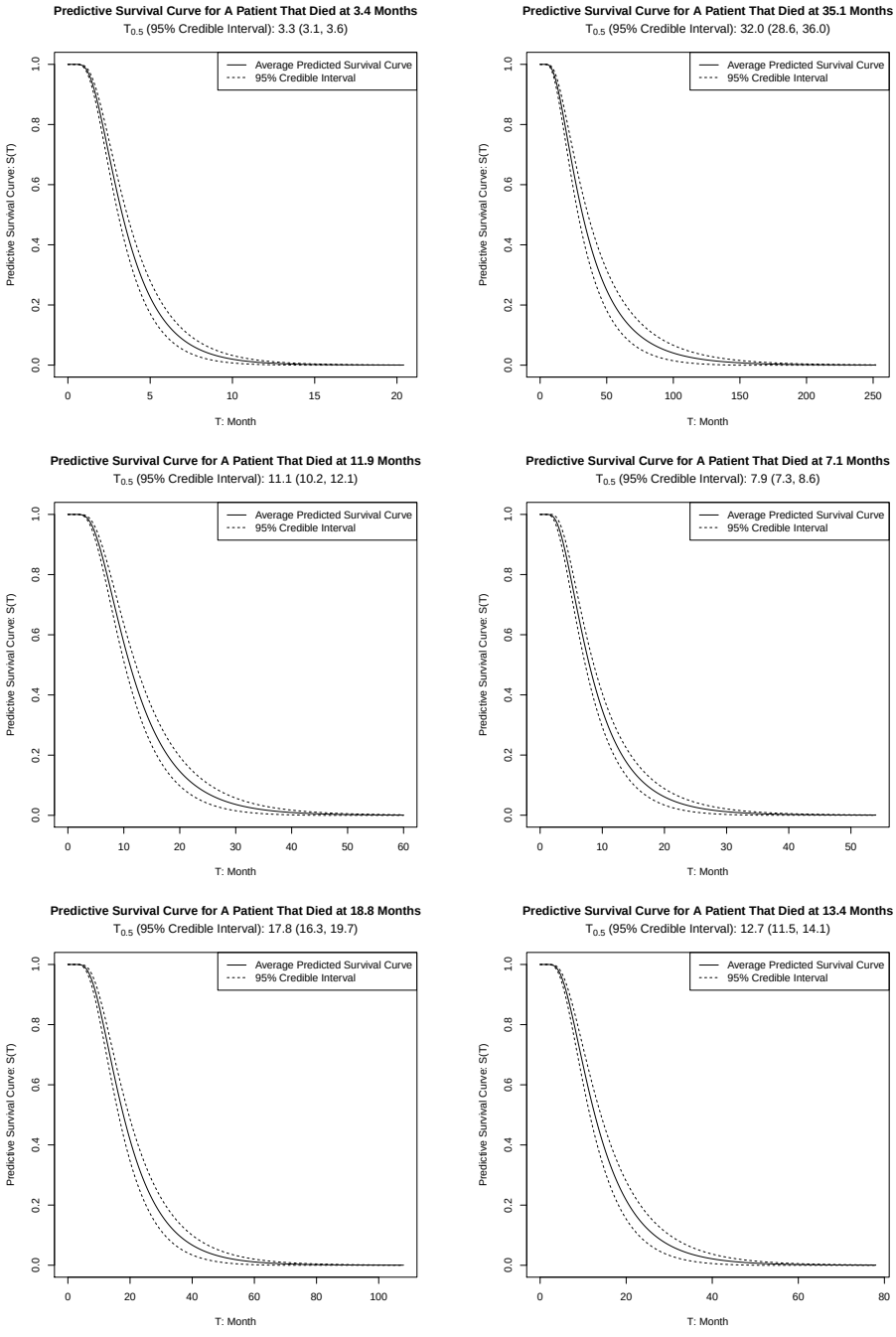

Figure 3.7: Predicted survival curves for patients who died

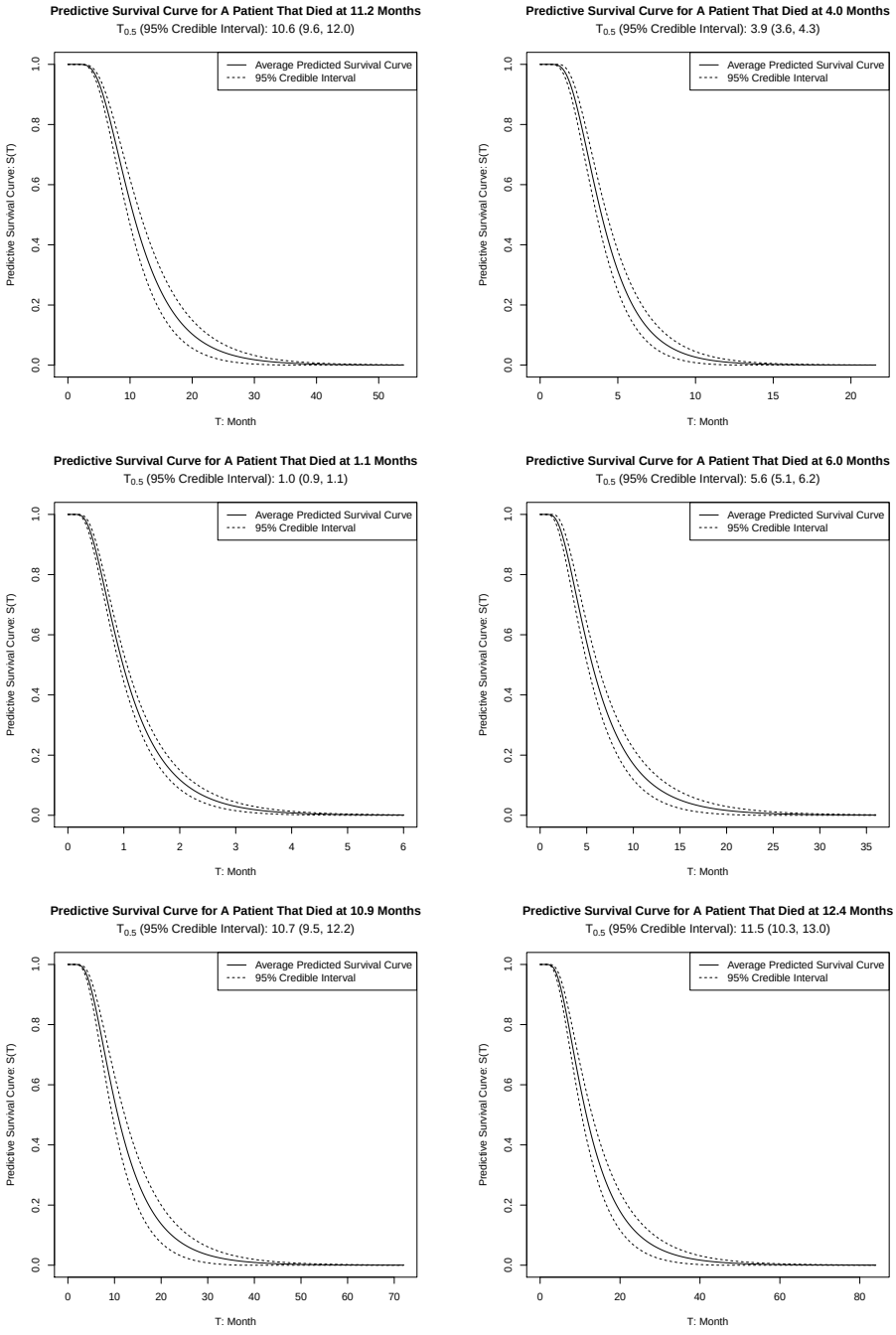

Figure 3.8: Predicted survival curves for patients who died

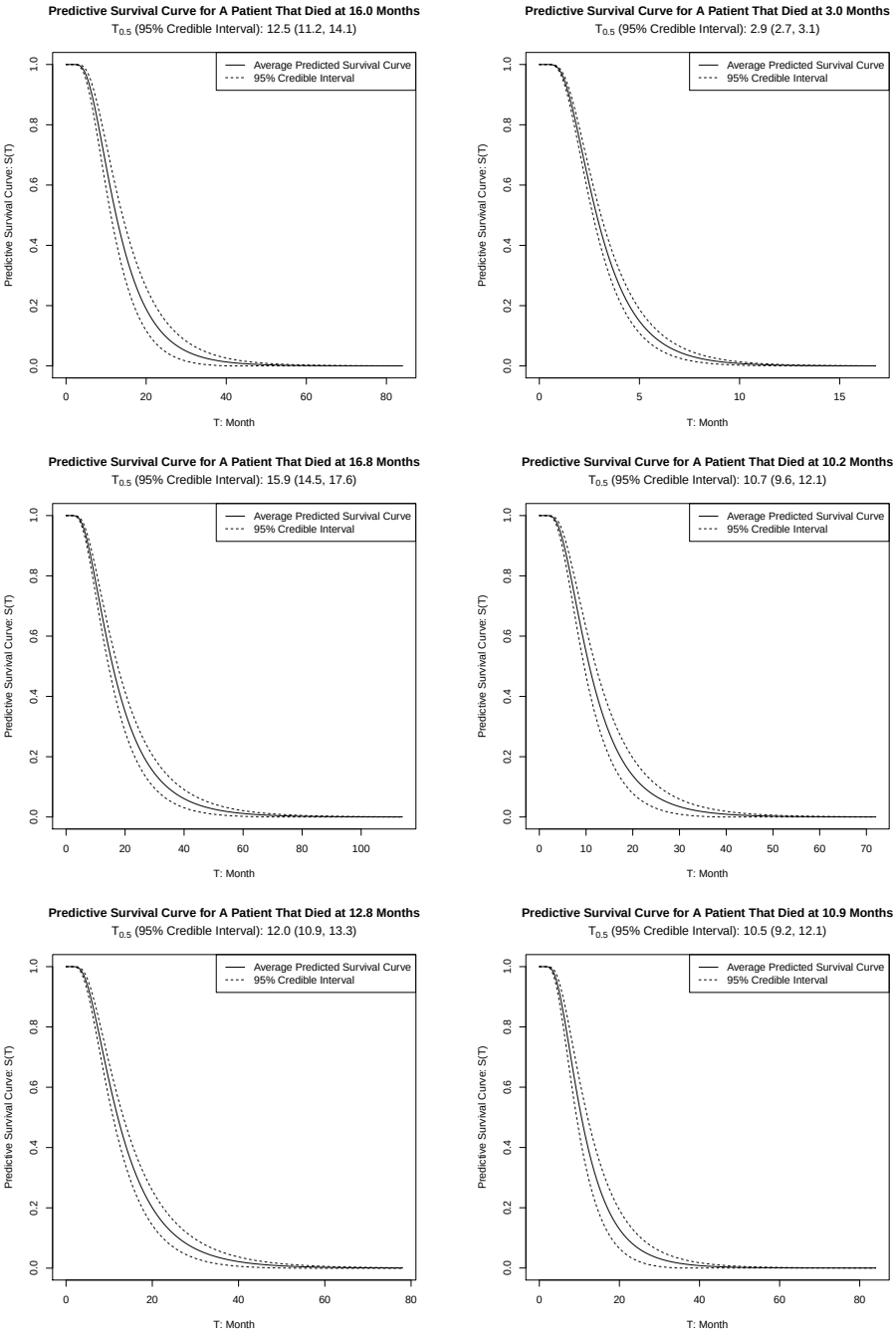

Figure 3.9: Predicted survival curves for patients who died

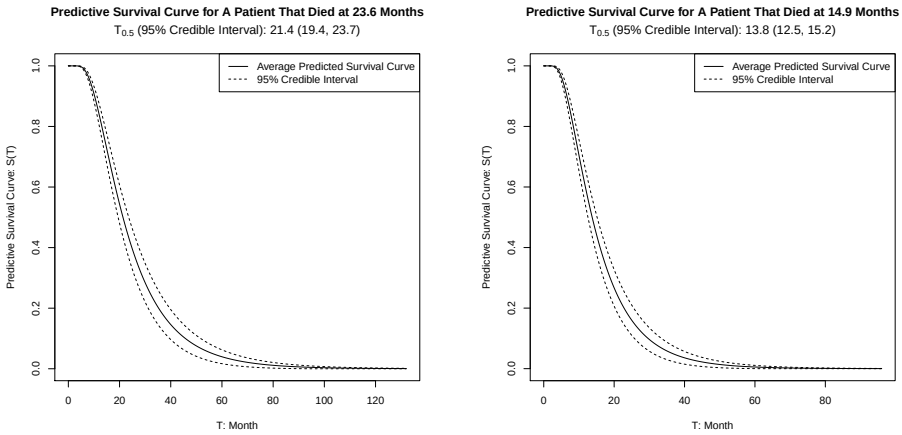

Figure 3.10: Predicted survival curves for patients who died

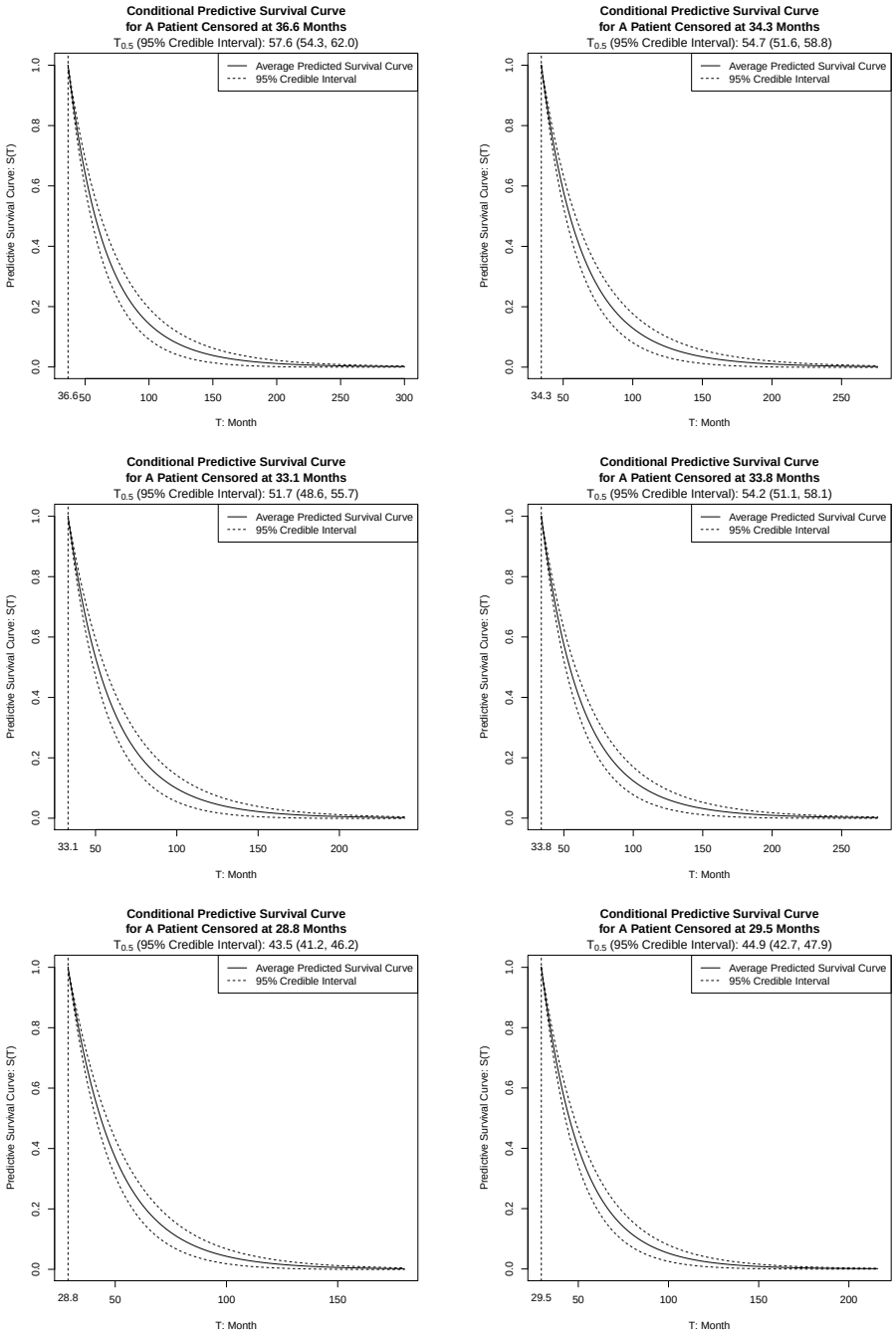

Figure 3.11: Conditional predicted survival curves for patients who censored, conditional on $\mathcal{T}_j > T_j$

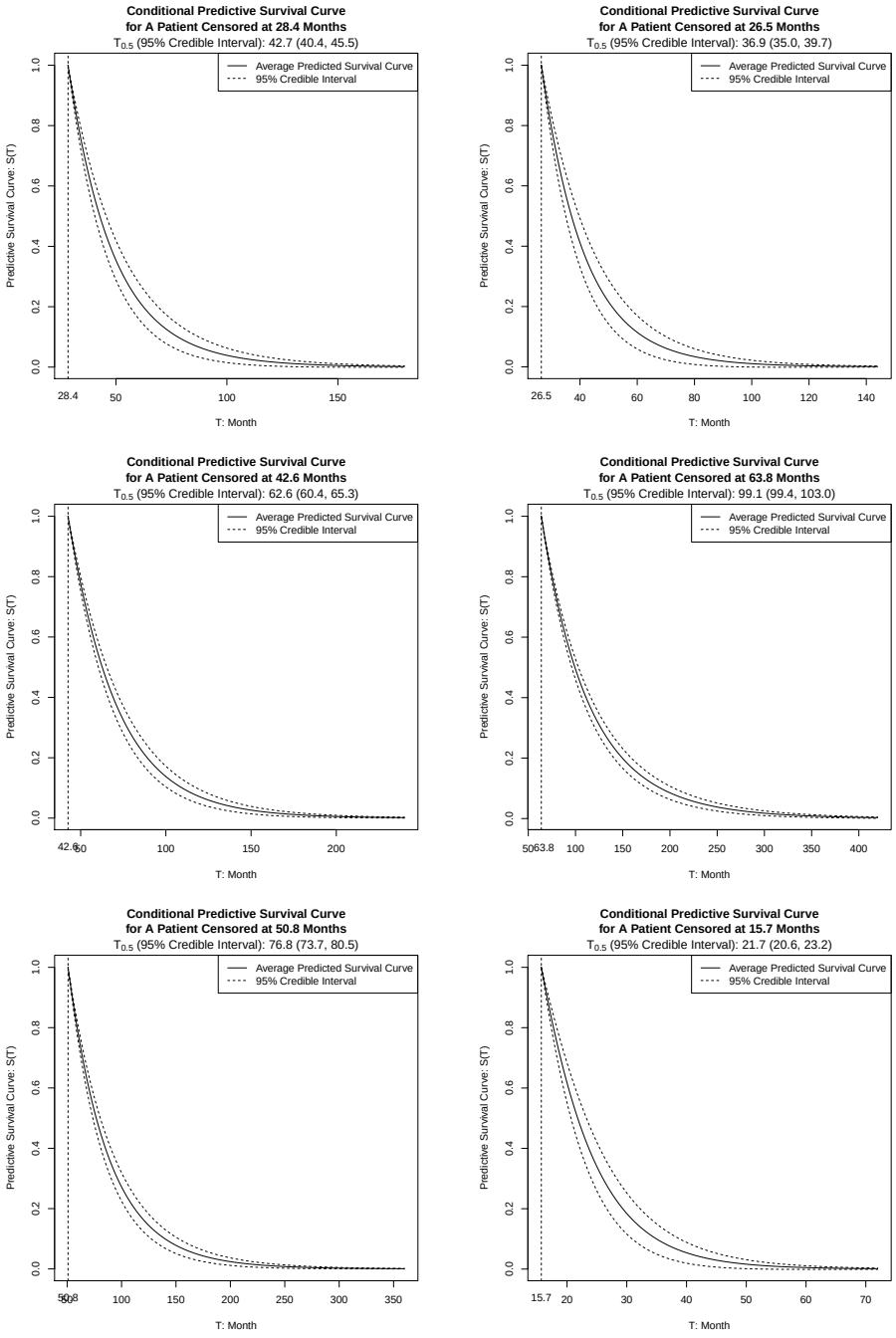

Figure 3.12: Conditional predicted survival curves for patients who censored, conditional on $\mathcal{T}_j > T_j$

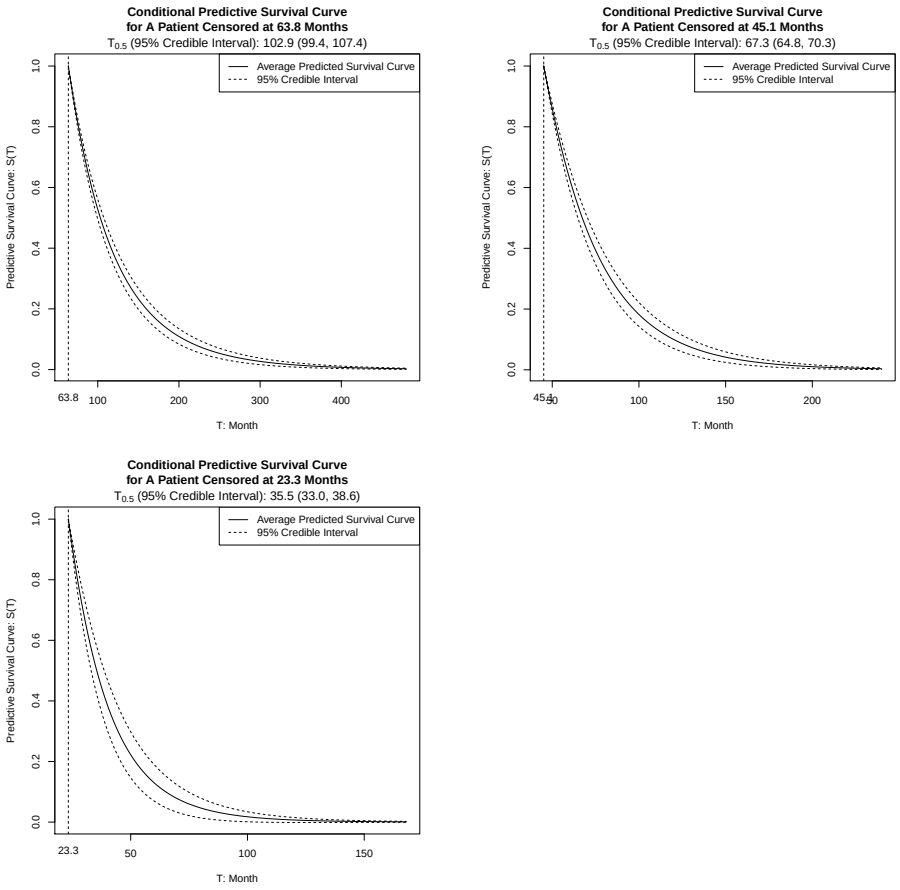

Figure 3.13: Conditional predicted survival curves for patients who censored, conditional on $\mathcal{T}_j > T_j$

CHAPTER IV

Joint Modeling of MRI and Polychotomous Disease Status Using Wavelets: An Application to Alzheimer's Disease

4.1 Introduction

Our work is motivated by a large-scale project: the Alzheimer's Disease Neuroimaging Initiative (ADNI) (adni.loni.ucla.edu). The project is primarily concerned with developing imaging measures and biomarkers for incipient Alzheimer's Disease (AD) and Mild Cognitive Impairment (MCI). AD is an irreversible neurodegenerative disease which results in a loss of mental function due to the deterioration of brain tissue. It is the most common cause of dementia among people over the age of 65, affecting over 24 million people worldwide and 5 million in the U.S. (Leow et al. (2009)). As the disease progresses, it gradually destroys a person's memory and ability to learn and carry out daily activities, and eventually, overall mental and physical function leading to death. Alzheimer's disease develops slowly over time. The symptoms begin to appear so gradually that it is often mistaken for normal aging. The transitional stage between normal aging and dementia is MCI. MCI differs from both AD and normal age-related memory change. People with MCI have ongoing memory problems but not to the point where their impairment interferes significantly with daily activities. Unfortunately, there is no cure for Alzheimer's disease and there is no way to predict how fast someone will progress through the

stages of the disease. However, early Alzheimer's (i.e. especially in the stage of MCI) diagnosis and treatment can slow the progression of Alzheimer's symptoms. Therefore, our study aims to build a predictive model that can aid in the diagnosis of AD and MCI, using MRI data.

Alzheimer's disease is conventionally considered a gray matter disease as its most prominent consequence is severe memory loss. Because neurons in the gray matter are vital to cognitive activity, scientists have generally assumed that Alzheimer's disease – with its memory loss and other cognitive debilitations – must begin in the gray matter. The overwhelming majority of Alzheimer's research has focused on gray matter, despite the fact that white matter makes up about 50 percent of total brain tissue and is substantially altered during Alzheimer's progression (De Leeuw et al. (2005)). Some researchers in recent MRI studies have found that the white matter changes (WMC) (e.g. white matter lesions) are associated with the cognitive decline in early-onset Alzheimer's disease (Kavcic et al. (2004), Bronge and Wahlund (2007), Kavcic et al. (2008), Anderson (2009)). However, none of the previous studies on WMC have developed statistically robust and efficient methods to classify AD and MCI patients from healthy controls. Moreover, the WMC may occur in different brain regions that may vary in size (Braccoa et al. (2005)). The heterogeneity of WMC also makes it difficult to use MRI as an accurate diagnostic tool.

In our study, we apply a three dimensional discrete wavelet transformation (DWT) on MRI data for each patient in stage I. The properties of the wavelet transformation can mitigate the inherent problems of WMC in MRI and allow us to separate the signal of WMC and the noise in the wavelet domain. We then employ Bayesian Lasso (Park and Casella (2008)) to denoise the wavelet transformed images. Summary statistics are derived based on the smoothed wavelet coefficients and are then

included in a Bayesian cumulative probit regression model with the patient's disease status (i.e. AD, MCI and normal) as the outcome. Here the MRI data and the categorical outcome variable are modeled jointly.

Wavelets are useful spectral image analysis tools capable of exploiting significant spatial correlations between voxels and have been used with great success in many areas, especially in imaging analysis. Wavelets have a number of advantages over Fourier analysis, as demonstrated in the literature, including the ability to represent finite and discrete signals and the ability to natively and easily accommodate multi-resolution analysis (Vidakovic (1999), Bullmore et al. (2004)).

MRI data usually have poor signal-to-noise ratio (SNR), arising from intrinsic biological heterogeneity and scanner-induced noise. The direct statistical analysis of these data in the spatial domain is problematic because of a poor SNR, the large number of voxels and strong spatial correlation among them. Wavelets are efficient for the representation of a wide variety of signals. If the signal to be detected is spatially localized, it can be represented by a small number of strong local coefficients, while the power of white noise is uniformly spread throughout the wavelet space. These are due to the three major features of the wavelet transformation: 1) wavelets allow an easy separation of the data by filtering the data with different supports; 2) wavelets disbalance energy in the data as the signal is represented by a small number of components in the wavelet domain; 3) wavelets whiten data as the orthogonal wavelets decorrelate the data in the wavelet domain. Another important advantage of using the wavelet transformation is the ability to perform shrinkage of the parameters to gain a better signal to noise ratio. This is also due to the multi-resolution nature of the transform in that the signal tends to be concentrated in a few coefficients while the noise is spread through all coefficients.

We adopt the Bayesian Lasso proposed by Park and Casella (2008) to further shrink the wavelet coefficients to denoise the images. By placing a Laplacian distribution as the prior for the expected value of the wavelet coefficients, we can shrink most of the non-important wavelet coefficients to zero. The Bayesian Lasso pulls the more weakly related parameters to 0 faster than ridge regression does, indicating a potential advantage of the Laplacian prior over a Gaussian (or a Student t) prior. The summary statistics are derived on the smoothed wavelet transformed images and are then included into a cumulative probit regression model in stage II. Reversible jump MCMC is utilized to select the variables.

In our data, there are 42 AD patients, 85 MCI subjects at high risk for conversion to AD, and 59 age matched controls. MRI scans at 3 Tesla are performed on all subjects at baseline. All MRI data are obtained from the ADNI project and have gone through image correction: gradwarping for correction of geometric distortion due to gradient non-linearity (Jovicich et al. (2006)); "B1-correction" to adjust for image intensity inhomogeneity due to B1 non uniformity using calibration scans (Jack Jr et al. (2008)); "N3" bias field correction, for reducing residual intensity inhomogeneity (Sled et al. (1998)); and also have been scaled for gradient drift using phantom data (Jack Jr et al. (2008)).

The 3-D images for all the subjects are first segmented into white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) using SPM8 (Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, London UK) in Matlab (www.mathworks.com/products/matlab/). To adjust for global differences in brain positioning and scale across individuals, the WM, GM and CSF segments are then further normalized to the tissue probability maps defined by the International Consortium for Brain Mapping (ICBM-53) (Mazziotta et al. (2001)) with a 9-parameter

transformation (3 translations, 3 rotations, 3 scales). Modulation is carried out to compensate for the effect of spatial normalization (Ashburner and Friston (2000)). Moreover, to adjust for intensity inhomogeneities across patients, we further carry out intensity standardization by using the dynamic programming method proposed by Cox et al. (1995). The images after segmentation, are of size $91 \times 109 \times 91$. To make the computation more efficient, we re-sliced the images to a size of $64 \times 64 \times 64$.

This chapter is organized as follows. In Section 4.2, we introduce notation and specify the model. In Section 4.3, we discuss posterior simulation. Results from the simulation studies and the real data analysis are shown in Sections 4.4 and 4.5. In Section 4.6, we discuss some challenges and future work.

4.2 Model

4.2.1 The Wavelet Transformation

The wavelet transform (WT) has been found to be particularly useful for analyzing signals which can best be described as aperiodic, noisy, intermittent, transient and so on. Its ability to examine the signal simultaneously in both time and space in a distinctly different way from the traditional short time Fourier transform has spawned a number of sophisticated wavelet-based methods for signal manipulation and interrogation. Mathematically speaking, the wavelet transform is just the convolution of the wavelet function with the signal. The WT can be done in a smooth continuous fashion for continuous wavelet transform (CWT) or in discrete steps for discrete wavelet transform (DWT). Our method focuses on the DWT for the brain image data.

For simplicity of the exposition, the 1-D case is considered first, which can be easily extended to multiple dimensions. An orthogonal wavelet transform is characterized by two continuously-defined functions: 1) the scaling function $\phi(x)$, and

2) its associated wavelet $\psi(x) = \sqrt{2}\sum_{k\in Z} g(k)\phi(2x-k)$, where $g(k)$ is a suitable weighting sequence. The scaling function ϕ is the solution of a two-scale equation

$$\phi(x) = \sqrt{2}\sum_{k\in Z} h(k)\phi(2x-k). \tag{4.1}$$

The sequence $h(k)$ is the so-called refinement filter. The wavelet basis functions are constructed by dyadic dilation (index j) and translation (index k) of the mother wavelet

$$\psi_{j,k} = 2^{-j/2}\psi(x/2^{-j}-k). \tag{4.2}$$

The sequences h and g – or, equivalently ϕ and ψ – can be selected such that $\{\psi_{j,k}\}_{(j,k)\in Z^2}$ constitutes an orthonormal basis of L_2, the space of finite energy functions. This orthogonality permits the wavelet coefficients $d_i(k)$ and approximate coefficients $c_j(k)$ for any function $f(x) \in L_2$ to be obtained by the inner product with the corresponding basis functions

$$d_j(k) = \langle f, \psi_{j,k}\rangle, \quad c_j(k) = \langle f, \phi_{j,k}\rangle, \tag{4.3}$$

where $\langle f, g\rangle = \int f(x)g(x)dx$ is the conventional L_2-inner product. In practice, the decomposition is only carried out over a finite number of scales J. The wavelet transform with a depth J is then given by

$$f(x) = \sum_{j=1}^{J}\sum_{k\in Z} d_j(k)\psi_{j,k} + \sum_{k\in Z} c_J(k)\phi_{J,k}, \tag{4.4}$$

where $d_j(k)$ and $c_j(k)$ are defined in (4.3)

Although the synthesis and expansion formulas (4.3) and (4.4) are usually given for continuous signals, equivalent expressions also exit for a purely discrete framework. In the discrete case, these formulas can be rewritten in the following matrix form:

$$\boldsymbol{f} = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{d}, \tag{4.5}$$

$$\boldsymbol{d} = \boldsymbol{W}\boldsymbol{f}, \tag{4.6}$$

where $\boldsymbol{f} = (\dots, f(k), \dots)$ is the signal (or image) vector, $\boldsymbol{W}$ the orthogonal wavelet transformation matrix, and $\boldsymbol{d} = (\dots, d_1(k), \dots, d_J(k), \dots, c_J(k), \dots)$ the wavelet coefficient vector. The wavelet transform (4.6) is therefore an orthonormal transformation of the signal vector $\boldsymbol{f}$.

Rather than defining the transform matrix $\boldsymbol{W}$ explicitly, it is much easier to describe the underlying decomposition algorithm, which uses the two complementary filters h and g. In the orthogonal case, the low-pass filter h satisfies the so-called quadrature mirror filter (QMF) conditions

$$H(z)H(z^{-1}) + H(-z)H(-z^{-1}) = 2, \tag{4.7}$$

$$H(1) = \sqrt{2}, H(-1) = 0, \tag{4.8}$$

where $H(z)$ is the transfer function (z-transform) of h. The high-pass filter g is the modulated version of h given by

$$G(z) = z \cdot H(-z^{-1}). \tag{4.9}$$

The lowpass filter lets through low signal frequencies and hence a smoothed version of the signal, while the highpass filter lets through the high frequencies corresponding to the signal details.

The wavelet decomposition is implemented iteratively by successive filtering and decimation using the QMF filterband. The iterative definition of $\boldsymbol{W}$ enables implementation of (4.5) and (4.6) for a signal vector of length N_0 by $O(N_0)$ operations, rather than $O(N_0^2)$. This makes the computation of the wavelet transform slightly more efficient than that of the standard fast Fourier Transform (FFT), which has complexity of $O(N_0 log N_0)$ (Bullmore et al. (2004)).

The decomposition (4.3) is easily extended to two-dimensional (2-D) or three-dimensional (3-D) signals by using tensor product basis functions, which amounts to

applying the 1-D decomposition algorithm successively along the separate dimensions of the data. The effect of one iteration of this splitting process is illustrated in Figure 4.1 for the 2-D case. In this way, one generates 2^q different types of basis functions in q dimensions. The corresponding q-D separable scaling functions with $\boldsymbol{x} = (x_1, \ldots, x_q)$ are given by

$$\phi_{j,\boldsymbol{k}}(\boldsymbol{x}) = \prod_{i=1}^{q} \phi_{j,k_i}(x_i), \tag{4.10}$$

where we use the vector integer index $\boldsymbol{k} = (k_1, \ldots, k_q)$. The other $2_q - 1$ types of wavelet basis functions are obtained in a similar fashion by replacing one or several factors in (4.10) by a wavelet term of the form $\psi_{j,k_i}(x_i)$. Let $\boldsymbol{b} = \{b_1, \ldots, b_q\}$ denote a binary vector with $b_i = 1$ if ϕ_{j,k_i} is replaced by ψ_{j,k_i}, otherwise $b_i = 0$. By defining

$$\varphi_{j,k_i} = \begin{cases} \psi_{j,k_i}, & \text{if } b_i = 1 \\ \phi_{j,k_i}, & \text{otherwise,} \end{cases}$$

the mixed tensor product wavelets are then

$$w^m_{j,\boldsymbol{k}}(\boldsymbol{x}) = \prod_{i=1}^{q} \varphi_{j,k_i}(x_i) \quad m = 1, \ldots, 2^q - 1 \tag{4.11}$$

$$\text{with} \quad m = \sum_{i=1}^{q} b_i 2^{i-1}. \tag{4.12}$$

Since ϕ is a low pass filter and ψ is a high pass filter, the mixed tensor product wavelets will typically have a preferential spatial orientation along one (or several if $q > 2$) of the spatial directions. In this view, m assumes the role of a spatial direction indicator.

In our study, we will use a 3-D Daubechies (Daubechies (1992)) wavelet transform. The Daubechies wavelets are a family of orthogonal wavelets defining a discrete wavelet transform and characterized by compact support and a maximal number of

vanishing moments for any given smoothness. These properties allow an effective and parsimonious representation of images with local behavior. Daubechies wavelets are extensively used in applications. For the following write-up, we use DWT to denote the 3-D Daubechies wavelet transform with 4 vanishing moments.

4.2.2 Joint Modeling MRI and Polychotomous Disease Outcome

Let $(\mathbf{Y}_i, Z_i)$ for $i = 1, \ldots, n$ indicate the observed data, with the vector $\mathbf{Y}_i$ denoting the 3-D MRI data, written as a $64^3 \times 1$ vector, for either GM or WM. Since we will analyze the segmented GM and WM images using the same model separately, we do not use different notation to distinguish two images in order to ease the notational burden. Variable Z_i is the categorical ordinal outcome with value 0 representing normal controls, value 1 MCI and value 2 AD patients.

From Mallat (1989), it is straightforward to calculate the wavelet basis matrix, $\mathbf{W}^3$. Let $\mathbf{W}^3$ contain all the bases of 3 dimensional DWT. For each subject i, we can write out the first model in regression form:

$$\mathbf{Y}_i = \mathbf{W}^3 \boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i \tag{4.13}$$

where $\boldsymbol{\beta}_i$ and $\boldsymbol{\epsilon}_i$ are vectors of the same length $(64^3 \times 1)$ as $\mathbf{Y}_i$. $\boldsymbol{\epsilon}_i$ is the error term, of which every element is assumed to be independent and identically distributed as $N(0, \sigma_i^2)$. $\mathbf{W}^3\boldsymbol{\beta}_i$ represents the 3-D Inverse DWT on $\boldsymbol{\beta}_i$. Tibshirani (1996) suggests that an independent and identical Laplacian prior for the regression coefficients can be viewed as a Bayesian Lasso method. Park and Casella (2008) represent the Laplacian prior as a scale mixture of normals. Hence, we assign a prior distribution for each element of $\boldsymbol{\beta}_i$: $[\beta_{il}|\sigma_i^2, \tau_{il}^2] \sim N(0, \sigma_i^2\tau_{il}^2)$ where all the elements are independent from each other. The hyper-prior distributions for the variance σ_i^2 and the τ_{il}^2,

$l = 1, \ldots, m$, where $m = 64 \times 64 \times 64$, are

$$\pi(\sigma_i^2) \propto 1/\sigma_i^2 \tag{4.14}$$

$$\pi(\tau_{il}^2) = 0.5\lambda_i^2 e^{-\lambda_i^2 \tau_{il}^2/2} \tag{4.15}$$

The parameter λ_i^2 controls the degree of shrinkage for all the wavelet coefficients. In our model, we employ the empirical Bayes method proposed by Park and Casella (2008) to estimate the value of λ_i^2. In our model, we have more parameters than the sample size, therefore, it may lead to an identifiability issue when calculating λ using the marginal maximum likelihood estimate. To solve this problem, we first apply the wavelet transformation directly on the data. Then 5% of the wavelet coefficients which are closest to zero are set to zero. By doing so, we force the number of parameters to be less than the sample size of the data. Then at each iteration, λ_i^2 is calculated from the sample of the previous iteration by using the EM algorithm. Specifically, iteration k uses the Gibbs sampler with the value of $\lambda^{(k-1)}$ from the iteration $k-1$:

$$\lambda^{(k)} = \sqrt{\frac{2(m-m_0)}{\sum_{j=1}^{(m-m_0)} E_{\lambda^{(k-1)}}[\tau_j^2|\mathbf{Y}]}} \tag{4.16}$$

where m_0 is the number of wavelet coefficients that are set to zero. The subscript j in τ_j^2 indexes the voxels without β_j set to zero.

The summary statistics are derived based on the smoothed wavelet coefficients. From exploratory analyses, we notice that the wavelet transformation allows differences among the three groups of subjects to be represented by a small number of wavelet coefficients, while the noise is uniformly spread through the wavelet space (Figure 4.2 and 4.3). It inspires us to use the wavelet coefficients with large absolute magnitudes. We calculate the 97.5^{th} and 2.5^{th} percentile of the smoothed wavelet coefficients for the GM and WM respectively. We denote the 97.5^{th} percentile as

UpGM for GM, UpWM for WM, and 2.5^{th} percentile as LowGM for GM, LowWM for WM. These percentiles are included into the model in stage II as the covariates, denoted as $\boldsymbol{X}_i$ for the i^{th} subject.

For classification purposes, we use a cumulative probit regression for the ordered outcome in stage II. Albert and Chib (1993) proposed a Bayesian approach that uses data augmentation and introduces latent variables into the model. Observed ordinal outcome variable $\mathbf{Z} = \{Z_1, \ldots, Z_n\}$ are coded as $0, 1, 2$, for 3 classes. Let G_i be a latent variable for each Z_i. Assume the latent variable G_i follows a normal distribution with mean $\mathbf{B}(\boldsymbol{X}_i)\boldsymbol{\alpha}$ and variance 1. Here $\mathbf{B}(\boldsymbol{X}_i)$ is the vector of multivariate adaptive regression spline basis functions (MARS, Friedman (1991), Denison et al. (1998) and Holmes and Denison (2003)) as defined in Chapters II and III of this dissertation. Regression coefficients are $\boldsymbol{\alpha} = \{\alpha_0, \ldots, \alpha_K\}$ where K represents the total number of basis functions and α_0 is the model intercept. We observe Z_i, where $Z_i = j - 1$ if $\gamma_{j-1} < G_i \leq \gamma_j$ for $j = 1, 2, 3$ and we define $\gamma_0 = -\infty$ and $\gamma_3 = \infty$. To ensure that the parameters are identifiable, it is necessary to impose one restriction on the bin boundary, without loss of generality, we take $\gamma_1 = 0$. This transforms the generalized linear regression problem into a linear regression problem. We denote the model in this stage as the GNLM-BMARS, standing for the Generalized Non-Linear Model with Bayesian MARS bases.

4.3 Implementation and Evaluation

By utilizing the orthonormality property of the wavelet basis, it is straightforward to derive the posterior distributions for σ_i^2 and τ_{il}^{-2}:

$$[\sigma_i^2 | \cdot\,] \sim \text{Inv-Gamma}\left(m,\ \frac{\sum_{l=1}^{m}\left(Y_{li} - [\mathbf{W}^3\boldsymbol{\beta}_i]_l\right)^2}{2} + \frac{\sum_{l=1}^{m}\beta_{il}^2}{2\tau_{il}^2}\right), \tag{4.17}$$

$$[\tau_{il}^{-2} | \cdot\,] \sim \text{Inv-Gaussian}\left(\sqrt{\frac{\lambda_i^2\sigma_i^2}{\beta_{il}^2}},\ \lambda_i^2\right) \quad l = 1, \ldots, m \tag{4.18}$$

where $\mathbf{W}^3\boldsymbol{\beta}_i$ denotes the inverse 3-D DWT on $\boldsymbol{\beta}_i$, and is the predictive value of the image denoted by $\hat{\mathbf{Y}}_i$. $[\mathbf{W}^3\boldsymbol{\beta}_i]_l$ represents the lth voxel of the image $\hat{\mathbf{Y}}_i$. Inv-Gamma stands for the inverse Gamma distribution. Inv-Gaussian is the inverse Gaussian distribution (μ', λ') with density:

$$f(x) = \sqrt{\frac{\lambda'}{2\pi}}x^{-3/2}\exp\left\{-\frac{\lambda'(x-\mu')^2}{2(\mu')^2 x}\right\}, \quad x > 0.$$

To derive the full conditional posterior distribution of β_{il}, we note that

$$\pi(\beta_{il} | \cdot\,) \propto \pi(\beta_{il} | \mathbf{Y}_i, \sigma_i^2, \tau_{il}^2)\pi(G_i | \boldsymbol{X}_i(\beta_{il}, \cdot\,), \boldsymbol{\Omega}_2) \tag{4.19}$$

where $\boldsymbol{\Omega}_2$ denotes all the parameters in stage II. And with the orthonormality property of the wavelet basis, the distribution of $\pi(\beta_{il} |\ \mathbf{Y}_i, \sigma_i^2, \tau_{il}^2\ )$ is:

$$[\beta_{il} |\ \mathbf{Y}_i, \sigma_i^2, \tau_{il}^2\ ] \sim N\left(\frac{Y_{li}^w}{\tau_{il}^{-2}+1},\ \frac{\sigma_i^2}{\tau_{il}^{-2}+1}\right) \quad l = 1, \ldots, m \tag{4.20}$$

where Y_{li}^w indicates the transformed value of Y_{li} in the wavelet domain. Therefore, we propose a new value β_{il}^{prop} from (4.20), and accept this draw with probability:

$$\alpha_\beta = \min\left\{1, \frac{\pi(G_i \mid \mathbf{X}_i(\beta_{il}^{prop}, \cdot\,), \boldsymbol{\Omega}_2)}{\pi(G_i \mid \mathbf{X}_i(\beta_{il}^{current}, \cdot\,), \boldsymbol{\Omega}_2)}\right\} \tag{4.21}$$

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

In stage II, we place priors over the MARS parameters in the same way as described in Chapters II and III. We assign prior distributions to regression parameters $\boldsymbol{\alpha}$ and ν

$$[\boldsymbol{\alpha}|\nu] \sim N(\mathbf{0}, \nu \boldsymbol{I}_{K+1}), \tag{4.22}$$

$$\left[\nu^{-1}\right] \sim \text{Gamma}(0.001, 0.001). \tag{4.23}$$

It is straightforward to derive the posterior distributions:

$$[\boldsymbol{\alpha}| \cdot] \sim N(\mathbf{m}, \mathbf{V}), \tag{4.24}$$

$$[G_i| \cdot] \sim N(\mathbf{B}\left(\boldsymbol{X}_i\right)^{\mathrm{T}} \boldsymbol{\alpha}, 1) I(Z_i = j-1), \tag{4.25}$$

$$\text{if } \gamma_{j-1} < G_i \leq \gamma_j \quad j = 1, 2, 3,$$

$$[\gamma_j| \cdot] \sim \text{Unif}\left[L_j, U_j\right] \quad j = 2, \tag{4.26}$$

$$\left[\nu^{-1}| \cdot \right] \sim G\left(0.001 + \frac{K+1}{2}, 0.001 + \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}\right) \quad j = 1, 2 \tag{4.27}$$

where $\mathbf{V} = \left(\mathbf{B}^{\mathrm{T}}\mathbf{B} + (\nu \boldsymbol{I}_{K+1})^{-1}\right)^{-1}$, $\hat{\boldsymbol{\alpha}} = \mathbf{V}\mathbf{B}^{\mathrm{T}}\mathbf{G}$, $\mathbf{B}$ is the matrix of MARS basis functions defined in Chapters II and III with the i^{th} row vector $\mathbf{B}\left(\boldsymbol{X}_i\right)$, $L_j = \max\{\max\left\{G_i : Z_i = j-1\right\}, \gamma_{j-1}\}$, $U_j = \min\left\{\min\left\{G_i : Z_i = j\right\}, \gamma_{j+1}\right\}$ and I is the indicator function. For the MARS parameters, we specify the same prior distributions as described in Chapters II and III.

We use a varying dimensional reversible jump sampler (Green (1995)) to sample appropriate models by adding or removing basis functions. This can be referred to as a "variable selection" approach. Details of the RJMCMC algorithm are given in the appendices of Chapters II and III.

To evaluate the model, we adopt a a leave-one-out cross-validation approach proposed by Gelfand et al. (1992). Their method uses the idea of importance sampling. Details of their method can also be found in the Appendix of Chapter II.

4.4 Simulation Studies

We perform a series of simulation studies to assess the performance of the model in stage I. We firstly construct a smoothed image in the wavelet domain of size $16 \times 16 \times 16$ and regard this image as the truth. The intensity of voxel l in the image is indexed as β_l^{True}. We generate $\beta_l^{True}, l = 1, \ldots, m = 16^3$ from a Gaussian random field with mean 2 and covariance function specified by a variance of 0.25, a nugget of 1 and a scale of 10. To introduce sparsity into the image, we assume that 80% of the voxels in the smoothed images are 0. Therefore, we randomly sample $0.8(16^3)$ number of voxels from the image and set them to 0. We calculate the 97.5^{th} percentile of $\boldsymbol{\beta}^{True}$ and denote it as Q^{True}. The observed image denoted by $\mathbf{Y}$ is constructed from formula (4.13) with $\epsilon_l \sim N(0, \sigma^2)$. We apply the Wavelet transformation with Bayesian Lasso to the simulated data $\mathbf{Y}$ and calculate the 97.5^{th} percentile of $\boldsymbol{\beta}$ and denote it by Q^{Lasso}. We vary the value of σ^2 and compare Q^{Lasso} with Q^{True}. Under each simulation scenario, we generate $N = 500$ simulated data sets and compute the average mean squared relative error (rMSE) and relative bias (rBias) of the statistics:

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

where $\overline{Q}_i^{Lasso}$ is the posterior mean from the ith simulation.

To illustrate the benefits of the Bayesian Lasso shrinkage, we calculate the 97.5^{th} percentile of the maximum likelihood estimate of β_l, which is the wavelet coefficient $\mathbf{Y}_l^w$. We denote the percentile as Q^{MLE} and compute the rMSE and rBias of Q^{MLE}. The results are shown in Table 4.1. We find that by using the Bayesian Lasso as the prior for the wavelet transformation, we can attain much lower rMSE and rBias

than those obtained by the MLE of β. As σ^2 increases, our method performs better. These indicate that the shrinkage by using Bayesian Lasso can improve the precision and bias of the results.

4.5 Results

We iteratively sample from the posterior distributions between the two stages. Stage I is computationally much more expensive than stage II due to the large sample size of the imaging data. We run the algorithm for 60,000 iterations and we over-sample draws (10:1) in stage II. It takes around 3 days for all 186 subjects. The algorithms are programmed in C and implemented on a 3.0 GHz Mac XServe. The first 30,000 draws are discarded as burn-in. The burn-in period is chosen by examining plots of all the parameters, which have converged by this point.

The results of the ADNI data are shown in Table 4.2. We calculated the overall cross-validated correct classification rate, CCR_{CV}, and the proportions of the predicted counts of each disease group among the the total observed counts of that group (i.e. total number of truths). Our proposed model yields an overall CCR_{CV} of 62.4%(116/186). From Table 4.2, we can see that our model can correctly classify 69.0% of the AD patients, 75.3% of the MCI patients and only 39.0% of the normal controls. Among 59 normal controls, our model mis-classifies 54.2% of subjects as MCI. It suggests the our model can not discriminate normal controls from MCI.

We then apply our proposed model to subsets of the ADNI data: the sub-data containing only two categories of the disease status (i.e. AD vs. NORM, AD vs. MCI and MCI vs. NORM). We compute the positive predictive value, PPV_{CV}, which is defined as the probability of being in group A conditional on a prediction of A when comparing group A vs. group B. And, the negative predictive value, NPV_{CV}

is defined as the probability of being in group B given a prediction of B. From Table 4.6, we can see that our proposed model gives a CCR_{CV} of 71.3% when comparing AD vs. NORM, with PPV_{CV} of 67.6% and NPV_{CV} of 73.4%. The CCR_{CV} of comparing AD vs. MCI and comparing MCI vs. NORM are 66.9% and 55.6% respectively. Therefore, we can see that the classification rate is poor when comparing MCI with the normal controls, which is consistent with what we have found by applying our model to the entire data set.

To check the performance of our proposed model, we compare the results with those obtained from three simpler models. In one model, we include the wavelet coefficients directly into the GNLM-BMARS model in stage II and we denote this model as Wavelet-BMARS. The results obtained from this model are shown in Table 4.3. This model yields an overall CCR_{CV} of 58.1%(108/186). Comparing with the results in Table 4.2, we can see that our proposed model can give us better overall classification results by shrinking the wavelet coefficients using the Bayesian Lasso. We also compare our results with results from a second model, in which we include the wavelet coefficients into a cumulative probit regression in stage II. This model is denoted as Wavelet-GLM and it yields an overall CCR_{CV} of 47.3%(88/186). In Table 4.4, we can see that this model performs poorly in classifying the three disease statuses. Comparing the results in Table 4.2, 4.4 and 4.3, we find that BMARS can provide extra flexibility to the model and improve prediction. In the third model, we calculate the summary statistics directly on the raw image data and then include the statistics as covariates into the cumulative probit regression model in stage II. We found that this model, denoted as Raw-GLM, gives the worst classification results with an overall CCR_{CV} of 43.0%(88/186) (Table 4.5). Therefore, we conclude that the wavelet transformation can also contribute to better classification results by

separating the signal from noise.

We also compute the CCR_{CV}, PPV_{CV} and NPV_{CV} after analyzing the subset of the ADNI data using the three simple models described in the previous paragraph: Wavelet-BMARS, Wavelet-GLM and Raw-GLM. The results are shown in Table 4.6. By comparing all four models, we find again that our joint model outperforms the simple models in terms of prediction.

4.6 Discussion

In this chapter, we propose a Bayesian joint classification model with discrete wavelet transformation to aid in the diagnosis of Alzheimer's Disease and the transitional stage: MCI. In stage I, we apply the DWT on the image data to de-correlate the voxels and to separate the signal from noise. We further shrink the wavelet coefficients by using the Bayesian Lasso to denoise the images. The summary statistics are derived based on the wavelet coefficients after shrinkage and are included as covariates into the model in stage II. In stage II, we build a cumulative probit regression model for the polychotomous outcome – disease status with three levels: AD, MCI and normal controls. We employ the BMARS basis functions to introduce more flexibility into the model. Through both simulation studies and model performance comparisons, we find that we can improve classification results by de-noising the images in the wavelet domain using the Bayesian Lasso and by allowing for a more complex and flexible decision boundary provided by the GNLM with MARS.

From our study, we find that the our method does not perform well in classifying the transitional stage MCI from the AD and especially from the normal control group. Compared with the large body of information about Alzheimers disease, research about MCI is at a relatively early stage. Scientists are still answering basic questions

about this disorder and there are significant inconsistencies in definitions of MCI. For example, some definitions involve problems with aspects of thinking other than memory. The different findings also point to the need for long-term studies that follow the progression of symptoms in people with differently defined MCI. Therefore, more work is needed on the biological changes associated with normal aging, MCI, and Alzheimers disease and other dementias. Moreover, in our study, we only consider MRI as a predictor for disease status. Clinical characteristics obtained from the cognitive assessments and physical examination can also be included in our model. In the ADNI project, besides the image and clinical data, there is also substantial genetic data available. Thus more genetics studies about Alzheimer's disease are also in great demand, in order to better understand the disease.

Our proposed model could also be applied to classify Multiple Sclerosis (MS) patients from normal subjects. MS is an inflammatory, chronic, degenerative disorder that affects nerves in the brain and spinal cord. Predominantly, it is a disease of the "white matter" tissue. In people affected by MS, patches of damage called plaques or lesions appear in seemingly random areas of the CNS white matter. MS plaques appear as irregular, sharply demarcated, gray areas randomly distributed in the white matter. There is no known cure for MS. Treatments attempt to return function after an attack, prevent new attacks, and prevent disability. Moreover, the prognosis of MS is difficult to predict. It has been shown that MRI is a sensitive biomarker for MS (Paty et al. (1988)). Old plaques are hyperintense on T2-weighted and FLAIR studies. The plaques in MRI are of different sizes, shapes, locations and are highly heterogeneous. Therefore, our proposed method should also be applicable to MS.

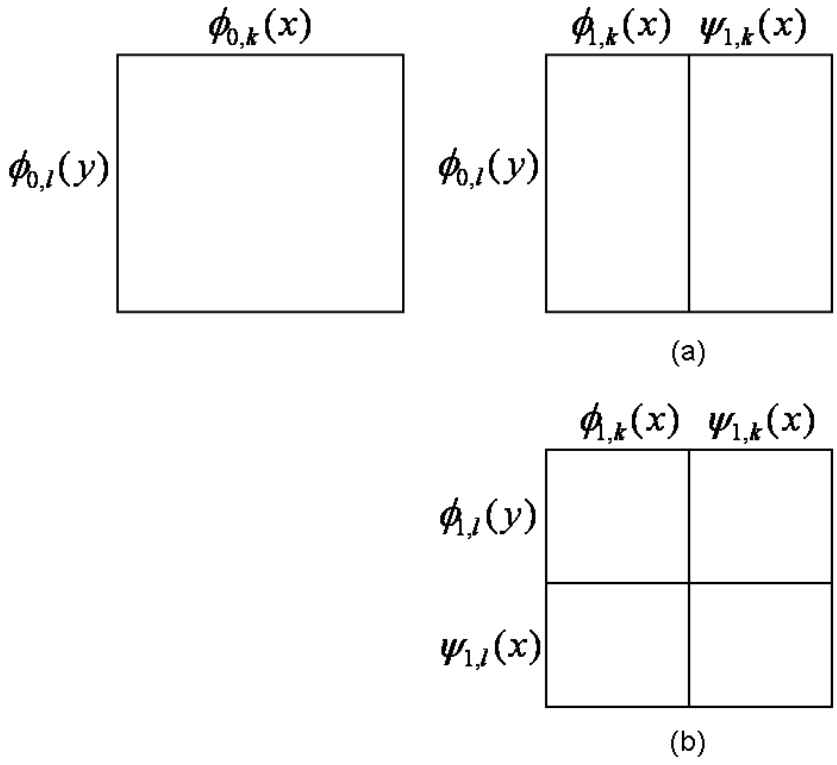

Figure 4.1: (a) One iteration of the separable wavelet transform in 2-D. First, the basic 1-D algorithm is applied in the x-direction, which splits the columns of the data into two halves. Second, it is applied in the y-direction with (a) as input, splitting the rows into two halves. (b) The basis functions for each quadrant are obtained from the product of the corresponding basis functions in x and y. The procedure is then iterated on the upper left quadrant in (b).

Table 4.1: Average rMSE , rBias comparing Q^{Lasso} and Q^{True} and Q^{MLE} and Q^{True}

	Q^{MLE}		Q^{Lasso}	
σ^2	rMSE $(SD)^1$	rBias $(SD)^2$	rMSE $(SD)^1$	rBias $(SD)^2$
1.0	12.9(6.0)	11.0(2.71)	0.94(1.36)	1.78(2.52)
0.9	4.68(3.26)	6.36(2.53)	0.66(1.03)	1.36(2.20)
0.8	1.49(2.11)	2.70(2.78)	0.56(0.97)	0.99(2.18)
0.5	0.56(0.84)	−0.65(2.32)	0.55(0.84)	−0.61(2.31)

[1] $\times 10^{-3}$. That is, all numbers are to be multiplied by .001
[2] $\times 10^{-2}$. That is, all numbers are to be multiplied by .01

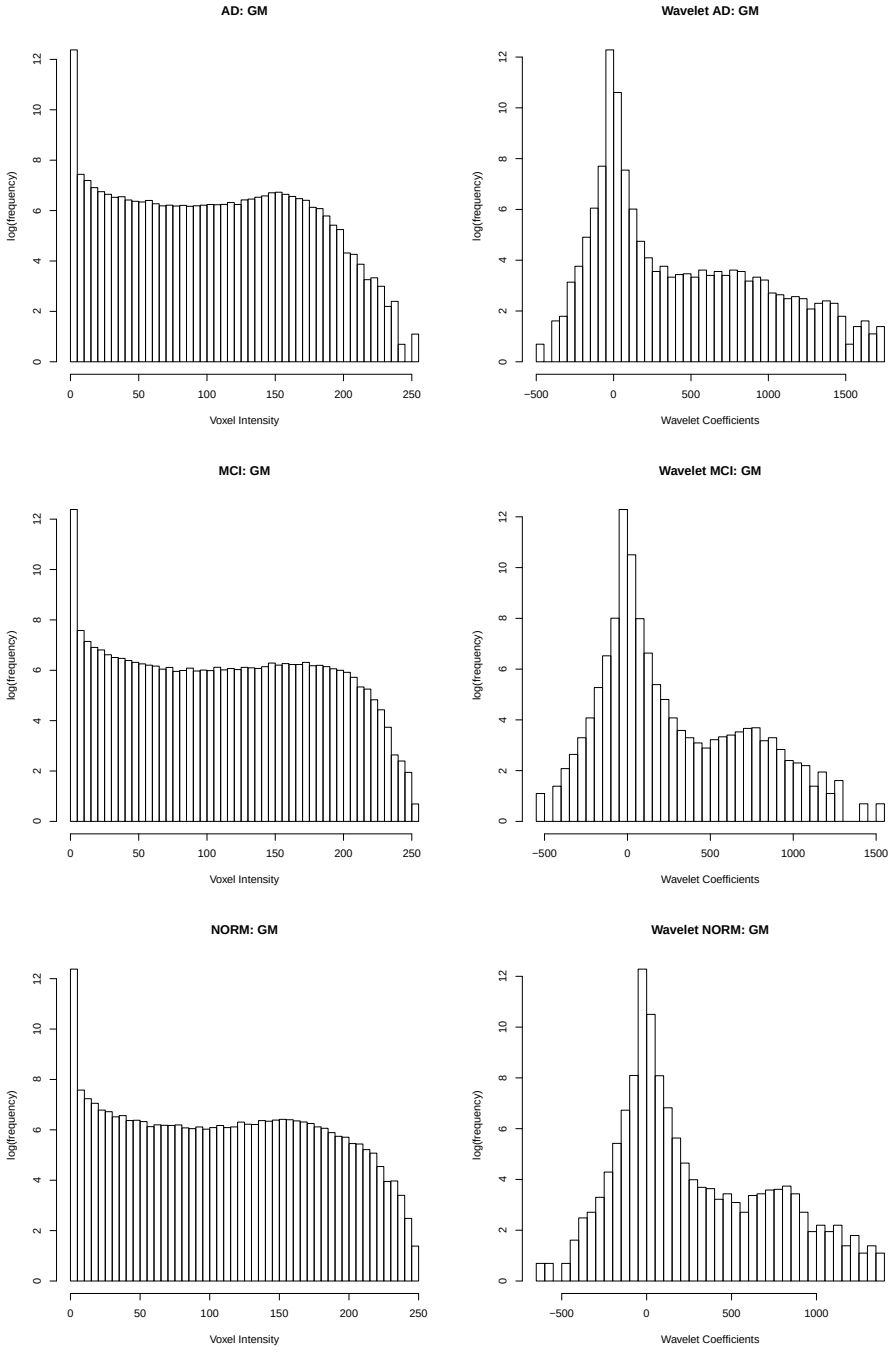

Figure 4.2: The figures in the left panel represent log-histograms of the raw segmented gray matter of AD, MCI and normal subjects, while in the right panel, the figures represent the log-histograms of the segmented gray matter in the wavelet domain for AD, MCI and normal subjects.

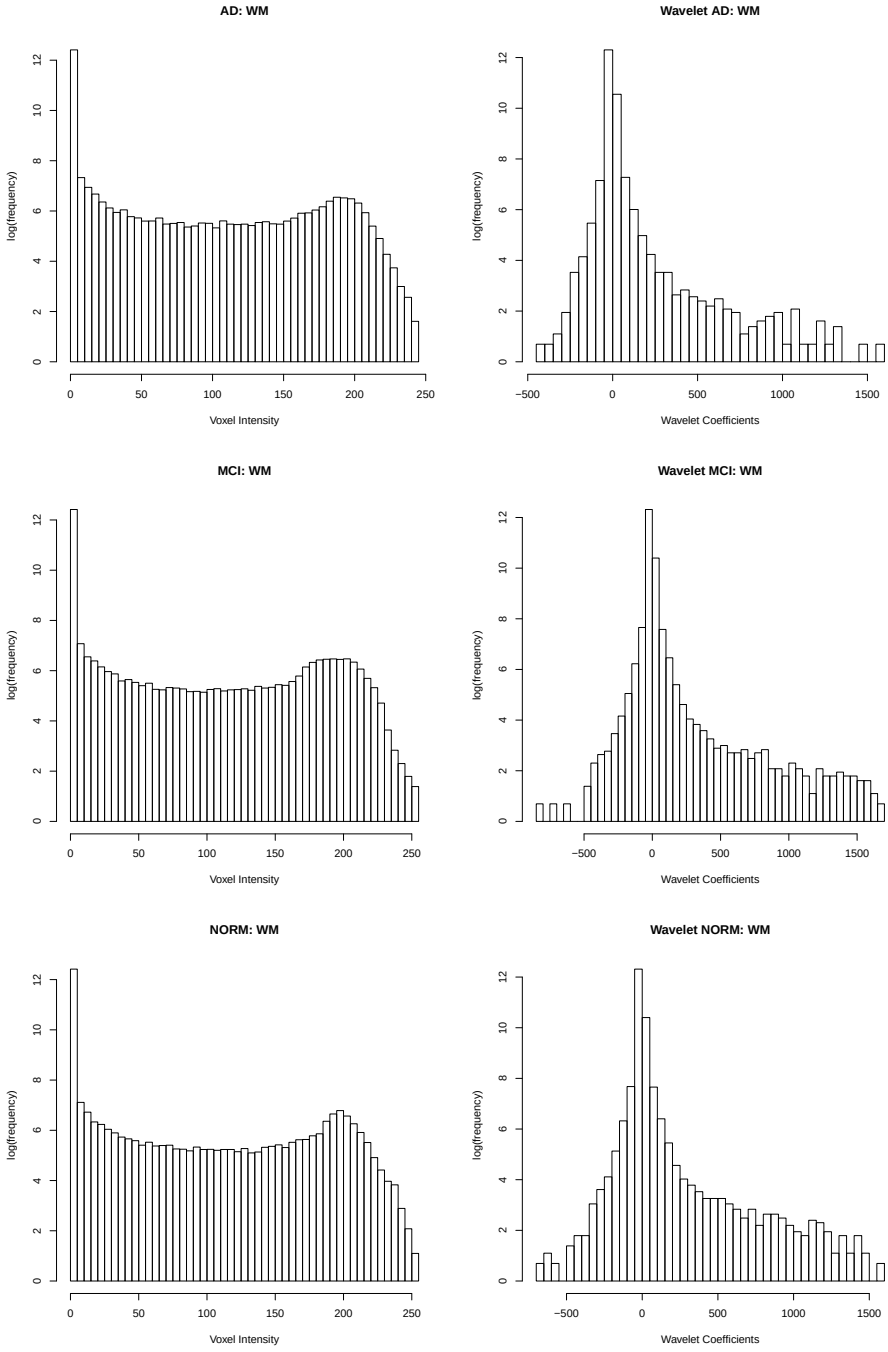

Figure 4.3: The figures in the left panel represent log-histograms of the raw segmented white matter of AD, MCI and normal subjects, while in the right panel, the figures represent the log-histograms of the segmented white matter in the wavelet domain for AD, MCI and normal subjects.

Table 4.2: Overall classification results obtained by our proposed method.

Truth	Prediction: Counts			Prediction: Proportion* %		
	AD	MCI	NORM	AD	MCI	NORM
AD	29	12	1	69.0	28.6	2.38
MCI	9	64	12	10.6	75.3	14.1
NORM	4	32	23	6.78	54.2	39.0

*The proportion is calculated as the proportion of the predicted counts of certain group among the the total observed counts of that group (i.e. total number of truths).

Table 4.3: Overall classification results obtained by using wavelet transformation without shrinkage via Bayesian Lasso in a GNLM with BMARS.

Truth	Prediction: Counts			Prediction: Proportion* %		
	AD	MCI	NORM	AD	MCI	NORM
AD	27	15	0	64.3	35.7	0
MCI	8	67	10	9.41	78.8	11.8
NORM	4	41	14	6.78	69.5	23.7

*The proportion is calculated as the proportion of the predicted counts of certain group among the the total observed counts of that group (i.e. total number of truths).

Table 4.4: Overall classification results obtained by using wavelet transformation without shrinkage via Bayesian Lasso in a cumulative probit regression model.

Truth	Prediction: Counts			Prediction: Proportion* %		
	AD	MCI	NORM	AD	MCI	NORM
AD	12	28	2	28.6	66.7	4.76
MCI	4	66	15	4.71	77.6	17.6
NORM	2	47	10	3.39	79.7	16.9

*The proportion is calculated as the proportion of the predicted counts of certain group among the the total observed counts of that group (i.e. total number of truths).

Table 4.5: Overall classification results obtained by including summary statistics derived directly from the raw image data in a cumulative probit regression model.

Truth	Prediction: Counts			Prediction: Proportion* %		
	AD	MCI	NORM	AD	MCI	NORM
AD	2	38	2	4.76	90.5	4.76
MCI	4	77	4	4.71	90.6	4.71
NORM	2	56	1	3.39	94.9	1.69

*The proportion is calculated as the proportion of the predicted counts of certain group among the the total observed counts of that group (i.e. total number of truths).

Table 4.6: Model comparisons based on subsets of ADNI data. WLasso-BMARS represents our proposed joint model with the wavelet coefficients after shrinkage using Bayesian Lasso, included into the GNLM with BMARS; WNLasso-BMARS represents the model with the wavelet coefficients directly included into the GNLM with BMARS; Wavelet-GLM represents the model with the wavelet coefficients included into a cumulative probit regression. Raw-GLM represents the model with the summary statistics derived directly from the raw image data and then included into a cumulative probit regression model.

Model		[1]CCR_{cv}	[2]PPV_{cv}	[3]NPV_{cv}
WLasso-BMARS	AD vs. NORM	0.713 (72/101)	0.676 (25/37)	0.734 (47/64)
	AD vs. MCI	0.669 (85/127)	0.500 (13/26)	0.713 (72/101)
	MCI vs. NORM	0.556 (80/144)	0.584 (73/125)	0.368 (7/19)
Wavelet-BMARS	AD vs. NORM	0.683 (69/101)	0.632 (24/38)	0.714 (45/63)
	AD vs. MCI	0.630 (80/127)	0.414 (12/29)	0.694 (68/98)
	MCI vs. NORM	0.535 (77/144)	0.574 (70/122)	0.318 (7/22)
Wavelet-GLM	AD vs. NORM	0.614 (62/101)	0.538 (21/39)	0.661 (41/62)
	AD vs. MCI	0.567 (72/127)	0.303 (10/33)	0.659 (62/94)
	MCI vs. NORM	0.424 (61/144)	0.510 (50/98)	0.239 (11/46)
Raw-GLM	AD vs. NORM	0.584 (59/101)	0.500 (6/12)	0.596 (53/89)
	AD vs. MCI	0.551 (70/127)	0.200 (5/25)	0.637 (65/102)
	MCI vs. NORM	0.451 (65/144)	0.526 (61/116)	0.143 (4/28)

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

CHAPTER V

Conclusion & Future Work

In this dissertation, we present three Bayesian joint models to predict survival or clinical outcomes based on high-dimensional brain imaging data. In Chapter I, we first smooth the images by using a multivariate pairwise difference prior which accounts for the spatio-temporal correlation in and between the images. We reduce the data dimension by proposing four novel summary statistics based on the smoothed images. A Bayesian generalized non-linear model (GNLM) with multivariate adaptive regression spline (MARS, Holmes and Denison (2003)) is adopted to predict patients' one year survival status. In Chapter II, we extend the joint model in Chapter I and we propose a Bayesian survival model in stage II by assuming that the health status of patients follows a latent Wiener process. In both chapters, the information of the image data are summarized by four novel summary statistics, which are then included in either the GNLM or survival model as predictors. In Chapter IV, we apply the three dimensional discrete wavelet transformation to decorrelate the image data. Bayesian Lasso is used to denoise the wavelet transformed images. Summary statistics are derived based on the wavelet images and included as covariates into a cumulative probit regression model in stage II with BMARS basis functions.

In Chapters II and III, we extracted information from the three dimensional image

data by proposing four summary statistics. From the results of our model, the four summary statistics capture information about the early changes in ADC and rCBF due to treatment and appear to be strong predictors for treatment efficacy. The four statistics may not capture the most, or even the best, information and information is always lost in data reduction. Therefore, better methods of data dimension reduction may be studied and pursued. Moreover, in our study, due to small sample size, we did not consider more than four statistics. A larger sample size would be necessary for better dimension reduction.

In Chapter III, we extend the FHT model proposed by Lee and Whitmore (2006) and Lee et al. (2004) into the Bayesian framework. In their original paper, they introduced the concept of operational time into the FHT model to distinguish from calendar time. The calendar time measures time in months and years, while the operational time measures the cumulative exposure of a system to aggregate physical effects that cause its deterioration, such as the wearout of a car, which is related more to usage than simple passage of time. And the aggregate effects may relate to multiple causes of death, such as different working exposures (Lee et al. (2004)). In our Bayesian FHT model, for the sake of model simplicity and interpretation, we assume a linear monotonic relationship between operational time and calendar time. More research may be conducted to see if results can be improved by modeling the operational time in our Bayesian FHT model.

Moreover, in the gliomas study, patients were actually imaged at more than two time points. Therefore, to study the changes of the intensities in qMRI data, a longitudinal model is required. Lee and Whitmore (2006) discussed the possibilities of extending the FHT model to a longitudinal analysis. However, not much has been done in this area and it remains an open topic for future research.

In Chapter IV, we only considered T1-weighted MRI. There are also T2-weighted MRI and proton density MRI data available in the ADNI project, both of which are shown to be better MRI methods for viewing white matter lesions. However, both the T2-weighted MRI and proton density MRI are raw images, which require several image correction steps before they can be analyzed. Therefore, more research can be done by investigating the quantitative differences among AD, MCI and normal controls in the T2-weighted MRI and proton density MRI.

Our proposed model in Chapter IV can also be applied to classify Multiple Sclerosis (MS) patients from normal subjects. Multiple sclerosis (MS) is a disease that affects the brain and spinal cord resulting in loss of muscle control, vision, balance, and sensation (such as numbness). MS is predominantly a disease of white matter in the central nervous system. White matter tracts are affected, including those of the cerebral hemispheres, infratentorium, and spinal cord. MS lesions, known as plaques, may form in CNS white matter in any location; thus, clinical presentations may be diverse. There is an important role for MRI in the diagnosis of MS. MRI is sensitive to the white matter lesions that characterize multiple sclerosis, since MRI can show multiple lesions as high intensity signals in the images (Paty et al. (1988)). The white matter lesions in MRI are also highly heterogeneous and differ in sizes, shapes, locations. Therefore, our proposed method could also be applicable to MS.

BIBLIOGRAPHY

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

Anderson, P. (2009). Changes to White-Matter Lesions Linked to Early Cognitive Impairment. *Neurology* **73,** 120–125.

Ashburner, J. and Friston, K. J. (2000). Voxel-based morphometrythe methods. *Neuroimage* **11,** 805–821.

Basser, P. J. and Jones, D. K. (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.

Braccoa, L., Piccinia, C., Morettib, M., Mascalchib, M., Sforzaa, A., Nacmiasa, B., Cellinia, E., Bagnolia, S., and Sorbi, S. (2005). Alzheimers disease: role of size and location of white matter changes in determining cognitive deficits. *Dementia Geriatric Cognitive Disorder* **20,** 358–366.

Bronge, L. and Wahlund, L. O. (2007). White matter changes in dementia: does radiology matter? *British Journal of Radiology* **80,** S115–S120.

Bullmore, E., Fadili, J., Maxim, V., Şendur, L., Whitcher, B., Suckling, J., Brammer, M., and Breakspear, M. (2004). Wavelets and functional magnetic resonance imaging of the human brain. *NeuroImage* **23,** 234–249.

Chenevert, T. L., Stegman, L. D., Taylor, J. M. G., Robertson, P. L., Greenberg, H. S., Rehemtulla, A., and Ross, B. D. (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., Prasad, U., Shankar, S., Hamstra, D. A., Shanaiah, M., Chenevert, T. L., Ross, B. D., and Rehemtulla, A. (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.

Chow, S. C. and Liu, J. P. (2004). *Design and analysis of clinical trials: concepts and methodologies*. Wiley-IEEE.

Cox, I. J., Roy, S., and Hingorani, S. L. (1995). Dynamic histogram warping of image pairs for constant image brightness. *International Conference on Image Processing* **2,** 2366.

Daubechies, I. (1992). *Ten lectures on wavelets*. Society for Industrial Mathematics.

De Leeuw, F. E., Barkhof, F., and Scheltens, P. (2005). Progression of cerebral white matter lesions in Alzheimers disease: a new window for therapy? *Journal of Neurology, Neurosurgery & Psychiatry* **76,** 1286–1288.

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

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

Doksum, K. A. and Hóyland, A. (1992). Models for variable-stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution. *Technometrics* pages 74–82.

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

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

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

Gelman, A., Meng, X. L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* **6,** 733–759.

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Journal of the Royal Statistical Society. Series C, Applied Statistics* **44,** 455–472.

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

Hamstra, D. A., Chenevert, T. L., Moffat, B. A., Johnson, T. D., Meyer, C. R., Mukherji, S. K., Quint, D. J., Gebarski, S. S., Fan, X., Tsien, C. I., Lawrence, T. S., Junck, L., Rehemtulla, A., and Ross, B. D. (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., Galbn, C. J., Meyer, C. R., Johnson, T. D., Sundgren, P. C., Tsien, C., Lawrence, T. S., Junck, L., Ross, D. J., Rehemtulla, A., Ross, B. D., and Chenevert, T. L. (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.

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

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

Jack Jr, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., Whitwell, J. L., Ward, C., et al. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of magnetic resonance imaging: JMRI* **27,** 685–691.

Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., MacFall, J., et al. (2006). Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage* **30,** 436–443.

Kavcic, V., Duffy, C. J., Ekholm, S., and Zhong, J. (2004). P2-204 White matter correlates of cognitive impairments in early Alzheimer's disease. *Neurobiology of Aging* page S288.

Kavcic, V., Ni, H., Zhu, T., Zhong, J., and Duffy, C. J. (2008). White matter integrity linked to functional impairments in aging and early Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **4,** 381–389.

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

Law, M., Young, R., Babb, J., Pollack, E., and Johnson, G. (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., Parney, I. F., Huang, W., Anderson, F., Morris, A. M., Asher, A., Lillehei, K. O., Bernstein, M., Brem, H., Sloan, A., 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.

Lee, M. L. T. and Whitmore, G. A. (2006). Threshold regression for survival analysis: Modeling event times by a stochastic process reaching a boundary. *Statistical Science* **21,** 501–513.

Lee, M. L. T., Whitmore, G. A., Laden, F., Hart, J. E., and Garshick, E. (2004). Assessing lung cancer risk in railroad workers using a first hitting time regression model. *Environmetrics* **15,** 501–512.

Leow, A. D., Yanovsky, I., Parikshak, N., Hua, X., Lee, S., Toga, A. W., Jack, C. R., Bernstein, M. A., Britson, P. J., Gunter, J. L., et al. (2009). Alzheimer's Disease Neuroimaging Initiative: A one-year follow up study using tensor-based morphometry correlating degenerative rates, biomarkers and cognition. *Neuroimage* **45,** 645–655.

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

Lu, C. J. and Meeker, W. Q. (1993). Using degradation measures to estimate a time-to-failure distribution. *Technometrics* **35,** 161–174.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transaction on Pattern Analysis Machine Intelligence.* **11,** 674–693.

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

Mason, R. (2006). Non-invasive assessment of kidney oxygenation: a role for BOLD MRI. *Kidney International* **70,** 10–11.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., et al. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society of London. Series B* **356,** 1293–1322.

Medhi, J. (1994). *Stochastic processes.* New Age International.

Meyer, C. R., Boes, J. L., Kim, B., Bland, P. H., Zasadny, K. R., Kison, P. V., Koral, K., Frey, K. A., and Wahl, R. L. (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., Chenevert, T. L., Lawrence, T. S., Meyer, C. R., Johnson, T. D., Dong, Q., Tsien, C., Mukherji, S., Quint, D. J., Gebarski, S. S., Robertson, P. L., Junck, L. R., Rehemtulla, A., and Ross, B. D. (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., Chenevert, T. L., Meyer, C. R., McKeever, P. E., Hall, D. E., Hoff, B. A., Johnson, T. D., Rehemtulla, A., and Ross, B. D. (2006). The functional diffusion map: an imaging biomarker for the early prediction of cancer treatment outcome. *Neoplasia* **8,** 259–267.

Park, C. and Padgett, W. J. (2005). Accelerated degradation models for failure based on geometric Brownian motion and gamma processes. *Lifetime Data Analysis* **11,** 511–527.

Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association* **103,** 681–686.

Paty, D. W., Oger, J. J. F., Kastrukoff, L. F., Hashimoto, S. A., Hooge, J. P., Eisen, A. A., Eisen, K. A., Purves, S. J., Low, M. D., Brandejs, V., et al. (1988). MRI in the diagnosis of MS: a prospective study with comparison of clinical evaluation, evoked potentials, oligoclonal banding, and CT. *Neurology* **38,** 180–185.

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

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

Rao, H., Wang, J., Giannetta, J., Korczykowski, M., Shera, D., Avants, B. B., Gee, J., Detre, J. A., and Hurt, H. (2007). Altered resting cerebral blood flow in adolescents with in utero cocaine exposure revealed by perfusion functional MRI. *Pediatrics* **120,** e1245–e1254.

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

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

Singpurwalla, N. D. (1995). Survival in dynamic environments. *Statistical Science* **10,** 86–103.

Sled, J. G., Zijdenbos, A. P., and Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging* **17,** 87–97.

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

Vidakovic, B. (1999). *Statistical modeling by wavelets.* Wiley New York.

Whitmore, G. A. (1975). The inverse Gaussian distribution as a model of hospital stay. *Health Services Research* **10,** 297–302.

Young, R., Babb, J., Law, M., Pollack, E., and Johnson, G. (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., Pipe, J. G., Bonnett, J., and Evelhoch, J. L. (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.