ADVANCED BAYESIAN MODELS FOR HIGH-DIMENSIONAL BIOMEDICAL DATA

Eunjee Lee

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2016

Approved by:

Joseph G. Ibrahim

Hongtu Zhu

Jan Hannig

Yufeng Liu

Sonia Davis

ABSTRACT

EUNJEE LEE: ADVANCED BAYESIAN MODELS FOR HIGH-DIMENSIONAL BIOMEDICAL DATA
(Under the direction of Joseph G. Ibrahim and Hongtu Zhu)

Alzheimer's Disease (AD) is a neurodegenerative and firmly incurable disease, and the total number of AD patients is predicted to be 13.8 million by 2050 [68]. Our motivation comes from needs to unravel a missing link between AD and biomedical information for a better understanding of AD. With the advent of data acquisition techniques, we could obtain more biomedical data with a massive and complex structure. Classical statistical models, however, often fail to address the unique structures, which hinders rigorous analysis.

A fundamental question this dissertation is asking is how to use the data in a better way. Bayesian methods for high-dimensional data have been successfully employed by using novel priors, MCMC algorithms, and hierarchical modeling. This dissertation proposes novel Bayesian approaches to address statistical challenges arising in biomedical data including brain imaging and genetic data. The first and second projects aim to quantify effects of hippocampal morphology and genetic variants on the time to conversion to AD within mild cognitive impairment (MCI) patients. We propose Bayesian survival models with functional/high-dimensional covariates. The third project discusses a Bayesian matrix decomposition method applicable to brain functional connectivity. It facilitates estimation of clinical covariates, the examination of whether functional connectivity is different among normal, MCI, and AD subjects.

Dedicated to my family

ACKNOWLEDGEMENTS

My Ph.D. journey could not be completed without my advisors Professor Hongtu Zhu and Professor Joseph Ibrahim. Professor Ibrahim introduced Bayesian statistics and survival analysis with his motivation, immense knowledge, and patience. He guides me from very small details to main idea on my dissertation. Professor Zhu was always there whenever I was in trouble. I was inspired by his expertise in (say everything but in particular) statistical methods for neuroimaging and genetic data. He cared and encouraged what I did, and it helped me to finish my dissertation.

I would like to thank my committee members, Professor Jan Hannig, Professor Yufeng Liu, and Professor Sonia Davis for their insightful comments and encouragement. Professor Hannig's insights helped me to think further on Bayesian methods. Professor Liu's comments were very helpful to understand high-dimensional and functional data. Professor Davis's (written and verbal) comments improved my dissertation a lot. Her expertise in Alzheimer's disease helped me to find a better direction in the data analysis.

Thank you Professor Ja-Yong Koo for encouraging me. You helped me to start and finish my Master's and Doctoral degrees and I learnt how to do research from you.

Thank you my husband Kibok for encouraging me all the time. I could not overcome hard times without your support. Thanks for being a very good listener when I practiced presentations.

Thank you my parents for being my parents. My life is full of love because you have always cheered me up. I hope that this achievement will be a small gift to you.

I met good friends during the five years in Chapel Hill.

- My first officemate, Soyoung, you helped me a lot when I settled down in Chapel Hill. I enjoyed my first year with you.

- Thanks to Sunyoung for everything. You were the first person that I could contact when I was in trouble (well, except for my husband).
- Thanks to Hana, for being my friend. I laughed a lot when I talked with you. Also, your advise for my interviews was excellent!
- Jung In, you were friendly to me. I was so happy when we went to the Duke garden this Spring. Hope that we could meet in another lovely place.
- Thanks to Sieun, for having all the nice times with me: I remember a small concert in Carrboro. I will miss you so much.
- Hyosun, I enjoyed talking with you. I will miss our happy times in Spanky's. We came to UNC in 2011 and made it finally! I am so proud of you.
- Mihye, I could smoothly start my research in the lab because you helped me a lot. We have had nice times in many cities and we will do more.
- Eunhwa, you were such a good friend. Our weekly routine to study is unforgettable. Thanks for being on my side all the time.
- Thanks to Hanseul for cheering me up. You have supported me and now I will do it for you. I miss our memory in NC.
- Soyoung, thanks for always cheering me up when I had hard times.
- Hyo Young and Duyeol, thanks for having good times with me in Chapel Hill. For my defense, Duyeol helped me to set-up the snacks and it was just perfect.
- Thanks to Wan Suk for helping me a lot. I was so thankful when you brought a cable before my defense. I got very relieved because I had what I needed!
- Hojin, thanks for your help when we were in the lab.
- Eric and Dongqing were awesome officemates! Dongqing, thanks for all the kindness.
- Qing, Washington, and Guan, my life in Hanes was fun because of you!

PREFACE

Chapter 3 previously appeared as an article in the Annals of Applied Statistics. The original citation is as follows: Lee, E., Zhu, H., Kong, D., Wang, Y., Giovanello, K. S., & Ibrahim, J. G. (2015). BFLCRM: A Bayesian functional linear Cox regression model for predicting time to conversion to Alzheimers disease. *The annals of applied statistics*, **9(4)**, 2153-2178.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Alzheimer's Disease (AD) is a neurodegenerative and firmly incurable disease, and the total number of AD patients is predicted to be 13.8 million by 2050 [68]. Our motivation comes from needs to unravel a missing link between AD and biomedical information for a better understanding of AD. With the advent of data acquisition techniques, we could obtain more biomedical data with a massive and complex structure. Classical statistical models, however, often fail to address the unique structures, which hinders rigorous analysis

A fundamental question this dissertation is asking is how to use the data in a better way. Bayesian methods for high-dimensional data have been successfully employed by using novel priors, MCMC algorithms, and hierarchical modeling. This dissertation proposes novel Bayesian approaches to address statistical challenges arising in biomedical data including brain imaging and genetic data.

The first part of this dissertation aims to develop Bayesian survival models to quantify the effects of hippocampal morphology and genetic variants on the time to conversion to AD within mild cognitive impairment (MCI) patients. MCI is a clinical syndrome characterized by the onset and evolution of cognitive impairments. It is often considered as a transitional stage to Alzheimers disease, since around 50% of MCI patients progress to AD during the first 5 years [125]. It has been revealed that some therapeutic interventions and a few disease-modifying drugs are more effective at MCI or early stage of AD than at the severe stage of AD [120, 32, 16]. Due to this pathological feature of AD, it is an ongoing quest for prediction of the time to conversion from MCI to AD and detection of biomarkers affecting

progression to AD. While hippocampus and genetic variants are important risk factors, many classical survival models are not theoretically and computationally suitable because they take the forms of functional and high dimensional data. Motivated by the limitations of classical survival models, this dissertation proposes survival models with functional or high dimensional covariates by taking a Bayesian approach. In order to make use of functional data as covariates to predict the time to event, a Bayesian functional linear Cox regression model is proposed in Chapter 3. In Chapter 4, we propose Bayesian bi-level variable selection to enable variable selection in both group and element levels for high-dimensional predictors. In the following sections, we discuss brief motivation and description of the two modern survival models.

The second part discusses a Bayesian matrix decomposition method applicable to brain functional connectivity. Functional connectivity is the connectivity between different brain areas sharing information and functions [153]. It can be estimated by pairwise temporal correlation between two spatially remote BOLD signals in rsfMRI. It is an important biomarker in psychiatric disorders because its abnormality has been observed in subjects with brain disorders including AD, schizophrenia, and ADHD [58]. One specific question will be"is there any relationship between altered functional connectivity and neurological disorders?". In Chapter 5, we propose a Bayesian hierarchical group spectral clustering model to analyze brain connectivity. It facilitates estimation of clinical covariates, which enables to examine if functional connectivity had group difference among normal, MCI, and AD subjects.

1.1 Bayesian Functional Cox Regression Model

Hippocampus is one of the first brain regions that suffer from damage in Alzheimers disease. Among various measures to quantify changes in hippocampus size, volume, or shape, the hippocampal radial distance [151] has gained attention recently, since it reflects different atrophy level in different subfields in the hippocampus rather than giving a single

global measure. It is the distance from medial core of the hippocampus to its surface, which takes the form of 2-dimensional curve quantifying the thickness of hippocampus relative to its centerline. By comparing the radial distance of different individuals, the relative atrophy against each other can be measured. Apostolova et al. [6] have shown that MCI patients who progress to AD have smaller hippocampi around CA1 region than MCI patients who does not develop AD. Also AD patients tend to have a higher annual atrophy rate on few regions of CA1 than MCI patients [51].

Since the hippocampal radial distance is a 2-dimensional curve, which can be considered as functional data, classical regression models should be adapted to incorporate functional covariates. In order to describe the nature of functional data as covariates in linear models, functional linear regression has been proposed and discussed in many literatures including Cardot et al. [19], Ramsay and Silverman [131], Müller and Stadtmüller [114]. Functional covariates can be estimated by employing functional principal component analysis (fPCA) on a continuous covariance function of the functional covariates [19, 114, 18]. In the generalized linear model framework, Müller and Stadtmüller [114] apply a Karhunen-Loeve expansion on the functional predictor, which enables dimension reduction to a finite number of components of the expansion. More reviews regarding functional regression models will be discussed in section 2.2.

In chapter 3, we propose a Bayesian functional linear Cox regression model (BFLCRM) to incorporate functional covariates in a Bayesian survival model by applying fPCA. The BFLCRM can be an important extension of various statistical models including parametric, semiparametric and nonparametric models for handling survival response data and scalar covariates. Second, BFLCRM can be an important extension of various functional linear models for handling discrete or continuous response data and functional covariates. Third, BFLCRM can be regarded as an important extension of high-dimensional survival models.

1.2 Bayesian Bi-level Variable Selection with Censored Outcomes

Since AD is highly heritable, genome-wide association studies (GWASs) have been conducted for the purpose of identifying genetic variants contributing to AD. Typically, it is of interest to elucidate the association between the traits and single-nucleotide polymorphisms (SNPs), a DNA sequence variation occurring commonly within a population. A genome-wide association study (GWAS) focuses on identifying important SNPs to relate to clinical outcomes in this context. Since the SNP data is ultra high-dimensional (half a million or more), the simple (and popular) GWASs conduct a number of marginal tests: examination of the effect of each SNP one by one. It makes GWAS to be theoretically and computationally feasible in the classical regression setup, $p < n$, where p is the number of covariates and n is the number of observations. But the simple GWASs face two main challenges: dealing with multiple testing issue and accounting for the dependency structure among SNPs.

One solution is to consider all the SNPs simultaneously in one model by applying penalization methods or Bayesian variable selection methods. To detect effective SNPs to time to conversion from MCI to AD, variable selection methods can be applied in survival models in order to incorporate a number of SNPs in one survival model. In fact, several variable selection schemes have been applied in the Cox proportional hazards model and accelerated failure time (AFT) models in the high dimensional setting. Lasso penalization [63, 171, 56], principled sure independence screening [173], a boosting procedure [95, 10], threshold gradient descent method [62] have been applied on the Cox regression. In the AFT models, Cai et al. [17], Datta et al. [34], Huang et al. [70], Wang et al. [159] apply LASSO regularization approaches, a nonlinear kernel based AFT model has been described in [100]. In the Bayesian framework, iterative Bayesian model averaging (BMA) [4], Bayesian ensemble method [12] have been proposed. Konrath et al. [88] take a Bayesian approach using a penalized Gaussian mixture error distribution, a structured additive predictor with Bayesian P-splines. In order to achieve sparsity, they also use Bayesian versions of ridge and LASSO, and a spike

and slab priors. Bayesian variable selection approach [138] and Bayesian LASSO [92] are considered in the AFT model. But these methods do not take into account any grouping information among predictors.

In order to tackle the limitations that the existing variable selection methods have in the context of GWAS, we propose a Bayesian bi-level variable selection (BBVS) method in the accelerated failure time model in Chapter 4. Our main goal is (1) identification of SNPs associated with time to conversion from MCI to AD (2) by considering all the SNPs simultaneously and (3) incorporating the grouping information of the SNP data. Our method has two hierarchical levels of variable selection: the first one is group-wise and the second level is element-wise variable selection. In the first step, we identify important groups of variables and update the censored event time from its predictive posterior distribution by data augmentation [150, 138]. The dimension of the whole SNP data can be significantly reduced by eliminating irrelevant groups to time to event. Since this step also provides a posterior sample of censored time to event, the posterior mean will be used as imputed censored event time in the second level of variable selection. It converts the AFT model in the second level into a usual log-normal regression model. In the second level, we only include variables in the selected groups as covariates in the AFT model with the imputed event time as an outcome. To conduct element-wise variable selection, shrinkage priors are employed on regression parameters. In particular, we extend Dirichlet-Laplace shrinkage priors proposed by [9] to incorporate the grouping information.

1.3 Bayesian Hierarchical Group Spectral Clustering

The Human brain has been analyzed from a network perspective with the advent of neuroimaging acquisition techniques and network theory. Functional magnetic resonance imaging (fMRI) is a non-invasive neuroimaging procedure to assess brain neuronal activity that can be measured by changes in blood oxygen level-dependent (BOLD) signal [101].

Functional connectivity is the connectivity between different brain areas sharing information and functions [153]. Unraveling a missing link between neurological disorder and brain network is an ongoing quest in various fields including statistics, epidemiology, and neuroscience. One specific question will be "is there any relationship between altered functional connectivity and neurological disorders?". There are two main standard methods to tackle the question: univariate approaches to see if each correlation (node) has group difference using Fisher's z-transformation, and graph theoretical approaches based on some summary statistic for a network structure (i.e., girth, diameter, modularity, small-worldness). The first method easily faces high-dimensional multiple testing problem. If there is a small group difference in each correlation, the univariate approach is likely to miss the signal, which results in low power. Also, it discards a spatial structure among close or related brain regions (voxels). The second methods often fail to detect local differences among subject groups, because the connectivity structures are too simplified by the summary statistics. Thus, alternative connectivity analysis methods are critically needed.

In this study, we aim to address the following four issues in connectivity studies: (1) high-dimensionality of connectivity matrices, (2) detection of local differences in brain, but at the same time, (3) need to utilize an underlying relational structure among brain areas. Regarding (3), while the correlation between two marginal regions has weak group difference, two linear combinations of some brain regions can have strongly different connection across groups. Here we call the linear combinations of brain regions underlying factors or eigenmaps. We can identify more significant group differences in organization maps among underlying factors instead among individual brain regions. The last thing to be considered is (4) heterogeneity of functional connectivity. FMRI itself is noisy in terms of spatial resolution and head motion, which leads estimated functional connectivity to be varied across subjects. Also, functional connectivity has individual and group variation, as Mueller et al. [113] suggests that functional connectivity is individually heterogeneous across the cortex due to the degree of evolutionary cortical expansion. The heterogeneity should be accordingly

taken into account to have meaningful signals.

We propose a Bayesian hierarchical group spectral clustering model to take a global approach to analyze brain connectivity. We decompose a correlation matrix (possibly it can be any symmetric matrix) with underlying common factors across subjects and the subject-specific coefficient matrix ($\mathbf{\Lambda}_i$). The $\mathbf{\Lambda}_i$ matrix preserves an individual network structure in the low-dimensional space spanned by the common factors. One more intriguing part is a hierarchical structure within a prior of $\mathbf{\Lambda}_i$ in order to estimate effects of clinical/demographic covariates. We take a Bayesian approach to estimate the parameters involved in the decomposition, and the regression model with clinical covariates. Our method automatically estimates the effects of covariates on the $\mathbf{\Lambda}_i$ matrices within the same MCMC iterations.

CHAPTER 2

LITERATURE REVIEWS

2.1 Survival Analysis

In survival analysis, the outcome of interest is the survival times of subjects. The survival times are denoted by T, a continuous nonnegative random variable, with cumulative distribution function $F(t)$ on the interval $[0, \infty)$. The probability that an individual survives till specified time point t is given by the survival function,

$$S(t) = 1 - F(t) = P(T > t).$$

The survival function $S(t)$ is a monotone decreasing function with $S(0) = 1$ and $\lim_{t\to\infty} S(t) = 0$. The hazard function $h(t)$ is defined as the ratio of the probability density function $f(t)$ to the survival function $S(t)$, which is instantaneous rate of occurrence of the event. It can be written as

$$h(t) = \lim_{t\to\infty} \frac{P(t \le T < t + dt | T \ge t)}{dt} = \frac{f(t)}{S(t)} \tag{2.1.1}$$

Note that $f(t) = -\frac{d}{dt}S(t)$. Thus, (2.1.1) implies that

$$h(t) = -\frac{d}{dt} \log S(t). \tag{2.1.2}$$

Integrating both sides of (2.1.2) and exponentiating give the following:

$$S(t) = \exp\left(-\int_0^t h(u)du\right) = \exp(-H(t)),$$

where $H(t)$ is is the cumulative hazard. From (2.1.2), it is derived that

$$f(t) = h(t) \exp\left(-\int_0^t h(u)du\right).$$

2.2 Proportional Hazards Model

Survival models consist of two ingredients: the baseline hazard function, denoted $h_0(t)$ and the effect covariates. The baseline hazard function $h_0(t)$ quantifies how the risk of event per time unit varies over time when levels of covariates are fixed at baseline. The proportional hazards condition (Cox 1972) is that effect of a unit increase in a covariate is multiplicatively related to the hazard rate. Under the proportional hazards assumption, the effect of covariates can be estimated without taking into account the hazard function. The Proportional hazards model takes this approach on survival data. The hazard function of a subject at time t can be specified as

$$h(t|\boldsymbol{x}) = h_0(t)\exp\{\boldsymbol{x}'\boldsymbol{\beta}\}, \tag{2.2.1}$$

where $\boldsymbol{x} = (x_1, \cdots, x_p)$ is a p covariates vector and $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$ is a vector of p regression coefficients, and $h_0(t)$ is called the baseline hazard function. The model (2.2.1) implies that the ratio of hazards between two subjects is time-invariant and actually depends on the difference between their linear predictors, $\eta = \boldsymbol{x}'\boldsymbol{\beta}$. In biomedical research regarding survival data, some individuals are still alive or normal at the end of the study so death or occurrence of a disease has not happened. Therefore we have right censored data, that is, some of the survival times exceed a certain value. The certain time point is called a censoring time. The survival times t_is are assumed to be independent and identically distributed with density $f(t)$ and survival function $S(t)$. Then survival time t_i for the i-th subject will be observed only when $t_i \leq c_i$, where c_i is a fixed censoring time. The data D with a size n random sample consists of $(y_i, \nu_i, x_i)_{i=1}^n$, where $y_i = \min(t_i, c_i)$, $\boldsymbol{x}_i$ is the $p \times 1$ vector of covariates and ν_i is an (event) indicator variable taking 1 if $t_i \leq c_i$ or 0 if $t_i > c_i$. Then the likelihood

function for $(\boldsymbol{\beta}, h_0(\cdot))$ is given by

$$\begin{aligned} L(\boldsymbol{\beta}, h_0(t)|D) &\propto \prod_{i=1}^{n} [h_0(y_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})]^{\nu_i} \left(S_0(y_i)^{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}\right) \\ &= \prod_{i=1}^{n} [h_0(y_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})]^{\nu_i} \exp\left\{\sum_{i=1}^{n} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) H_0(y_i)\right\}, \end{aligned}$$

where $S_0(t)$ is the baseline survivor function such that $S_0(t) = \exp\left(-\int_0^t h_0(u)du\right) = \exp(H_0(t))$.

One of the most popular semiparametric survival models is a piecewise constant hazard model. On the time axis, we set a finite partition as $0 < s_1 < \cdots < s_J$ with $s_J > y_i$ for all $i = 1, \cdots, n$, where $s_0 = 0$. In the j-th interval, we assume a constant baseline hazard $h_0(y) = \lambda_i$ for $s_{j1} < y \leq s_j$. Then the likelihood function can be written as

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\lambda}|D) &= \prod_{i=1}^{n}\prod_{j=1}^{J} [\lambda_j \exp(\boldsymbol{x}_i'\boldsymbol{\beta})]^{u_{ij}\nu_i} \\ &\quad \times \prod_{i=1}^{n}\prod_{j=1}^{J} \exp\left[-u_{ij}\left\{\lambda_j(y_i - s_{j-1})\sum_{g=1}^{j-1}\lambda_g(s_g - s_{g-1})\right\}\exp(\boldsymbol{x}_i{}'\boldsymbol{\beta})\right], \end{aligned}$$

where $u_{ij} = 1$ if the i-th subject is right censored in the j-th interval and 0 otherwise.

2.3 Accelerated Failure Time Model

An accelerated failure time model (AFT model) is an alternative to the proportional hazards model. Since the proportional hazards model assumes a multiplicative effect on the hazard function, it is hard to interpret the estimates of regression parameters. The AFT model assumes a multiplicative effect on the time to event. For the i-th subject, its probability model is given by

$$Y_i = \exp(-\boldsymbol{x}_i'\boldsymbol{\beta})\nu_i, i = 1, \cdots, n,$$

which becomes the linear model in log scale

$$\log Y_i = -x_i'\boldsymbol{\beta} + \epsilon_i, i = 1, \cdots, n,$$

where $Y_1, Y_2, \cdots, Y_n$ are failure times, $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{ip})'$ is a vector of known explanatory variables for the i-th individual, $\boldsymbol{\beta}$ is a vector of p unknown regression coefficients, and $\epsilon_i = \log \nu_i$ is the error term. Usually the error term is assumed to follow a parametric distribution, such as the Normal distribution.

2.4 Functional Linear Regression

2.4.1 Functional Data and Smoothing

With the development of new technologies, data objects of interest are measured in a continuous time fashion. Functional data includes daily temperature data [131], Spectrometric curves data, log-periodograms data [47], and many other objects in various fields. Since those data objects are curves, functional methods should be taken to analyze their global behavior. In practice, functional data is measured at (fine) discrete grid points. For a random function $X(s), s \in \mathcal{S}$, we measure a discretized observation $\{W(s_l)\}_{l=1}^L$ at grid points $\{s_l\}_{l=1}^L \in \mathcal{S}$ with measurement errors $\epsilon(s_l)$. Assume that we have a random sample $\{X_i(\cdot)\}_{i=1}^n$, where the random functions $X_i(\cdot)$s are independently and identically distributed with the same distribution as $X_i(\cdot)$. For the i-th observation,

$$W_i(s_l) = X_i(s_l) + \epsilon_i(s_l) = \mu(s_l) + \bar{X}_i(s_l) + \epsilon_i(s_l),$$

where $X_i(s)$ characterizes individual functional variations from $\mu(s)$. The $\epsilon_i(s)$s are measurement errors with mean zero and variance $\sigma_\epsilon^2(s)$ at each s and independent of each other for $s \neq s'$. Moreover, $\mu(s)$ can be consistently estimated by $\bar{\mu}(s) = \sum_{i=1}^n W_i(s)/n$. To proceed with functional data analysis, one needs to obtain a continuous smooth functional data by smoothing the discretized observations. We apply a cubic smoothing spline, which finds a minimizer $f(x)$ of the following penalized residual sum of squares

$$\sum_{i=1}^{n} \sum_{l=1}^{L} (w_{il} - f_i(s_l))'(w_{il} - f_i(s_l)) + \lambda \int \{f''(t)\}^2 dt.$$

The second penalty term is the integrated squared second derivative and penalizes curvature in the function. Then a unique minimizer is a natural cubic spline with knots at the unique values of s_l. In practice, we use the smoothed observations in the functional data analysis.

2.4.2 Model Setup

It is of interest to explain variation in a (scalar) response variable by relating with a functional covariate. We consider a linear regression setting, where the response variable Y is a real-valued continuous random variable and the covariate $X(s)$, $s \in \mathcal{S}$, is a square integrable random function observed on $\mathcal{S}$ such that $E(X(s)) = \mu(s)$. The "functional" linear regression model can be expressed as

$$Y = \beta_0 + \int_{\mathcal{S}} \beta(s)X(s)dt + \epsilon,$$

where ϵ_is are iid with $E(\epsilon|X(s), t \in \mathcal{S}) = 0$, $\text{Var}(\epsilon|X(s), t \in \mathcal{S}) = \sigma^2$. A constant intercept β_0 is given by $E(Y) - \int_{\mathcal{S}} \beta(s)\mu(s)ds$. Without loss of generality, we assume that the variables $X(\cdot)$ and Y are centered so that we have a simplified regression model

$$Y = \int_{\mathcal{S}} \beta(s)X(s)dt + \epsilon.$$

The coefficient function $\beta(\cdot)$ quantifies how much $X(s)$ has an effect on Y across $\mathcal{S}$, which is a major parameter in the functional linear model.

2.4.3 Estimation of Functional Coefficients

There are two main methods to estimate the functional coefficient $\beta(\cdot)$. One is using basis expansions on the functional covariates and functional coefficients. Ramsay and Silverman [131] introduced regularization methods: using the truncated number of basis functions to expand the functional coefficient and using roughness penalties. Cardot et al. [20] expressed the functional coefficient as a B-splines expansion and used a penalty proportional to the

squared norm of a derivative given the order of the functional coefficient in the estimation step. The other widely used method is employing functional principal component analysis (fPCA) on a continuous covariance function of $X(s)$ [19, 114, 18]. In the generalized linear model framework, Müller and Stadtmüller [114] apply a Karhunen-Loeve expansion on the functional predictor, which enables dimension reduction to a finite number of components of the expansion.

We review the fPCA approach in the functional linear regression setting. It is assumed that $X_i(s)$ and $\epsilon_i(s)$ are independent of each other and the covariance function of $\{X_i(s) : s \in \mathcal{S}\}$, denoted by $K(s, s') = \mathbb{E}\{X(s)X(s')\}$, is continuous on $\mathcal{S} \times \mathcal{S}$. According to Mercer's theorem, $K(s, s')$ also admits a spectral decomposition

$$K(s, s') = \sum_{j=1}^{\infty} \psi_j \phi_j(s) \phi_j(s'),$$

where $(\psi_j, \phi_j(s))$'s are the eigenvalue-eigenfunction pairs of $K(s, s')$ such that $\{\psi_j : j \geq 1\}$ are the eigenvalues in decreasing order with $\sum_{j=1}^{\infty} \psi_j^2 < \infty$. Thus, $X_i(s)$ admits the Karhunen-Loeve expansion as

$$X_i(s) = \sum_{j=1}^{\infty} \xi_{ij} \phi_j(s),$$

where the ξ_{ij} are referred to functional principal component (fPC) scores and the $\xi_{ij} = \int_{\mathcal{S}} X(s)\phi_j(s)ds$ are uncorrelated random variables with mean zero and variance $\psi_j = \mathbb{E}(\xi_{ij}^2)$. To estimate ξ_{ij} based on the observed covariate functions $W_i(s)$, we first employ the cubic smoothing spline [65] to estimate the underlying signal $X_i(s)$. We then use the sample mean and covariance functions of the estimated $X_i(s)$ to estimate $\mu(s)$ and $K(s, s')$. Subsequently, we estimate $\psi_j(s)$ and ξ_{ij} for all $i, j \leq n$. Since the eigenfunctions $\psi_j(\cdot)$ form a complete orthonormal system on the space of square-integrable functions on $\mathcal{S}$, the covariate function $\gamma(s)$ can be expanded as

$$\beta(s) = \sum_{j=1}^{\infty} \phi_j(s)\beta_j \quad \text{with} \quad \sum_{j=1}^{\infty} \beta_j^2 < \infty. \tag{2.4.1}$$

Therefore, we have

$$\int_{\mathcal{S}} X_i(s)\beta(s)ds = \sum_{i=1}^{\infty} \xi_{ij}\beta_j \tag{2.4.2}$$

and with truncated linear predictors

$$Y_i = \sum_{i=1}^{\infty} \xi_{ij}\beta_j + \epsilon_i \approx \sum_{i=1}^{q_n} \xi_{ij}\beta_j + \epsilon_i, \tag{2.4.3}$$

where q_n is a sufficiently large integer that may depend on n. One can control smoothness of the functional coefficient by truncating the infinite summation in the equation (2.4.2) at the first q_n summation. Since the approximation error of the truncated model is bounded by the variance of the truncated part and is controlled by a sequence of eigenvalues, ψ_j, the truncation step would allow nice approximation for the infinite summation. Practically, it is common to choose q_n such that the percentage of variance explained by the first q_n fPCA components is 70%, 85%, or 95%. Alternatively, we may formulate it as a model selection procedure and choose it by using some model selection criterion, such as the deviance information criterion (DIC) [146, 76].

CHAPTER 3

BAYESIAN FUNCTIONAL LINEAR COX REGRESSION MODELS

3.1 Introduction

Alzheimer's Disease (AD) is a firmly incurable and progressive disease [36]. In the pathology of AD, mild cognitive impairment (MCI) is a clinical syndrome characterized by insidious onset and gradual progression, and commonly arising as a result of underlying neurodegenerative pathology [53]. Since MCI is considered as a risk state for AD, a major research focus in recent years has been to delineate a set of biomarkers that provide evidence of such a neurodegenerative pathology in living individuals, with the goal of specifying the likelihood that the pathophysiological process is due to Alzheimer's disease (MCI due to AD; MCI-AD) and will lead to dementia within a few years [1]. Accordingly, increasing attention has been devoted to investigate the utility of various imaging, genetic, clinical, behavioral, and fluid data to predict the conversion from MCI to AD.

Several studies have utilized a small subset of biosignatures and then assessed the relative importance of different modalities in predicting the diagnostic change from MCI to AD [31, 128, 170, 133, 45]. For example, in [31], the authors simultaneously examined multiple features from different modalities, including structural magnetic resonance imaging (MRI) morphometry, cerebrospinal fluid biomarkers, and neuropsychological measures to assess an optimal set of predictors of conversion from MCI to AD. They observed that structural changes within the medial temporal lobe (MTL), particularly the hippocampus, as well as performance on cognitive tests that rely on MTL integrity (i.e., episodic memory), were good predictors of MCI to AD conversion.

Recently, most researchers have turned to the analysis of longitudinal data to assess the dynamic changes of various biomarkers associated with the MCI-to-AD transition across time. To begin, a prominent neural correlate of MCI-AD is volume loss within the MTL, especially within the hippocampus and entorhinal cortex [42], with increasing atrophy in these structures from normal aging to MCI to AD [123]. Longitudinal studies of individuals with MCI-AD have also highlighted the importance of assessing MTL changes in tracking the progression of MCI to AD. For example, several studies have documented diminished baseline hippocampal and entorhinal volumes that are associated with an increased likelihood of progressing to clinical dementia [85, 59]. Additionally, several modalities of disease indicators have been studied to assess progression to AD, including neuroimaging biomarkers [154, 166, 133], biomedical markers [139], and neuropsychological assessments [124]. Finally, a number of structural MRI studies, covering region of interest (ROI), volume of interest, voxel-based morphometry, and shape analysis have reported that the degree of atrophy in several brain regions, such as the hippocampus and entorhinal cortex, is not only sensitive to disease progression, but also predicts MCI conversion [28, 108, 40].

Despite the importance of these investigations, a central question remains. Namely, how do we accurately predict the time to conversion in individuals who harbor AD pathology, as well as determine the optimal early markers of conversion? In [149], 148 MCI subjects were used to identify the most predictive neuropsychological measures. In [97], 139 MCI subjects in ADNI-1 were used to evaluate the predictive power of brain volume, ventricular volume, hippocampus volume, APOE status, cerebrospinal fluid (CSF) biomarkers, and behavioral scores. Their results show a moderately accurate prediction with the value of an area under the curve of 0.757 at 36 months, whereas they found that baseline volumetric MRI and behavioral scores were selectively predictive. Finally, in [33], 381 MCI subjects from ADNI 1 were examined to evaluate several biomarkers for predicting MCI to AD conversion including spatial patterns of brain atrophy, ADAS-Cog, APOE genotype, and cerebrospinal fluid (CSF) biomarkers. Their findings suggest that a combination of spatial patterns of brain

atrophy and ADAS-Cog offers good predictive power of conversion from MCI to AD, whereas APOE genotype did not significantly improve prediction. To the best of our knowledge, no prior study has examined the role of functional covariates including hippocampus surface morphology in predicting time to conversion from MCI to AD with/without adjusting for low-dimensional behavioral and clinical measures.

To assess the predictability of hippocampus surface morphology in survival models, we develop a Bayesian functional linear Cox regression model (BFLCRM) with both functional and scalar covariates. The BFLCRM integrates a Cox proportional hazard regression and functional linear model into a single framework. First, BFLCRM can be an important extension of various statistical models including parametric, semiparametric and nonparametric models for handling survival response data and scalar covariates. See overviews of various survival models in [84, 75, 48] and the references therein. Recent advances in computation and prior elicitation have made Bayesian analyses of these survival models with scalar covariates feasible. For instance, nonparametric prior processes including the gamma process prior, the Beta process model, the correlated gamma process, and the Dirichlet process prior have been developed as the prior distribution of the baseline cumulative hazard function [144, 75]. Second, BFLCRM can be an important extension of various functional linear models for handling discrete or continuous response data and functional covariates. The existing literature focuses on the development of frequentist methods for functional linear models. Some examples include [165, 47, 131, 81, 132] and the references therein. Third, BFLCRM can be regarded as an important extension of high-dimensional survival models. However, most high-dimensional survival models focus on the identification of a small set of covariates and their overall effect on time-to-event outcomes [82, 96, 11]. These approaches can be sub-optimal for high-dimensional imaging data, since the effect of imaging data on clinical data and other imaging data is often *non-sparse*, which makes it notoriously difficult for many existing regularization methods [44, 152].

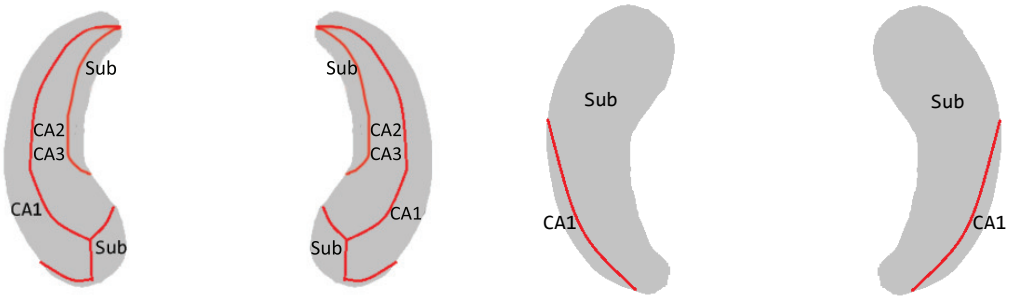
In Section 2, we will introduce BFLCRM and its associated Bayesian estimation procedure. In Section 3, we will introduce the NIH Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset and illustrate the use of BFLCRM in the prediction of time to conversion from MCI to AD by using both functional and scalar covariates. In Section 4, we conduct simulation studies to examine the finite sample performance of BFLCRM. In Section 5, we interpret the findings obtained from the analysis of ADNI dataset.

3.2 Bayesian Functional Linear Cox Regression Models

3.2.1 Setup

Consider imaging, genetic, and clinical data from $n = 346$ independent MCI patients in ADNI-1. For the i-th MCI patient, we observe a possibly right censored time to conversion to AD, denoted by y_i. Specifically, $y_i = T_i \wedge C_i$ is the minimum of the censoring time C_i and the transition time T_i and $\nu_i = \mathbf{1}(y_i = T_i)$, where $\mathbf{1}(\cdot)$ is an indicator function. Moreover, we observe a $p \times 1$ vector of scalar covariates, denoted by $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})^T$, and a functional covariate, denoted by $Z_i(\cdot)$, on a compact set $\mathcal{S}$. In this paper, we focus on the noninformative censoring setting such that T_i and C_i are conditionally independent given all covariates of interest. The scalar covariates of interest include age at baseline, length of education, gender, handedness, marital status, retirement, and the well-known Apolipoprotein E (APOE) SNPs. The APOE has three major forms, $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$, where $\epsilon 3$ is the most common form. The functional covariate of interest is the hippocampus surface morphology. Figure 3.1 on page 19 shows the example hippocampus surface morphology data in ADNI-1 data.

Our problems of interest are to establish the likelihood of conversion to Alzheimer's disease (AD) in 346 MCI patients enrolled in the ADNI-1 and to select the optimal early markers of conversion from both the scalar covariates and the functional covariate. With the



(a) Hippocampal subfields

(b) Top (c) Bottom (d) Color bar

Figure 3.1: ADNI data: panel (a) is hippocampal subfields mapped onto a representative hippocampal surface [5], and panels (b) and (c), respectively, show the top and bottom views of the first subject's hippocampal surface data where the corresponding radial distances are color-coded by the colorbar in panel (d).

sole presence of $\mathbf{x}_i$, it is common to consider Cox's proportional hazards model [29], which assumes that the conditional hazard function of y_i given $\mathbf{x}_i$ is given by

$$h(y|\mathbf{x}_i) = h_0(y)\exp(\mathbf{x}_i^T\boldsymbol{\beta}) = h_0(y)\exp\left(\sum_{k=1}^{p} x_{ik}\beta_k\right), \tag{3.2.1}$$

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^T$ is a $p \times 1$ vector of regression coefficients and $h_0(\cdot)$ is an unknown baseline hazard function. However, the Cox proportional hazards model (3.2.1) does not incorporate the effect of the functional covariate $Z_i(\cdot)$ on the time to conversion.

3.2.2 Model Formulation

We propose a Bayesian functional linear Cox regression model with three main ingredients for handling both functional and scalar covariates as a natural extension of (3.2.1). Based on this formulation, we take a Bayesian approach to estimate the parameters of interest.

In the first component of BFLCRM, it is assumed that the hazard function of y_i given $(\mathbf{x}_i, Z_i(\cdot))$ is given by

$$h(y|\mathbf{x}_i, Z_i(\cdot)) = h_0(y) \exp\left(\sum_{k=1}^{p} x_{ik}\beta_k + \int_{\mathcal{S}} \gamma(s)(Z_i(s) - \mu(s))ds\right), \tag{3.2.2}$$

where $\mu(s)$ is the mean function of $Z_i(s)$ and $\gamma(\cdot)$ is an unknown coefficient function, a square integrable function on $\mathcal{S}$.

The second component of BFLCRM is the functional principal component analysis (fPCA) model of the $Z_i(\cdot)$'s. It is assumed that the $Z_i(s)$'s are square integrable random functions and $W_i(s)$ is measured at a set of grid points in $\mathcal{S}$ with measurement errors such that

$$W_i(s) = Z_i(s) + \epsilon_i(s) = \mu(s) + \bar{Z}_i(s) + \epsilon_i(s), \tag{3.2.3}$$

where $\bar{Z}_i(s)$ characterizes individual functional variations from $\mu(s)$. The $\epsilon_i(s)$'s are measurement errors with mean zero and variance $\sigma^2_\epsilon(s)$ at each s and independent of each other for $s \neq s'$. Moreover, $\mu(s)$ can be consistently estimated by $\hat{\mu}(s) = \sum_{i=1}^{n} W_i(s)/n$.

It is assumed that $Z_i(s)$ and $\epsilon_i(s)$ are independent of each other and the covariance function of $\{\bar{Z}_i(s) : s \in S\}$, denoted by $K(s, s') = E\{\bar{Z}(s)\bar{Z}(s')\}$, is continuous on $\mathcal{S} \times \mathcal{S}$. According to Mercer's theorem, $K(s, s')$ also admits a spectral decomposition $K(s, s') = \sum_{j=1}^{\infty} \psi_j \phi_j(s)\phi_j(s')$, where $(\psi_j, \phi_j(s))$'s are the eigenvalue-eigenfunction pairs of $K(s, s')$ such that $\{\psi_j : j \geq 1\}$ are the eigenvalues in decreasing order with $\sum_{j=1}^{\infty} \psi_j^2 < \infty$. Thus, $\bar{Z}_i(s)$ admits the Karhunen-Loeve expansion as $\bar{Z}_i(s) = \sum_{j=1}^{\infty} \xi_{ij}\phi_j(s)$, where the ξ_{ij} are referred to functional principal component (fPC) scores and the $\xi_{ij} = \int_{\mathcal{S}} \bar{Z}(s)\phi_j(s)ds$ are uncorrelated

random variables with mean zero and variance $\psi_j = E(\xi_{ij}^2)$. To estimate ξ_{ij} based on the observed covariate functions $W_i(s)$, we first employ the cubic smoothing spline [65] to estimate the underlying signal $Z_i(s)$. We then use the sample mean and covariance functions of the estimated $Z_i(s)$ to estimate $\mu(s)$ and $K(s, s')$. Subsequently, we estimate $\phi_j(s)$ and ξ_{ij} for all $i, j \leq n$.

The third component of the BFLCRM is an approximation of $\int_{\mathcal{S}} \gamma(s)\bar{Z}(s)ds$. Since the eigenfunctions $\psi_j(\cdot)$ form a complete orthonormal system on the space of square-integrable functions on $\mathcal{S}$, the covariate function $\gamma(s)$ can be expanded as

$$\gamma(s) = \sum_{j=1}^{\infty} \phi_j(s)\gamma_j \quad \text{with} \quad \sum_{j=1}^{\infty} \gamma_j^2 < \infty. \tag{3.2.4}$$

Therefore, we have

$$\int_{\mathcal{S}} \bar{Z}_i(s)\gamma(s)ds = \sum_{j=1}^{\infty} \xi_{ij}\gamma_j \tag{3.2.5}$$

and approximate $h(y|\mathbf{x}_i, Z_i(\cdot))$ as

$$h_0(y) \exp\left(\sum_{k=1}^{p} x_{ik}\beta_k + \sum_{j=1}^{\infty} \xi_{ij}\gamma_j\right) \approx h_0(y) \exp\left(\sum_{k=1}^{p} x_{ik}\beta_k + \sum_{j=1}^{q_n} \xi_{ij}\gamma_j\right), \tag{3.2.6}$$

where q_n is a sufficiently large integer that may depend on n. As shown in the literature, such an approximation is accurate under some conditions on the decay rate of the γ_j's. Practically, it is common to choose q_n such that the percentage of variance explained by the first q_n fPCA components is 70%, 85%, or 95%. Alternatively, we may formulate it as a model selection procedure and choose it by using some model selection criterion, such as the deviance information criterion (DIC) [146, 76].

3.2.3 Priors

To carry out a Bayesian analysis of model (3.2.6), we specify joint prior distributions for all unknown parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}, h_0)$, where $h_0(\cdot)$ is the baseline hazard function. We first set $p(\boldsymbol{\beta}, \boldsymbol{\gamma}, h_0) = p(\boldsymbol{\beta}, \boldsymbol{\gamma})p(h_0)$ and assume $(\boldsymbol{\beta}, \boldsymbol{\gamma}) \sim N(\boldsymbol{\mu}_0, \Sigma_0)$, where $N(\boldsymbol{\mu}_0, \Sigma_0)$ is the

multivariate Normal distribution with a $(p+q_n)\times 1$ mean vector $\boldsymbol{\mu}_0$ and a $(p+q_n)\times(p+q_n)$ covariance matrix Σ_0.

We may specify different prior distributions for $h_0(y)$. The most convenient and popular distribution for $h_0(y)$ is the piecewise constant hazard model. Specifically, we first construct a finite partition of the time axis, $0 < s_1 < s_2 < \ldots < s_J$, with $s_J > y_i$ for all i, which leads to J intervals $(0, s_1]$, ..., $(s_{J-1}, s_J]$. In the j-th interval, we set $h_0(y) = \lambda_j$ for $y \in I_j = (s_{j-1}, s_j]$. A common prior of the baseline hazard $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_J)^T$ is the independent gamma prior $\lambda_j \sim \mathcal{G}(\alpha_{0j}, \alpha_{1j})$ for $j = 1, \ldots, J$, where α_{0j} and α_{1j} are prior hyperparameters. Another approach is to build prior correlation among the λ_j's using a prior $\boldsymbol{\psi} \sim N(\boldsymbol{\psi}_0, \Sigma_J)$, where $\psi_j = \log(\lambda_j)$ for $j = 1, \ldots, J$ and $\psi_j = (\psi_1, \cdots, \psi_J)$. For notational simplicity, we focus on the piecewise constant hazard model with the independent gamma prior from here on.

We consider the strategy of choosing the hyperparameters Σ_0, α_{0j} and α_{1j} as follows. We can tune the eigenvalues of Σ_0 in order to control the prior information for the regression coefficients. If the smallest eigenvalue $\lambda_{min}(\Sigma_0)$ converges to ∞, then $N(\mu_0, \Sigma_0)$ tends to be an improper prior. In contrast, if the largest eigenvalue $\lambda_{max}(\Sigma_0)$ is very small, then $N(\mu_0, \Sigma_0)$ tends to be a strongly informative prior. In order to use a noninformative prior for the λ_j's, the shape and scale parameters of the gamma distributions are set to be $\alpha_{0j} = 0.2$ and $\alpha_{1j} = 0.4$ for all $j = 1, \cdots, J$ [143]. Also setting either $(\alpha_{0j}, \alpha_{1j}) = (0.5, 1)$ or $(\alpha_{0j}, \alpha_{1j}) = (0.2, 1)$ would make the gamma distribution flat as well.

3.2.4 Posterior Computation

The log-posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ (unnormalized) is given by

$$\sum_{i=1}^{n}\sum_{j=1}^{J}[u_{ij}\nu_i(\log\lambda_j + \boldsymbol{z}_i^T\boldsymbol{\theta}) - u_{ij}\{\lambda_j(y_i - s_{j-1}) + \sum_{g=1}^{j-1}\lambda_g(s_g - s_{g-1})\}\exp(\boldsymbol{z}_i^T\boldsymbol{\theta})$$
$$-\{\log|\boldsymbol{\Sigma}_0| + (\boldsymbol{\theta}-\boldsymbol{\mu}_0)^T\Sigma_0^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu}_0)\}/2$$
$$+\sum_{j=1}^{J}\{(\alpha_{0j}-1)\log\lambda_j - \lambda_j\alpha_{1j} + \alpha_{0j}\log(\alpha_{1j}) - \log\Gamma(\alpha_{0j})\}, \qquad (3.2.7)$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$, $\boldsymbol{z}_i = (\boldsymbol{x}_i^T, \xi_{i1}, \cdots, \xi_{iq_n})^T$, and $s_0 = 0$. Moreover, $u_{ij} = 1$ if the i-th subject fails or is right censored in the j-th interval and 0 otherwise. We propose a Gibbs sampler for posterior computation after truncating the sum of the infinite series to have $q_n < \infty$ terms. The Gibbs sampler is computationally efficient and mixes rapidly. We first specify the hyperparameters $\boldsymbol{\mu}_0, \Sigma_0, \alpha_{0j}$ and α_{1j} for all j at appropriate values to represent prior opinion. Starting from the initiation step, the Gibbs sampler for model (3.2.6) with the truncated term q_n proceeds as follows:

1. Update $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ according to their full conditional distribution in (3.2.7). Specifically, we employ the random walk Metropolis-Hastings (M-H) [66, 106] algorithm and choose a multivariate Normal proposal density yielding an average acceptance rate of 23.4% [54]. The mean of the proposal density is the posterior sample $(\boldsymbol{\beta}^{t-1}, \boldsymbol{\gamma}^{t-1})$ from the previous iteration. The covariance matrix is the inverse of the Fisher information matrix of the posterior distribution evaluated at $(\boldsymbol{\beta}^{t-1}, \boldsymbol{\gamma}^{t-1})$.

2. Update λ_j from its full conditional distribution

$$p(\lambda_j|\boldsymbol{\lambda}_0^{(-j)}, -) \sim \text{Gamma}(\alpha_{0j} + \sum_{i=1}^{n} u_{ij}\nu_i, \tilde{\alpha}_{1j}),$$

where $\boldsymbol{\lambda}_0^{(-j)}$ is the vector $\boldsymbol{\lambda}_0$ without the j-th element and $\tilde{\alpha}_{1j}$ is given by

$$\tilde{\alpha}_{1j} = \begin{cases} \alpha_{1j} + \sum_{i=1}^{n} \{u_{ij}(y_i - s_{j-1}) + (s_j - s_{j-1}) \sum_{k=j+1}^{J} u_{ik}\} \exp(\boldsymbol{z}_i^T \boldsymbol{\theta}), \\ \qquad \text{if } j \leq J-1; \\ \alpha_{1J} + \sum_{i=1}^{n} \{u_{iJ}(y_i - s_{J-1}) \exp(\boldsymbol{z}_i^T \boldsymbol{\theta})\}, \\ \qquad \text{if } j = J. \end{cases}$$

3.3 Alzheimer's Disease Neuroimaging Initiative Data Analysis

3.3.1 Alzheimer's Disease Neuroimaging Initiative

The development of the BFLCRM is motivated by the analysis of imaging, genetic, and clinical data collected by ADNI. "Data used in the preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (www.loni.usc.edu/ADNI). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (aMCI) and early Alzheimers disease (AD). In specific, a subject will be classified as a MCI subject if the following conditions are met: mini-mental state exam (MMSE) scores are between 24-30, a subject has a memory complaint, objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II, a Clinical Dementia Rating (CDR) of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia. A subject will be classified as an AD subject if the following conditions are met: MMSE scores between 20-26 (in- clusive), CDR of 0.5 or

1.0, and meets NINCDS/ADRDA criteria for probable AD. The NINCDS/ADRDA criteria, the most used ones for the diagnosis of AD, were developed by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimers Association.

Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, M.D., VA Medical Center and University of California - San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with aMCI to be followed for 3 years, and 200 people with early AD to be followed for 2 years. For up-to-date information see www.adni-info.org."

3.3.2 Data Description

The aim of this ADNI data analysis is to examine the predictability of clinical, genetic, and imaging data for the time to conversion to AD in MCI patients. Conversion is established if the diagnosis has been changed from MCI at the baseline to AD at some visit. We focused on 346 MCI patients at baseline of the ADNI-1 database. Among the 346 MCI patients, 151 of them are converters and 195 are non-converters at 48 months.

For each MCI patient, we included his/her clinical, genetic, and imaging variables at baseline. The clinical characteristics include Gender (0=Male; 1=Female), Handedness (0=Right; 1=Left), Marital Status (1=Married; 2=Widowed; 3=Divorced; 4=Never married), Education length (in years), Retirement (1=Yes; 0=No), Age, and Alzheimer's Disease

Assessment Scale-Cognition (ADAS-Cog) score. Marriage status is coded using 3 dummy variables: "Widowed", "Divorced", "Never married". The ADAS-Cog test has been widely used to assess the severity of dysfunction in adults [134], where the larger the ADAS-Cog score, the greater the dysfunction.

The genetic variables include the APOE genetic covariates, since it is well known that mutations in APOE raise the risk of progression from amnestic MCI to AD [126]. The Apolipoprotein E (APOE) SNPs, rs429358 and rs7412 were, separately, genotyped in ADNI-1. These two SNPs together define a 3 allele haplotype, namely the $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$ variants and the presence of each of these variants was available in the ADNI database for all the individuals. Among these variants, APOE-$\epsilon 3$ is known to be most common allele, while APOE-$\epsilon 4$ has been turned out to be a risk factor for early onset of AD [119]. In this model, we consider the presence of APOE-$\epsilon 4$ as a covariate to incorporate its effect on the time to conversion from MCI to AD. In addition, we selected 7 regions of interest (ROIs) that may significantly influence MCI progression among the 93 ROI volume data [14, 80, 46]. These 7 ROIs are bilateral hippocampal formation, bilateral amygdala, posterior limb of internal capsule, bilateral thalamus. In total, we have 17 scalar covariates. The imaging data include the hippocampal radial distances of 30,000 surface points on the left and right hippocampal surfaces. The hippocampal radial distance is a distance from its medial core to the hippocampal surface and measures hippocampal thickness.

In the demographic information, 220 participants are male, and 126 are female; 316 are right-handed, and 30 are left-handed. For Marital Status, 283 were married, 40 were widowed, 19 were divorced, and 4 were never married at baseline. Among these individuals, 276 were retired and 70 were not. On average, the subjects had 15.7 years of education with standard deviation 3.0 years, the minimum 6 years, and the maximum 20 years. The average age of subjects was 75.0 years with standard deviation of 7.3 years. The youngest person was 55 years old, while the oldest person was 90 years old. For the genetics information

on the first allele of APOE, 25 subjects had genotype 2, 277 subjects had genotype 3, and 44 subjects had genotype 4. For the second allele, 156 subjects had genotype 3, while 190 subjects had genotype 4. The average ADAS-cog score was 11.5, with standard deviation of 4.4. The lowest score was 2 and the highest score was 27.67.

3.3.3 Hippocampus Image Preprocessing

We used a hippocampal image analysis package to calculate hippocampal surface data as follows [140, 161, 141, 142, 111, 26, 103]. Given the 3D MRI scans, we used FIRST [121] to segment hippocampal substructures and then applied the marching cube method [102] to automatically reconstructing hippocampal surfaces. Then, an automatic algorithm, called topology optimization, was used to introduce two cuts on a hippocampal surface in order to convert it into a genus zero surface with two open boundaries. The two cuts, whose locations were at the front and back of the hippocampal surface, represent its anterior junction with the amygdala and its posterior limit as it turns into the white matter of the fornix. We then computed holomorphic 1-form basis functions [160]. It allows us to induce conformal grids of the hippocampal surfaces which were consistent across subjects. The conformal representation of the surface was computed with this conformal grid [140]. We also computed the "feature image" of a surface by combining the conformal factor and mean curvature and linearly scaling the dynamic range into [0, 255]. Finally, the feature image of each surface in the dataset was registered to a common template by using an inverse consistent fluid registration algorithm [142]. It establishes high-order correspondences between 3D surfaces. Finally, we computed various surface statistics based on the registered surface, such as multivariate tensor-based morphometry (mTBM) statistics [160].

3.3.4 Data Analysis

We focused on 346 MCI patients in the ADNI-1 data in order to examine the predictability of clinical, genetic, and imaging covariates for the time to conversion to AD from MCI. The patients consist of 151 converters and 195 non-converters. We fit BFLCRM with time to conversion to AD as the response y_i, the clinical, genetic, and ROI volume data as scalar covariates in $\boldsymbol{x}_i$, and the hippocampus surface data based on radial distance as functional covariates in $Z_i(\cdot)$ for the i-th subject. In all posterior computations, we centered the scalar covariates $\boldsymbol{x}_i$ using their mean. We chose the first 14 eigenfunctions of hippocampal surface data, which explain about 73.48% of the variance in the hippocampus surface data. The first 14 largest eigenfunctions projected on the hippocampal surfaces were presented in Figure 3.8 in the supplementary section. For the piecewise constant hazards model of $h_0(\cdot)$, we chose $J = 70$ intervals so that each interval contains at least one failure or censored observation. The full BFLCRM model (3.2.6) contains 19 scalar covariates and the first 14 fPC scores.

Due to the lack of prior information, all hyperparameters were chosen to reflect nearly noninformative priors. For regression coefficients, the hyperparameters of the multivariate Normal priors were set as $\boldsymbol{\mu}_0 = (0, \cdots, 0)$ and $\Sigma_0 = \text{diag}(5, \cdots, 5)$. For the λ_j's, the shape and scale parameters of the Gamma priors were set to be $\alpha_{0j} = 0.2$ and $\alpha_{1j} = 0.4$ for $j = 1, \cdots, 70$ [143]. We ran the Gibbs sampler for 25,000 iterations after 5,000 burn-in iterations. Based on the 20,000 MCMC samples, we calculated various posterior quantities of $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$. For the full BFLCRM model, in section 3.3.5, we also conducted sensitivity analyses in order to investigate the influence of different choices of hyperparameters in the prior distributions. From the results shown in Tables 3.11-3.14 in the supplementary section, we found that the proposed priors were robust to various choices of the hyperparameters in the prior distributions. The computational time (in C/C++ using an 8-cores 2.80 GHz Intel processors) was 350 seconds for running the Gibbs sampler for the full BFLCRM model with 25,000 iterations.

Table 3.1: ADNI data analysis results for the full BFLCRM model: the posterior quantities of 19 regression coefficients β_ks, that correspond to $\boldsymbol{x}_i$ =(Gender, Handedness, Widowed, Divorced, Never married, Length of Education, Retirement, Age, APOE-ϵ4 carrier, ADAS-cog Score, posterior limb of internal capsule, Right hippocampal formation, Left hippocampal formation, Left thalamus, Left amygdala, Right amygdala, and Right thalamus). Mean denotes 'posterior mean', SD denotes 'posterior standard deviation', and lower and upper, respectively, represent the 'lower and upper limits' of a 95% highest posterior density interval.

	β_1 Female	β_2 Left	β_3 Widowed	β_4 Divorced	β_5 Never-married	β_6 Years of education	β_7 Retired	β_8 Age in years
Mean	0.4344	0.2255	0.3119	0.2729	0.7203	-0.0874	0.3455	-0.0519
SD	0.2513	0.3647	0.3827	0.4663	0.7867	0.0367	0.2482	0.0178
lower	-0.0495	-0.5248	-0.4632	-0.6789	-0.9383	-0.1691	-0.0919	-0.0873
upper	0.9478	0.8628	1.0138	1.1195	2.2009	-0.0244	0.8608	-0.0188

	β_9 ApoE4	β_{10} ADAS	β_{11} PLIC	β_{12} RHF	β_{13} LHF	β_{14} LT	β_{15} LA	β_{16} RA	β_{17} RT
Mean	0.5550	0.1568	0.0008	0.0006	-0.0011	-0.0004	0.0018	-0.0012	0.0003
SD	0.2341	0.0265	0.0005	0.0004	0.0004	0.0004	0.0009	0.0005	0.0004
lower	0.1258	0.1030	-0.0002	-0.0002	-0.0019	-0.0012	0.0000	-0.0023	-0.0004
upper	1.0258	0.2075	0.0019	0.0014	-0.0004	0.0003	0.0036	-0.0002	0.0010

Table 3.1 shows the posterior means of the regression coefficients $\boldsymbol{\beta}$ and their standard deviations, as well as the lower and upper limits of the 95% highest posterior density (HPD) intervals based on the full BFLCRM model. Six scalar covariates including "Length of Education", "Age", "APOE-ϵ4 carrier", "ADAS-cog score", "Left Hippocampal formation", and "Right amygdala" have 95% HPD intervals that do not contain 0. This implies that we can expect a worse prognosis of AD for MCI patients with lower ROI volume in the left hippocampal formation and the right amygdala. This finding supports the finding that atrophy of the hippocampal formation is a significant diagnostic marker [86, 78]. It also confirms the previous finding that the amygdala volume tends to be reduced in the early stage of AD [109, 127]. Moreover, the 95% HPD intervals of the 1st, 7th, and 14th fPCs do not contain 0. This may indicate that the hippocampal radial distance is an important functional covariate for predicting the time to conversion to AD in MCI subjects.

(a) Top (b) Bottom (c) Color bar

Figure 3.2: ADNI data analysis results for the full BFLCRM model: panels (a) and (b), respectively, show the top and bottom views of the estimated coefficient function associated with the hippocampal surface data color-coded by the colorbar in panel (c).

We estimated the coefficient function $\gamma(\cdot)$ by using $\hat{\gamma}(s) = \sum_{j=1}^{14} \hat{\phi}_j(s)\hat{\gamma}_j$, where $\hat{\gamma}_j$ is the posterior mean of γ_j for each j. Figure 3.2 shows the estimated coefficient function $\hat{\gamma}(\cdot)$ associated with the hippocampal surface data. When hippocampal atrophy in a red region is greater, a risk of progressing from MCI to AD is expected to be increased. A blue region suggests that the thicker the area is on the hippocampus, the shorter the time to conversion to AD is. Inspecting Figure 3.2 reveals that the subfields of CA1 and subiculum on the hippocampi have positive effects on the hazard function, indicating that the thinner these areas on the hippocampus are, the shorter the time is to conversion to AD.

Figure 3.3 on page 31 shows the estimated survival functions of the APOE-ϵ4 carriers and non-carriers, when the values of the continuous covariates are set at their mean values and the categorical variables are set at their reference levels. The dotted lines show the 95% HPD intervals of survival functions. The APOE-ϵ4 carriers are expected to convert from MCI to AD faster than non-carriers. These results are consistent with several prior studies suggesting that the presence of the APOE-ϵ4 allele increases the risk of developing AD [148, 135, 27].

We compared the full BFLCRM model with three reduced models in terms of their

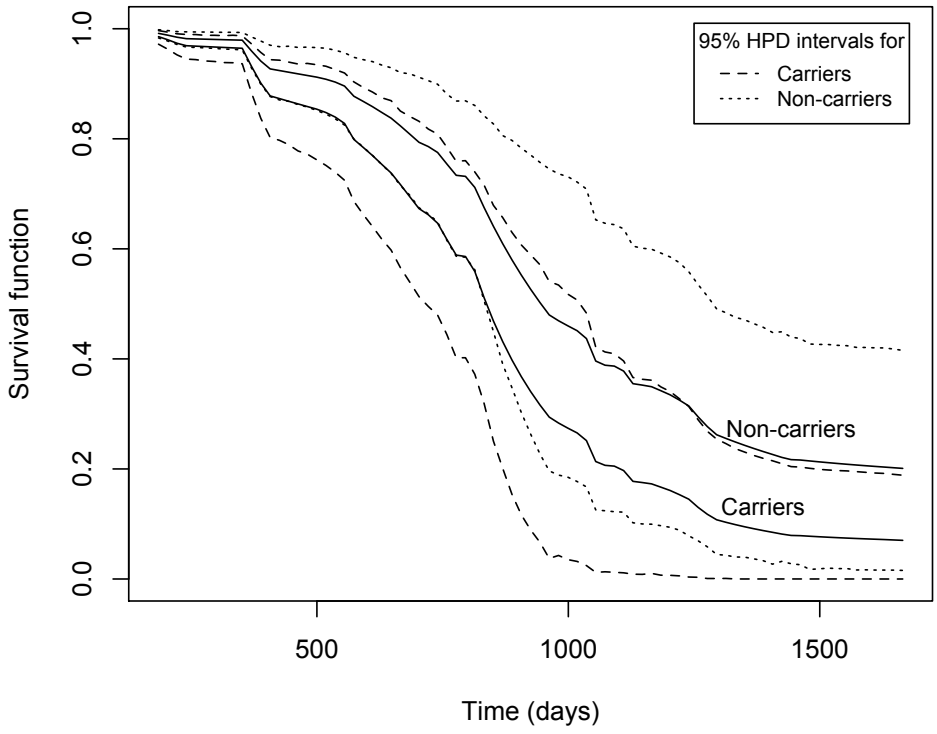


Figure 3.3: ADNI data analysis results: the estimated survival curves of APOE-ϵ4 carriers and non-carriers under the full BFLCRM model. Other continuous or categorical covariates are fixed at the mean values or reference levels. The dotted lines show the 95% HPD intervals of the estimated survival functions.

predictive performance. For Model 1, we excluded the ROI volume covariates from the full BFLCRM model. For Model 2, we only included all the scalar covariates. For Model 3, we only included the clinical covariates, APOE-ϵ4 status, and the ADAS-cog score. For all three reduced models, we set $J = 70$ intervals so that each interval contains at least one failure or censored observation. For the regression coefficients, the hyperparameters of the multivariate Normal priors were set as $\boldsymbol{\mu}_0 = (0, \cdots, 0)$ and $\Sigma_0 = \text{diag}(5, \cdots, 5)$. We set $\alpha_{0j} = 0.2$ and $\alpha_{1j} = 0.4$ for $j = 1, \cdots, 70$. We ran the Gibbs sampler for 25,000 iterations after 5,000 burn-in iterations. We also calculated the DIC and integrated AUC (iAUC) [74] for all four models, where AUC denotes the area under the Receiver Operating Characteristic (ROC) curve. The DIC can be estimated within the MCMC iterations. More details can be

Table 3.2: ADNI data analysis results under the four models: DICs and the empirical means of iAUC values and their corresponding standard errors in the parenthesis calculated from the Monte Carlo cross-validation (MCCV).

	the full BFLCRM model	Model 1	Model 2	Model 3
DIC	427.19	413.08	417.04	438.22
iAUC	0.840 (0.003)	0.836 (0.003)	0.809 (0.003)	0.751 (0.004)

Table 3.3: ADNI data analysis results for Model 1: the posterior quantities of 12 regression coefficients β_ks, that correspond to $\boldsymbol{x}_i$ =(Gender, Handedness, Widowed, Divorced, Never married, Length of Education, Retirement, Age, APOE-ϵ4 carrier, and ADAS-cog Score). Mean denotes 'posterior mean', SD denotes 'posterior standard deviation', and lower and upper, respectively, represent the 'lower and upper limits' of a 95% highest posterior density interval.

	β_1 Female	β_2 Left	β_3 Widowed	β_4 Divorced	β_5 Never-married	β_6 Years of edu	β_7 Retired	β_8 Age	β_9 ApoE4	β_{10} ADAS
Mean	0.3943	0.2826	0.2422	0.1409	0.5174	-0.0577	0.4045	-0.0406	0.5143	0.1572
SD	0.2016	0.3317	0.3121	0.4377	0.7682	0.0341	0.2442	0.0156	0.2190	0.0239
lower	-0.0157	-0.3488	-0.4151	-0.7356	-0.9930	-0.1260	-0.0610	-0.0699	0.1069	0.1104
upper	0.7842	0.9094	0.8308	0.9196	1.9468	0.0045	0.8900	-0.0089	0.9634	0.2030

found in [146, 76].

Table 3.2 shows the values of DIC and iAUC for the four models. The full BFLCRM model yields the DIC value of 427.19, which is smaller than that of Models 3, but larger than those of Model 1 and 2. The Model 1 had the smallest DIC value as 413.08. For Model 1, as shown in Table 3.8, "Age", "APOE-ϵ4 carrier", and "ADAS-cog score" have their 95% HPD intervals that do not contain 0. See additional estimation results associated with Models 1-3 in Tables 3.8, 3.9, and 3.10 and Figures 3.6 and 3.7 of the supplementary section. Based on the iAUC values, however, the full model achieves the best predictive performance. Moreover, the full model and Model 1, which include the hippocampal surface data as functional covariates, provide better predictive performance than Models 2 and 3. This may indicate that the hippocampal surface data contributes significantly to the time of conversion from MCI to AD. We estimated the iAUC by using a Monte Carlo cross-validation (MCCV) method. Specifically, the full data set was randomly split into a training set with

200 subjects and a test set with 146 subjects. For each such split, we fitted each model to the training set and then calculated iAUC based on the test set. This random split was repeated 100 times yielding the estimated iAUC values for all models.

3.3.5 Sensitivity Analysis

We examine the effects of varying hyperparameters on the posterior estimation. In our main paper, we first set $p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = p(\boldsymbol{\beta}, \boldsymbol{\gamma})p(\boldsymbol{\lambda})$ and assume $(\boldsymbol{\beta}, \boldsymbol{\gamma}) \sim N(\boldsymbol{\mu}_0, \Sigma_0)$, where $N(\boldsymbol{\mu}_0, \Sigma_0)$ is the multivariate normal distribution with a $(p + q_n) \times 1$ mean vector $\boldsymbol{\mu}_0$ and a $(p+q_n) \times (p+q_n)$ covariance matrix Σ_0. For the piecewise constant baseline hazard function, $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_J)^T$ follows the independent gamma prior such that the λ_j are independent and $\lambda_j \sim \mathcal{G}(\alpha_{0j}, \alpha_{1j})$ for $j = 1, \ldots, J$, where α_{0j} and α_{1j} are prior hyperparameters. We consider different choices of Σ_0 and $(\alpha_{0j}, \alpha_{1j})$ and evaluate their effects on the estimation of $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$.

First, we varied the hyperparameter Σ_0, while fixing $(\alpha_{0j}, \alpha_{1j})$ at (0.2, 0.4) for all $j = 1, \cdots, 70$. The mean vector $\boldsymbol{\mu}_0$ was set as $(0, \cdots, 0)$ in order to reflect an insufficient prior information. For the covariance matrix Σ_0, we considered three scenarios including $\text{diag}(5, \cdots, 5)$, $\text{diag}(25, \cdots, 25)$, and $\text{diag}(100, \cdots, 100)$. Particularly, $\Sigma_0 = \text{diag}(100, \cdots, 100)$ represents an approximately non-informative prior. Tables 3.11 and 3.13 show that all posterior estimates are quite stable as Σ_0 is varied.

Second, we varied the hyperparameters of the gamma priors on the piecewise constant baseline hazard function, when $\boldsymbol{\mu}_0$ and Σ_0 were, respectively, set to be $(0, \cdots, 0)$ and $\text{diag}(100, \cdots, 100)$. We considered three scenarios. To use non-informative priors for λ_j's, we set the shape and scale parameters of the gamma distributions to be $\alpha_{0j} = 0.2$ and $\alpha_{1j} = 0.4$ for all $j = 1, \cdots, 70$ [143]. We then set either $(\alpha_{0j}, \alpha_{1j}) = (0.2, 1)$ or $(\alpha_{0j}, \alpha_{1j}) = (0.5, 1)$ in order to make the gamma distribution flat. For the sake of space, we only report the first

50 estimated λ_j's. Tables 3.12 and 3.14 show that all posterior estimates are quite stable as the hyper parameters $(\alpha_{0j}, \alpha_{1j})$ vary. Therefore, we may conclude that the proposed priors for a large range of hyper-parameters can yield stable posterior estimates.

3.4 Simulation Studies

We conduct Monte Carlo simulations to evaluate the proposed BFLCRM across different censoring rates and sample sizes. Moreover, we will evaluate the predictability of our BFLCRM compared to proportional hazards models without the use of functional covariates.

3.4.1 Setup

We generated all simulated data sets according to model (3.2.1). The $\boldsymbol{x}_i$ is a 4×1 vector and its corresponding elements were independently generated from $N(0, 0.5)$. We set the true value of $\boldsymbol{\beta}$ to be $(0.7, 0.2, -0.5, -1)^T$. The functional covariate $Z_i(s)$ was generated from model (3.2.3), where its underlying function follows the standard Gaussian process with the covariance function $K(s,t) = \exp(-3(s-t)^2)$. The observed functional covariate data $W_i(s)$ consists of noisy observations evaluated at 100 equally spaced grids in the interval [-4,4] with some measurement errors. Specifically, the measurement errors $\epsilon_i(s)$ were independently generated from a $N(0, 0.5)$ across s. The functional coefficient $\gamma(s)$ was generated from the standard Gaussian process with covariance function $K_\gamma(s,t) = \exp(-2(s-t)^2)$. To generate the survival time, we considered two different baseline hazard functions $h_{01}(\cdot)$ and $h_{02}(\cdot)$ as follows.

$$h_{01}(t) = 1 \quad \text{if} \quad t > 0, \tag{3.4.1}$$

$$h_{02}(t) = \begin{cases} \kappa\omega & \text{if } 0 < t \leq 2; \\ \kappa\omega(t-1)^{\omega-1} & \text{if } 2 < t \leq 3; \\ \kappa\omega 2^{\omega-1} & \text{if } t > 3. \end{cases} \tag{3.4.2}$$

The first baseline hazard function $h_{01}(\cdot)$ assumes that it is constant over time. As a more general form of hazard function, we consider a mixed form of baseline hazard functions for the exponential and Weibull distributions. The hazard function $h_{02}(\cdot)$ depends on κ and ω. In this simulation study, we set $\kappa = 1/3$ and $\omega = 2$. Finally, the censoring times were independently generated from a uniform distribution with parameter chosen to achieve a desired censoring rate of 30% or 50%. We considered sample sizes of $n = 200$ and $n = 500$ for each censoring rate and simulated 100 data sets for each case.

3.4.2 Simulation Results

We used the piecewise constant hazard model for $h_0(s)$, in which we set $J = 5$ and subintervals $(s_{j-1}, s_j]$ so that each interval contains at least one failure or censored observation. We set $(\alpha_{0j}, \alpha_{1j}) = (0.2, 0.4)$ for all j, $\boldsymbol{\Sigma}_0 = \text{diag}(5, \cdots, 5)$, and $\boldsymbol{\mu}_0 = (0, \cdots, 0)^T$. We calculated the first 12 fPC scores explaining 95% of the variation of the functional covariates, and then compared the estimation results using the first 12 PC scores in order to investigate the efficacy of using fPCA. For each simulated data set, we ran the Gibbs sampler for 20,000 iterations with 5,000 burn-in iterations.

To examine the estimation and prediction performance of BFLCRM, we calculated mean squared errors (MSEs) and time-dependent integrated area under the curve (iAUC) [74] based on 100 simulated data sets for each scenario. The computational time (in C/C++ using an 8-cores 2.80 GHz Intel processor) was 50.3 seconds for BFLCRM with sample size 200 for one repetition. We let $\hat{\boldsymbol{\beta}}$ denote the posterior mean of $\boldsymbol{\beta}$. The MSE of $\hat{\boldsymbol{\beta}}$ is defined by $\text{MSE}_{\hat{\beta}} = \sum_{j=1}^{p}(\hat{\beta}_j - \beta)^2$, whereas the MSE for $\gamma(\cdot)$ is defined by $\text{MSE}_{\hat{\gamma}} = \int_{-4}^{4}\{\hat{\gamma}(s) - \gamma(s)\}^2 ds$, where $\hat{\gamma}(s)$ denotes the posterior mean of γ at time s. A smaller MSE implies better estimation accuracy, and a large value of iAUC implies a better predictive model. To evaluate the predictive value of the functional covariate to the hazard function, we calculated iAUC for two nested models including a reduced BFLCRM model with solely scalar covariates in $\boldsymbol{x}_i$

Table 3.4: Simulation results corresponding to $h_{01}(\cdot)$ under different censoring rates and sample sizes: the deviance information criterion (DIC), the mean squared errors (MSE) of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ and the estimated integrated area under the curve (iAUCs) and their standard errors in parentheses calculated from the 100 simulated data sets. The Gibbs sampler was run for 20,000 iterations with 5,000 burn-in iterations for each simulated data set.

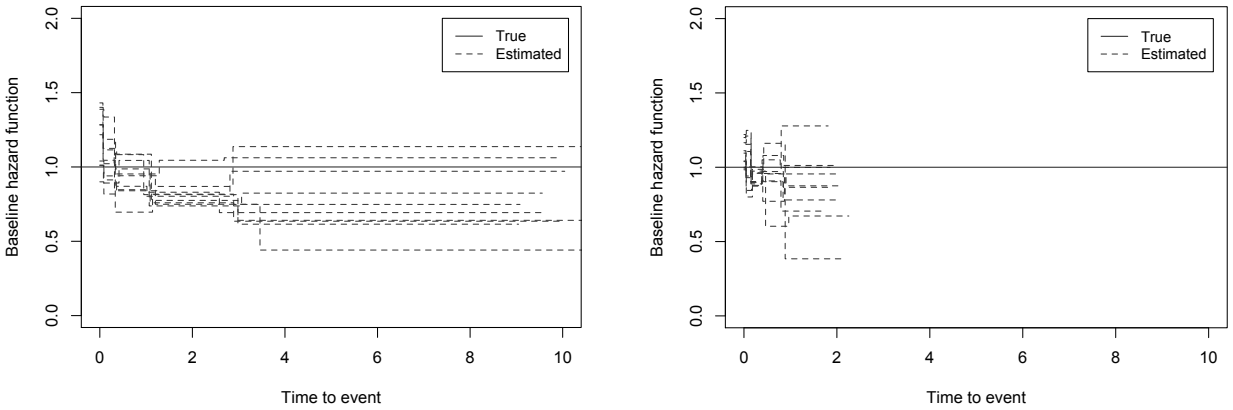
sample size	censoring rate		$\text{MSE}_{\hat{\beta}}$	$\text{MSE}_{\hat{\gamma}}$	iAUC	DIC
200	0.3	FPCA	0.109 (0.009)	0.614 (0.016)	0.935 (0.001)	42.99 (3.46)
		PCA	0.113 (0.010)	0.847 (0.020)	0.934 (0.001)	44.66 (3.58)
	0.5	FPCA	0.181 (0.014)	0.696 (0.021)	0.933 (0.002)	-93.80 (3.44)
		PCA	0.186 (0.014)	0.913 (0.025)	0.932 (0.002)	-92.21 (3.46)
500	0.3	FPCA	0.045 (0.003)	0.445 (0.012)	0.932 (0.001)	83.52 (4.52)
		PCA	0.047 (0.003)	0.581 (0.015)	0.930 (0.001)	85.50 (4.57)
	0.5	FPCA	0.052 (0.004)	0.454 (0.013)	0.928 (0.001)	-260.93 (4.58)
		PCA	0.052 (0.004)	0.600 (0.015)	0.927 (0.001)	-259.37 (4.63)

Table 3.5: Simulation results corresponding to $h_{01}(\cdot)$: the mean iAUC and the corresponding standard error in the parenthesis calculated from the 100 simulated data sets for each scenario. The Gibbs sampler was run for 20,000 iterations with 5,000 burn-in iterations for each simulated data set.

n	200		500	
Censoring rate	0.3	0.5	0.3	0.5
reduced	0.675 (0.004)	0.612 (0.006)	0.668 (0.002)	0.666 (0.002)
full	0.935 (0.001)	0.933 (0.002)	0.932 (0.001)	0.928 (0.001)

and a full BFLCRM model with both $W_i(\cdot)$ and $\boldsymbol{x}_i$.

Table 3.4 presents the estimation results with $h_{01}(\cdot)$ in (3.4.1) based on 100 simulated data sets for each scenario. The MSE values of both $\hat{\boldsymbol{\beta}}$ and $\hat{\gamma}(\cdot)$ are fairly small in all the cases. The values of iAUC indicate reasonable predictive performance of our BFLCRM. The MSE value decreases as either the sample size gets larger or the censoring rate gets smaller. Also, estimation results of fPCA are better than those of PCA in both MSE and iAUC. When the functional covariate has moderate measurement noise, fPCA will lead better estimation and prediction results since the use of smoothing step in fPCA can dramatically reduce measurement errors. Table 3.5 presents the means and standard errors of iAUC for the reduced and full BFLCRM models under each scenario. The iAUC value of the full



(a) Censoring rate=0.3

(b) Censoring rate=0.5

Figure 3.4: Simulation results corresponding to $h_{01}(\cdot)$: panels (a) and (b) respectively show the first 10 estimated baseline hazard functions with 0.3 and 0.5 censoring rates based on the size 500 samples. The solid line is the true baseline hazard function, $h_{01}(\cdot)$.

BFLCRM model is generally larger than that of the reduced model in all scenarios. This may indicate that the use of functional covariates can improve predictability of the hazard function. Figure 3.4 shows the baseline hazard functions estimated by the full BFLCRM from the first 10 data sets in the sample size 500 cases. The dotted lines show the estimated baseline hazard functions, and the true baseline hazard function $h_{01}(\cdot)$ is plotted as a solid line on each plot. When the true baseline hazard function is constant, our model estimates the true function well in low to moderate censoring cases.

Table 3.6 presents the estimation results with $h_{02}(\cdot)$ in (3.4.2) based on 100 simulated data sets for each scenario. Table 3.7 shows iAUC values for the two nested models. These results in Tables 3.6 and 3.7 are consistent with those based on $h_{01}(\cdot)$. The estimated baseline hazard functions are presented in Figure 3.5 on on page 39 for the sample size 500 cases. We plot the estimated baseline hazard functions of the first 10 data sets using the full BFLCRM model. The solid line shows the true baseline hazard function $h_{02}(\cdot)$ on each plot. When the true baseline hazard function is not piecewise constant, it is well approximated by the

Table 3.6: Simulation results corresponding to $h_{02}(\cdot)$ under different censoring rates and sample sizes: the deviance information criterion (DIC), the mean squared errors (MSE) of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ and the estimated integrated area under the curve (iAUCs) and their standard errors in parentheses calculated from the 100 simulated data sets. The Gibbs sampler was run for 20,000 iterations with 5,000 burn-in iterations for each simulated data set.

sample size	censoring rate		$\text{MSE}_{\hat{\beta}}$	$\text{MSE}_{\hat{\gamma}}$	iAUC	DIC
200	0.3	FPCA	0.112 (0.009)	0.618 (0.018)	0.934 (0.001)	128.21 (3.57)
		PCA	0.112 (0.010)	0.854 (0.021)	0.933 (0.001)	129.02 (3.66)
	0.5	FPCA	0.183 (0.017)	0.698 (0.023)	0.933 (0.002)	-15.26 (3.37)
		PCA	0.189 (0.018)	0.913 (0.023)	0.932 (0.002)	-13.70 (3.38)
500	0.3	FPCA	0.048 (0.003)	0.453 (0.012)	0.931 (0.001)	306.94 (4.46)
		PCA	0.049 (0.005)	0.586 (0.015)	0.930 (0.001)	308.62 (4.49)
	0.5	FPCA	0.054 (0.004)	0.457 (0.013)	0.927 (0.001)	-69.55 (4.50)
		PCA	0.054 (0.004)	0.611 (0.016)	0.926 (0.001)	-68.00 (4.54)

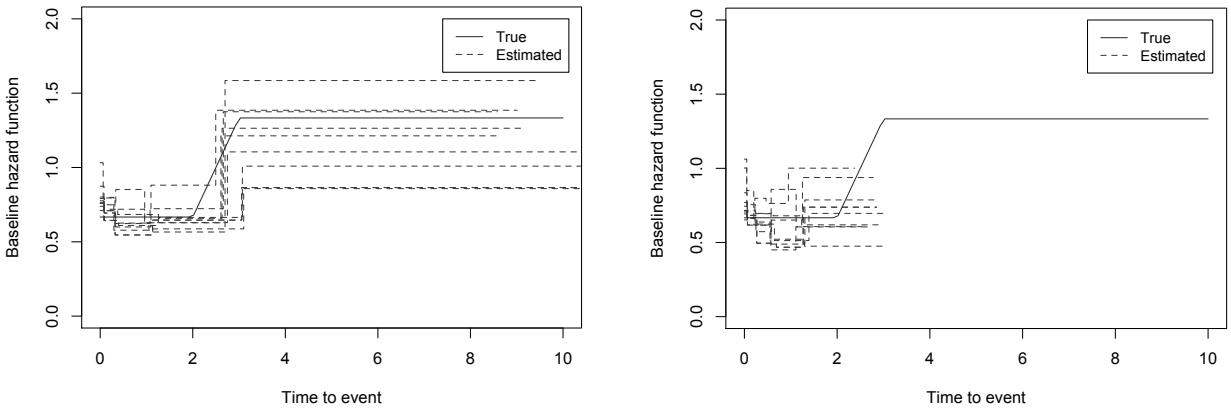
Table 3.7: Simulation results corresponding to $h_{02}(\cdot)$: the mean iAUC and the corresponding standard error in the parenthesis calculated from the 100 simulated data sets for each scenario. The Gibbs sampler was run for 20,000 iterations with 5,000 burn-in iterations for each simulated data set.

n	200		500	
Censoring rate	0.3	0.5	0.3	0.5
reduced	0.673 (0.004)	0.612 (0.006)	0.668 (0.002)	0.665 (0.002)
full	0.934 (0.001)	0.933 (0.002)	0.931 (0.001)	0.927 (0.001)

estimated baseline hazard function in the low censoring case. In the moderate censoring case, our model captures the general pattern of the true baseline hazard function. It may indicate that our BFLCRM approximates the general form of the baseline hazard function fairly well and therefore it is applicable for most of the practical settings.

3.5 Discussion

The BFLCRM was developed to predict the time of conversion from MCI to AD, as well as to determine the optimal set of predictors at baseline that effect the time of conversion. We obtained estimation and prediction results for functional and scalar predictors. This study has examined a very large set of predictors for predicting the time of conversion from



(a) Censoring rate=0.3

(b) Censoring rate=0.5

Figure 3.5: Simulation results corresponding to $h_{02}(\cdot)$: panels (a) and (b) respectively show the first 10 estimated baseline hazard functions with 0.3 and 0.5 censoring rates based on the size 500 samples. The solid line is the true baseline hazard function, $h_{02}(\cdot)$.

MCI to AD. We observed several important predictors including (i) length of education, (ii) age, (iii) APOE-ϵ4 carrier, (iv) ADAS-cog score, (v) the left hippocampal formation volume, (vi) the right amygdala volume, and (vii) surface morphology changes with the right and left hippocampi. These findings highlight the importance of including not only demographic and clinical information, but also high-dimensional imaging data, in statistical analyses of MCI-AD conversion. These results are also consistent with newly published clinical research criteria which incorporate the use of an array of biomarkers in research settings and clinical trials [1].

Several prior studies have highlighted the importance of hippocampal changes in the context of AD-related neurodegeneration and prediction of MCI-AD conversion [43]. These studies, however, commonly assess changes to hippocampal volume rather than surface morphology. The current analysis includes both measures of volume and surface area, with the changes in surface morphology adding additional predictive value.

As shown in Figure 3.2, the changes in surface area occur more prominently on the

anterior portion of the long axis of the hippocampus. Functional MRI studies in healthy adults suggest that anterior portions of the hippocampus are critical for the mnemonic binding processes that are engaged in tasks of episodic (day-to-day) memory. Since episodic memory tasks, particularly those that require binding operations, are some of the earliest cognitive impairments observed in MCI-AD [2], the anterior surface changes identified in the current analysis may underlie these early memory changes and serve as an important predictor of time of conversion.

From Figure 3.2, we observed that when hippocampal atrophy was greater in the CA1 subfield and subiculum of the hippocampi, the hazard rate to develop AD was increased. Similar to our finding, it was reported that greater atrophy of CA1 and subicular subfields in hippocampi was related to increased risk for conversion from MCI to AD [6]. The subregional atrophy rate in the CA1 and subicular subfields was also turned out to be the best predictor to explain the progression to AD from MCI [51]. Also, it was revealed that left hippocampal body volume was associated with delayed verbal memory [21], where the delayed verbal memory was one of important predictor for determining whether a subject was a MCI converter or not [57]. Thus, our finding in Figure 3.2 supports these research results.

Beyond the important effects on hippocampal surface morphology, we observed important volumetric changes in the left hippocampal formation and the right amygdala. There has been extensive research to diagnose Alzheimer's disease by using atrophy of various brain regions [136, 79, 38, 41]. In particular, it was reported that the hippocampal formation volume showed significant reduction in patients with clinically diagnosed Alzheimer's disease [86, 78]. It was also found that the amygdala volume was reduced in very early AD, which suggested that MRI-based amygdaloid volumetric measurement was a relevant marker [109]. Also, as shown in [127], the level of amygdala atrophy is related to global illness severity in the early stage of AD. Our ADNI-1 data analysis results agree with these findings in that volumetric changes in the hippocampal formation is an important variable to predict the

time to conversion from MCI to AD.

The analysis also shows that APOE status exerts important effects on the time of conversion. Our results also agree with several prior studies that have documented that the presence of the APOE $\epsilon 4$ allele increases the risk of developing Alzheimer's disease. Particularly, if a subject has APOE $\epsilon 4$, then MCI progression more likely occurs.

We have demonstrated the utility of BFLCRM as a valuable method for identifying optimal early markers of conversion to AD in patients with MCI. The early markers identified from our analysis could be used in case selection for various clinical trials for evaluating drug/therapeutic efficiency in slowing or modifying AD-related pathophysiology, when such drugs and therapeutic treatments are available.

There are some limitations to our analysis. Our findings survived internal cross validation, but they need replication in an independent community-based sample. We did not include measures of pathology (e.g. beta-amyloid) in our models since CSF and amyloid-PET were available only in a small subset of individuals in ADNI-1. However, a study of ADNI-2 subjects has shown a robust correlation between the APOE $\epsilon 4$ allele and cortical amyloid burden [115], suggesting that APOE $\epsilon 4$ may have served as a surrogate for cortical amyloid plaque load in our analysis.

We have developed BFLCRM for the use of functional and scalar covariates to predict time-to-event outcomes. Several important methodological issues need to be addressed in future research. First, it would be interesting to investigate the theoretical properties of our Bayesian procedure, including the support of the prior and truncation approximation bounds q_n. Second, it is interesting to develop a new Bayesian method to automatically determine the distribution of q_n. Third, it is interesting to incorporate high-dimensional scalar covariates (e.g., genetic markers in the whole genome) in BFLCRM and develop its associated estimation and testing procedures. Developing such statistical methods poses

many new challenges both computationally and theoretically.

3.6 Supplementary: Posterior Estimation Results for the Reduced Models

We have introduced three reduced models (Models 1-3) in our real data analysis in order to compare the full BFLCRM model with the reduced models. In this section, we discuss the estimation results for the three reduced models.

For regression coefficients, the hyperparameters of the multivariate Normal priors were set as $\boldsymbol{\mu}_0 = (0, \cdots, 0)$ and $\Sigma_0 = \text{diag}(5, \cdots, 5)$. We set $J = 70$ intervals so that each interval contains at least one failure or censored observation. For λ_j's, the shape and scale parameters of gamma priors were set to be $\alpha_{0j} = 0.2$ and $\alpha_{1j} = 0.4$ for $j = 1, \cdots, 70$ [143]. We ran the Gibbs sampler for 25,000 iterations after 5,000 burn-in iterations. We plotted the estimated survival functions for patients with different APOE status for all four models in Figure 3.6, where the values of continuous covariates were set as their mean values and categorical variables were set at their reference levels.

Model 1 includes all the scalar and functional covariates except for the ROI volume covariates. We applied BFLCRM to obtain the estimation results for the 26 covariates. Among them, "Age", "APOE-ϵ4 carrier", and "ADAS-cog score" have their 95% HPD intervals that do not contain 0. Moreover, the 1st, 3rd, and 14th fPCs have their 95% HPD intervals that do not contain 0. It suggests that the three scalar covariates and the functional covariate are important predictors in Model 1. Inspecting Figure 3.7 reveals that the subfields of CA1, CA2, CA3, and subiculum on the hippocampi have positive effects on the hazard function, indicating that the thinner these areas on the hippocampus, the shorter the time to conversion to AD is. Compared with the results obtained from the full BFLCRM model, the red areas are broader for Model 1. Inspecting Figure 3.6(b) reveals that the subjects having APOE-ϵ4 are more likely to progress to AD than the non-carriers. The dotted lines show

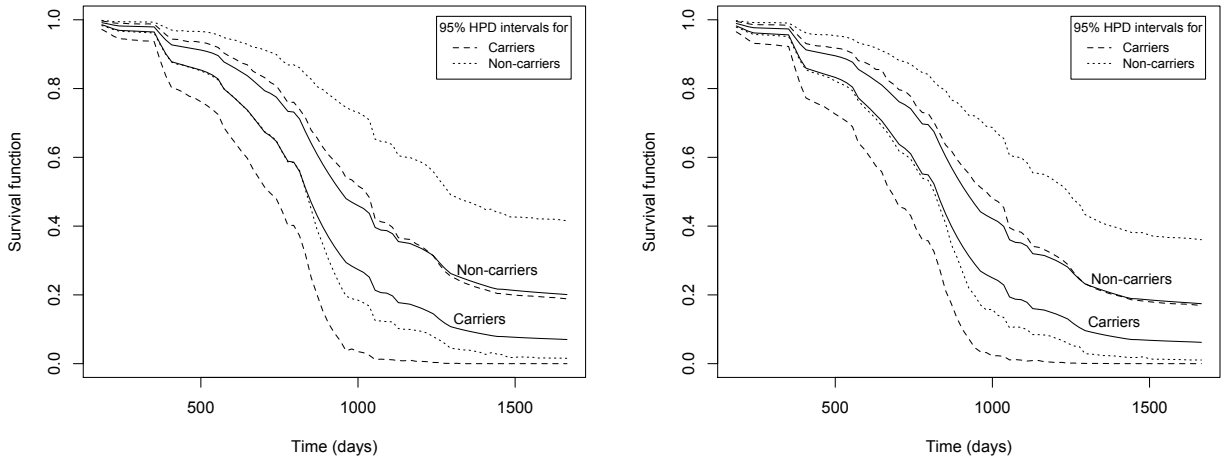
the 95% HPD intervals of the survival functions.

For Model 2, we excluded the hippocampal surfaces covariates from the full BFLCRM. The BFLCRM was fitted to obtain the estimation results for the 19 covariates. Among them, "Age", "APOE-ϵ4 carrier", and "ADAS-cog score", and "Left hippocampal formation" have their 95% HPD intervals that do not contain 0, which implies that they play an important role in predicting the conversion time to AD. Figure 3.6(c) shows that the APOE-ϵ4 carriers tend to be more hazardous than the other subjects.

Model 3 only includes the clinical and genetic covariates, and ADAS-cog score. The BFLCRM was applied to estimate the regression coefficients for the 12 covariates. There are important predictors "APOE-ϵ4 carrier" and "ADAS-cog score" whose 95% HPD intervals do not contain 0. Figure 3.6(d) shows the estimated survival curves for average patients with different APOE status. It shows the similar pattern with the other models.

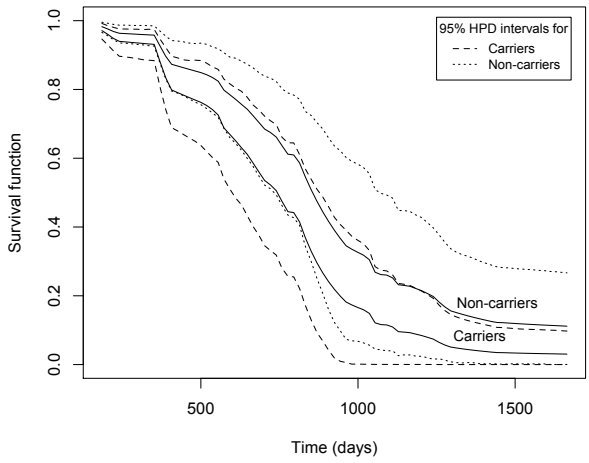
Table 3.8: ADNI data analysis results for Model 1: the posterior quantities of 10 regression coefficients β_ks, that correspond to $\boldsymbol{x}_i$ =(Gender, Handedness, Widowed, Divorced, Never married, Length of Education, Retirement, Age, APOE-ϵ4 carrier, and ADAS-cog Score). Mean denotes 'posterior mean', SD denotes 'posterior standard deviation', and lower and upper, respectively, represent the 'lower and upper limits' of a 95% highest posterior density interval.

	β_1 Female	β_2 Left	β_3 Widowed	β_4 Divorced	β_5 Never-married	β_6 Years of edu	β_7 Retired	β_8 Age	β_9 ApoE4	β_{10} ADAS
Mean	0.3943	0.2826	0.2422	0.1409	0.5174	-0.0577	0.4045	-0.0406	0.5143	0.1572
SD	0.2016	0.3317	0.3121	0.4377	0.7682	0.0341	0.2442	0.0156	0.2190	0.0239
lower	-0.0157	-0.3488	-0.4151	-0.7356	-0.9930	-0.1260	-0.0610	-0.0699	0.1069	0.1104
upper	0.7842	0.9094	0.8308	0.9196	1.9468	0.0045	0.8900	-0.0089	0.9634	0.2030

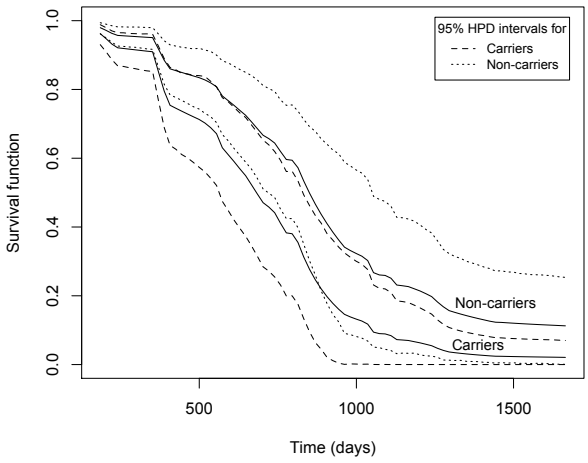


(a) The full BFLCRM model

(b) Model 1

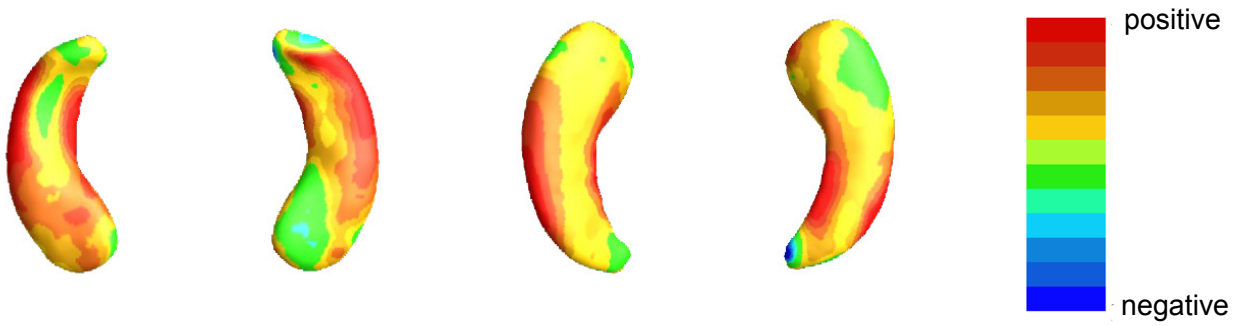


(c) Model 2



(d) Model 3

Figure 3.6: ADNI data analysis results: the estimated survival curves of APOE-ϵ4 carriers and non-carriers under the full and reduced BFLCRM models. Other continuous or categorical covariates are fixed at the mean values or reference levels. The dotted lines are the 95% HPD intervals of the estimated survival functions.



(a) Top (b) Bottom (c) Color bar

Figure 3.7: ADNI data analysis results for Model 1: panels (a) and (b), respectively, show the top and bottom views of the estimated coefficient function associated with the hippocampal surface data color-coded by the colorbar in panel (c).

Table 3.9: ADNI data analysis results for Model 2: the posterior quantities of 17 regression coefficients β_ks, that correspond to $\boldsymbol{x}_i$ =(Gender, Handedness, Widowed, Divorced, Never married, Length of Education, Retirement, Age, APOE-ϵ4 carrier, ADAS-cog Score, posterior limb of internal capsule, Right hippocampal formation, Left hippocampal formation, Left thalamus, Left amygdala, Right amygdala, and Right thalamus). Mean denotes 'posterior mean', SD denotes 'posterior standard deviation', and lower and upper, respectively, represent the 'lower and upper limits' of a 95% highest posterior density interval.

	β_1 Female	β_2 Left	β_3 Widowed	β_4 Divorced	β_5 Never-married	β_6 Years of education	β_7 Retired	β_8 Age in years
Mean	0.2929	0.0524	0.0711	0.3212	0.9195	-0.0456	0.1739	-0.0344
SD	0.2249	0.3488	0.3209	0.4731	0.7498	0.0323	0.2386	0.0167
lower	-0.1426	-0.6263	-0.5757	-0.6099	-0.5290	-0.1108	-0.2810	-0.0669
upper	0.7190	0.7409	0.6871	1.2075	2.3582	0.0162	0.6708	-0.0019

	β_9 ApoE4	β_{10} ADAS	β_{11} PLIC	β_{12} RHF	β_{13} LHF	β_{14} LT	β_{15} LA	β_{16} RA	β_{17} RT
Mean	0.5183	0.1296	0.0008	-0.0001	-0.0009	-0.0005	0.0015	-0.0007	0.0004
SD	0.2149	0.0231	0.0005	0.0003	0.0003	0.0004	0.0008	0.0005	0.0004
lower	0.1052	0.0867	-0.0002	-0.0007	-0.0015	-0.0012	0.0000	-0.0016	-0.0003
upper	0.9639	0.1757	0.0017	0.0005	-0.0004	0.0002	0.0033	0.0002	0.0011

Table 3.10: ADNI data analysis results for Model 3: the posterior quantities of 10 regression coefficients β_ks, that correspond to $\boldsymbol{x}_i$ =(Gender, Handedness, Widowed, Divorced, Never married, Length of Education, Retirement, Age, APOE-ϵ4 carrier, and ADAS-cog Score). Mean denotes 'posterior mean', SD denotes 'posterior standard deviation', and lower and upper, respectively, represent the 'lower and upper limits' of a 95% highest posterior density interval.

	β_1 Female	β_2 Left	β_3 Widowed	β_4 Divorced	β_5 Never-married	β_6 Years of edu	β_7 Retired	β_8 Age	β_9 ApoE4	β_{10} ADAS
Mean	0.3214	0.2696	0.0853	0.1423	0.6547	-0.0094	0.1612	0.0066	0.6369	0.1475
SD	0.1942	0.3268	0.3069	0.4368	0.7398	0.0311	0.2259	0.0136	0.1931	0.0214
lower	-0.0647	-0.3921	-0.5072	-0.7098	-0.7606	-0.0702	-0.2575	-0.0190	0.2790	0.1064
upper	0.6940	0.8669	0.6927	1.0022	2.0798	0.0551	0.6290	0.0339	1.0212	0.1896

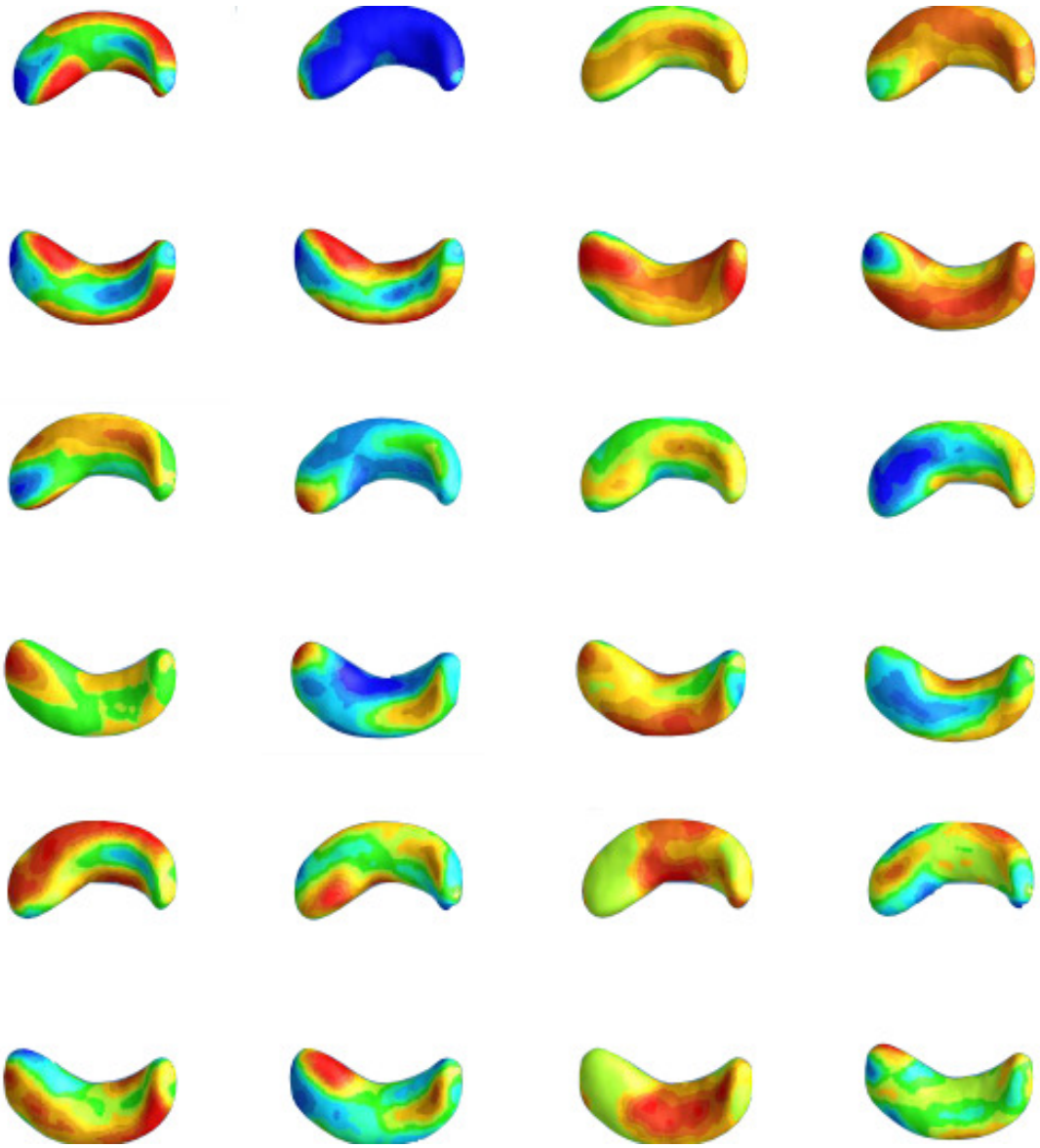


Figure 3.8: ADNI data analysis results: the first 12 largest estimated eigenfunctions projected on the hippocampal surface

Table 3.11: Sensitivity analysis of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ for ADNI-1 data with different values of hyperparameters in the Normal priors on the regression coefficients.

	σ^2=5				σ^2=25				σ^2=100			
	Mean	SD	lower	upper	Mean	SD	lower	upper	Mean	SD	lower	upper
β_1	0.434	0.251	-0.049	0.948	0.462	0.250	-0.004	0.952	0.451	0.246	-0.069	0.893
β_2	0.225	0.365	-0.525	0.863	0.216	0.360	-0.446	0.943	0.223	0.388	-0.652	0.903
β_3	0.312	0.383	-0.463	1.014	0.362	0.337	-0.253	1.029	0.344	0.368	-0.402	1.070
β_4	0.273	0.466	-0.679	1.119	0.276	0.474	-0.702	1.162	0.297	0.513	-0.769	1.225
β_5	0.720	0.787	-0.938	2.201	0.792	0.773	-0.732	2.256	0.817	0.823	-0.798	2.364
β_6	-0.087	0.037	-0.169	-0.024	-0.088	0.036	-0.155	-0.015	-0.090	0.035	-0.157	-0.021
β_7	0.345	0.248	-0.092	0.861	0.370	0.255	-0.128	0.855	0.391	0.244	-0.090	0.832
β_8	-0.052	0.018	-0.087	-0.019	-0.053	0.018	-0.091	-0.020	-0.052	0.019	-0.087	-0.016
β_9	0.555	0.234	0.126	1.026	0.576	0.224	0.180	1.029	0.554	0.228	0.125	1.003
β_{10}	0.157	0.027	0.103	0.208	0.160	0.027	0.111	0.215	0.156	0.027	0.106	0.215
β_{11}	0.001	0.001	0.000	0.002	0.001	0.001	0.000	0.002	0.001	0.001	0.000	0.002
β_{12}	0.001	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.001	0.000	0.000	0.001
β_{13}	-0.001	0.000	-0.002	0.000	-0.001	0.000	-0.002	0.000	-0.001	0.000	-0.002	0.000
β_{14}	0.000	0.000	-0.001	0.000	0.000	0.000	-0.001	0.000	0.000	0.000	-0.001	0.000
β_{15}	0.002	0.001	0.000	0.004	0.002	0.001	0.000	0.004	0.002	0.001	0.000	0.003
β_{16}	-0.001	0.001	-0.002	0.000	-0.001	0.001	-0.002	0.000	-0.001	0.001	-0.002	0.000
β_{17}	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001
γ_1	0.007	0.002	0.003	0.011	0.007	0.002	0.003	0.011	0.007	0.002	0.002	0.011
γ_2	0.003	0.003	-0.002	0.008	0.003	0.003	-0.002	0.008	0.004	0.003	-0.002	0.009
γ_3	0.005	0.003	-0.002	0.012	0.005	0.003	-0.002	0.011	0.004	0.003	-0.002	0.011
γ_4	-0.002	0.003	-0.009	0.004	-0.002	0.003	-0.009	0.004	-0.002	0.003	-0.009	0.005
γ_5	-0.003	0.003	-0.009	0.004	-0.003	0.003	-0.009	0.004	-0.003	0.003	-0.010	0.003
γ_6	0.000	0.004	-0.008	0.009	0.000	0.004	-0.008	0.009	0.000	0.004	-0.009	0.008
γ_7	-0.009	0.005	-0.018	0.000	-0.009	0.005	-0.018	0.002	-0.008	0.005	-0.017	0.002
γ_8	0.003	0.006	-0.008	0.015	0.003	0.006	-0.008	0.014	0.003	0.006	-0.009	0.014
γ_9	-0.009	0.005	-0.019	0.002	-0.010	0.005	-0.021	0.000	-0.009	0.005	-0.020	0.001
γ_{10}	0.008	0.006	-0.004	0.020	0.007	0.006	-0.005	0.019	0.007	0.006	-0.004	0.019
γ_{11}	-0.014	0.007	-0.028	0.001	-0.015	0.007	-0.029	0.000	-0.015	0.007	-0.029	-0.001
γ_{12}	0.001	0.007	-0.013	0.015	0.001	0.007	-0.013	0.015	0.001	0.007	-0.013	0.016
γ_{13}	0.000	0.007	-0.014	0.014	0.000	0.007	-0.014	0.015	0.000	0.007	-0.013	0.016
γ_{14}	-0.028	0.007	-0.043	-0.015	-0.028	0.008	-0.045	-0.013	-0.029	0.007	-0.042	-0.015

Table 3.12: Sensitivity analysis of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ for ADNI-1 data with different values of hyperparameters in the Gamma priors on the piecewise constant baseline hazard function.

	(α_0, α_1)=(0.2,0.4)				(α_0, α_1)=(0.2,1)				(α_0, α_1)=(0.5,1)			
	Mean	SD	lower	upper	Mean	SD	lower	upper	Mean	SD	lower	upper
β_1	0.434	0.251	-0.049	0.948	0.461	0.260	-0.031	0.966	0.482	0.256	-0.011	1.014
β_2	0.225	0.365	-0.525	0.863	0.235	0.401	-0.564	0.948	0.232	0.396	-0.566	0.940
β_3	0.312	0.383	-0.463	1.014	0.380	0.396	-0.383	1.137	0.397	0.366	-0.370	1.064
β_4	0.273	0.466	-0.679	1.119	0.276	0.559	-0.743	1.457	0.254	0.513	-0.761	1.207
β_5	0.720	0.787	-0.938	2.201	0.709	0.861	-1.041	2.279	0.714	0.820	-0.911	2.235
β_6	-0.087	0.037	-0.169	-0.024	-0.100	0.041	-0.175	-0.016	-0.102	0.041	-0.183	-0.025
β_7	0.345	0.248	-0.092	0.861	0.414	0.267	-0.149	0.896	0.398	0.272	-0.135	0.962
β_8	-0.052	0.018	-0.087	-0.019	-0.055	0.019	-0.092	-0.020	-0.055	0.019	-0.093	-0.020
β_9	0.555	0.234	0.126	1.026	0.656	0.265	0.171	1.233	0.669	0.246	0.184	1.138
β_{10}	0.157	0.027	0.103	0.208	0.167	0.027	0.115	0.218	0.168	0.026	0.118	0.217
β_{11}	0.001	0.001	0.000	0.002	0.001	0.001	0.000	0.002	0.001	0.001	0.000	0.002
β_{12}	0.001	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.001	0.000	0.000	0.002
β_{13}	-0.001	0.000	-0.002	0.000	-0.001	0.000	-0.002	-0.001	-0.001	0.000	-0.002	-0.001
β_{14}	0.000	0.000	-0.001	0.000	0.000	0.000	-0.001	0.000	0.000	0.000	-0.001	0.000
β_{15}	0.002	0.001	0.000	0.004	0.002	0.001	0.000	0.004	0.002	0.001	0.000	0.004
β_{16}	-0.001	0.001	-0.002	0.000	-0.001	0.001	-0.002	0.000	-0.001	0.001	-0.003	0.000
β_{17}	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001
γ_1	0.007	0.002	0.003	0.011	0.007	0.002	0.003	0.012	0.007	0.002	0.003	0.012
γ_2	0.003	0.003	-0.002	0.008	0.004	0.003	-0.002	0.010	0.004	0.003	-0.002	0.009
γ_3	0.005	0.003	-0.002	0.012	0.004	0.004	-0.003	0.011	0.004	0.004	-0.002	0.012
γ_4	-0.002	0.003	-0.009	0.004	-0.002	0.004	-0.009	0.005	-0.002	0.004	-0.010	0.005
γ_5	-0.003	0.003	-0.009	0.004	-0.004	0.004	-0.011	0.003	-0.003	0.003	-0.010	0.004
γ_6	0.000	0.004	-0.008	0.009	0.000	0.005	-0.009	0.009	0.000	0.005	-0.009	0.009
γ_7	-0.009	0.005	-0.018	0.000	-0.009	0.005	-0.019	0.001	-0.009	0.005	-0.019	0.001
γ_8	0.003	0.006	-0.008	0.015	0.004	0.006	-0.007	0.017	0.004	0.006	-0.008	0.016
γ_9	-0.009	0.005	-0.019	0.002	-0.010	0.006	-0.023	0.001	-0.011	0.006	-0.023	-0.001
γ_{10}	0.008	0.006	-0.004	0.020	0.009	0.007	-0.005	0.021	0.008	0.007	-0.005	0.020
γ_{11}	-0.014	0.007	-0.028	0.001	-0.017	0.008	-0.033	0.000	-0.016	0.008	-0.031	-0.002
γ_{12}	0.001	0.007	-0.013	0.015	0.000	0.008	-0.015	0.017	-0.001	0.008	-0.017	0.015
γ_{13}	0.000	0.007	-0.014	0.014	-0.002	0.008	-0.016	0.014	-0.001	0.008	-0.016	0.014
γ_{14}	-0.028	0.007	-0.043	-0.015	-0.032	0.008	-0.047	-0.017	-0.031	0.008	-0.047	-0.015

Table 3.13: Sensitivity analysis of $\boldsymbol{\lambda}$ for ADNI-1 data with different values of hyperparameters in the Normal priors on the regression coefficients.

	σ^2=5				σ^2=25				σ^2=100			
	Mean	SD	lower	upper	Mean	SD	lower	upper	Mean	SD	lower	upper
λ_1	0.004	0.003	0.000	0.010	0.004	0.003	0.000	0.010	0.004	0.003	0.000	0.010
λ_2	0.067	0.063	0.000	0.191	0.066	0.062	0.000	0.190	0.067	0.063	0.000	0.194
λ_3	1.087	0.514	0.242	2.104	1.080	0.521	0.230	2.117	1.091	0.522	0.235	2.121
λ_4	0.145	0.103	0.004	0.345	0.144	0.103	0.004	0.345	0.146	0.103	0.005	0.350
λ_5	0.146	0.085	0.015	0.312	0.145	0.085	0.015	0.312	0.147	0.086	0.018	0.321
λ_6	0.166	0.086	0.028	0.333	0.164	0.087	0.029	0.338	0.166	0.087	0.027	0.336
λ_7	0.020	0.019	0.000	0.057	0.020	0.019	0.000	0.057	0.020	0.019	0.000	0.058
λ_8	0.016	0.009	0.002	0.034	0.016	0.009	0.002	0.033	0.016	0.009	0.002	0.034
λ_9	0.038	0.026	0.002	0.089	0.037	0.027	0.001	0.089	0.038	0.027	0.002	0.090
λ_{10}	0.488	0.204	0.133	0.890	0.485	0.205	0.133	0.886	0.487	0.206	0.139	0.891
λ_{11}	0.249	0.171	0.010	0.584	0.248	0.171	0.008	0.575	0.249	0.171	0.010	0.581
λ_{12}	0.033	0.074	0.000	0.175	0.033	0.074	0.000	0.174	0.033	0.074	0.000	0.174
λ_{13}	3.999	1.788	0.927	7.461	3.978	1.788	0.986	7.549	3.991	1.791	0.974	7.528
λ_{14}	0.143	0.131	0.000	0.406	0.142	0.131	0.000	0.403	0.143	0.132	0.000	0.404
λ_{15}	1.097	0.552	0.206	2.191	1.091	0.552	0.187	2.166	1.095	0.554	0.191	2.174
λ_{16}	0.617	0.351	0.073	1.304	0.614	0.351	0.070	1.298	0.616	0.351	0.074	1.311
λ_{17}	1.212	0.549	0.329	2.330	1.206	0.549	0.275	2.268	1.208	0.549	0.270	2.271
λ_{18}	0.143	0.098	0.005	0.332	0.142	0.098	0.006	0.335	0.142	0.098	0.005	0.333
λ_{19}	0.143	0.132	0.000	0.401	0.142	0.131	0.000	0.400	0.142	0.131	0.000	0.400
λ_{20}	0.273	0.157	0.036	0.590	0.272	0.157	0.032	0.582	0.272	0.157	0.032	0.583
λ_{21}	0.338	0.194	0.032	0.712	0.336	0.194	0.037	0.716	0.336	0.194	0.037	0.718
λ_{22}	0.177	0.101	0.020	0.372	0.176	0.101	0.021	0.373	0.176	0.101	0.020	0.372
λ_{23}	0.150	0.085	0.022	0.322	0.149	0.085	0.019	0.320	0.149	0.085	0.020	0.319
λ_{24}	0.018	0.012	0.001	0.043	0.018	0.012	0.001	0.043	0.018	0.012	0.001	0.043
λ_{25}	0.552	0.315	0.065	1.169	0.551	0.315	0.049	1.160	0.550	0.314	0.054	1.156
λ_{26}	1.630	0.735	0.417	3.102	1.625	0.736	0.391	3.074	1.623	0.731	0.398	3.067
λ_{27}	0.473	0.323	0.019	1.109	0.472	0.321	0.019	1.108	0.471	0.321	0.022	1.109
λ_{28}	0.546	0.308	0.057	1.141	0.545	0.308	0.065	1.155	0.545	0.308	0.065	1.155
λ_{29}	0.811	0.401	0.159	1.606	0.809	0.400	0.149	1.587	0.809	0.399	0.151	1.591
λ_{30}	1.148	0.508	0.292	2.159	1.145	0.508	0.277	2.141	1.145	0.509	0.304	2.174
λ_{31}	0.458	0.314	0.013	1.070	0.457	0.313	0.016	1.068	0.457	0.313	0.017	1.067
λ_{32}	0.712	0.644	0.000	1.992	0.711	0.642	0.000	2.002	0.711	0.643	0.000	1.990
λ_{33}	0.188	0.108	0.021	0.401	0.187	0.107	0.025	0.404	0.187	0.108	0.017	0.396
λ_{34}	0.065	0.045	0.002	0.151	0.065	0.044	0.002	0.151	0.065	0.044	0.001	0.151
λ_{35}	0.075	0.051	0.003	0.175	0.074	0.051	0.002	0.173	0.074	0.051	0.002	0.174
λ_{36}	0.088	0.050	0.011	0.187	0.088	0.050	0.009	0.184	0.088	0.050	0.009	0.185
λ_{37}	0.281	0.259	0.000	0.790	0.280	0.258	0.000	0.793	0.280	0.258	0.000	0.790
λ_{38}	0.688	0.465	0.033	1.603	0.687	0.464	0.031	1.600	0.688	0.465	0.032	1.607
λ_{39}	1.224	0.606	0.221	2.407	1.223	0.605	0.235	2.433	1.224	0.605	0.239	2.426
λ_{40}	1.274	0.858	0.049	2.965	1.274	0.857	0.038	2.955	1.276	0.859	0.060	2.988

Table 3.14: Sensitivity analysis of $\boldsymbol{\lambda}$ for ADNI-1 data with different values of hyperparameters in the Gamma priors on the piecewise constant baseline hazard function.

	(α_0, α_1)=(0.2,0.4)				(α_0, α_1)=(0.2,1)				(α_0, α_1)=(0.5,1)			
	Mean	SD	lower	upper	Mean	SD	lower	upper	Mean	SD	lower	upper
λ_1	0.004	0.003	0.000	0.010	0.003	0.002	0.000	0.008	0.003	0.002	0.000	0.008
λ_2	0.067	0.063	0.000	0.191	0.051	0.048	0.000	0.147	0.051	0.048	0.000	0.145
λ_3	1.087	0.514	0.242	2.104	0.783	0.378	0.152	1.528	0.775	0.370	0.183	1.519
λ_4	0.145	0.103	0.004	0.345	0.112	0.080	0.005	0.270	0.110	0.078	0.003	0.262
λ_5	0.146	0.085	0.015	0.312	0.115	0.068	0.012	0.250	0.113	0.067	0.010	0.242
λ_6	0.166	0.086	0.028	0.333	0.131	0.069	0.018	0.265	0.129	0.067	0.022	0.261
λ_7	0.020	0.019	0.000	0.057	0.016	0.015	0.000	0.046	0.016	0.015	0.000	0.045
λ_8	0.016	0.009	0.002	0.034	0.013	0.007	0.002	0.028	0.013	0.007	0.001	0.027
λ_9	0.038	0.026	0.002	0.089	0.031	0.022	0.001	0.073	0.031	0.022	0.001	0.072
λ_{10}	0.488	0.204	0.133	0.890	0.391	0.166	0.109	0.725	0.387	0.162	0.120	0.724
λ_{11}	0.249	0.171	0.010	0.584	0.197	0.136	0.011	0.464	0.195	0.134	0.006	0.454
λ_{12}	0.033	0.074	0.000	0.175	0.026	0.057	0.000	0.135	0.026	0.056	0.000	0.133
λ_{13}	3.999	1.788	0.927	7.461	2.491	1.109	0.621	4.701	2.478	1.097	0.572	4.599
λ_{14}	0.143	0.131	0.000	0.406	0.113	0.104	0.000	0.318	0.112	0.103	0.000	0.316
λ_{15}	1.097	0.552	0.206	2.191	0.818	0.413	0.145	1.625	0.811	0.407	0.159	1.631
λ_{16}	0.617	0.351	0.073	1.304	0.476	0.272	0.056	1.010	0.473	0.268	0.056	1.002
λ_{17}	1.212	0.549	0.329	2.330	0.920	0.418	0.237	1.760	0.913	0.413	0.232	1.737
λ_{18}	0.143	0.098	0.005	0.332	0.117	0.081	0.005	0.273	0.116	0.080	0.004	0.269
λ_{19}	0.143	0.132	0.000	0.401	0.114	0.105	0.000	0.322	0.113	0.104	0.000	0.319
λ_{20}	0.273	0.157	0.036	0.590	0.222	0.129	0.023	0.477	0.220	0.127	0.027	0.474
λ_{21}	0.338	0.194	0.032	0.712	0.272	0.157	0.037	0.587	0.270	0.155	0.027	0.571
λ_{22}	0.177	0.101	0.020	0.372	0.145	0.084	0.016	0.309	0.145	0.083	0.017	0.305
λ_{23}	0.150	0.085	0.022	0.322	0.124	0.071	0.016	0.266	0.123	0.070	0.015	0.262
λ_{24}	0.018	0.012	0.001	0.043	0.015	0.011	0.001	0.036	0.015	0.010	0.000	0.036
λ_{25}	0.552	0.315	0.065	1.169	0.432	0.247	0.050	0.919	0.432	0.246	0.046	0.914
λ_{26}	1.630	0.735	0.417	3.102	1.210	0.546	0.291	2.279	1.209	0.543	0.279	2.267
λ_{27}	0.473	0.323	0.019	1.109	0.366	0.250	0.014	0.858	0.366	0.249	0.015	0.856
λ_{28}	0.546	0.308	0.057	1.141	0.432	0.245	0.057	0.922	0.433	0.244	0.042	0.903
λ_{29}	0.811	0.401	0.159	1.606	0.638	0.317	0.123	1.268	0.639	0.316	0.119	1.259
λ_{30}	1.148	0.508	0.292	2.159	0.894	0.397	0.213	1.674	0.896	0.397	0.222	1.679
λ_{31}	0.458	0.314	0.013	1.070	0.359	0.246	0.013	0.840	0.359	0.246	0.013	0.839
λ_{32}	0.712	0.644	0.000	1.992	0.479	0.434	0.000	1.343	0.480	0.434	0.000	1.342
λ_{33}	0.188	0.108	0.021	0.401	0.157	0.091	0.014	0.337	0.157	0.090	0.017	0.335
λ_{34}	0.065	0.045	0.002	0.151	0.055	0.038	0.002	0.128	0.055	0.038	0.002	0.129
λ_{35}	0.075	0.051	0.003	0.175	0.063	0.043	0.001	0.147	0.063	0.043	0.002	0.148
λ_{36}	0.088	0.050	0.011	0.187	0.075	0.043	0.008	0.158	0.076	0.043	0.008	0.159
λ_{37}	0.281	0.259	0.000	0.790	0.220	0.203	0.000	0.623	0.220	0.203	0.000	0.623
λ_{38}	0.688	0.465	0.033	1.603	0.520	0.352	0.025	1.214	0.521	0.352	0.023	1.207
λ_{39}	1.224	0.606	0.221	2.407	0.934	0.462	0.192	1.863	0.938	0.463	0.175	1.856
λ_{40}	1.274	0.858	0.049	2.965	0.867	0.584	0.045	2.031	0.869	0.585	0.038	2.022

CHAPTER 4

BAYESIAN BI-LEVEL VARIABLE SELECTION IN SURVIVAL MODEL

4.1 Introduction

In genetic epidemiology, there has been increased interest to identify genetic variants associated with traits. Typically, it is of interest to elucidate the association between the traits and single-nucleotide polymorphisms (SNPs), a DNA sequence variation occurring commonly within a population. A genome-wide association study (GWAS) focuses on identifying important SNPs to relate to clinical outcomes in this context. Since the SNP data is ultra high-dimensional (half a million or more), the simple (and popular) GWASs conduct a number of marginal tests: examination of the effect of each SNP one by one. It makes the GWASs to be theoretically and computationally feasible in the classical regression setup, when the number of covariates is less than the sample size. But the simple GWASs face two main challenges: dealing with multiple testing issue and accounting for the intricate dependency structure among SNPs. Due to the multiple comparison problem, GWAS requires appropriate control for population stratification and false discovery rate. Also it does not incorporate a joint structure within SNPs by only considering the marginal effect of each SNP. Since SNPs are naturally grouped by genes or linkage disequilibrium (LD) blocks, discarding the grouping information will result in the poor performance of GWAS.

In order to resolve the multiple testing issue, one may consider all the SNPs simultaneously in one model. It is a high-dimensional problem, which can be tackled by penalization methods, sure independence screening (SIS) strategy, Bayesian variable selection methods. After prescreening SNPs using marginal screening procedures, the LASSO and elastic-net

penalties were used respectively in [163] and [24]. But these approaches are likely to remove important features marginally uncorrelated with an outcome because the prescreening procedure is only conducted on their marginal correlations. Liu et al. [99] used a combination of two penalties: the minimax concave penalty (MCP) and an L2 penalty that encourages a smaller difference in genetic effect at adjacent SNPs that are in stronger LD. He and Lin [67] proposed a modified ISIS including one marginal SIS and two rounds of conditional SIS in order to capture important SNPs that are not marginally correlated with the outcome. Guan and Stephens [60] used a Bayesian variable selection method by introducing a binary variable indicating if a variable is important or not. These methods select relevant SNPs without incorporating the natural grouping information such as LD blocks or genes. These methods select relevant SNPs without using the grouping information of SNPs, while Zhang et al. [172] enables bi-level selection by incorporating the grouping information to select genetic variants in both gene and SNP level simultaneously. Zhang et al. [172] proposed integrative Bayesian Variable Selection (iBVS) by introducing two binary variables to indicate the selection of genes and SNPs respectively into a Bayesian variable selection model. Although they used group information, they prescreened SNPs using marginal testing, which may lead to discard important features marginally uncorrelated with an outcome in their screening step. In order to resolve those limitations in the GWASs, we propose a Bayesian bi-level variable selection (BBVS) method. This method aims to detect SNPs associated with (continuous) clinical outcomes (1) by considering all the SNPs simultaneously and (2) incorporating the grouping information of the SNP data. In particular, this paper focuses on time to event as an outcome, which is motivated by Alzheimers disease (AD) studies. In the pathology of AD, mild cognitive impairment (MCI) is a clinical syndrome characterized by the onset and evolution of cognitive impairments, which is often considered as a transitional stage to AD. Since therapeutic interventions and a few disease-modifying drugs are more effective at MCI or early stage of AD than at the severe stage of AD, early diagnosis of AD is a critical concern. It is an ongoing quest to predict the time to conversion from MCI to AD

and detect biomarkers affecting progression to AD. Although the effect of genetic factors on the progression of AD is substantial, up to our knowledge, however, the time to conversion from MCI to AD has not been addressed in the context of GWASs.

Our method has two hierarchical levels of variable selection: the first one is group-wise and the second level is element-wise variable selection. In the first step, we identify important groups of variables and update the censored event time from its predictive posterior distribution by data augmentation [150, 138]. The dimension of the whole SNP data can be significantly reduced by eliminating irrelevant groups to time to event. Since this step also provides a posterior sample of censored time to event, the posterior mean will be used as imputed censored event time in the second level of variable selection. It converts the AFT model in the second level into a usual log-normal regression model. In the second level, we only include variables in the selected groups as covariates in the AFT model with the imputed event time as an outcome. To conduct element-wise variable selection, shrinkage priors are employed on regression parameters. In particular, we extend Dirichlet-Laplace shrinkage priors proposed by Bhattacharya et al. [9] to incorporate the grouping information. The rest of this paper is organized as follows. In section 4.2, a basic AFT model will be discussed. In section 4.3, we discuss our Bayesian bi-level variable selection (BBVS) in the AFT model. In section 4.4, we conduct simulation studies to validate and compare the performance of BBVS with other group/bi-level selection methods. In section 4.5, the ADNI data will be analyzed by BBVS to detect a set of SNPs associated with the time to conversion from MCI to AD.

4.2 Accelerated Failure Time Model

While Cox regression postulates that covariates are multiplicatively related to the hazard, accelerated failure time (AFT) model assumes a direct relationship between failure time and covariates. AFT model can be very appealing, since it is straight-forward to interpret the

effect of the estimated regression coefficients. For the i-th subject, its probability model is given by

$$Y_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\nu_i, i = 1, \cdots, n,$$

which becomes the linear model in log scale

$$\log Y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i, i = 1, \cdots, n,$$

where $Y_1, Y_2, \cdots, Y_n$ are failure times, $\boldsymbol{x}_i = (1, x_{i1}, x_{i2}, \cdots, x_{ip})'$ is a vector of known explanatory variables (including the intercept) for the i-th individual, $\boldsymbol{\beta}$ is a vector of $p+1$ unknown regression coefficients, and $\epsilon_i = \log \nu_i$ is the error term. Usually the error term is assumed to follow the parametric distribution, such as Normal distribution. A parametric approach was discussed by Kalbfleisch and Prentice [84]. There are extensive discussion about parametric AFT model in [91, 105, 116] in the frequentist framework. A parametric Bayesian analysis was presented by Bedrick et al. [7], while a semiparametric Bayesian approaches have been described in Christensen and Johnson [25] where they model v_i using a Dirichlet process. A fully Bayesian analysis is developed for the estimation of regression coefficients in Johnson and Christensen [83]. Kuo and Mallick [90] consider a Dirichlet process mixture prior for $\boldsymbol{\beta}_i$, which provides a more flexible model than a Dirichlet. Instead of the Dirichlet process, a Polya tree prior is considered in [155]. In this paper, we consider the parametric approach to model the error term ϵ_i with a Normal distribution.

4.3 Bayesian Bi-level Variable Selection in Accelerated Failure Time Model

We propose a Bayesian bi-level variable selection method on the AFT model. This method has two hierarchical levels of variable selection: the first one is group-wise and the second level is element-wise variable selection. It is motivated by natural grouping structures of SNPs, which can be captured by genes, or LD blocks. The dimension of SNP data can be significantly reduced by eliminating irrelevant groups to disease susceptibility. By making

use of the block structure in the model frame, we can efficiently select a small number of SNPs associated with clinical outcomes, such as time to event. With predefined G blocks we can write our model as follows.

$$\log Y_i = \boldsymbol{x}'_{i,0}\boldsymbol{\beta}_0 + \sum_{g=1}^{G} \gamma_g \boldsymbol{x}'_{i,g}\boldsymbol{\beta}_g + \epsilon_i, i = 1, \cdots, n, \tag{4.3.1}$$

where $\boldsymbol{x}_{i,0} = (1, x^0_{i,1}, \cdots, x^0_{i,p_0-1})$, $\boldsymbol{\beta}_0 = (\beta_{0,0}, \beta_{0,1}, \cdots, \beta_{0,p_0-1})$. For each g-th group of variables, $\boldsymbol{x}_{i,g} = (x^g_{i,1}, x^g_{i,2}, \cdots, x^g_{i,k_g})$, $\boldsymbol{\beta}_1 = (\beta_{1,1}, \beta_{1,2}, \cdots, \beta_{1,k_g})$. Denote $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_G)$, $\boldsymbol{\gamma}_=(\gamma_1, \cdots, \gamma_G)$, where γ_g is an indicator variable having 0 or 1. When $\gamma_g = 1$, the g-th set of variables will be included in the model. If $\gamma_g = 0$, we remove the g-th group in the model construction. The covariates $x^0_{i,1}, \cdots, x^0_{i,p_0-1}$ are included in the model to address their effects on the time to event. They can be some clinical/demographic characteristics of subjects. The error term ϵ_i are i.i.d. $N(0, \sigma^2)$ so that the failure time Y_i follows a log-Normal distribution. When y_i is possibly right censored, we only observe $t_i = \min(y_i, c_i)$ and $\nu_i = I\{y_i < c_i\}$, where c_i is the censoring time. Here $w_i = \log(y_i)$ can be considered as the augmented data such that

$$\begin{aligned} w_i &= \log(t_i) \text{ if } \nu_i = 1, \\ w_i &> \log(t_i) \text{ if } \nu_i = 0. \end{aligned} \tag{4.3.2}$$

Our bi-level variable selection method addresses issues that the model (4.3.1) has: the selection of the relevant groups of SNPs and the imputation of the censored time to event y_i. In the first step, we identify important groups of variables by only updating the group inclusion vector $\boldsymbol{\gamma}$ and the censored time y_i from their posterior distributions. In the second step, the model (4.3.1) can be reduced by

$$\log Y_i = \boldsymbol{x}'_{i,0}\boldsymbol{\beta}_0 + \sum_{g=1}^{Q} \gamma_g \boldsymbol{x}^{*}_{\boldsymbol{i,g}}{}'\boldsymbol{\theta}_g + \epsilon_i, i = 1, \cdots, n, \tag{4.3.3}$$

where $\boldsymbol{x}^{*}_{\boldsymbol{i,g}}{}'$, $g = 1, 2, \cdots, Q$ are the Q selected groups in the first step, and $\boldsymbol{\theta}_g, g = 1, 2, \cdots, Q$ are the corresponding regression coefficient vectors. The censored time to event y_i is imputed

by the mean of the posterior samples of w_i collected in the first step. It converts the AFT model to a usual log-linear regression problem. We employ a shrinkage prior on the regression parameters $\boldsymbol{\theta}_g$ to enable the element-wise variable selection within $\boldsymbol{x}^{*}_{\boldsymbol{i,g}}{}', g = 1, 2, \cdots, Q$. We consider a Dirichlet-Laplace prior proposed by Bhattacharya et al. [9] on the regression parameters and extend it to incorporate grouping information. Since the regression parameters $\boldsymbol{\beta}_0,\ \boldsymbol{\beta}$ and the standard deviation σ of the error term are not of interests, the computational burden in the first step can be reduced by integrating out the irrelevant parameters, $\boldsymbol{\beta}_0, \boldsymbol{\beta}, \sigma$ from the full posterior distribution. This kind of strategy has been employed in Sha et al. [138], although their variable selection has been conducted only in an element-wise fashion.

4.3.1 The First Step: Groupwise Variable Selection

In the first step, we consider the following conjugate priors.

$$\begin{aligned}
\boldsymbol{\beta}_0|\sigma^2 &\sim N(0, \sigma^2 h_0 \boldsymbol{I}_{p_0}) \\
\boldsymbol{\beta}_g|\sigma^2 &\sim N(0, c_0\sigma^2 \boldsymbol{\Sigma}_g), g = 1, \cdots, G \\
\sigma^2 &\sim IG(\nu_0/2, \nu_0\sigma_0^2/2) \\
\gamma_j &\sim \text{Bernoulli}(p_j) \\
p_j &\sim \text{Beta}(a, b),
\end{aligned}$$

where $\boldsymbol{X}_0 = [\boldsymbol{x}_{1,0}, \cdots, \boldsymbol{x}_{n,0}]'$, $\boldsymbol{X}_g = [\boldsymbol{x}_{1,g}, \cdots, \boldsymbol{x}_{n,g}]'$, $\boldsymbol{X} = [\boldsymbol{X}_0, \boldsymbol{X}_1, \cdots, \boldsymbol{X}_G]$, $\boldsymbol{\Sigma}_g = (\boldsymbol{X}'_g\boldsymbol{X}_g)^{-1}$ when $k_g \leq n$ and $\boldsymbol{\Sigma}_g = (\boldsymbol{X}'_g\boldsymbol{X}_g + \lambda\boldsymbol{I}_{k_g})^{-1}$ when $k_g > n$ for the g-th group with size k_g. The priors on $\boldsymbol{\beta}_g$ are Information Matrix (IM) or Information Matrix Ridge (IMR) priors proposed by Gupta and Ibrahim [64]. They are a generalization of Zellners g-prior [168], while the IM prior is equal to the Zellners g-prior in the Gaussian linear regression setting. The

full posterior distribution of $(\boldsymbol{\beta}_0, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2)$ is given by

$$
\begin{aligned}
L(\boldsymbol{\beta}_0, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2 | \boldsymbol{w}, \boldsymbol{X}) &\propto L(\boldsymbol{w}|\boldsymbol{X}, \boldsymbol{\beta}_0, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}) \pi(\boldsymbol{\beta}_0|\sigma^2) \pi(\boldsymbol{\beta}|\sigma^2) \pi(\boldsymbol{\gamma}) \pi(\sigma^2) \\
&\propto (\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(w_i - \boldsymbol{x}'_{i,0}\boldsymbol{\beta}_0 - \sum_{g=1}^{G} \gamma_g \boldsymbol{x}'_{i,g} \boldsymbol{\beta}_g \right)^2 \right\} \times \exp\left\{ -\frac{1}{2h_0\sigma^2} \boldsymbol{\beta}'_0 \boldsymbol{\beta}_0 \right\} \\
&\quad \times \prod_{g=1}^{G} \exp\left\{ -\frac{1}{2c_0\sigma^2} \boldsymbol{\beta}'_g \boldsymbol{\Sigma}_g^{-1} \boldsymbol{\beta}_g \right\} \times (\sigma^2)^{-\nu_0/2-1} \exp\left(-\frac{\nu_0 \sigma_0^2}{2\sigma^2} \right) \\
&\quad \times \prod_{g=1}^{G} p_g^{\gamma_g} (1-p_g)^{1-\gamma_g} \times \prod_{g=1}^{G} \frac{1}{B(a,b)} p_g^{a-1} (1-p_g)^{b-1}.
\end{aligned}
$$

By integrating out $\boldsymbol{\beta}_0, \boldsymbol{\beta}, \sigma^2$, we can obtain the posterior distribution of $\boldsymbol{\gamma}$:

$$
L(\boldsymbol{\gamma}|\boldsymbol{w}, \boldsymbol{X}) \propto \left\{ \nu_0 \sigma_0^2 + \boldsymbol{w}' \left(\boldsymbol{I} + h_0 \boldsymbol{X}_0 \boldsymbol{X}'_0 + c_0 \sum_{g=1}^{G} \gamma_g \boldsymbol{X}_g \boldsymbol{\Sigma}_g \boldsymbol{X}'_g \right)^{-1} \boldsymbol{w} \right\}^{-\frac{n+\nu_0}{2}} \times \prod_{g=1}^{G} p_g^{\gamma_g} (1-p_g)^{1-\gamma_g}.
$$

When $\boldsymbol{\gamma}_{(g)} = (\gamma_1, \cdots, \gamma_{g-1}, \gamma_{g+1}, \cdots, \gamma_G)$ is given, the posterior distribution of γ_g is the Bernoulli distribution with success probability $\frac{A}{A+B}$, where $A_{\gamma_{(g)}} = \boldsymbol{I} + h_0 \boldsymbol{X}_0 \boldsymbol{X}'_0 + c_0 \sum_{k \neq g}^{G} \gamma_k \boldsymbol{X}_k \boldsymbol{\Sigma}_k \boldsymbol{X}'_k$ and

$$
\begin{aligned}
A &= f_t(\boldsymbol{w}|\nu_0, \sigma_0 (A_{\gamma_{(g)}} + c_0 \gamma_g \boldsymbol{X}_g \boldsymbol{\Sigma}_g \boldsymbol{X}'_g)) \times p_g, \\
B &= f_t(\boldsymbol{w}|\nu_0, \sigma_0 A_{\gamma_{(g)}}) \times (1 - p_g),
\end{aligned}
$$

Then update p_g from its posterior distribution Beta$(a+\gamma_g, b+1-\gamma_g)$. The marginal likelihood of the augmented data $\boldsymbol{w}$ can be derived as

$$
L(\boldsymbol{w}|\boldsymbol{X}, \boldsymbol{\gamma}) \propto \left\{ 1 + \frac{1}{\nu_0 \sigma_0^2} \boldsymbol{w}' \left(\boldsymbol{I} + h_0 \boldsymbol{X}_0 \boldsymbol{X}'_0 + c_0 \sum_{g=1}^{G} \gamma_g \boldsymbol{X}_g \boldsymbol{\Sigma}_g \boldsymbol{X}'_g \right)^{-1} \boldsymbol{w} \right\}^{-\frac{\nu_0+n}{2}},
$$

which is proportional to the truncated n-dimensional multivariate t-distribution such that

$$
\boldsymbol{w}|\boldsymbol{X}, \boldsymbol{\gamma} \sim t_n \left[\nu_0, \boldsymbol{0}, \sigma_0^2 \left(\boldsymbol{I} + h_0 \boldsymbol{X}_0 \boldsymbol{X}'_0 + c_0 \sum_{g=1}^{G} \gamma_g \boldsymbol{X}_g \boldsymbol{\Sigma}_g \boldsymbol{X}'_g \right) \right]
$$

with truncation given by (4.3.2). By using the full conditional distribution of w_i for a censored case $\nu_i = 0$, the censored time w_i can be imputed by its posterior mean. Denote

$\boldsymbol{H}_\gamma = \boldsymbol{I} + h_0 \boldsymbol{X}_0 \boldsymbol{X}_0' + c_0 \sum_{g=1}^{G} \gamma_g \boldsymbol{X}_g \boldsymbol{\Sigma}_g \boldsymbol{X}_g'$, where $h_{i,j}$ is a scalar element in i-th row, j-th column of $\boldsymbol{H}_\gamma$ and $\boldsymbol{H}_{(i,j)}$ is the matrix $\boldsymbol{H}_\gamma$ without its i-th row and j-th column, and $\boldsymbol{h}_i^{(i)}$ is the i-th row of $\boldsymbol{H}_\gamma$ without its i-th element. Similarly, let $\boldsymbol{w}^{(i)}$ be the vector $\boldsymbol{w}$ without its i-th element. When w_i is censored, its full conditional distribution can be written as a truncated t location-scale distribution such that

$$w_i | \boldsymbol{w}^{(i)}, \boldsymbol{X}, \boldsymbol{\gamma} \sim t_{n+\nu_0-1}(\mu_{w_i}, s_{w_i}), \quad w_i > \log(t_i) \tag{4.3.4}$$

where μ_{w_i}, s_{w_i}, and $n + \nu_0 - 1$ are respectively the location, scale, and shape parameters. The location and scale parameters are give by

$$\begin{aligned} \mu_{w_i} &= \boldsymbol{h}_i^{(i)} \boldsymbol{H}_{(i,j)}^{-1} \boldsymbol{w}_i^{(i)}, \\ s_{w_i} &= \sqrt{\left(h_{(i,i)} - \boldsymbol{h}_i^{(i)} \boldsymbol{H}_{(i,i)}^{-1} \boldsymbol{h}_i^{(i)\prime} \right) \left(\nu_0 \sigma_0^2 + \boldsymbol{w}_i^{(i)} \boldsymbol{H}_{(i,i)}^{-1} \boldsymbol{w}_i^{(i)} \right) / (n + \nu_0 - 1)}. \end{aligned}$$

The censored w_i will be updated from (4.3.4) at each iteration and it will be imputed as their posterior mean in the element-wise selection step.

After running Gibbs sampling with M iterations, posterior inclusion probability can be calculated from the posterior sample of $\boldsymbol{\gamma}$ as their posterior mean,

$$\hat{p}_g = \frac{1}{M} \sum_{m=1}^{M} \gamma_g^{(m)}.$$

The posterior inclusion probability $1 - \hat{p}_g$ can be considered as Bayesian q-values, or estimates of the local false discovery rate (FDR) [117, 147], because they measure the probability of a false positive if the g-th group is "decided" to be included in the model. In order to select important groups, for some threshold p^*, we consider that any group with $\hat{p}_g \geq p^*$ is relevant and will include them in our model. We determine the threshold p^* to control the average Bayesian FDR by following the proposed method by Morris et al. [112].

4.3.2 The Second Step: Element-wise Variable Selection

In this step, we include all the variables of the Q selected groups in the first step and assume shrinkage priors on the regression parameters $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_Q$ to achieve further sparsity in the element-wise level in the reduced model (4.3.3). As a shrinkage prior, Dirichlet-Laplace (DL) prior is assumed and extended to incorporate grouping information. The DL prior has been proposed in Bhattacharya et al. [9] as a novel form of shrinkage prior. Under the normal means setting

$$y_i = \theta_i + \epsilon_i,\ \epsilon_i \sim N(0,1), \quad 1 \leq i \leq p,$$

the true signal θ_i has a DL prior, which has a hierarchical structure such that

$$\theta_j|\psi_j, \phi_j, \tau \sim N(\psi_j\phi_j^2\tau^2), \quad \psi_j \sim Exp(1/2),\ \boldsymbol{\phi} \sim \text{Dir}(a, \cdots, a), \quad \tau \sim \text{Gamma}(pa, 1/2), \tag{4.3.5}$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2, \cdots, \phi_p)$. In order to efficiently control the global shrinkage, they introduced global (τ) and local ($\boldsymbol{\phi}$) scales, where the local scales have a joint structure such that they lie in the $(p-1)$ dimensional simplex. Under the moderate-sized coefficients with sparse signal setting, their simulation study has shown that the DL prior outperforms Lasso, Bayesian Lasso, empirical Bayes median, and point mass prior, while its performance is similar with that of Horseshoe prior.

In our model framework, we have prespecified grouping information. In order to get more flexibility depending on the grouping structure, we allow the hyperparameters ψ_j, ϕ_j, and τ in (4.3.5) to be group index(g)-dependent. In the selected group g, there are q_g variables and the total number of selected variables in the model (4.3.3) is $q = \sum_{g=q}^{Q} q_g$. Here, we impute $\boldsymbol{w}$ by the posterior mean $\widetilde{\boldsymbol{w}}$ obtained from the group-wise selection step. For $g = 1, 2, \cdots, Q$,

the priors are set to be

$$\begin{aligned}
\boldsymbol{\theta}_g|\sigma^2,\boldsymbol{\psi}_g,\boldsymbol{\phi}_g,\tau_g &\sim N(0,\sigma^2\boldsymbol{\Sigma}_g^*)\\
\sigma^2 &\sim \text{IG}(\nu_0/2,\nu_0\sigma_0^2/2)\\
\psi_{gj} &\sim \text{Exp}(1/2),\ j=1,\cdots,q_g\\
(\phi_{g1},\cdots,\phi_{gq_g}) &\sim \text{Dir}(a_g,\cdots,a_g)\\
\tau_g &\sim \text{gamma}(q_g a_g,1/2)\\
a_g &\sim \text{Discrete uniform from } \frac{1}{q_g} \text{ to } 1/2 \text{ with length } 50, \qquad (4.3.6)
\end{aligned}$$

where $\boldsymbol{\Sigma}_g^* = \text{diag}(\psi_{g1}\phi_{g1}^2\tau_g^2,\cdots,\psi_{gq_g}\phi_{gq_g}^2\tau_g^2)$. Here $\text{IG}(a,b)$ denotes the inverse gamma distribution with shape parameter a and the rate parameter b.

Denote $\boldsymbol{X}_* = [\boldsymbol{X}_1^*,\cdots,\boldsymbol{X}_Q^*]$, $\boldsymbol{\phi} = (\boldsymbol{\phi}_1,\cdots,\boldsymbol{\phi}_Q), \boldsymbol{\psi} = (\boldsymbol{\psi}_1,\cdots,\boldsymbol{\psi}_Q)$, and $\boldsymbol{\tau} = (\tau_1,\cdots,\tau_Q)$. $\boldsymbol{\Sigma}^*$ is a block diagonal matrix with element matrices Q. By combining (4.3.3) and (4.3.6), the posterior distribution can be obtained as

$$\begin{aligned}
&L(\boldsymbol{\beta}_0,\boldsymbol{\theta},\sigma^2,\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\tau}|\widetilde{\boldsymbol{w}},\boldsymbol{X}_*)\\
&\quad\propto L(\widetilde{\boldsymbol{w}}|\boldsymbol{X}_*,\boldsymbol{\beta}_0,\boldsymbol{\theta},\sigma^2)\pi(\boldsymbol{\beta}_0|\sigma^2)\pi(\boldsymbol{\theta}|\sigma^2,\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\tau})\pi(\sigma^2)\pi(\boldsymbol{\phi})\pi(\boldsymbol{\psi})\pi(\boldsymbol{\tau})\\
&\quad\propto (\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\widetilde{w}_i-\boldsymbol{x}_{i0}'\boldsymbol{\beta}_0-\boldsymbol{x}_{i*}'\boldsymbol{\theta})^2\right\}\times\exp\left\{-\frac{1}{2h_0\sigma^2}\boldsymbol{\beta}_0'\boldsymbol{\beta}_0\right\}\\
&\quad\times\prod_{g=1}^{G}\exp\left\{-\frac{1}{2c_0\sigma^2}\boldsymbol{\theta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\right\}\times(\sigma^2)^{-\nu_0/2-1}\exp\left(-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right)\times\prod_{g=1}^{G}p_g^{\gamma_g}(1-p_g)^{1-\gamma_g}\\
&\quad\times\prod_{g=1}^{G}\frac{1}{B(a,b)}p_g^{a-1}(1-p_g)^{b-1}\times\exp\left(-\frac{\sum_{g=1}^{Q}\sum_{j=1}^{q_g}\psi_{gj}}{2}\right)\\
&\quad\times\prod_{g=1}^{Q}\left(\frac{1}{B(\boldsymbol{\phi}_g)}\prod_{j=1}^{q_g}\phi gj^{a_g-1}\right)\times\prod_{g=1}^{Q}\left\{\tau_g^{q_g a_g-1}\exp\left(-\frac{\tau_g}{2}\right)\right\}, \qquad (4.3.7)
\end{aligned}$$

where $B(\boldsymbol{\phi}_g)$ denotes a multivariate Beta function. We propose a Gibbs sampler for posterior computation, which enables parameter estimation and variable selection simultaneously. The Gibbs sampler is computationally efficient and mixes rapidly. We first specify the hyperparameters $h_0,\sigma_0,\nu_0,a_1,\cdots,a_g$ at appropriate values. Starting from the initiation step,

the Gibbs sampler for the model (4.3.3) and (4.3.7) proceeds as follows:

1. Update $\boldsymbol{\beta}_0$ according to its full conditional distribution

$$p(\boldsymbol{\beta}_0|-) \sim N_{p_0}\left(\left(\boldsymbol{X}'_0\boldsymbol{X}_0 + \frac{1}{h_0}\boldsymbol{I}\right)^{-1}\boldsymbol{X}'_0(\tilde{\boldsymbol{w}} - \boldsymbol{X}^*\boldsymbol{\theta}), \sigma^2\left(\boldsymbol{X}'_0\boldsymbol{X}_0 + \frac{1}{h_0}\boldsymbol{I}\right)^{-1}\right).$$

2. Update $\boldsymbol{\theta}_g$ from its full conditional distribution $N_{q_g}(\widetilde{\boldsymbol{\mu}}_g, \widetilde{\boldsymbol{\Sigma}}_g)$, where

$$\begin{aligned}\widetilde{\boldsymbol{\mu}}_g &= (\boldsymbol{X}_g^{*\prime}\boldsymbol{X}_g{}^* + \boldsymbol{\Sigma}^{-1})^{-1}\boldsymbol{X}_{\boldsymbol{g}}^{*\prime}(\tilde{\boldsymbol{w}} - \boldsymbol{X}_0\boldsymbol{\beta}_0 - \boldsymbol{X}^*_{(g)}\boldsymbol{\theta}_{(g)}),\\ \widetilde{\boldsymbol{\Sigma}}_g &= \sigma^2(\boldsymbol{X}_g^{*\prime}\boldsymbol{X}_g^* + \boldsymbol{\Sigma}_g^{-1})^{-1}.\end{aligned}$$

3. Let $N = n + q + p_0 + \nu_0$ and $\boldsymbol{\eta} = \boldsymbol{X}'_0\boldsymbol{\beta}_0 + \boldsymbol{X}'_*\boldsymbol{\theta}$. Update σ^2 from

$$p(\sigma^2|-) \sim IG\left(\frac{N}{2}, \frac{1}{2}\left\{\nu_0\sigma_0^2 + ||\tilde{\boldsymbol{w}} - \boldsymbol{\eta}||^2 + \frac{\boldsymbol{\beta}'_0\boldsymbol{\beta}_0}{h_0} + \boldsymbol{\theta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\right\}\right).$$

4. Independently sample ψ_{gj} from its full conditional distribution

$$p(\psi_{gj}|-) \sim IG\left(\frac{\phi_{gj}\tau_g\sigma}{|\theta_{gj}|}, 1\right).$$

5. Update τ_g from its full conditional distribution

$$p(\tau_g|-) \sim \text{giG}\left(q_g \times a_g - q_g, 1, 2\sum_{j=1}^{q_g}\frac{|\theta_{gj}|}{\phi_{gj}\sigma}\right).$$

6. Update ϕ_{gj}, where $\phi_{gj} = T_{gj}/T_g$ such that

$$p(T_{gj}|-) \sim \text{giG}\left(a_g - 1, 1, 2\frac{|\theta_{gj}|}{\sigma}\right).$$

7. Update a_g from $\text{MN}(1, \tilde{p}_1/\tilde{p}, \cdots, \tilde{p}_{50}/\tilde{p})$, where $\tilde{p} = \sum_{l=1}^{50}\tilde{p}_l$ and

$$\tilde{p}_l = \exp\left((u_l - 1)\sum_{j=1}^{q_g}\log(\phi_{gj}) + (q_g u_l - 1)\log(\tau_g) - \log 50\right).$$

Because the DL prior does not give exactly zero coefficient value, an additional step is needed to select relevant variables. We follow a simple approach proposed by Bhattacharya et al. [9] to choose important variable by using k-means clustering. There can exist two clusters of $|\theta_j|$s, where (1) one cluster has nearly zero coefficient values while (2) the another cluster has relatively bigger absolute coefficients away from zero. The clusters (1) and (2) can be considered as noise and signal, respectively. We cluster $|\theta_j|$s at each MCMC iteration using k-means with $k = 2$ clusters. At each i-th iteration, the number of important variables h_i is set to be the smaller cluster size out of the two clusters. Then the number of important variables is finally estimated by taking the mode from the whole MCMC iterations, i.e., H = mode$\{h_i\}$. The H largest elements of the absolute values of posterior medians $|\boldsymbol{\theta}|$ are identified as the important variables.

4.4 Simulation Study

We conduct simulation studies to examine the performance of the Bayesian bi-level variable selection in the AFT model. Without censoring, the AFT model is the log-normal regression model, we compare the performance of groupwise and element- wise variable selection with other variable selection methods implemented based on the regression models.

4.4.1 Setup

In order to convey the correlation structure of SNP data in practice, our SNP data is simulated from the Hapmap projects 2009 phase III data [77]. For each subject, we randomly combined two haplotypes from the CEPH population to form its genotypes and used PLINK [130] to form SNP-sets by determining LD blocks. Among blocks whose size are larger than 30, we randomly selected 2000 SNP-sets in each block, which results in about 86,000 total number of SNPs. Since the SNP data has duplicated columns, we removed those SNPs. The

final total number of SNPs is about 45,000. Then the time to event outcome was generated from the model (4.3.1), where $\gamma_j = 1, j = 1, \cdots, 10$ and $\gamma'_j = 1, j' = 11, \cdots, 2000$. Within the 10 relevant blocks, we randomly selected 10 SNPs and assumed an additive model. The additive model assumes that there is a uniform, linear increase in risk for each copy of the minor allele. The corresponding non-zero regression coefficients were generated from $N(-1, 0.5)$, which mimics the situation in which a single copy of the minor allele decreases the time to event in relation to major allele. The outcome variables were generated by the model (4.3.1), where censored event times were independently generated from a uniform distribution from 0 to c^*. The value of c^* was set to achieve a desired censoring rate, here 30%. We replicated the simulation 50 times under the same setting. We assumed the inclusion indicator $\gamma_g \sim \text{Beta}(10, 190)$ which gives average 5% of inclusion probability to reflect prior information that the important signal is sparse in the GWAS.

4.4.2 Simulation Results

A. Non-censored case

After running the group-level selection, our method perfectly selected the 10 relevant blocks where most of the posterior inclusion probabilities were 1. But there were 2 simulation data sets that have identified 11 important groups including the important 10 groups. In order to compare the performance of our group selection step with other group selection methods [72], we ran the same simulation studies with non-censored responses, since all the group selection methods are based on regression settings with fully observed responses. We ran regression models with log-transformed time to event with group penalties such as the group Lasso (grLasso)[167], the group MCP (grMCP)[169], the group bridge (gBridge)[71], the group exponential lasso (gel) [13] , the composite MCP (cMCP)[13]. The cMCP, gel, and gBridge penalties carry out bi-level selection, meaning that they carry out variable selection at the group level and at the level of individual covariates (i.e., they select important groups

Table 4.1: When the group-level variable selection performs perfectly, TP=10, FP=0, TPR=TNR=PPV=NPV=1.

	TP	FP	TPR	TNR	PPV	NPV
BBVS	10.00 (0.00)	0.04 (0.03)	1.000 (0.000)	1.000 (0.000)	0.996 (0.003)	1.000 (0.000)
gBridge	9.86 (0.05)	0.00 (0.00)	0.986 (0.005)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
gel	9.80 (0.07)	0.00 (0.00)	0.980 (0.007)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
grMCP	9.98 (0.02)	0.34 (0.12)	0.998 (0.002)	1.000 (0.000)	0.972 (0.009)	1.000 (0.000)
grSCAD	10.00 (0.00)	11.46 (0.37)	1.000 (0.000)	1.000 (0.000)	0.472 (0.007)	1.000 (0.000)
grLASSO	10.00 (0.00)	19.20 (0.49)	1.000 (0.000)	0.990 (0.000)	0.348 (0.007)	1.000 (0.000)
cMCP	10.00 (0.00)	73.06 (4.30)	1.000 (0.000)	0.963 (0.002)	0.144 (0.011)	1.000 (0.000)

as well as important members of those groups). The grLasso, grMCP, and grSCAD penalties carry out group selection, meaning that within a group, coefficients will either all be zero or all nonzero. We use BIC to select the tuning parameter value for each method.

We consider the following performance measurements: true positive rate (TPR or sensitivity), true negative rate (TNR or specificity), positive predictive value (PPV), and negative predictive value (NPV). They are defined as follows.

$$TPR = \frac{TP}{10},\ TNR = \frac{TN}{1990},\ PPV = \frac{TP}{TP+FP},\ NPV = \frac{TN}{TN+FN},$$

where the TP and TN are the number of correctly identified significant variables and the number of correctly rejected non-significant variables respectively. The FP and FN denote the number of identified non-significant variables and the number of rejected significant variables respectively. Under the true model, $TP = 10$, $TN = 1990$, and $FP = FN = 0$, which implies that all the four rates are equal to one.

Table 4.1 shows the group level variable selection results. Our method achieves the highest values of all the criteria, TPR, TNR, NPV, and PPV compared with other group penalty methods by not only removing the irrelevant groups consistently but also selecting important groups very well. Since the group penalties with only group-level selection especially grSCAD, grLasso tend to select groups more generously, they select important groups perfectly while the numbers of true positive cases are much bigger than other methods. The

Table 4.2: When the element-wise variable selection performs perfectly, TPR=TNR= PPV=NPV=1.

	TPR	TNR	PPV	NPV
BBVS	0.686 (0.012)	0.999 (0.000)	0.616 (0.009)	1.000 (0.000)
gBridge	0.643 (0.011)	0.999 (0.000)	0.503 (0.008)	1.000 (0.000)
gel	0.651 (0.012)	0.999 (0.000)	0.441 (0.009)	1.000 (0.000)
grMCP	0.306 (0.009)	0.998 (0.000)	0.165 (0.009)	0.999 (0.000)

bi-level selection penalties, gBridge and gel show comparative performance to our proposed method.

Table 4.2 shows the element-wise variable selection results. Our method yields the highest values of all the criteria, TPR, TNR, NPV, and PPV compared with other group penalty methods enabling bi-level selection. Since the important signals are very sparse, all the bi-level methods perform very well in terms of removing irrelevant signals. The BBVS shows satisfactory performance in terms of selecting important variables, as well.

B. Censored case

Table 4.3: Group-level selection results. When the group-level variable selection performs perfectly, TP=10, FP=0, TPR=TNR=PPV=NPV=1.

	TP	FP	TPR	TNR	PPV	NPV
BBVS	9.70(0.09)	0.00(0.00)	0.970(0.009)	1.000(0.000)	1.000(0.000)	1.000(0.000)

Table 4.4: Element-wise selection results.When the element-wise variable selection performs perfectly, TPR=TNR=PPV=NPV=1.

	TPR	TNR	PPV	NPV
BBVS	0.634(0.012)	0.999(0.000)	0.589(0.011)	1.000(0.000)

Tables 4.3 and 4.4 show the group-level and element-wise variable selection results. Compared with non-censored cases, the performance of BBVS is satisfactory in the censored case as well.

4.5 ADNI-1 Data Analysis

We applied BBVS on the ADNI-1 data. In order to reveal SNPs associated with the time of conversion to AD, we ran our model using clinical and SNP covariates. The variable selection was only performed on the SNP data. We performed a stringent quality control (QC) step on the raw genotype data to ensure that only high-quality data were included in the final analysis. QC procedures include (i) call rate check per subject and per SNP marker, (ii) gender check, (iii) sibling pair identification, (iv) the Hardy-Weinberg equilibrium test, (v) marker removal by the minor allele frequency, and (vi) population stratification. The second line preprocessing steps include removal of SNPs with (i) more than 5% missing values, (ii) minor allele frequency smaller than 5%, and (iii) Hardy-Weinberg equilibrium p-value $< 10^{-6}$. Remaining missing genotype variables were imputed as the modal value. After the QC procedures, 347 subjects and 494,564 SNPs remained in the current study. The above procedures were carried out in PLINK version 1.9 with visualization performed in R (http://www.r-project.org/) using the qqman package (http://cran.r-project.org/web/packages/qqman/). We also calculated the LD blocks to form the SNP-sets and remove SNP-sets with a single SNP. Eventually, 421,823 SNPs were left in our analysis grouped into 16,084 SNP-sets. Other than the whole SNPs data, we also included gender, age, and the first 5 principle components calculated by PLINK into the analysis. The 5 principle components would adjust for population stratification in the model[129].

We determined the threshold α to control the average Bayesian FDR proposed by Morris et al. [112] and consider any group whose posterior inclusion probability is greater than α. In the ADNI data, the threshold is calculated by 0.941 (Figure 4.4). In total, 19 SNP sets were detected as important groups and 106 SNPs were identified by the elementwise-level selection. Figure 4.2 shows the estimated coefficient values for 795 SNPs included in the 19 SNP-sets. Figure 4.3 shows trace and ACF plots of the regression coefficients of the first selected SNP-set $\theta_{1,1}, \theta_{1,2}$, and $\theta_{1,3}$ are respectively plotted in panels (a), (b), and (c) for 5000

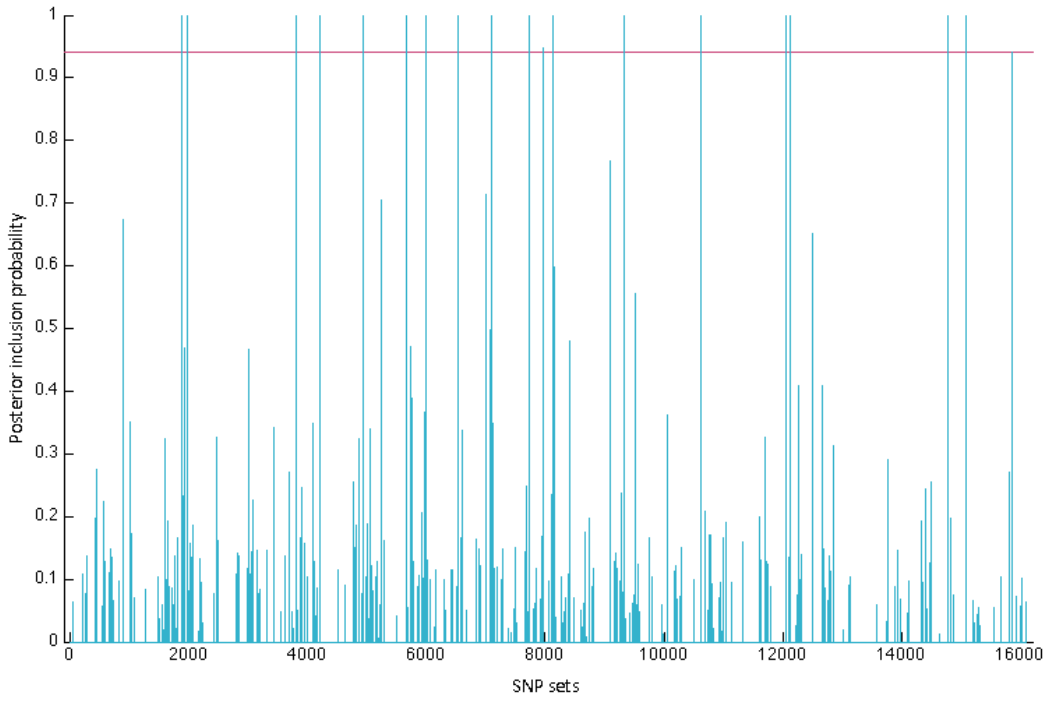


Figure 4.1: Posterior inclusion probabilities of 16,106 SNP-sets. Our proposed method identified 19 important SNP-sets after Bayesian FDR correction. The solid line shows the FDR criteria, 0.941 in this data.

iterations of the MCMC algorithm. They show fast convergence of the algorithm, indicating its good mixing properties. We highlighted 106 selected SNPs in the elementwise-selection step. We summarized the variable selection results of BBVS to present which genes are involved in Table 4.5.

Among them, 4 genes have been reported in other studies to be related to AD directly or indirectly. Dipeptidyl-Peptidase 10 (DPP10) is known to modulate the electrophysiological properties, cell-surface expression and subcellular localization of voltage-gated potassium channels [8]. Chen et al. [22] demonstrated that aggregation of DPP10 was related to neurodegenerative disorders including Alzheimer's, diffuse Lewy body disease and fronto-temporal dementia. Also DPP10 had robust reactivity within neurofibrillary tangles and plaque-associated dystrophic neurites in AD brains, which suggested that it is involved in the

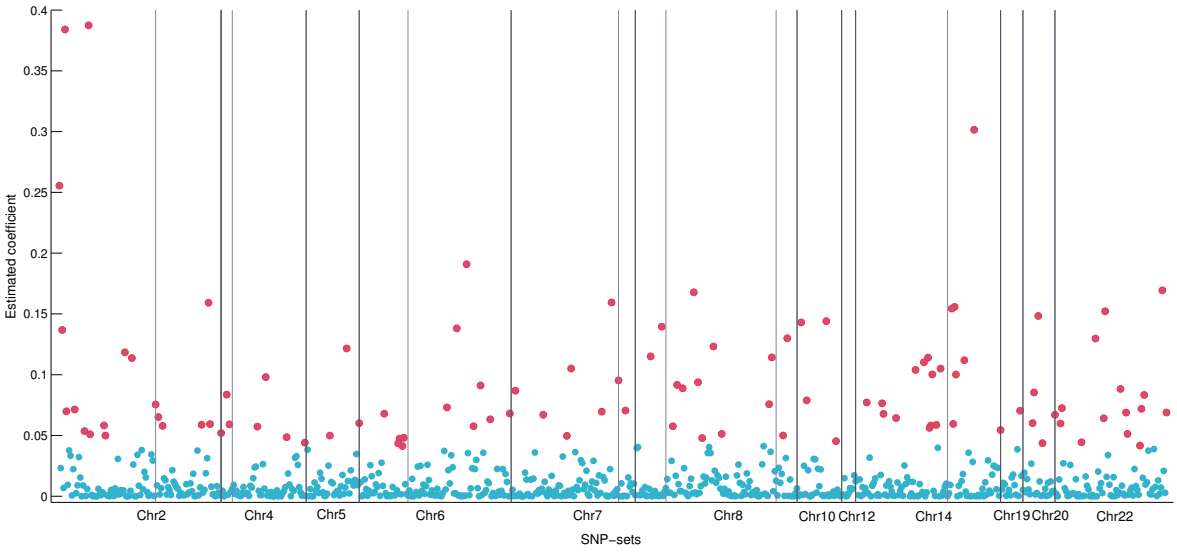
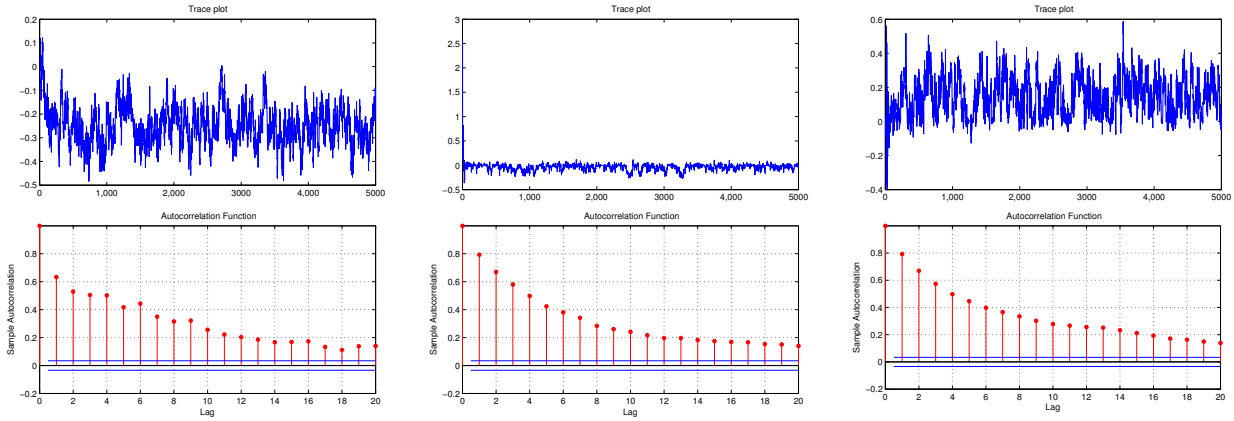


Figure 4.2: Estimated coefficient values for 795 SNPs included in the 19 SNP-sets. We highlighted 106 selected SNPs in the elementwise-selection step.



(a) $\theta_{1,1}$ (b) $\theta_{1,2}$ (c) $\theta_{1,3}$

Figure 4.3: Trace and ACF plots of the regression coefficients of the first selected SNP-set $\theta_{1,1}$, $\theta_{1,2}$, and $\theta_{1,3}$ are respectively plotted in panels (a), (b), and (c) for 5000 iterations of the MCMC algorithm. The trace plots show fast convergence of the algorithm, indicating its good mixing properties.

pathology of AD [23]. Furthermore, its mutations have been associated with asthma[107, 162] and autism spectrum disorders (ASD) [55], while adult asthma is a potential risk factor of developing AD [122] and ASD patients might have better progression in age-related cognitive decline and dementia. All the findings indicate that DPP10 is associated with a risk to

Table 4.5: Bilevel selection results on ADNI-1 data: It shows the list of selected SNP-sets associated with time to conversion from MCI to AD. For each selected group, the corresponding bp ranges, the number of SNPs, and gene names are shown.

Chr	Begin bp	End bp	#(SNPs)	#(Selected)	Genes
2	50596	50665	70	13	**DPP10**
2	53530	53576	47	6	**THSD7B**
4	104778	104785	8	2	ATP8A1
4	117728	117780	53	4	**FREM3**, LOC101927636, GYPA
5	135218	135255	38	3	
6	154825	154859	35	5	
6	166701	166774	74	7	
7	181879	181955	77	7	**SDK1**
7	197664	197675	12	1	
8	216741	216762	22	2	PREX2
8	224172	224250	79	11	TNFRSF11B, **COLEC10**
8	228413	228427	15	2	**TRAPPC9**
10	261007	261038	32	4	SRGN, **VPS26A**, SUPV3L1, HKDC1
12	294754	294763	10	0	CLEC2A , KLRF2
14	332351	332416	66	12	**HEATR5A**, DTD2, **NUBPL**
14	334919	334956	38	7	
19	394149	394164	16	1	CPAMD8, HAUS8, MYO9B
20	399491	399513	23	5	
22	22468984	22671741	80	14	VPREB1, **BMS1P20**

develop AD in a direct/indirect manner. THSD7B has been reported to be related with age-related cognitive decline based on repeated measures of 17 cognitive tests in the Religious Orders Study by [35]. Also, several linkage mapping have identified VPS26A is associated AD [61]. Sidekick Cell Adhesion Molecule 1 (SDK1) was reported as a susceptibility gene for hypertension in Japanese individuals [118], where hypertension moderately increased risk of AD [145].

For comparison purposes, we conducted two different types of GWASs: (1) a simple GWAS, multiple testing on each SNP and (2) kernel-machine SNP set GWAS proposed by Lin et al. [98]. Figure 4.4 shows a manhattan plot with -log 10(p-value) for the simple GWAS. It only identified 6 SNPs at the 1×10^{-5} suggestive significance level, where none of them had been reported in previous GWASs. For the kernel-machine method, we used LD blocks

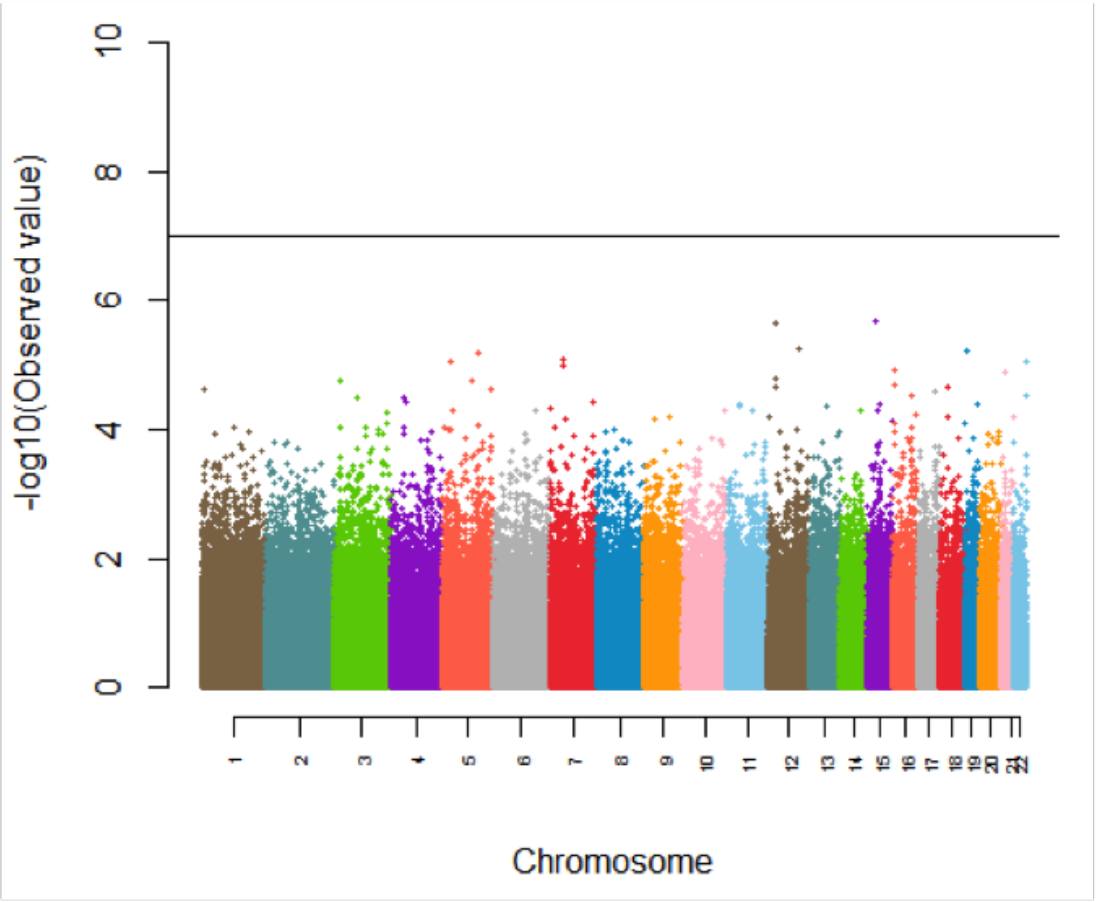


Figure 4.4: A manhattan plot with -log 10(p-value) for the simple GWAS. The solid line shows the 5×10^{-8} significance level.

information and it identified two genes CAMTA1 and RBFOX1 at the 5×10^{-8} significance level. CAMTA1 gene was detected under the linear kernel, where Huentelman et al. [73] identified SNPs within the CAMTA1 gene that were significantly related to memory performance and memory-related regions on human brain, which could be considered as potential biomarkers of AD. RBFOX1 was identified under the quadratic kernel function. Hooli et al. [69] reported that the gene co-segregates with disease status within Early-onset familial AD (EO-FAD) and early/ mixed-onset AD families. While the kernel-machine method uses the natural grouping information, it is a basically multiple testing method discarding possible correlation among the groups.

4.5.1 Sensitivity Analysis

We explored how varying hyperparameters of the IM/IMR priors affects posterior estimation. Since the IM/IMR priors were only used for the group level selection, we focused on how the group selection results were varied as changing the values of hyperparameters in the IM/IMR priors. There are two hyperparameteres involved in the IM/IMR priors: c_0 and λ, which are respectively dispersion and 'ridge' parameters. The 'ridge' parameter is usually used to deal with high-dimensional and/or collinear covariates in regression settings. We tested $0.001, 0.05, 0.1, 0.3, 0.6$, and 1.0 for the parameter λ while fixing $c_0 = 1$ and tried $0.5, 1, 2$, and 4 for the parameter c_0 with a fixed λ value as .1.

Tables 4.6 and 4.7 show that group selection results are quite stable as the hyper parameters vary. Therefore, we can conclude that the proposed priors can yield stable posterior estimates in terms of group-level selection.

4.6 Discussion

The BBVS was developed to enable bi-level variable selection as incorporating grouping information within covariates in the high-dimensional setting. In the context of GWAS, our method addressed the challenging issues by making use of natural grouping information of SNPs in the group-level variable selection step. In addition, DL priors were adapted to reflect the grouping information in the element-wise variable selection. The simulation studies showed that our proposed method outperformed other bi-level and group-level variable selection methods in the GWAS setting. We applied BBVS on the ADNI-1 data to identify relevant SNP sets associated with the time to develop AD within MCI patients. We identified 106 informative SNPs located within 10 genes, where 4 genes were directly and indirectly related to AD, while the simple form of GWAS only detected 3 SNPs that had not been reported in the literature. As a future work, we need to analyze other AD data sets to

Table 4.6: Sensitivity analysis of λ for ADNI-1 data with different values of hyperparameters in the IM/IMR priors within the group-level selection.

	λ					
	0.001	0.050	0.100	0.300	0.600	1.000
1876		•	•			•
1974	•	•	•	•	•	•
2605	•			•	•	
3656	•					
3814	•	•	•	•	•	•
4213	•	•	•	•	•	•
4928	•	•	•	•	•	•
5207				•		
5669	•	•	•	•	•	•
5990	•	•	•		•	
6527		•	•			•
7088	•	•	•	•	•	•
7154		•				
7725	•					
7653				•	•	•
7742	•	•	•	•	•	•
7945	•					
7974		•	•			
8137	•	•	•	•	•	•
8403				•	•	•
8458				•		
9320	•	•	•	•	•	•
9937				•		
10618	•	•	•	•	•	•
10789					•	
12050	•	•	•	•	•	•
12133	•	•	•	•	•	•
12673				•		
14785	•	•	•	•	•	•
15093	•	•	•	•	•	•
15865			•			•

Table 4.7: Sensitivity analysis of c_0 for ADNI-1 data with different values of hyperparameters in the IM/IMR priors within the group-level selection.

	c_0			
	0.5	1	2	4
1142		•		
1647			•	
1741	•			
1876	•	•		
1922	•			
1974	•	•	•	•
3106			•	
3814	•	•	•	•
4213	•	•	•	•
4214	•	•	•	•
4928	•			
5669	•	•	•	•
5990		•		
6527		•		
7088	•	•	•	•
7653	•			
7742	•	•	•	•
7974		•		
8137	•	•	•	•
8458	•			
8964	•			
9320	•	•	•	•
10618	•	•	•	•
12050	•	•	•	•
12133	•	•	•	•
12673	•			
14785	•	•	•	•
15093	•	•	•	•
15865		•		

see if the implicated genes are reproducible when we used different subjects.

Our BBVS method was established in the AFT model. Since the AFT model is a (log) linear regression model without any censoring, the method can be directly used for any continuous outcomes, such as brain MRI volumes, behavior score for AD patients, and so on. As an extension, our method can be adapted in a logistic regression model with a binary response variable. It is often of interest to detect SNPs that are associated with

hippocampal morphology. In this case, we need to use functional regression models with functional response.

CHAPTER 5

BAYESIAN HIERARCHICAL GROUP SPECTRAL CLUSTERING

5.1 Introduction

The Human brain has been analyzed from a network perspective with the advent of neuroimaging acquisition techniques and network theory. Functional magnetic resonance imaging (fMRI) is a non-invasive neuroimaging procedure to assess brain neuronal activity that can be measured by changes in blood oxygen level-dependent (BOLD) signal [101]. In particular, resting state fMRI (rsfMRI) observes regional brain activity when a subject is not performing any explicit task. The rsfMRI is useful for understanding how different brain regions interact with each other, which can be called functional connectivity. Functional connectivity is the connectivity between different brain areas sharing information and functions [153]. It can be estimated by calculating pairwise temporal correlation between two spatially remote BOLD signals in rsfMRI. It can be an important biomarker in psychiatric disorders because its abnormality has been observed in subjects with brain disorders including AD, schizophrenia, and ADHD [58].

Unraveling a missing link between neurological disorders and brain network is an ongoing quest in various fields including statistics, epidemiology, and neuroscience. One specific question will be "is there any relationship between altered functional connectivity and neurological disorders?". There are two main standard methods used to tackle the question: univariate approaches to see if each correlation (node) has group difference using Fisher's z-transformation, and graph theoretical approaches based on some summary statistic for a network structure (i.e., girth, diameter, modularity, small-worldness). The first method

easily faces a high-dimensional multiple testing problem. If there is a small group difference in each correlation, the univariate approach is likely to miss the signal, which results in low power. Also, it discards a spatial structure among close or related brain regions (voxels). The second methods often fail to detect local differences among subject groups, because the connectivity structures are too simplified by the summary statistics. Thus, alternative connectivity analysis methods are critically needed.

In this study, we aim to address the following four issues in connectivity studies: (1) high-dimensionality of connectivity matrices, (2) detection of local differences in brain, but at the same time, (3) need to utilize an underlying relational structure among brain areas. Regarding (3), while the correlation between two marginal regions has weak group difference, two linear combinations of some brain regions can have strongly different connection across groups. Here we call the linear combinations of brain regions "underlying factors" or "eigenmaps". We can identify more significant group differences among underlying factors than among individual brain regions. The last thing to be considered is (4) heterogeneity of functional connectivity. FMRI itself is noisy in terms of spatial resolution and head motion, which leads estimated functional connectivity to be varied across subjects. Also, functional connectivity has individual and group variation, as Mueller et al. [113] suggests that functional connectivity is individually heterogeneous across the cortex due to the degree of evolutionary cortical expansion. The heterogeneity should be accordingly taken into account to have meaningful signals.

In order to address the four issues, we propose a Bayesian group spectral clustering model by taking a global approach to analyze brain connectivity. We decompose a correlation matrix (possibly it can be any symmetric matrix) with underlying common factors across subjects and the subject-specific coefficient matrix ($\mathbf{\Lambda}_i$). The subject-specific coefficient matrix preserves an individual network structure in the low-dimensional space spanned by the common factors. One more intriguing part is that we incorporate a hierarchical

structure within a prior of $\mathbf{\Lambda}_i$ in order to estimate effects of clinical/demographic covariates. We take a Bayesian approach to estimate the parameters involved in the decomposition, and the regression model with clinical covariates. Our method automatically estimate the effects of covariates on the $\mathbf{\Lambda}_i$ matrices while decomposing the connectivity structure within MCMC iterations. A similar matrix decomposition method has been proposed by Wang et al. [156], while they took a frequentist approach to estimate the parameters of interest. But our method could enable to estimate the regression coefficient during the estimation steps without explicit modelling.

In the next section, we describe the proposed method in detail, and discuss how our method can deal with the previous four issues. In section 5.3, we present simulation results to show validity of the method. We applied our method to Alzheimer's disease data (ADNI) to explore functional connectivity across subjects and see any group difference. The results are presented in section 5.4. In the last section, we discuss how our finding can be utilized in a clinical application perspective.

5.2 Bayesian Hierarchical Group Spectral Clustering

5.2.1 Model Formulation

For the i-th subject, denote a connectivity (or any symmetric data) matrix as $L_i = (L_i(g, g'))_{g,g'=1,\cdots,V}$ with V vertices. For the relatively small number $R << V$, we propose a *group spectral clustering model* as follows:

$$L_i(g, g') = \sum_{r=1}^{R} \sum_{s=1}^{R} \beta_r(g) \lambda_{i,(r,s)} \beta_s(g') + \epsilon_i(g, g'),$$

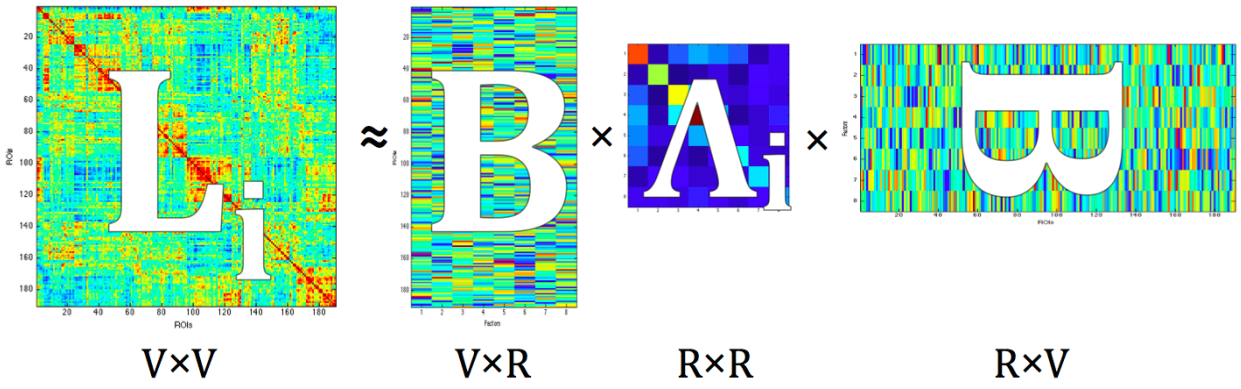
where $\epsilon_i(g, g')$ are measurement errors. Eigenmaps $\beta_s(g)$ are orthonormal underlying common factors (bases) for all subjects, and $\lambda_{i,(r,s)}$ are individual coefficients and can be non-zero even for $r \neq s$. The element $\lambda_{i,(r,s)}$ globally determines the amount of contribution for connection between r-th and s-th factors to the i-th connectivity structure. The subject-specific

coefficient matrix $\boldsymbol{\Lambda_i} = \left[\lambda_{i,(r,s)}\right]_{r,s=1,\cdots,R}$ preserves an individual network structure in the low-dimensional space spanned by the common factors $\boldsymbol{\beta}_r$. To express the proposed model in a matrix form, let's denote

$$\begin{aligned} \boldsymbol{L_i}_{V\times V} &= [L_i(g,g')]_{g,g'=1,\cdots,V} \\ \boldsymbol{B}_{V\times R} &= [\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_R] \\ \boldsymbol{\Lambda_i}_{R\times R} &= \left[\lambda_{i,(r,s)}\right]_{r,s=1,\cdots,R}, \end{aligned}$$

where $R << V$. The group spectral clustering of $\boldsymbol{L}$ can be expressed by

$$\boldsymbol{L}_i = \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}' + \boldsymbol{\epsilon}_i. \tag{5.2.1}$$



Model (5.2.1) has at least four unique features. First, it achieves substantial dimension reduction by reducing from $V(V-1)/2$ to around $VR + R(R+1)/2$ for each subject to deal with high-dimensionality of connectivity matrices. Second, the R eigenmaps $\boldsymbol{\beta}_r = \{\beta_r(g) : g \in \mathcal{G}\}$ can be considered as independent networks of vertices that characterize the latent organization of connectivity structures across subjects at a system level. Then, the subject-specific coefficient matrix $\boldsymbol{\Lambda_i}$ preserves an individual network structure in the low-dimensional space spanned by the eigenmaps. It enables us to construct an underlying relational structure among brain areas and to reduce heterogeneity of functional connectivity

in the low-dimensional space. Third, $\boldsymbol{\eta}_{i,r} = \{\eta_{i,r}(g) = \sum_{s=1}^{R} \beta_s(g)\lambda_{i,(s,r)} : g \in \mathcal{G}\}$ delineate individual organization of connection maps between vertices across R different latent networks. One more intriguing part is that we assume a hierarchical structure within a prior of Λ_i in order to estimate effects of clinical/demographic covariates. Our method facilitates estimation of the effects of covariates on the Λ_i matrices while decomposing the connectivity structure within MCMC iterations. By considering a diagnostic indicator as a covariate, e.g., Alzheimer's disease or cognitively normal, local differences in functional connectivity can be detected. Details will be followed in the next section.

5.2.2 Bayesian Approach

In order to estimate the parameters of interest, we take a Bayesian approach. For parameters of interest, the full posterior distribution is given by

$$\pi(\boldsymbol{\Lambda}, \boldsymbol{B}, \sigma^2, \sigma_0^2 | \boldsymbol{L}; \boldsymbol{X}) \propto p(\boldsymbol{L}|\boldsymbol{\Lambda}, \boldsymbol{B}, \sigma^2)\pi(\boldsymbol{\Lambda}|\boldsymbol{\Gamma}, \sigma_0^2; \boldsymbol{X})\pi(\boldsymbol{\Gamma}|\sigma_\gamma^2)\pi(\boldsymbol{B})\pi(\sigma^2)\pi(\sigma_0^2)\pi(\sigma_\gamma^2).$$

We assume that the measurement errors $\epsilon_i(g, g')$ follow i.i.d. symmetric Normal distribution [137] with the mean 0 and the variance σ^2. For the pre-specified number of factors R, the likelihood of $\boldsymbol{L}_1, \cdots, \boldsymbol{L}_n$ can be written by

$$\begin{aligned}
p(\boldsymbol{L}_1, \cdots, \boldsymbol{L}_n | \boldsymbol{B}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_n) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{nV(V+1)/2} \prod_{i=1}^{n} \exp\left[-\frac{1}{2\sigma^2}\mathrm{tr}\left(\left(\boldsymbol{L}_i - \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}'\right)^2\right)\right] \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{nV(V+1)/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\mathrm{tr}\left(\left(\boldsymbol{L}_i - \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}'\right)^2\right)\right] \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{nV(V+1)/2} \exp\left[-\frac{1}{2\sigma^2}\left\{\mathrm{tr}\left(\sum_{i=1}^{n}\boldsymbol{L}_i^2\right) + \mathrm{tr}\left(\sum_{i=1}^{n}\left(\boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}'\right)^2\right) - 2\mathrm{tr}\left(\sum_{i=1}^{n}\boldsymbol{L}_i\boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}'\right)\right\}\right].
\end{aligned}$$

We assume following priors:

$$\begin{aligned}
\boldsymbol{\beta}_r &\sim N\left(0, \frac{1}{V}\boldsymbol{I}_V\right), \\
\text{i.e., } \pi(\boldsymbol{B}) &\propto \exp\left(-\frac{V}{2}\text{tr}(\boldsymbol{B}'\boldsymbol{B})\right) \\
\pi(\boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_n | \boldsymbol{\Delta}_1, \cdots, \boldsymbol{\Delta}_n, \sigma_0^2) &\propto \prod_{i=1}^{n} \exp\left[-\frac{1}{2\sigma_0^2}\text{tr}\left((\boldsymbol{\Lambda}_i - \boldsymbol{\Delta}_i)^2\right)\right] \\
\pi(\boldsymbol{\Gamma}) &\propto \exp\left[-\frac{1}{2\sigma_\gamma^2}\text{tr}\left(\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\right)\right] \\
\sigma^2 &\sim \text{Inverse-Gamma}(a_0, a_1) \\
\sigma_0^2 &\sim \text{Inverse-Gamma}(b_0, b_1) \\
\sigma_\gamma^2 &\sim \text{Inverse-Gamma}(c_0, c_1),
\end{aligned}$$

where $vech(\boldsymbol{\Delta}_i) = \boldsymbol{\Gamma}'\boldsymbol{x}_i = \left[\boldsymbol{x}_i'\boldsymbol{\gamma}_1, \boldsymbol{x}_i'\boldsymbol{\gamma}_2, \cdots, \boldsymbol{x}_i'\boldsymbol{\gamma}_q\right]'$, $q = R(R+1)/2$. The regression covariates for the i-th subject are denoted by $\boldsymbol{x}_i = (x_0, \cdots, x_{p-1})'$ and the regression coefficients are given by $\boldsymbol{\Gamma} = [\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_q] = [\boldsymbol{\gamma}_j]_{j=1}^q$, where $\boldsymbol{\gamma}_j = (\gamma_{0j}, \cdots, \gamma_{(p-1)j})'$. For $\boldsymbol{\beta}_k$'s, the assumed prior yields approximate orthogonality of the columns in $\boldsymbol{B}$ when the value of R is large.

Let's denote $\boldsymbol{Q} = \boldsymbol{B}'\boldsymbol{B}$, $\boldsymbol{P}_i = \frac{1}{\sigma^2}\boldsymbol{B}'\boldsymbol{L}_i\boldsymbol{B} + \frac{1}{\sigma_\gamma^2}\boldsymbol{\Delta}_i$, and the matrix $\boldsymbol{D}$ is a diagonal matrix with elements

$$(\ \underbrace{1, \sqrt{2}, \cdots, \sqrt{2}}_{R \text{ elements}}, \underbrace{1, \sqrt{2}, \cdots, \sqrt{2}}_{(R-1) \text{ elements}}, \cdots, \underbrace{1}_{1 \text{ element}}\ ).$$

The posterior distribution for each parameter is given by

$$\pi(\boldsymbol{B}|\bullet) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\text{tr}\left((\boldsymbol{L}_i - \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}')^2\right) - \frac{V}{2}\text{tr}(\boldsymbol{B}'\boldsymbol{B})\right], \tag{5.2.2}$$

$$vech(\boldsymbol{\Lambda}_i|\bullet) \sim N\left(\boldsymbol{\mu}_i, \boldsymbol{S}\right), \tag{5.2.3}$$

where

$$\boldsymbol{\mu}_i = \boldsymbol{S}\boldsymbol{D}svec\left(\boldsymbol{P}_i\right), \quad \boldsymbol{S} = \left(\frac{\boldsymbol{D}(\boldsymbol{Q} \otimes_s \boldsymbol{Q})\boldsymbol{D}}{\sigma^2} + \frac{\boldsymbol{D}^2}{\sigma_0^2}\right)^{-1}.$$

The regression parameters are sampled from the following posterior:

$$vec(\boldsymbol{\Gamma}|\bullet) \sim N\left(\boldsymbol{\mu}_\gamma, \boldsymbol{S}_\gamma\right), \tag{5.2.4}$$

where

$$\boldsymbol{\mu}_\gamma = \frac{1}{\sigma_0^2}\boldsymbol{S}_\gamma \sum_{i=1}^{n} vec\left(\boldsymbol{x}_i svec(\boldsymbol{\Lambda}_i)'\boldsymbol{D}\right), \quad \boldsymbol{S}_\gamma = \left(\frac{1}{\sigma_0^2}\sum_{i=1}^{n}\left(\boldsymbol{D}^2 \otimes \boldsymbol{x}_i\boldsymbol{x}_i'\right) + \frac{1}{\sigma_\gamma^2}\boldsymbol{I}\right)^{-1}.$$

The hyper-parameters are sampled based on the following posterior distributions:

$$\begin{aligned}
\sigma^2 &\propto \exp\left\{-\frac{1}{\sigma^2}\left(\frac{1}{2}\mathrm{tr}\left(\sum_{i=1}^{n}(\boldsymbol{L}_i - \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}')^2\right) + a_1\right)\right\}\left(\frac{1}{\sigma^2}\right)^{a_0+\frac{nV(V+1)}{4}-1} \qquad (5.2.5)\\
&\sim IG\left(a_0 + \frac{nV(V+1)}{4}, \frac{1}{2}\mathrm{tr}\left(\sum_{i=1}^{n}(\boldsymbol{L}_i - \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}')^2\right) + a_1\right)\\
&\triangleq IG\left(\widetilde{a}_0, \widetilde{a}_1\right)\\
\sigma_0^2 &\propto \exp\left\{-\frac{1}{\sigma_0^2}\left(\frac{1}{2}\mathrm{tr}\left(\sum_{i=1}^{n}\boldsymbol{\Lambda}_i^2\right) + b_1\right)\right\}\left(\frac{1}{\sigma_0^2}\right)^{b_0+\frac{nR(R+1)}{4}-1} \qquad (5.2.6)\\
&\sim IG\left(b_0 + \frac{nR(R+1)}{4}, \frac{1}{2}\mathrm{tr}\left(\sum_{i=1}^{n}\boldsymbol{\Lambda}_i^2\right) + b_1\right)\\
&\triangleq IG\left(\widetilde{b}_0, \widetilde{b}_1\right)\\
\sigma_\gamma^2 &\propto \exp\left\{-\frac{1}{\sigma_\gamma^2}\left(\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\right) + c_1\right)\right\}\left(\frac{1}{\sigma_\gamma^2}\right)^{c_0+\frac{nR(R+1)}{4}-1} \qquad (5.2.7)\\
&\sim IG\left(c_0 + \frac{pR(R+1)}{4}, \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\right) + c_1\right)\\
&\triangleq IG\left(\widetilde{c}_0, \widetilde{c}_1\right).
\end{aligned}$$

In summary, the posterior sampling proceeds as follows.

1. Employ slice sampling to get a posterior sample from $\pi(\boldsymbol{\beta}_g|\boldsymbol{B}_{(-g)}, \bullet)$ in (5.2.2) with burn-in 100.
2. Update $\boldsymbol{\Lambda}_i$ generated from $N\left(\boldsymbol{\mu}_i, \boldsymbol{S}\right)$ in (5.2.3).
3. Update $\boldsymbol{\Gamma}$ generated from $N\left(\boldsymbol{\mu}_\gamma, \boldsymbol{S}_\gamma\right)$ in (5.2.4).
4. Update σ^2 generated from $IG(\widetilde{a}_0, \widetilde{a}_1)$ in (5.2.5).
5. Update σ_0^2 generated from $IG(\widetilde{b}_0, \widetilde{b}_1)$ in (5.2.6).
6. Update σ_γ^2 generated from $IG(\widetilde{c}_0, \widetilde{c}_1)$ in (5.2.7).

In terms of estimation of $\boldsymbol{B}$ and $\boldsymbol{\Lambda}_i$, we cannot simply use posterior means of their posterior samples, because our model does not assume any constraints for identifiability of $\boldsymbol{B}$ and $\boldsymbol{\Lambda}_i$. The identifiability issue comes from permutations of the factors, and arbitrary matrix rotations, which makes the posterior means meaningless. [110] But in our method, estimation of $\boldsymbol{B}$, $\boldsymbol{\Lambda}_i$, and $\boldsymbol{\Gamma}$ is essential for low-dimensional representation of $\boldsymbol{L}_i$ and test of group difference in $\boldsymbol{L}_i$. In order to get meaningful estimates, we follow simple approaches discussed by Mohamed et al. [110]. First, we estimate $\boldsymbol{B}$ as $\widehat{\boldsymbol{B}}$ such that $\widehat{\boldsymbol{B}} = \underset{\boldsymbol{B}^{(m)}}{\arg\max}\, \pi(\boldsymbol{B}^{(m)}, \bullet | \boldsymbol{L}, ; \boldsymbol{X})$, where m denotes each MCMC iteration. We need variability information of $\boldsymbol{\Gamma}$ to find which brain areas have group difference by using their HPD intervals. After fixing $\boldsymbol{B}$ as $\widehat{\boldsymbol{B}}$, we rerun the above steps and estimate $\boldsymbol{\Lambda}_i$ and $\boldsymbol{\Gamma}$ as their posterior means $\widehat{\boldsymbol{\Lambda}}_i$ and $\widehat{\boldsymbol{\Gamma}}$. They are now meaningful because we add a constraint on $\boldsymbol{B}$ to deal with the permutation and rotation issues.

In order to identify any group difference of functional connectivity, a diagnostic indicator can be included as a covariate, for example, $x_i = 1$ if the i-th subject has AD. Denote the corresponding regression coefficient $\boldsymbol{\Gamma}_1$ such that $vech(\boldsymbol{\Gamma}_1) = (\gamma_{11}, \gamma_{12}, \cdots, \gamma_{1q})'$. Then $\boldsymbol{\Gamma}_1$ quantifies a disease effect on any pair of two underlying factors. In order to find specific brain areas that have group difference, we need to estimate regression coefficients for $\boldsymbol{L}_i$ as a response variable instead of $\boldsymbol{\Lambda}_i$. We can map the estimated coefficient matrix $\widehat{\boldsymbol{\Gamma}}_1$ from the factor space to the original node space by calculating $\widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Gamma}}_1\widehat{\boldsymbol{B}}'$.

The number of factors R is chosen by a Bayesian information criterion (BIC) and the reconstruction error is measured by the ratio of Frobenius norms as follows.

$$\text{error} = \frac{1}{n}\sum_{i=1}^{n} \frac{||\boldsymbol{L}_i - \widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Lambda}}_i\widehat{\boldsymbol{B}}'||_F}{||\boldsymbol{L}_i||_F}. \tag{5.2.8}$$

5.3 Simulation study

In this section, we use simulation studies to illustrate the performance of the proposed Bayesian hierarchical group spectral clustering method. We assume that there are 50 subjects $i = 1, \cdots, 50$ and their connectivity data are simulated from the following underlying model:

$$\boldsymbol{L}_i = \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B} + \boldsymbol{\epsilon}_i,$$

where $\boldsymbol{L}_i$ denotes the network data. We applied our proposed method to recover $\boldsymbol{B}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_n$ and $\boldsymbol{L}_1, \cdots, \boldsymbol{L}_n$. We run 5,000 MCMC iterations with 1,000 burn-in. We repeated the simulation 20 times under the following four different scenarios.

In the section 5.3.1, we examine if the true number of underlying factors is correctly chosen by BIC. We use two different scenarios: (1) there is a common basis $\boldsymbol{B}$ with rank 3, and (2) there are two basis $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ with rank 3 respectively. In order to measure how the raw data matrices are recovered accurately, we calculate the estimation error in 5.2.8. We don't include any covariates $\boldsymbol{x}_i$ in this study. Section 5.3.2 discusses how accurately regression coefficients are estimated by our method. We consider binary (case 1) and continuous (case 2) covariates with an intercept.

5.3.1 Simulation 1

A. Scenario 1

We generate all simulation data sets from $\boldsymbol{L}_i = \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}' + \boldsymbol{\epsilon}_i$. There exist 3 true underlying common eigenmaps ($R = 3$). We assume that $\boldsymbol{B}$ and $\boldsymbol{\Lambda}_i$s are 50×3 and 3×3 matrices, respectively. The eigenmaps, subject-specific coefficient matrices, and measurement errors

Table 5.1: In order to select the number of common factors, we used BIC. TP and FP denote the number of true positive and the number of false positive, respectively.

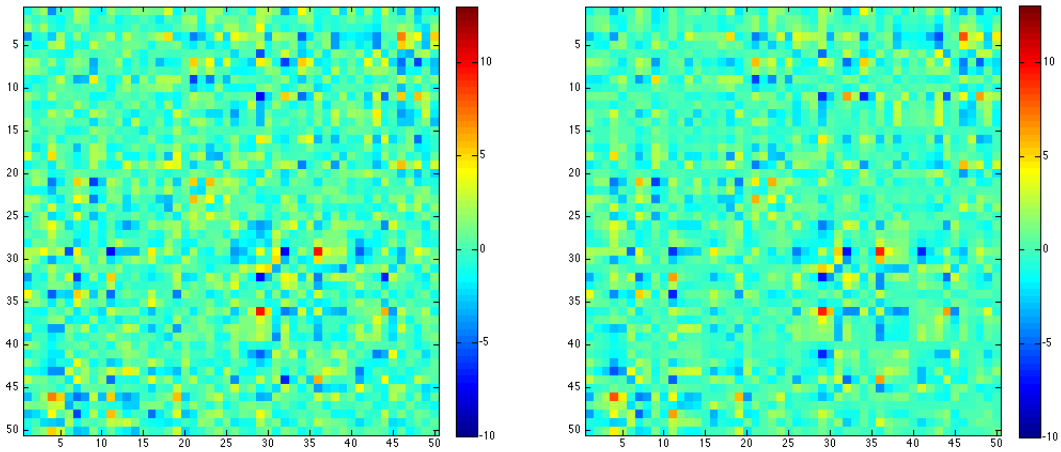
	Error	TP	FP
BIC-selected	0.025 (0.001)	3.00 (0.00)	0.00 (0.00)

are generated according to:

$$\begin{aligned} \boldsymbol{\beta}_k &\sim N(\mathbf{0}, \boldsymbol{I}_V), \text{ i.e., } \pi(\boldsymbol{\beta}_k) = \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{B}'\boldsymbol{B})\right) \\ \pi(\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_n) &= \prod_{i=1}^{n} \exp\left[-\frac{1}{2\sigma^2}\mathrm{tr}\left(\boldsymbol{\epsilon}_i^2\right)\right] \end{aligned}$$

where $\sigma = 1$. Set $\sigma_0 = 1$ and $\boldsymbol{\Lambda}_i$s are generated from

$$\pi(\boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_n) = \prod_{i=1}^{n} \exp\left[-\frac{1}{2\sigma_0^2}\mathrm{tr}\left(\boldsymbol{\Lambda}_i^2\right)\right].$$



(a) Raw structure (b) Approximated structure

Figure 5.1: Figure (a) shows the first raw data matrix $\boldsymbol{L}_1$ in the first simulation data set, while figure (b) shows the approximated matrix by the proposed method.

Table 5.1 shows the simulation results. In order to select the number of common factors, we used BIC. TP and FP denote the number of true positive and the number of false positive, respectively. The BIC selects the correct number of factors consistently. Reconstruction error and Figure 5.1 demonstrate that our decomposition method can approximate the raw matrix well.

Table 5.2: In order to select the number of common factors, we used BIC. TP and FP denote the number of true positive and the number of false positive, respectively.

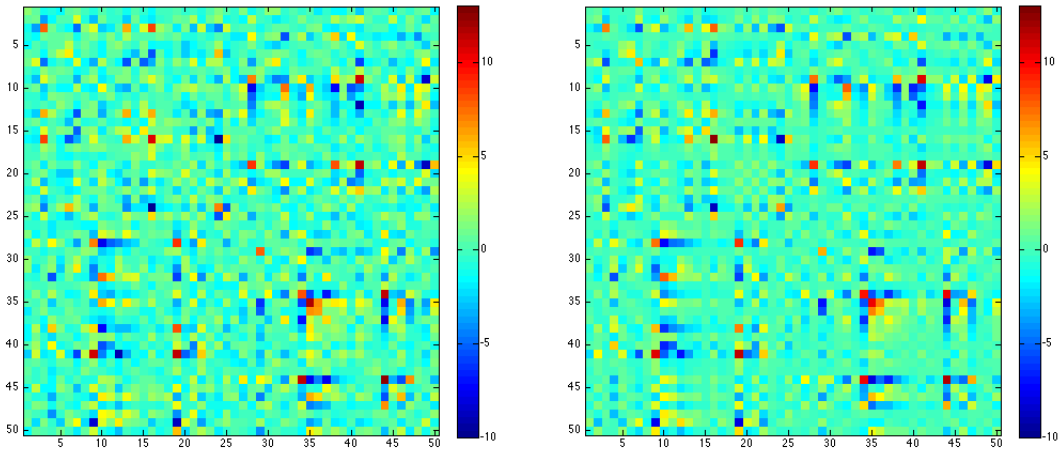
	Error	TP	FP
BIC-selected	0.059 (0.002)	6.00 (0.00)	0.00 (0.00)

B. Scenario 2

We generate other simulation data sets from $\boldsymbol{L}_i^g = \boldsymbol{B}_g \boldsymbol{\Lambda}_i \boldsymbol{B_g}' + \boldsymbol{\epsilon}_i,\ g = 1, \cdots, G,\ i = 1, \cdots, n_g$, where $n = \sum_{g=1}^{G} n_g$. We consider $\sigma = 1$, $R = 3$ and two groups, i.e., $G = 2$. $\boldsymbol{B}_g = [\boldsymbol{\beta}_1^g, \cdots, \boldsymbol{\beta}_R^g]$ and $\boldsymbol{\Lambda}_i$s are generated from

$$\begin{aligned} \boldsymbol{\beta}_k^g &\sim N(\mathbf{0}, \boldsymbol{I}_V), \text{ i.e., } k = 1, \cdots, R \\ \pi(\boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_n) &= \prod_{i=1}^{n} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Lambda}_i^2\right)\right]. \end{aligned}$$

Because we have different (independent) eigenmaps for two groups, the total number of common factors across subjects must be at most 6.



(a) Raw structure (b) Approximated structure

Figure 5.2: Figure (a) shows the first raw data matrix $\boldsymbol{L}_i$ in the first simulation data set, while figure (b) shows the approximated matrix by the proposed method.

As Table 5.2 presents, the true number of underlying factors was estimated correctly. Reconstruction error in Table 5.2 and Figure 5.2 demonstrate that our decomposition method

approximates the raw matrix well.

5.3.2 Simulation 2

This simulation study aims to examine if the coefficient values are estimated well by the proposed method. We measure the estimation error for $\boldsymbol{\Gamma}_j$ $(j = 1, 2)$ in the original space, i.e., estimation error is defined by

$$\text{error}_{\gamma_j} = \frac{||\boldsymbol{B}\boldsymbol{\Gamma}_j\boldsymbol{B}' - \widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Gamma}}_j\widehat{\boldsymbol{B}}'||_F}{||\boldsymbol{B}\boldsymbol{\Gamma}_j\boldsymbol{B}'||_F}.$$

Also we check convergence of MCMC outputs by using trace and autocorrelation plots.

We generate all simulation data sets from $\boldsymbol{L}_i = \boldsymbol{B}\boldsymbol{\Lambda}_i\boldsymbol{B}' + \boldsymbol{\epsilon}_i$, where $R = 3$. Set $\sigma = 1$ $\boldsymbol{B} = [\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_R]$ and $\boldsymbol{\Lambda}_i$s are generated from

$$\begin{aligned} \boldsymbol{\beta}_k &\sim N(\mathbf{0}, \boldsymbol{I}_V) \\ \pi(\boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_n) &= \prod_{i=1}^{n} \exp\left[-\frac{1}{2}\text{tr}\left((\boldsymbol{\Lambda}_i - \boldsymbol{\Delta}_i)^2\right)\right]. \end{aligned}$$

A. Scenario 3

The scenario 3 includes a binary covariate such that $vech(\boldsymbol{\Delta}_i) = \boldsymbol{\Gamma}'\boldsymbol{x}_i$, $\boldsymbol{\Gamma} = [vech(\boldsymbol{\Gamma}_1),\ vech(\boldsymbol{\Gamma}_2)]$, $\boldsymbol{x}_i = [1,\ x_{1i}]$, and

$$x_{1i} \sim B\left(\frac{1}{2}\right), \quad \boldsymbol{\Gamma}_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{\Gamma}_2 = \begin{bmatrix} 0 & 4 & 0 \\ 4 & 0 & 4 \\ 0 & 4 & 0 \end{bmatrix}.$$

B. Scenario 4

In this scenario we consider a continuous covariate, where $vech(\boldsymbol{\Delta}_i) = \boldsymbol{\Gamma}'\boldsymbol{x}_i$, $\boldsymbol{\Gamma} =$

Table 5.3: Estimation error for $\mathbf{\Gamma}_0$ and $\mathbf{\Gamma}_1$ in the two different cases. They are mean estimation error with 20 replications and the corresponding standard errors are recorded in the parentheses.

	Scenario 3	Scenario 4
error_{γ_0}	0.074 (0.004)	0.038 (0.003)
error_{γ_1}	0.141 (0.009)	0.105 (0.010)

$[vech(\mathbf{\Gamma}_1),\ vech(\mathbf{\Gamma}_2)]$, $\boldsymbol{x}_i = [1,\ x_{1i}]$, and

$$x_{1i} \sim N(0.5, 1), \quad \mathbf{\Gamma}_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{\Gamma}_2 = \begin{bmatrix} 0 & 4 & 0 \\ 4 & 0 & 4 \\ 0 & 4 & 0 \end{bmatrix}.$$

Table 5.3 shows estimation error for $\mathbf{\Gamma}_0$ and $\mathbf{\Gamma}_1$ in the two different cases. They are mean estimation error with 20 replications with the corresponding standard errors in the parentheses. One can see that the coefficients are estimated satisfactorily.

5.4 Application to Alzheimer's Disease

Altered brain connectivity has been considered as a critical factor to explain cognitive decline in Alzheimer's disease (AD) [37, 39]. It has been reported that some regions in mild AD brains have abnormal functional connectivity with other brain regions including medial prefrontal cortex (MPFC), ventral anterior cingulate cortex (vACC), right inferotemporal cortex, right cuneus extending into precuneus, left cuneus, right superior and middle temporal gyrus and posterior cingulate cortex (PCG or PCC) through a seed based approach using Fisher's z-transformation and t-tests [158]. Wang et al. [157] showed that AD patients had decreased connectivity between prefrontal and parietal lobes and increased within-lobe functional connectivity by whole brain ROI based t-tests. But they conducted 6670 (116*115/2) multiple testing at 0.01 significance level, their results thus suffer from severe false positive problems.

To overcome limitations of the current methods, we applied our proposed method and

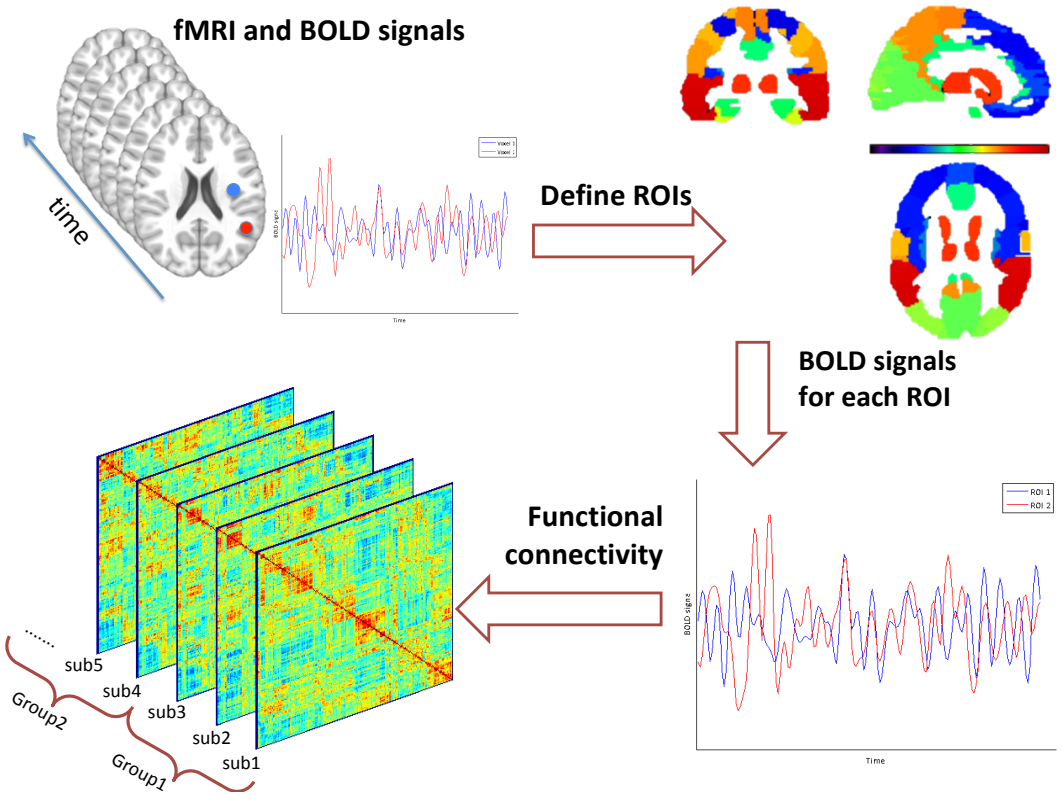


Figure 5.3: It shows process to estimate functional connectivity from resting-state fMRI data.

tested group differences among normal control (NC), mild cognitive impairment (MCI), and AD patient groups. We considered 4 covariates other than the intercept: gender (γ_1), age(γ_2), MCI=1(γ_3), and AD=1(γ_4). By using HPD intervals of γ_3, γ_4, $\gamma_4 - \gamma_3$, we determine if there is any group difference in their connectivity patterns.

5.4.1 Data Acquisition and Pre-processing

For resting state fMRI, the imaging protocol is Field Strength=3.0 tesla; Flip Angle=80.0 degree; Manufacturer=Philips Medical Systems; Matrix X=64.0 pixels; Matrix Y=64.0 pixels; Mfg Model=Intera; Pixel Spacing X=3.3125 mm; Pixel Spacing Y=3.3125 mm; Pulse Sequence=GR; Slices=6720.0; Slice Thickness=3.312999963760376 mm; TE=30.000999450683594

ms; TR=3000.0 ms; to obtain 140 volumes.

The fMRI data was pre-processed with the following steps: 1) discarding the first 10 time points, 2) slice timing, 3) head motion correction, 4) intensity scaling of each fMRI scan after motion correction to yield a whole-brain mean value of 10000, 5) temporally band-pass filtering (0.01 Hz-0.08 Hz), 6) regression out of a set of nuisance signals including signal averaged over the white matter, signal averaged over the cerebrospinal fluid, global signal averaged over the whole-brain, and six motion parameters, 7) nonlinear normalization into Montreal Neurological Institute (MNI) space with resolution 333mm^3 using SPM8; 8) spatially smoothing with a 6 mm full width at half maximum Gaussian kernel. The nonlinear normalization of fMRI data was implemented using DARTEL of SPM8 with the deformation fields of their co-registered T1-weighted images.

Figure 5.3 shows the overall procedure to calculate the resting-state functional connectivity from fMRI data. We used an Automated Anatomical Labeling (AAL) atlas, a widely used manual macroanatomical parcellation, and finally got 116 ROIs for a single subject. We used AFNI package [30] to compute the average BOLD signal over a ROI of all voxel values.

5.4.2 Data Analysis Results

In order to find the optimal number of factors, we used BIC values (Figure 5.4). We set the number of factors, R to be 15. Figure 5.5 shows the estimated $\boldsymbol{B}$ matrix. For better presentation, we plotted the standardized $\boldsymbol{B}$ matrix and the thresholded matrix having the absolute standardize values greater than 2.1. Interestingly, the factor 9 mainly consists of the left superior frontal gyrus, the left Cerebelum.7b, the right angular gyrus, and the left/right precuneus(PCUN), while all the areas are in the default mode network or positively correlated with PCUN/PCG [50, 52]. Also, the factor 11 mainly consists of the right angular

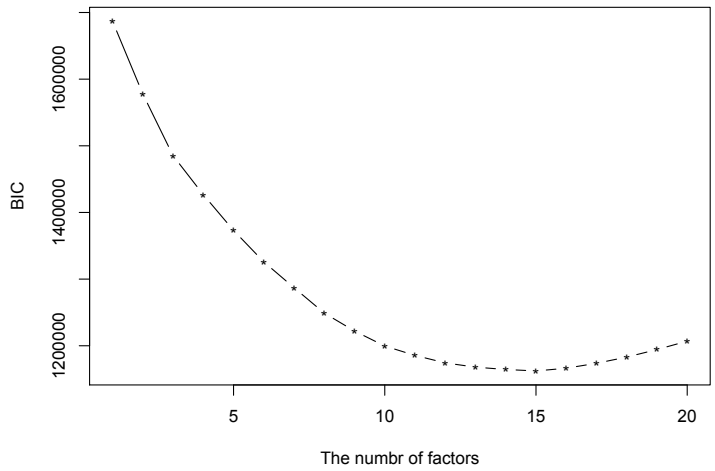
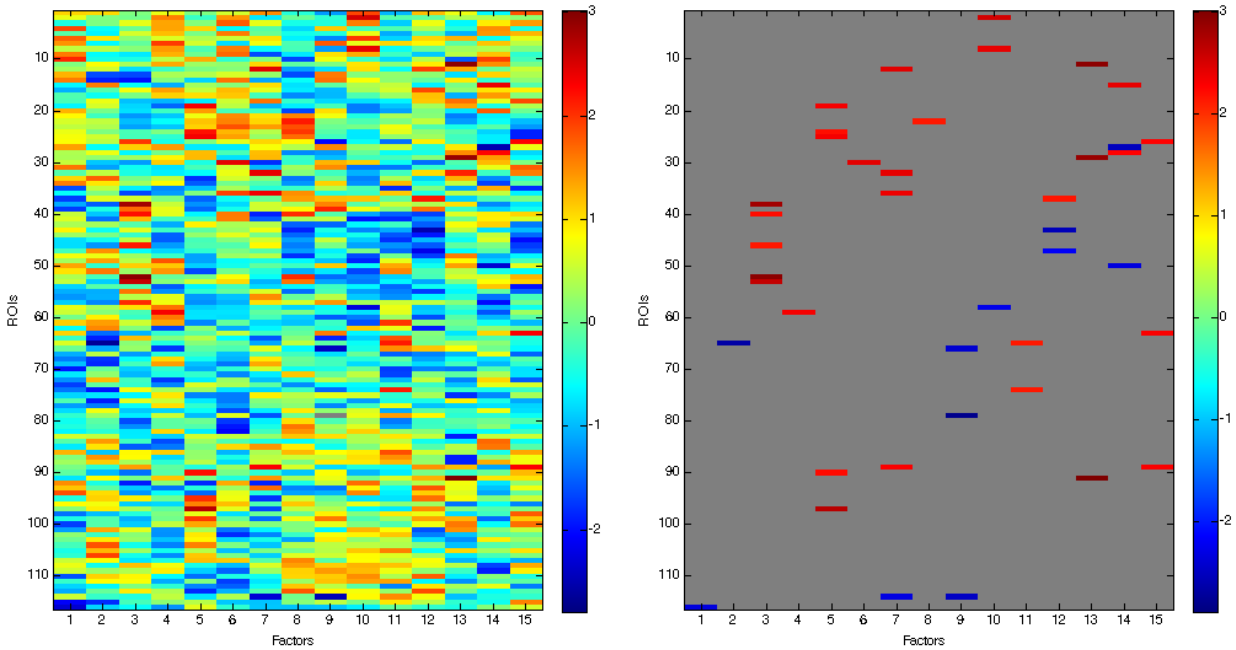


Figure 5.4: It shows BIC values for the different number of underlying factors. The optimal number can be selected as 15.



(a) Standardized

(b) Thresholded

Figure 5.5: Figure (a) shows the estimated B matrix with standardized scale, while figure (b) shows the standardized B matrix after thresholding the absolute values at 2.1.

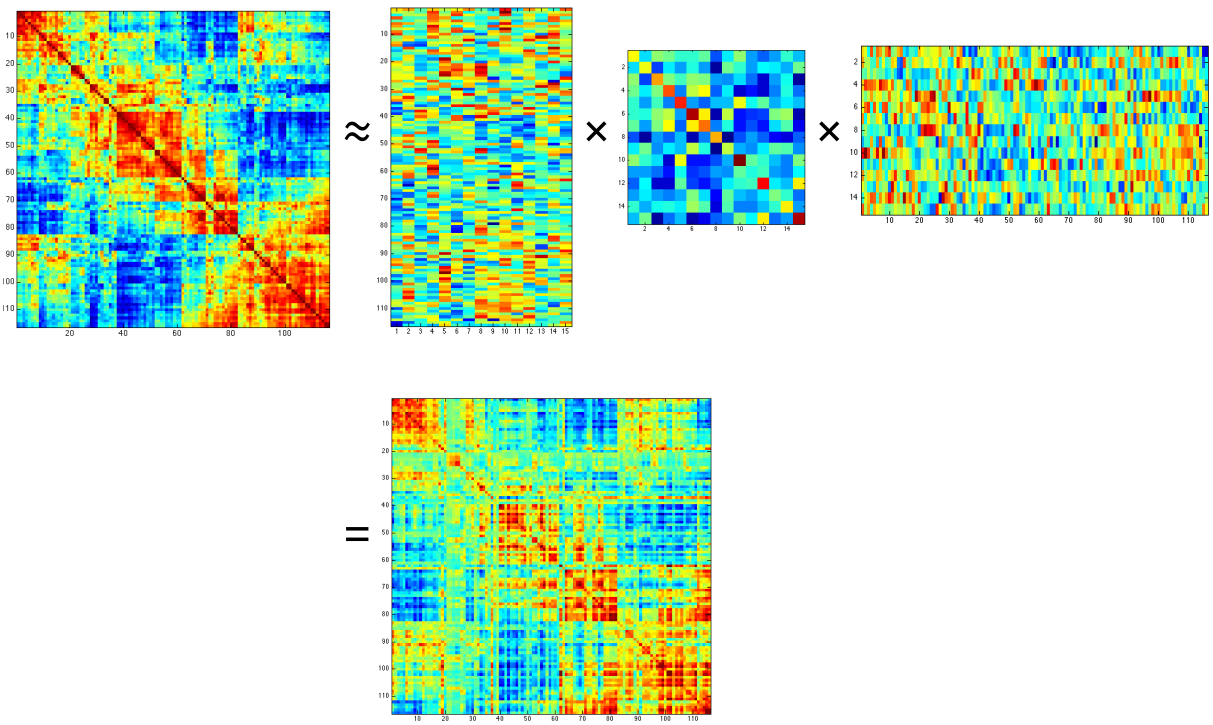


Figure 5.6: It shows how the raw correlation matrix of the first subject can be decomposed by the BGSC.

gyrus, and the right posterior cingulate gyrus (PCG) that are in the default mode network or positively correlated with PCUN/PCG [50, 52]. PCG/PCUN and angular gyrus are functional hubs of default mode network, which is disrupted in people with AD, and autism spectrum disorder [3]. For example of the factor 10, the corresponding ROIs are the left/right paracentral lobule, the left calcarine fissure and surrounding cortex, mostly representing the paracentral lobule. Figure 5.6 shows how well the raw correlation matrix of the first subject is approximated by the proposed method. Figures 5.9, 5.10, and 5.11 show fast convergence of the algorithm for some parameters, indicating its good mixing properties.

In order to see which ROI areas have functional connectivity differences among NC, MCI, and AD, we map estimated coefficients $\widehat{\mathbf{\Gamma}}$ from the factor space to the original ROI space. We use "truncated/sparse" $\widehat{\boldsymbol{B}}$ and $\widehat{\mathbf{\Gamma}}$ to calculate the coefficient matrix by the following steps:

Table 5.4: AAL parcellation of the entire brain and their abbreviations used in this paper.

Abbreviation	Name	Classification
SFG	Superior frontal gyrus	Prefrontal Lobe
SFGO	Superior frontal gyrus, orbital	Prefrontal Lobe
MFG	Middle frontal gyrus	Prefrontal Lobe
IFGOP	Inferior frontal gyrus, opercular	Prefrontal lobe
PreCG	Precentral gyrus	Other frontal
ROL	Rolandic operculum	Other frontal
ANG	Angular gyrus	Parietal lobe
PCG	Posterior cingulate gyrus	Parietal lobe
PCL	Paracentral lobule	Parietal lobe
PCUN	Precuneus	Parietal lobe
CAL	Calcarine fissure and surrounding cortex	Occipital lobe
IOG	Inferior occipital gyrus	Occipital lobe
LING.L	Lingual gyrus	Occipital lobe
MOG	Middle occipital gyrus	Occipital lobe
INS	Insula	Insula
FFG	Fusiform gyrus	Temporal lobe
HES	Heschl gyrus	Temporal lobe
STG	Superior temporal gyrus	Temporal lobe
STGP	Superior temporal gyrus, temporal pole	Temporal lobe
CRBL	Cerebellum	Cerebellum

1. Estimate $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ by the proposed method.
2. Calculate 95% HPD interval of all the elements in $\boldsymbol{\Gamma}$ and find unimportant elements (pairs of brain regions) whose HPD intervals include 0.
3. Set the corresponding coefficient values to be 0.
4. Threshold the $\widehat{\boldsymbol{B}}$ matrix by truncating their standardized values at 2.1. Let's denote the thresholded one as $\widetilde{\boldsymbol{B}}$.
5. Use $\widetilde{\boldsymbol{B}}\widehat{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{B}}'$ as estimated coefficients for the ROI space.

We found group difference between NC/MCI, NC/AD, and MCI/AD groups on several locations in $\boldsymbol{\Lambda}_i$ by examining HPD intervals of γ_3, γ_4 and $\gamma_4 - \gamma_3$. By using the above steps, we mapped the estimated coefficients from the $\boldsymbol{\Lambda}_i$ space to the $\boldsymbol{L}_i$ space. The values are summarized in Tables 5.5 for the MCI-AD group comparison. Then we plotted the

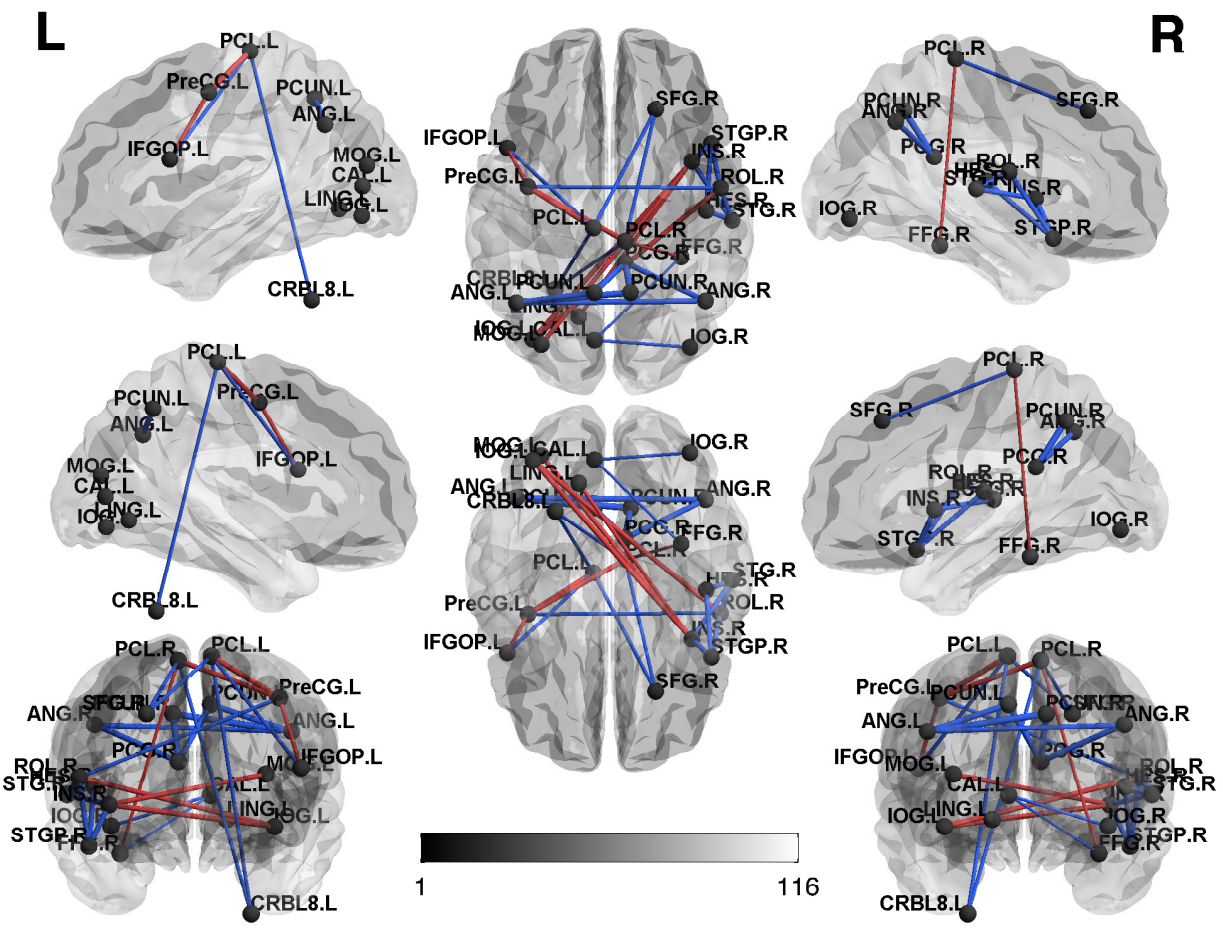


Figure 5.7: It shows which connectivity is different between MCI and AD groups. The red line represents the estimated coefficient value $\gamma_4 - \gamma_3 > 0$, which implies that AD patients have a stronger positive connection than MCI subjects between the corresponding two ROIs. The blue line represents the estimated coefficient value $\gamma_4 - \gamma_3 < 0$ indicating that AD patients have a weaker positive connection than MCI subjects between the corresponding two ROIs.

coefficients on the brain template using BrainNet Viewer [164]. Figures 5.12, 5.14, and 5.7 respectively show which brain regions have different positive (for each reference group) connectivity between NC/MCI groups, NC/AD groups, and MCI/AD groups. Figures 5.13, 5.15, and 5.8 respectively show which brain regions have different negative (for each reference group) connectivity between NC/MCI groups, NC/AD groups, and MCI/AD groups. For example of Figure 5.7, the red line represents the estimated coefficient value $\gamma_4 - \gamma_3 > 0$, which implies that AD patients have a stronger positive connection than MCI subjects between the corresponding two ROIs. The blue line represents the estimated coefficient

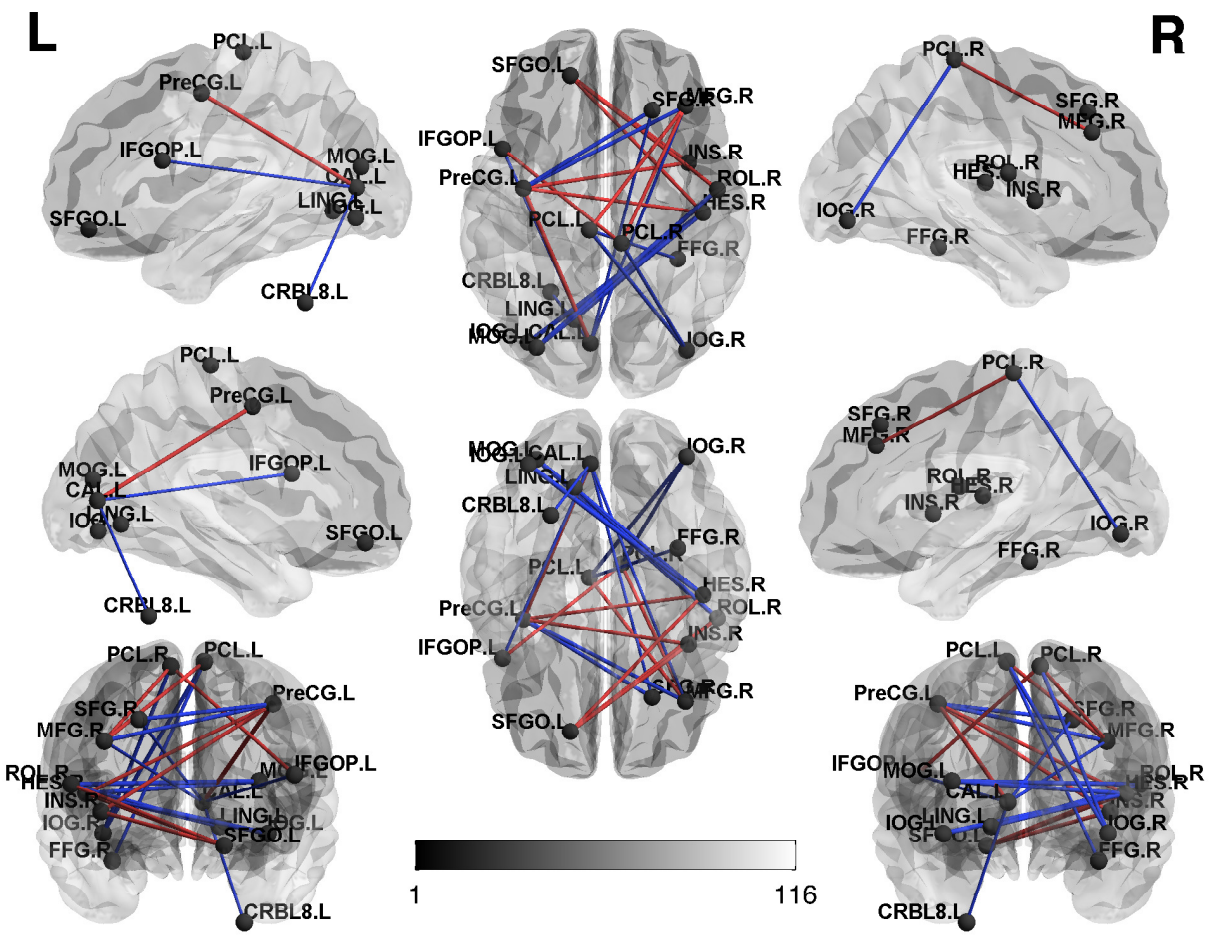
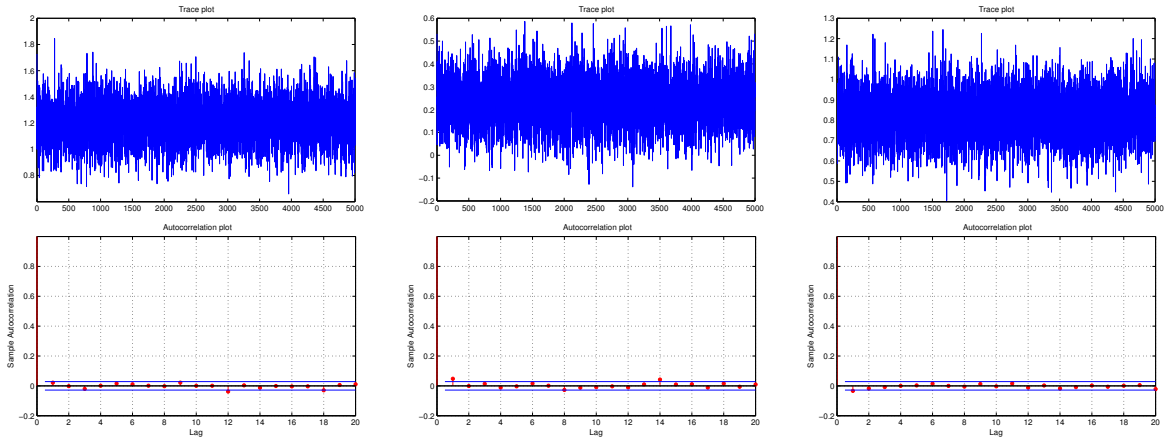


Figure 5.8: It shows which connectivity is different between MCI and AD groups. The red line represents the estimated coefficient value $\gamma_4 - \gamma_3 < 0$, which implies that AD patients have a stronger negative connection than MCI subjects between the corresponding two ROIs. The blue line represents the estimated coefficient value $\gamma_4 - \gamma_3 > 0$ indicating that AD patients have a weaker negative connection than MCI subjects between the corresponding two ROIs.

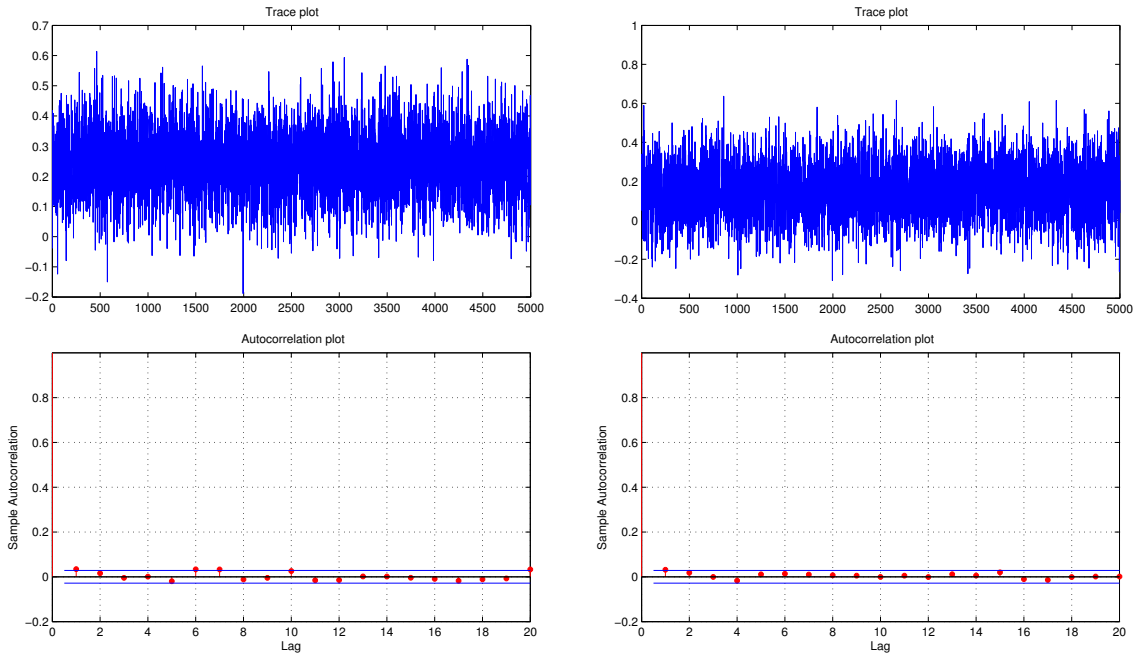
value $\gamma_4 - \gamma_3 < 0$ indicating that AD patients have a weaker positive connection than MCI subjects between the corresponding two ROIs. For the negative connectivity, in Figure 5.8, the red line implies that AD patients have a stronger negative connection than MCI subjects between the corresponding two ROIs. The blue line indicates that AD patients have a weaker negative connection than MCI subjects between the corresponding two ROIs.

In order to examine functional connectivity difference between MCI and AD patients, we focus on Figures 5.7, 5.8 and Table 5.5. The bilateral precuneus (PCUN) has weaker positive



(a) σ^2 (b) σ_0^2 (c) σ_γ^2

Figure 5.9: Trace and ACF plots of $\Lambda_1(1,1)$, $\Lambda_1(1,2)$, and $\Lambda_1(1,3)$ are respectively plotted in panels (a), (b), and (c) for 5000 iterations of the MCMC algorithm. The trace plots show fast convergence of the algorithm, indicating its good mixing properties.



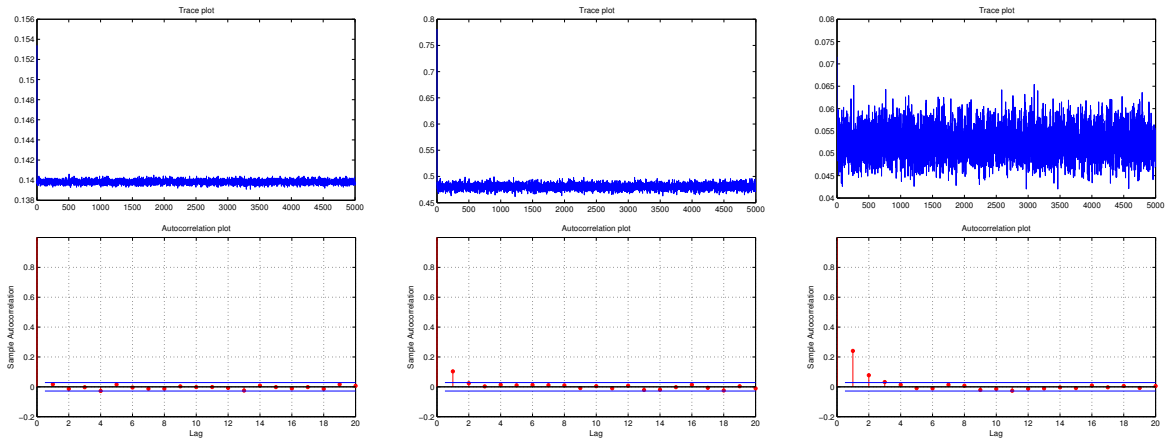
(a) $\Gamma_3(1,1)$ (b) $\Gamma_4(1,1)$

Figure 5.10: Trace and ACF plots of $\Gamma_3(1,1)$ and $\Gamma_4(1,1)$ are respectively plotted in panels (a) and (b) for 5000 iterations of the MCMC algorithm. The trace plots show fast convergence of the algorithm, indicating its good mixing properties.

Table 5.5: It shows group difference between NC and MCI subjects.

Region1	Region2	Coeff	Region1	Region2	Coeff
PCL.L	SFG.R	-0.036	SFG.R	CAL.L	0.029
PCL.L	PreCG.L	0.051	IOG.L	HES.R	0.046
PCL.L	CRBL8.L	-0.035	IOG.L	INS.R	0.049
PCL.L	IOG.R	0.038	STG.R	HES.R	-0.046
PCL.L	FFG.R	0.032	STG.R	INS.R	-0.048
PCL.L	IFGOP.L	-0.037	HES.R	PreCG.L	-0.035
PCL.L	MFG.R	-0.037	HES.R	MOG.L	0.043
PCL.R	SFG.R	-0.035	HES.R	LING.L	0.034
PCL.R	PreCG.L	0.049	PreCG.L	CAL.L	-0.041
PCL.R	CRBL8.L	-0.034	PreCG.L	IFGOP.L	0.039
PCL.R	IOG.R	0.036	PreCG.L	MFG.R	0.039
PCL.R	FFG.R	0.031	PreCG.L	INS.R	-0.037
PCL.R	IFGOP.L	-0.035	CRBL8.L	CAL.L	0.028
PCL.R	MFG.R	-0.035	IOG.R	CAL.L	-0.030
ROL.R	STGP.R	-0.049	FFG.R	CAL.L	-0.026
ROL.R	SFGO.L	-0.037	MOG.L	INS.R	0.045
ROL.R	IOG.L	0.047	CAL.L	IFGOP.L	0.030
ROL.R	STG.R	-0.047	CAL.L	MFG.R	0.030
ROL.R	PreCG.L	-0.036	LING.L	INS.R	0.036
ROL.R	MOG.L	0.044	ANG.L	ANG.R	-0.054
ROL.R	LING.L	0.035	ANG.L	PCUN.R	-0.062
STGP.R	HES.R	-0.048	ANG.L	PCUN.L	-0.055
STGP.R	INS.R	-0.051	ANG.R	PCG.R	-0.054
SFGO.L	HES.R	-0.036	PCG.R	PCUN.R	-0.062
SFGO.L	INS.R	-0.038	PCG.R	PCUN.L	-0.055
SFG.R	PreCG.L	0.039			

connection between right posterior cingulate gyrus and left angular gyrus for AD patients than MCI patients. In fact, PCG is often affected by neurodegenerative disease [15]. In early Alzheimer's disease, functional connectivity within the default mode network (DMN) is reduced, which subsequently alters the connection between the PCG and the hippocampus reflecting ApoE genetic status [93]. On the other hand, our study suggests that there is connectivity difference of paracentral gyrus with other brain regions including superior, middle, inferior frontal gyri (see Table 5.5). A fMRI study by Mason et al. [104] showed that, when stimuli from some senses are deliberately reduced or removed, the mind recruits some brain



(a) σ^2 (b) σ_0^2 (c) σ_γ^2

Figure 5.11: Trace and ACF plots of σ^2, σ_0^2, and σ_γ^2 are respectively plotted in panels (a), (b), and (c) for 5000 iterations of the MCMC algorithm. The trace plots show fast convergence of the algorithm about 200 iterations, indicating its good mixing properties.

regions including the medial posterior cingulate, the precuneus (PCUN), paracentral lobule (PCL), inferior parietal regions, the angular gyri, the inferior frontal cortices, superior and middle frontal gyri, and a cluster spanning dorsal medial frontal regions. These distributed foci have temporal coherence and constitute a tightly coupled, organized neural network. It may indicate that brain network difference between MCI and AD patients can be explained by certain brain reaction to sensory deprivation. These study results agree with our findings by the proposed method.

5.5 Discussion

In this study, we proposed a Bayesian group spectral clustering model by taking a global approach to analyze brain functional connectivity. It decomposes any symmetric data matrices with underlying common factors across subjects and the subject-specific coefficient matrix. We see that the subject-specific coefficient matrix preserves an individual network structure in the low-dimensional space spanned by the common factors. We took a Bayesian approach to estimate the underlying factors, individual coefficient matrix, and some param-

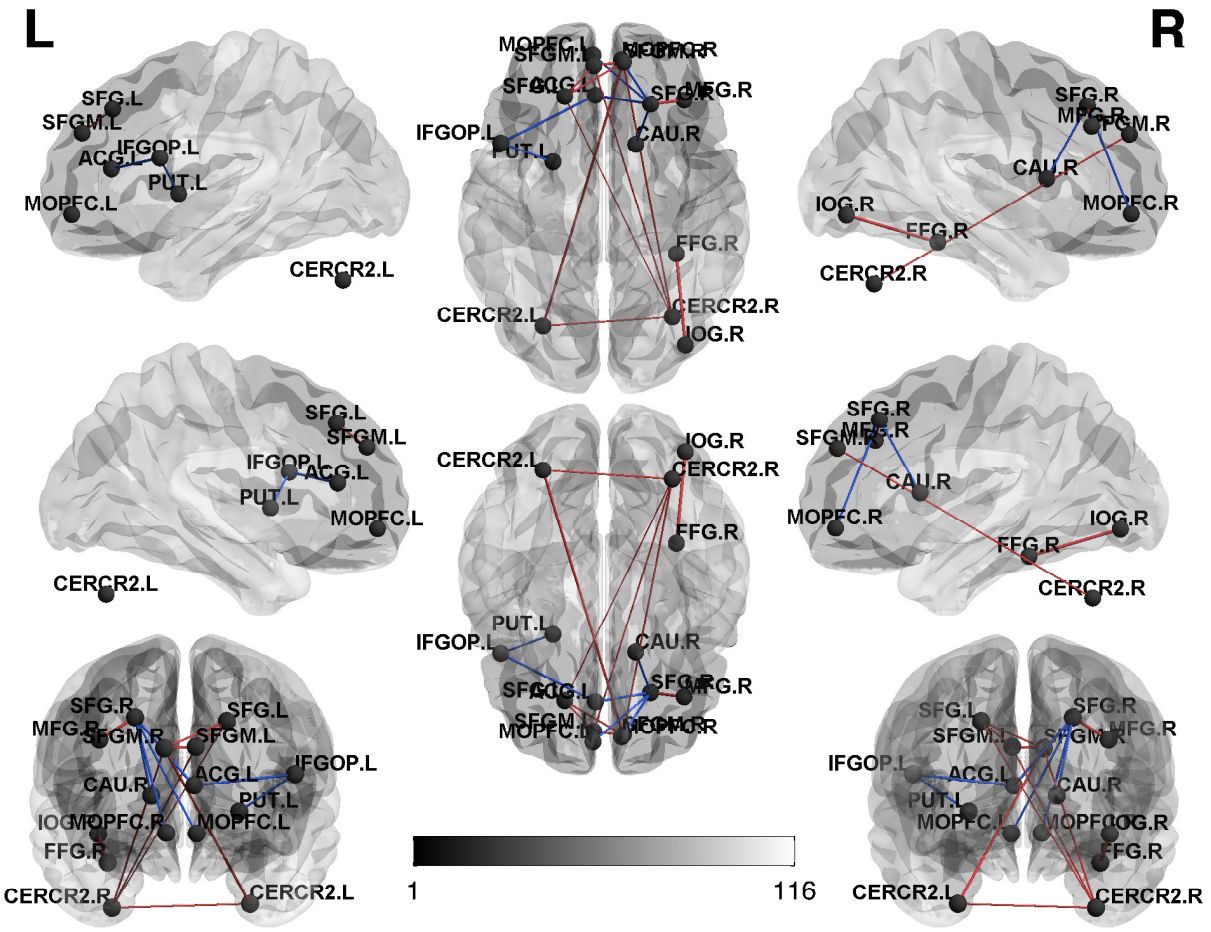


Figure 5.12: It shows which connectivity is different between NC and MCI groups. The red line represents the estimated coefficient value $\gamma_3 > 0$, which implies that MCI patients have a stronger positive connection than NC subjects between the corresponding two ROIs. The blue line represents the estimated coefficient value $\gamma_3 < 0$ indicating that MCI patients have a weaker positive connection than NC subjects between the corresponding two ROIs.

eters involved in the prediction model for clinical outcomes. We assumed a hierarchical structure within the prior of $\mathbf{\Lambda}_i$, so that we could automatically estimate the effects of covariates on the $\mathbf{\Lambda}_i$ matrices within MCMC iterations. The simulation studies demonstrated that our method efficiently approximated the raw symmetric data by the proposed decomposition with good MCMC mixing properties. Also the regression parameters for $\boldsymbol{L}_i$ could be recovered satisfactorily. The ADNI real data analysis revealed that the bilateral precuneus had weaker connection between right posterior cingulate gyrus (PCG) and left angular gyrus for AD patients than MCI patients. It implies that AD patients have more disrupted default

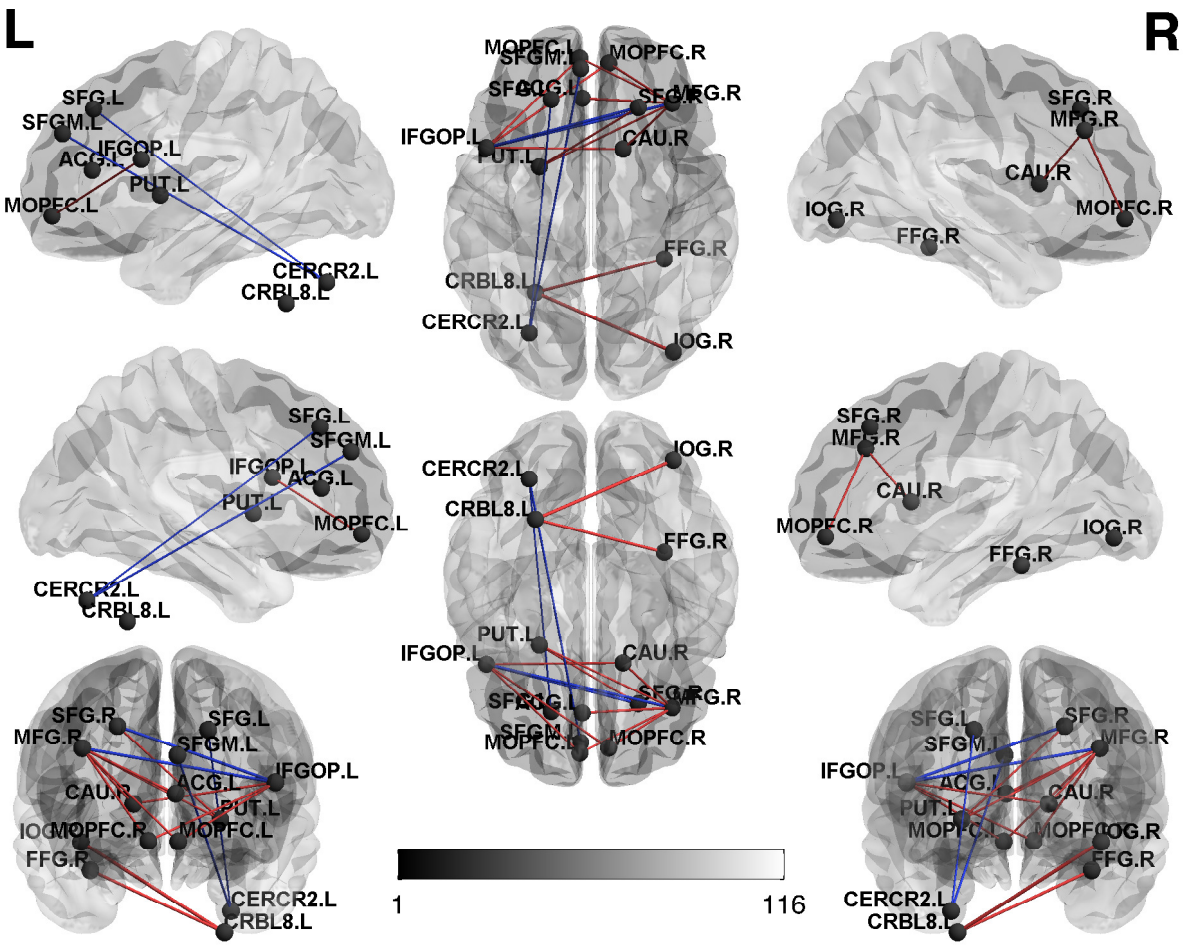


Figure 5.13: It shows which connectivity is different between NC and MCI groups. The red line represents the estimated coefficient value $\gamma_3 < 0$, which implies that MCI patients have a stronger negative connection than NC subjects between the corresponding two ROIs. The blue line represents the estimated coefficient value $\gamma_3 > 0$ indicating that MCI patients have a weaker negative connection than NC subjects between the corresponding two ROIs.

mode network than MCI subjects. Also, there was connectivity difference of paracentral gyrus with other brain regions including superior, middle, inferior frontal gyri. This finding agreed with other functional connectivity studies [15, 93, 104, 158].

This study could provide a guideline to elucidate hidden pathology of neurological disorders in a brain connectivity perspective. In clinical application, the functional connectivity can be used to (1) examine if normal subjects and patients (or among disease subtypes) have a different functional connectivity structure and where the difference comes from. This ex-

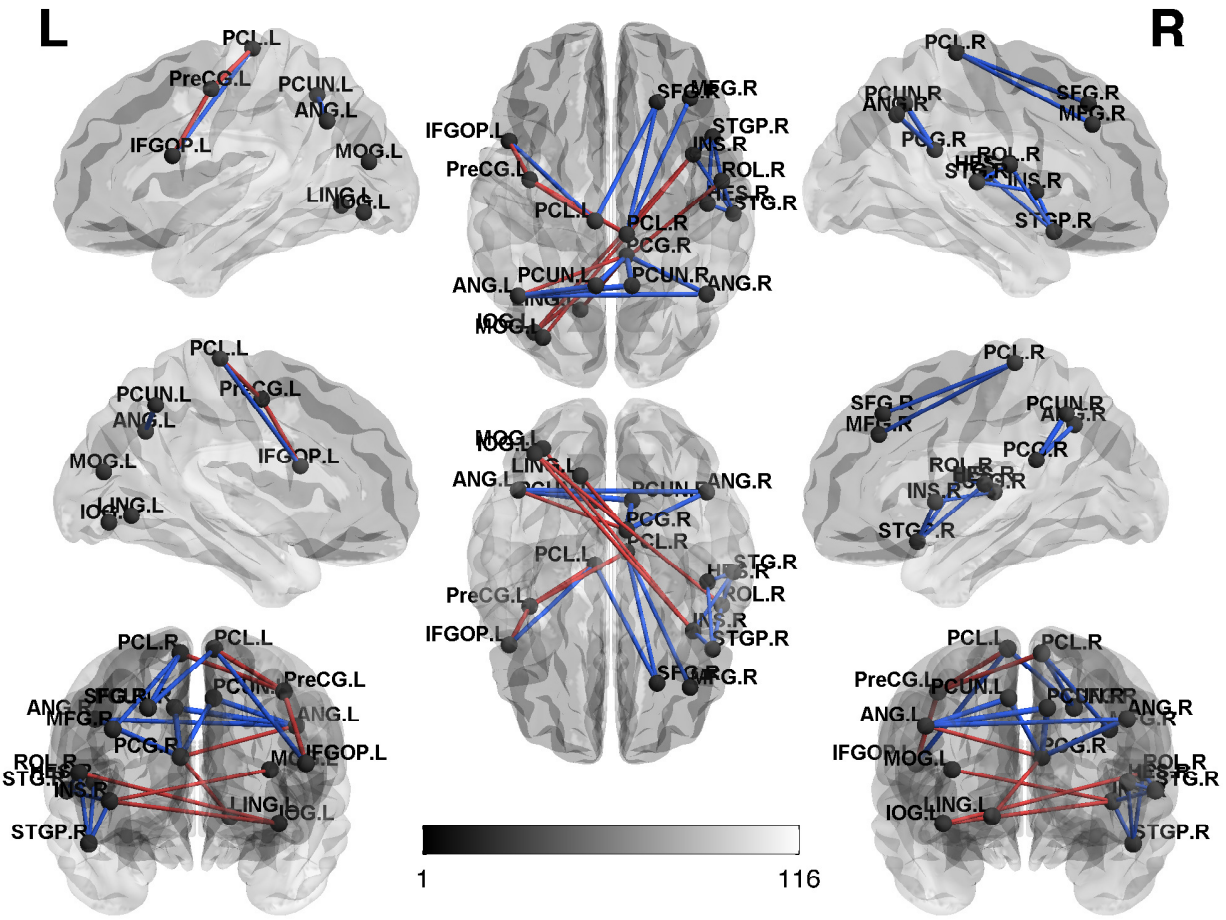


Figure 5.14: It shows which connectivity is different between NC and AD groups. The red line represents the estimated coefficient value $\gamma_4 > 0$, which implies that AD patients have a stronger positive connection than NC subjects between the corresponding two ROIs. The blue line represents the estimated coefficient value $\gamma_4 < 0$ indicating that AD patients have a weaker positive connection than NC subjects between the corresponding two ROIs.

ploratory analysis allows better understanding of the underlying mechanism and pathology of a disorder, which further may help to develop future treatments targeting some identified brain regions. Also, altered connectivity can be used as (2) diagnostic and prognostic information by relating with clinical outcomes [49], such as time to event, a binary outcome (the presence of disease), or a categorical outcome indicating disease subtypes. Diagnosis of psychiatric disorder is hard and subjective. For example ADHD, a subject is diagnosed as ADHD if the subject meets the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria. The test result can be varied depending on interview environment, examiners

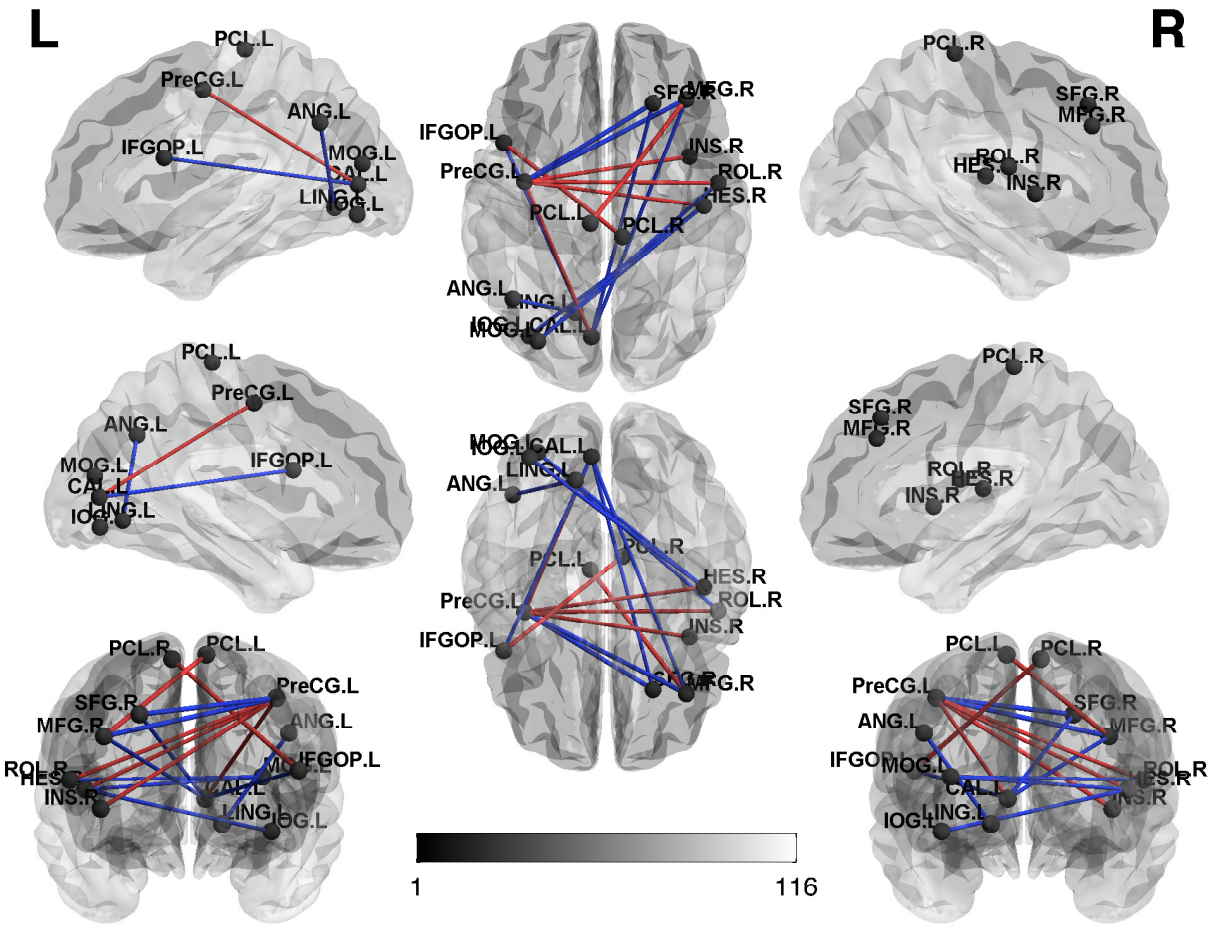


Figure 5.15: It shows which connectivity is different between NC and AD groups. The red line represents the estimated coefficient value $\gamma_4 < 0$, which implies that AD patients have a stronger negative connection than NC subjects between the corresponding two ROIs. The blue line represents the estimated coefficient value $\gamma_4 > 0$ indicating that AD patients have a weaker negative connection than NC subjects between the corresponding two ROIs.

and examinees. And it is still subjective criteria. If abnormality of functional connectivity is a very early sign to develop a disorder or a biomarker characterizing prognosis, it will be promising to use the information as an objective diagnostic/prognostic tool. One more intriguing suggestion is that (3) abnormal functional connectivity will become prominent in neurogenetic studies [87]. While genetic factors are emerging in psychiatric disorder research, their effect size is very small and their working mechanism still needs to be elucidated. Because brain organization and function are influenced by genetic factors [89, 94] and brain information has relatively large impact on disease progression, functional connectivity can

be a mediator to explain connection among genetic mutations and disorders.

Our proposed method also provides a general framework to work on any symmetric form of the data. It does not have any constraint on the $\boldsymbol{L}_i$ other than symmetricity, for example, this method can be applied on adjacency, weighted graph, and any distance matrices. There is a remaining work to improve thresholding on the coefficient matrix $\widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Gamma}}_j\widehat{\boldsymbol{B}}'$ to efficiently present it on the brain region space. Although the estimated coefficient matrix $\widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Gamma}}_j\widehat{\boldsymbol{B}}'$ is invariant with respect to the choice of $\boldsymbol{B}$ matrix, the thresholded $\boldsymbol{B}$ matrix (using their z-scores) and the thresholded coefficient matrix $\boldsymbol{\Gamma}_j$(based on their HPD intervals) will be varied. When we choose different $\boldsymbol{B}$ matrix, it will make it different which brain regions have "more" connectivity difference than the others. It urges us to study a more rigorous way of thresholding $\boldsymbol{B}$ and $\boldsymbol{\Gamma}_j$ matrices.

BIBLIOGRAPHY

[1] M. Albert, S. DeKosky, D. Dickson, B. Dubois, H. Feldman, N. Fox, A. Gamst, D. Holtzman, W. Jagust, R. Petersen, P. Snyder, M. Carrillo, B. Thies, and C. Phelps. The diagnosis of mild cognitive impairment due to alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & Dementia*, 7(3):270–279, 2011.

[2] N. D. Anderson, P. L. Ebert, J. M. Jennings, C. L. Grady, R. Cabeza, and S. J. Graham. Recollection-and familiarity-based memory in healthy aging and amnestic mild cognitive impairment. *Neuropsychology*, 22(2):177–187, 2008.

[3] J. R. Andrews-Hanna, J. S. Reidler, J. Sepulcre, R. Poulin, and R. L. Buckner. Functional-anatomic fractionation of the brain's default network. *Neuron*, 65(4):550–562, 2010.

[4] A. Annest, R. E. Bumgarner, A. E. Raftery, and K. Y. Yeung. Iterative bayesian model averaging: a method for the application of survival analysis to high-dimensional microarray data. *BMC bioinformatics*, 10(1):1, 2009.

[5] L. G. Apostolova, I. D. Dinov, R. A. Dutton, K. Hayashi, A. Toga, J. Cummings, and P. Thompson. 3d comparison of hippocampal atrophy in amnestic mild cognitive impairment and alzheimer's disease. *Brain*, 129(11):2867–2873, 2006.

[6] L. G. Apostolova, R. A. Dutton, I. D. Dinov, K. M. Hayashi, A. W. Toga, J. L. Cummings, and P. M. Thompson. Conversion of mild cognitive impairment to alzheimer disease predicted by hippocampal atrophy maps. *Archives of neurology*, 63(5):693–699, 2006.

[7] E. J. Bedrick, R. Christensen, and W. O. Johnson. Bayesian accelerated failure time analysis with application to veterinary epidemiology. *Statistics in medicine*, 19(2): 221–237, 2000.

[8] G. A. Bezerra, E. Dobrovetsky, A. Seitova, S. Fedosyuk, S. Dhe-Paganon, and K. Gruber. Structure of human dipeptidyl peptidase 10 (dppy): a modulator of neuronal kv4 channels. *Scientific reports*, 5, 2015.

[9] A. Bhattacharya, D. Pati, N. S. Pillai, and D. B. Dunson. Dirichlet–laplace priors for optimal shrinkage. *Journal of the American Statistical Association*, 110(512):1479–1490, 2015.

[10] H. Binder and M. Schumacher. Allowing for mandatory covariates in boosting estimation of sparse high-dimensional survival models. *BMC bioinformatics*, 9(1):14, 2008.

[11] A. Biswas, S. Datta, J. P. Fine, and M. R. Segal. *Statistical Advances in the Biomedical Sciences: Clinical Trials, Epidemiology, Survival analysis, and Bioinformatics*, volume 630. John Wiley & Sons, 2007.

[12] V. Bonato, V. Baladandayuthapani, B. M. Broom, E. P. Sulman, K. D. Aldape, and K.-A. Do. Bayesian ensemble methods for survival prediction in gene expression data. *Bioinformatics*, 27(3):359–367, 2011.

[13] P. Breheny. The group exponential lasso for bi-level variable selection. *Biometrics*, 71 (3):731–740, 2015.

[14] C. Bryant, K. S. Giovanello, J. G. Ibrahim, J. Chang, D. Shen, B. S. Peterson, H. Zhu, and ADNI. Mapping the genetic variation of regional brain volumes as explained by all common snps from the adni study. *PloS One*, 8(8):e71723, 2013.

[15] R. L. Buckner, J. R. Andrews-Hanna, and D. L. Schacter. The brain's default network. *Annals of the New York Academy of Sciences*, 1124(1):1–38, 2008.

[16] V. C. Buschert, U. Friese, S. J. Teipel, P. Schneider, W. Merensky, D. Rujescu, H.-J. Möller, H. Hampel, and K. Buerger. Effects of a newly developed cognitive intervention in amnestic mild cognitive impairment and mild alzheimer's disease: a pilot study. *Journal of Alzheimer's disease*, 25(4):679–694, 2011.

[17] T. Cai, J. Huang, and L. Tian. Regularized estimation for the accelerated failure time model. *Biometrics*, 65(2):394–404, 2009.

[18] T. T. Cai, P. Hall, et al. Prediction in functional linear regression. *The Annals of Statistics*, 34(5):2159–2179, 2006.

[19] H. Cardot, F. Ferraty, and P. Sarda. Functional linear model. *Statistics & Probability Letters*, 45(1):11–22, 1999.

[20] H. Cardot, F. Ferraty, and P. Sarda. Spline estimators for the functional linear model. *Statistica Sinica*, pages 571–591, 2003.

[21] K. H. Chen, L. Y. Chuah, S. K. Sim, and M. W. Chee. Hippocampal region-specific contributions to memory performance in normal elderly. *Brain and Cognition*, 72(3): 400–407, 2010.

[22] T. Chen, X. Shen, F. Chegini, W. Gai, and C. Abbott. Molecular characterisation of a novel dipeptidyl peptidase like protein: its pathological link to alzheimers disease. *Clin Chem Lab Med*, 46:A13, 2008.

[23] T. Chen, W.-P. Gai, and C. A. Abbott. Dipeptidyl peptidase 10 (dpp10 789): A voltage gated potassium channel associated protein is abnormally expressed in alzheimers and other neurodegenerative diseases. *BioMed research international*, 2014, 2014.

[24] S. Cho, H. Kim, S. Oh, K. Kim, and T. Park. Elastic-net regularization approaches for genome-wide association studies of rheumatoid arthritis. In *BMC proceedings*, volume 3, page 1. BioMed Central, 2009.

[25] R. Christensen and W. Johnson. Modelling accelerated failure time with a dirichlet process. *Biometrika*, 75(4):693–704, 1988.

[26] R. Colom, J. L. Stein, P. Rajagopalan, K. Martínez, D. Hermel, Y. Wang, J. Álvarez-Linera, M. Burgaleta, M. Quiroga, P. C. Shih, and P. M. Thompson. Hippocampal structure and human cognition: Key role of spatial processing and evidence supporting the efficiency hypothesis in females. *Intelligence*, 41(2):129–140, 2013.

[27] E. Corder, A. Saunders, W. Strittmatter, D. Schmechel, P. Gaskell, G. Small, A. Roses, J. Haines, and M. A. Pericak-Vance. Gene dose of apolipoprotein e type 4 allele and the risk of alzheimer's disease in late onset families. *Science*, 261(5123):921–923, 1993.

[28] S. Costafreda, I. Dinov, Z. Tu, Y. Shi, C. Liu, I. Kloszewska, P. Mecocci, H. Soininen, M. Tsolaki, B. Vellas, L. Wahlund, C. Spenger, A. Toga, S. Lovestone, and A. Simmons. Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment. *Neuroimage*, 56(1):212–219, 2011.

[29] D. R. Cox. Regression models and life tables. *Journal of Royal Statistical Society. Series B.*, 34(2):187–220, 1972.

[30] R. W. Cox. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3):162–173, 1996.

[31] Y. Cui, B. Liu, S. Luo, X. Zhen, M. Fan, T. Liu, W. Zhu, M. Park, T. Jiang, J. Jin, and ADNI. Identification of conversion from mild cognitive impairment to alzheimer's disease using multivariate predictors. *PloS One*, 6(7):e21896, 2011.

[32] J. L. Cummings, R. Doody, and C. Clark. Disease-modifying therapies for alzheimer disease challenges to early intervention. *Neurology*, 69(16):1622–1634, 2007.

[33] X. Da, J. Toledo, J. Zee, D. Wolk, S. Xie., Y. Ou, A. Shacklett, P. Parmpi, L. Shaw, J. Trojanowski, and C. Davatzikos. Integration and relative value of biomarkers for prediction of mci to ad progression: spatial patterns of brain atrophy, cognitive scores, apoe genotype, and csf markers. *Neuroimage: Clinical*, 4:164–173, 2014.

[34] S. Datta, J. Le-Rademacher, and S. Datta. Predicting patient survival from microarray data by accelerated failure time modeling using partial least squares and lasso. *Biometrics*, 63(1):259–271, 2007.

[35] P. L. De Jager, J. M. Shulman, L. B. Chibnik, B. T. Keenan, T. Raj, R. S. Wilson, L. Yu, S. E. Leurgans, D. Tran, C. Aubin, et al. A genome-wide scan for common variants affecting the rate of age-related cognitive decline. *Neurobiology of aging*, 33 (5):1017–e1, 2012.

[36] J. de la Torre. Alzheimer's disease is incurable but preventable. *Journal of Alzheimer's Disease*, 20(3):861–870, 2010.

[37] M.-C. de LaCoste and C. L. White. The role of cortical connectivity in alzheimer's disease pathogenesis: a review and model system. *Neurobiology of Aging*, 14(1):1–16, 1993.

[38] M. De Leon, A. George, J. Golomb, C. Tarshish, A. Convit, A. Kluger, S. De Santi, T. Mc Rae, S. Ferris, B. Reisberg, C. Ince, H. Rusinek, M. Bobinski, B. Quinn, D. C. Miller, and H. M. Wisniewski. Frequency of hippocampal formation atrophy in normal aging and alzheimer's disease. *Neurobiology of Aging*, 18(1):1–11, 1997.

[39] X. Delbeuck, M. Van der Linden, and F. Collette. Alzheimer'disease as a disconnection syndrome? *Neuropsychology review*, 13(2):79–92, 2003.

[40] R. S. Desikan, H. J. Cabral, B. Fischl, C. R. G. Guttmann, D. Blacker, B. T. Hyman, M. S. Albert, and R. J. Killiany. Temporoparietal mr imaging measures of atrophy in subjects with mild cognitive impairment that predict subsequent diagnosis of alzheimer disease. *American Journal of Neuroradiology*, 30(3):532–538, 2009.

[41] D. Devanand, G. Pradhaban, X. Liu, A. Khandji, S. De Santi, S. Segal, H. Rusinek, G. Pelton, L. Honig, R. Mayeux, Y. Stern, M. H. Tabert, and M. J. de Leon. Hippocampal and entorhinal atrophy in mild cognitive impairment prediction of alzheimer disease. *Neurology*, 68(11):828–836, 2007.

[42] B. Dickerson, I. Goncharova, M. Sullivan, C. Forchetti, R. Wilson, D. Bennett, L. Beckett, and L. deToledo Morrell. Mri-derived entorhinal and hippocampal atrophy in incipient and very mild alzheimer's disease. *Neurobiology of Aging*, 22(5):747–754, 2001.

[43] B. Dickerson, D. A. Wolk, and ADNI. Biomarker-based prediction of progression in mci: comparison of ad signature and hippocampal volume with spinal fluid amyloid-β and tau. *Frontiers in Aging Neuroscience*, 5, 2013.

[44] J. Fan and J. Lv. A selective overview of variable selection in high dimen- sional feature space. *Statistica Sinica*, 20:101–148, 2010.

[45] Y. Fan, N. Batmanghelich, C. M. Clark, and C. Davatzikos. Spatial patterns of brain atrophy in mci patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage*, 39(4):1731–1743, 2008.

[46] C. Fennema-Notestine, D. Hagler, L. McEvoy, A. S. Fleisher, E. H. Wu, D. Karow, and A. Dale. Structural mri biomarkers for preclinical and mild alzheimer's disease. *Human Brain Mapping*, 30(10):3238–3253, 2009.

[47] F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementation*. Springer, New York, 2006.

[48] T. R. Fleming and D. P. Harrington. *Counting processes and survival analysis*, volume 169. John Wiley & Sons, 2011.

[49] M. D. Fox and M. Greicius. Clinical applications of resting state functional connectivity. *Frontiers in systems neuroscience*, 4:19, 2010.

[50] M. D. Fox, A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen, and M. E. Raichle. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9673–9678, 2005.

[51] E. Frankó, O. Joly, and ADNI. Evaluating alzheimers disease progression using rate of regional hippocampal atrophy. *PloS one*, 8(8):e71354, 2013.

[52] P. Fransson. Spontaneous low-frequency bold signal fluctuations: An fmri investigation of the resting-state default mode of brain function hypothesis. *Human brain mapping*, 26(1):15–29, 2005.

[53] S. Gauthier, B. Reisberg, M. Zaudig, R. Petersen, K. Ritchie, K. Broich, S. Belleville, H. Brodaty, D. Bennett, H. Chertkow, J. Cummings, M. de Leon, H. Feldman, M. Ganguli, H. Hampel, P. Scheltens, M. Tierney, P. Whitehouse, and B. Winblad. Mild cognitive impairment. *The Lancet*, 367(9518):1262–1270, 2006.

[54] A. Gelman, W. R. Gilks, and G. O. Roberts. Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, 7:110–120, 1997.

[55] S. Girirajan, M. Y. Dennis, C. Baker, M. Malig, B. P. Coe, C. D. Campbell, K. Mark, T. H. Vu, C. Alkan, Z. Cheng, et al. Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *The American Journal of Human Genetics*, 92(2):221–237, 2013.

[56] J. J. Goeman. L1 penalized estimation in the cox proportional hazards model. *Biometrical journal*, 52(1):70–84, 2010.

[57] J. J. Gomar, M. T. Bobes-Bascaran, C. Conejero-Goldberg, P. Davies, T. E. Goldberg, and ADNI. Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to alzheimer disease in patients in the alzheimer's disease neuroimaging initiative. *Archives of General Psychiatry*, 68 (9):961–969, 2011.

[58] M. Greicius. Resting-state functional connectivity in neuropsychiatric disorders. *Current opinion in neurology*, 21(4):424–430, 2008.

[59] M. Grundman, D. Sencakova, C. Jack, R. C. Petersen, H. Kim, A. Schultz, M. Weiner, C. DeCarli, S. DeKosky, C. van Dyck, R. G. Thomas, L. J. Thal, and ADCS. Brain mri hippocampal volume and prediction of clinical status in a mild cognitive impairment trial. *Journal of Molecular Neuroscience*, 19(1-2):23–27, 2002.

[60] Y. Guan and M. Stephens. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *The Annals of Applied Statistics*, pages 1780–1815, 2011.

[61] R. J. Guerreiro, D. R. Gustafson, and J. Hardy. The genetic architecture of alzheimer's disease: beyond app, psens and apoe. *Neurobiology of aging*, 33(3):437–456, 2012.

[62] J. Gui and H. Li. Threshold gradient descent method for censored data regression with applications in pharmacogenomics. In *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pages 272–283, 2004.

[63] J. Gui and H. Li. Penalized cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. *Bioinformatics*, 21(13):3001–3008, 2005.

[64] M. Gupta and J. G. Ibrahim. An information matrix prior for bayesian analysis in generalized linear models with high dimensional data. *Statistica Sinica*, 19(4):1641–1663, 2009.

[65] T. J. Hastie and R. J. Tibshirani. *Generalized additive models*, volume 43. CRC Press, 1990.

[66] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[67] Q. He and D.-Y. Lin. A variable selection method for genome-wide association studies. *Bioinformatics*, 27(1):1–8, 2011.

[68] L. E. Hebert, J. Weuve, P. A. Scherr, and D. A. Evans. Alzheimer disease in the united states (2010–2050) estimated using the 2010 census. *Neurology*, 80(19):1778–1783, 2013.

[69] B. Hooli, Z. M. Kovacs-Vajna, K. Mullin, M. Blumenthal, M. Mattheisen, C. Zhang, C. Lange, G. Mohapatra, L. Bertram, and R. Tanzi. Rare autosomal copy number variations in early-onset familial alzheimer's disease. *Molecular psychiatry*, 19(6):676–681, 2014.

[70] J. Huang, S. Ma, and H. Xie. Regularized estimation in the accelerated failure time model with high-dimensional covariates. *Biometrics*, 62(3):813–820, 2006.

[71] J. Huang, S. Ma, H. Xie, and C.-H. Zhang. A group bridge approach for variable selection. *Biometrika*, 96(2):339–355, 2009.

[72] J. Huang, P. Breheny, and S. Ma. A selective review of group selection in high-dimensional models. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 27(4):481–499, 2012.

[73] M. J. Huentelman, A. Papassotiropoulos, D. W. Craig, F. J. Hoerndli, J. V. Pearson, K.-D. Huynh, J. Corneveaux, J. Hänggi, C. R. Mondadori, A. Buchmann, et al. Calmodulin-binding transcription activator 1 (camta1) alleles predispose human episodic memory performance. *Human molecular genetics*, 16(12):1469–1477, 2007.

[74] H. Hung and C.-T. Chiang. Estimation methods for time-dependent auc models with survival data. *Canadian Journal of Statistics*, 38(1):8–26, 2010.

[75] J. G. Ibrahim, M.-H. Chen, and D. Sinha. *Bayesian Survival Analysis*. Wiley Online Library, 2005.

[76] J. G. Ibrahim, M. Chen, and S. Kim. Bayesian variable selection for the cox regression model with missing covariates. *Lifetime Data Analysis*, 14(4):496–520, 2008.

[77] International HapMap 3 Consortium et al. Integrating common and rare genetic variation in diverse human populations. *Nature*, 467(7311):52–58, 2010.

[78] C. R. Jack, R. C. Petersen, P. C. O'Brien, and E. G. Tangalos. Mr-based hippocampal volumetry in the diagnosis of alzheimer's disease. *Neurology*, 42(1):183–183, 1992.

[79] C. R. Jack, R. C. Petersen, Y. C. Xu, S. C. Waring, P. C. O'Brien, E. G. Tangalos, G. E. Smith, R. J. Ivnik, and E. Kokmen. Medial temporal atrophy on mri in normal aging and very mild alzheimer's disease. *Neurology*, 49(3):786–794, 1997.

[80] C. R. Jack Jr, D. S. Knopman, W. J. Jagust, L. M. Shaw, P. S. Aisen, M. W. Weiner, R. C. Petersen, and J. Q. Trojanowski. Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128, 2010.

[81] G. James. Generalized linear models with functional predictors. *Journal of Royal Statistical Society. Series B.*, 64:411–432, 2002.

[82] J.Huang, T. Sun, Z. Ying, Y. Y, and C. Zhang. Oracle inequalities for the lasso in the cox model. *Ann. Statist.*, 41:1142–1165, 2013.

[83] W. Johnson and R. Christensen. Nonparametric bayesian analysis of the accelerated failure time model. *Statistics & Probability Letters*, 8(2):179–184, 1989.

[84] J. D. Kalbfleisch and R. L. Prentice. *The statistical analysis of failure time data*, volume 360. John Wiley & Sons, 2011.

[85] J. Kaye, M. Moore, A. Dame, J. Quinn, R. Camicioli, D. Howieson, E. Corbridge, B. Care, G. Nesbit, and G. Sexton. Asynchronous regional brain volume losses in presymptomatic to moderate ad. *Journal of Alzheimer's Disease*, 8(1):51–56, 2005.

[86] J. P. Kesslak, O. Nalcioglu, and C. W. Cotman. Quantification of magnetic resonance scans for hippocampal and parahippocampal atrophy in alzheimer's disease. *Neurology*, 41(1):51–54, 1991.

[87] K. Konrad and S. B. Eickhoff. Is the adhd brain wired differently? a review on structural and functional connectivity in attention deficit hyperactivity disorder. *Human brain mapping*, 31(6):904–916, 2010.

[88] S. Konrath, L. Fahrmeir, and T. Kneib. Bayesian accelerated failure time models based on penalized mixtures of gaussians: regularization and variable selection. *AStA Advances in Statistical Analysis*, 99(3):259–280, 2015.

[89] J. W. Koten, G. Wood, P. Hagoort, R. Goebel, P. Propping, K. Willmes, and D. I. Boomsma. Genetic contribution to variation in cognitive function: an fmri study in twins. *Science*, 323(5922):1737–1740, 2009.

[90] L. Kuo and B. Mallick. Bayesian semiparametric inference for the accelerated failure-time model. *Canadian Journal of Statistics*, 25(4):457–472, 1997.

[91] J. F. Lawless. *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons, 2011.

[92] K. H. Lee. *Bayesian variable selection in parametric and semiparametric high dimensional survival analysis*. PhD thesis, University of Missouri–Columbia, 2011.

[93] R. Leech and D. J. Sharp. The role of the posterior cingulate cortex in cognition and disease. *Brain*, 137(1):12–32, 2014.

[94] R. K. Lenroot, J. E. Schmitt, S. J. Ordaz, G. L. Wallace, M. C. Neale, J. P. Lerch, K. S. Kendler, A. C. Evans, and J. N. Giedd. Differences in genetic and environmental influences on the human cerebral cortex associated with development during childhood and adolescence. *Human brain mapping*, 30(1):163–174, 2009.

[95] H. Li and Y. Luan. Boosting proportional hazards models using smoothing splines, with applications to high-dimensional microarray data. *Bioinformatics*, 21(10):2403–2409, 2005.

[96] J. Li and S. Ma. *Survival Analysis in Medicine and Genetics*. Chapman & Hall/CRC, 2013.

[97] S. Li, O. Okonkwo, M. Albert, and M. C. Wang. Variation in variables that predict progression from mci to ad dementia over duration of follow-up. *American Journal of Alzheimer's Disease*, 1:12–28, 2013.

[98] X. Lin, T. Cai, M. C. Wu, Q. Zhou, G. Liu, D. C. Christiani, and X. Lin. Kernel machine snp-set analysis for censored survival outcomes in genome-wide association studies. *Genetic epidemiology*, 35(7):620–631, 2011.

[99] J. Liu, K. Wang, S. Ma, and J. Huang. Regularized regression method for genome-wide association studies. In *BMC proceedings*, volume 5, page 1. BioMed Central, 2011.

[100] Z. Liu, D. Chen, M. Tan, F. Jiang, and R. B. Gartenhaus. Kernel based methods for accelerated failure time model with ultra-high dimensional data. *BMC bioinformatics*, 11(1):606, 2010.

[101] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fmri signal. *Nature*, 412(6843):150–157, 2001.

[102] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *ACM Siggraph Computer Graphics*, volume 21, pages 163–169. ACM, 1987.

[103] E. Luders, P. M. Thompson, F. Kurth, J. Hong, O. Phillips, Y. Wang, B. Gutman, Y. Chou, K. L. Narr, and A. Toga. Global and regional alterations of hippocampal anatomy in long-term meditation practitioners. *Human Brain Mapping*, 34(12):3369–3375, 2013.

[104] M. F. Mason, M. I. Norton, J. D. Van Horn, D. M. Wegner, S. T. Grafton, and C. N. Macrae. Wandering minds: the default network and stimulus-independent thought. *Science*, 315(5810):393–395, 2007.

[105] W. Q. Meeker and L. A. Escobar. *Statistical methods for reliability data*. John Wiley & Sons, 2014.

[106] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21 (6):1087–1092, 2004.

[107] S. Michel, L. Liang, M. Depner, N. Klopp, A. Ruether, A. Kumar, M. Schedel, C. Vogelberg, E. von Mutius, A. von Berg, et al. Unifying candidate gene and gwas approaches in asthma. *PloS one*, 5(11):e13894, 2010.

[108] C. Misra, Y. Fan, and C. Davatzikos. Baseline and longitudinal patterns of brain atrophy in mci patients, and their use in prediction of short-term conversion to ad: results from adni. *Neuroimage*, 44(4):1415–1422, 2009.

[109] K. Mizuno, M. Wakai, A. Takeda, and G. Sobue. Medial temporal atrophy and memory impairment in early stage of alzheimers disease: an mri volumetric and memory assessment study. *Journal of the neurological sciences*, 173(1):18–24, 2000.

[110] S. Mohamed, Z. Ghahramani, and K. A. Heller. Bayesian exponential family pca. In *Advances in Neural Information Processing Systems*, pages 1089–1096, 2009.

[111] M. Monje, M. Thomason, L. Rigolo, Y. Wang, D. Waber, S. Sallan, and A. J. Golby. Functional and structural differences in the hippocampus associated with memory deficits in adult survivors of acute lymphoblastic leukemia. *Pediatric Blood & Cancer*, 60(2):293–300, 2013.

[112] J. S. Morris, P. J. Brown, R. C. Herrick, K. A. Baggerly, and K. R. Coombes. Bayesian analysis of mass spectrometry proteomic data using wavelet-based functional mixed models. *Biometrics*, 64(2):479–489, 2008.

[113] S. Mueller, D. Wang, M. D. Fox, B. T. Yeo, J. Sepulcre, M. R. Sabuncu, R. Shafee, J. Lu, and H. Liu. Individual variability in functional connectivity architecture of the human brain. *Neuron*, 77(3):586–595, 2013.

[114] H. Müller and U. Stadtmüller. Generalized functional linear models. *Annals of Statistics*, pages 774–805, 2005.

[115] K. R. Murphy, S. M. Landau, K. R. Choudhury, C. A. Hostage, K. S. Shpanskaya, H. I. Sair, J. R. Petrella, T. Z. Wong, and P. M. Doraiswamy. Mapping the effects of apoe4, age and cognitive status on 18f-florbetapir pet measured regional cortical patterns of beta-amyloid density and growth. *Neuroimage*, 78:474–480, 2013.

[116] W. B. Nelson. *Accelerated testing: statistical models, test plans, and data analysis*, volume 344. John Wiley & Sons, 2009.

[117] M. A. Newton, A. Noueiry, D. Sarkar, and P. Ahlquist. Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics*, 5(2): 155–176, 2004.

[118] M. Oguri, K. Kato, K. Yokoi, T. Yoshida, S. Watanabe, N. Metoki, H. Yoshida, K. Satoh, Y. Aoyagi, Y. Nozawa, et al. Assessment of a polymorphism of sdk1 with hypertension in japanese individuals. *American journal of hypertension*, 23(1):70–77, 2010.

[119] K. Okuizumi, O. Onodera, H. Tanaka, H. Kobayashi, S. Tsuji, H. Takahashi, K. Oyanagi, K. Seki, M. Tanaka, S. Naruse, T. Miyatake, H. Mizusawa, and I. Kanazawa. Apoe–ϵ4 and early–onset alzheimer's. *Nature genetics*, 7(1):10–11, 1994.

[120] J. Olazaran, R. Muniz, B. Reisberg, J. Peña-Casanova, T. Del Ser, A. Cruz-Jentoft, P. Serrano, E. Navarro, M. G. de la Rocha, A. Frank, et al. Benefits of cognitive-motor intervention in mci and mild to moderate alzheimer disease. *Neurology*, 63(12): 2348–2353, 2004.

[121] B. Patenaude, S. M. Smith, D. N. Kennedy, and M. Jenkinson. A bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage*, 56(3):907–922, 2011.

[122] Y.-H. Peng, B.-R. Wu, C.-H. Su, W.-C. Liao, C.-H. Muo, T.-C. Hsia, and C.-H. Kao. Adult asthma increases dementia risk: a nationwide cohort study. *Journal of epidemiology and community health*, 69(2):123–128, 2015.

[123] C. Pennanen, M. Kivipelto, S. Tuomainen, P. Hartikainen, T. Hänninen, M. Laakso, M. Hallikainen, M. Vanhanen, A. Nissinen, E. Helkala, P. Vainio, R. Vanninen, K. Partanen, and H. Soininen. Hippocampus and entorhinal cortex in mild cognitive impairment and early ad. *Neurobiology of Aging*, 25(3):303–310, 2004.

[124] R. Perri, L. Serra, G. A. Carlesimo, and C. Caltagirone. Amnestic mild cognitive impairment: difference of memory profile in subjects who converted or did not convert to alzheimer's disease. *Neuropsychology*, 21(5):549–558, 2007.

[125] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen. Mild cognitive impairment: clinical characterization and outcome. *Archives of neurology*, 56(3):303–308, 1999.

[126] R. C. Petersen, R. Thomas, M. Grundman, D. Bennett, R. Doody, S. Ferris, D. Galasko, S. Jin, J. Kaye, A. Levey, E. Pfeiffer, M. Sano, C. van Dyck, L. Thal, and ADCSG. Vitamin e and donepezil for the treatment of mild cognitive impairment. *New England Journal of Medicine*, 352(23):2379–2388, 2005.

[127] S. P. Poulin, R. Dautoff, J. C. Morris, L. F. Barrett, B. C. Dickerson, and ADNI. Amygdala atrophy is prominent in early alzheimer's disease and relates to symptom severity. *Psychiatry Research: Neuroimaging*, 194(1):7–13, 2011.

[128] A. Prestia, A. Caroli, W. M. van der Flier, R. Ossenkoppele, B. Van Berckel, F. Barkhof, C. E. Teunissen, A. E. Wall, S. F. Carter, M. Schöll, I. Choo, A. Nordberg, P. Scheltens, and G. Frisoni. Prediction of dementia in mci patients based on core diagnostic markers for alzheimer disease. *Neurology*, 80(11):1048–1056, 2013.

[129] A. L. Price, N. A. Zaitlen, D. Reich, and N. Patterson. New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, 11(7):459–463, 2010.

[130] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. De Bakker, M. J. Daly, et al. Plink: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3):559–575, 2007.

[131] J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition, 2005. ISBN 978-0387-40080-8; 0-387-40080-X.

[132] P. T. Reiss and R. T. Ogden. Functional generalized linear models with images as predictors. *Biometrics*, 66:61–69, 2010.

[133] S. Risacher, A. Saykin, J. D. Wes, L. Shen, H. A. Firpi, and B. McDonald. Baseline mri predictors of conversion from mci to probable ad in the adni cohort. *Current Alzheimer Research*, 6(4):347–361, 2009.

[134] W. G. Rosen, R. C. Mohs, and K. L. Davis. A new rating scale for alzheimer's disease. *The American Journal of Psychiatry*, 141:1356–1364, 1984.

[135] A. Saunders, W. Strittmatter, D. Schmechel, P. S. George-Hyslop, M. Pericak-Vance, S. Joo, B. Rosi, J. Gusella, D. Crapper-MacLachlan, M. Alberts, C. Hulette, B. Crain, D. Goldgaber, and A. D. Roses. Association of apolipoprotein e allele 4 with late-onset familial and sporadic alzheimer's disease. *Neurology*, 43(8):1467–1467, 1993.

[136] P. Scheltens, D. Leys, F. Barkhof, D. Huglo, H. Weinstein, P. Vermersch, M. Kuiper, M. Steinling, E. C. Wolters, and J. Valk. Atrophy of medial temporal lobes on mri in" probable" alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *Journal of Neurology, Neurosurgery & Psychiatry*, 55(10):967–972, 1992.

[137] A. Schwartzman. Random ellipsoids and false discovery rates: Statistics for diffusion tensor imaging data. *Ph.D. thesis, Stanford University*, 2006.

[138] N. Sha, M. G. Tadesse, and M. Vannucci. Bayesian variable selection for the analysis of microarray data with censored outcomes. *Bioinformatics*, 22(18):2262–2268, 2006.

[139] L. Shaw, H. Vanderstichele, M. Knapik-Czajka, C. Clark, P. Aisen, R. Petersen, K. Blennow, H. Soares, A. Simon, P. Lewczuk, R. Dean, E. Siemers, W. Potter, V. Lee, J. Trojanowski, and ADNI. Cerebrospinal fluid biomarker signature in alzheimer's disease neuroimaging initiative subjects. *Annals of Neurology*, 65(4):403–413, 2009.

[140] J. Shi, P. Thompson, B. Gutman, and Y. Wang. Surface fluid registration of conformal representation: Application to detect disease burden and genetic influence on hippocampus. *NeuroImage*, 78:111–134, 2013.

[141] J. Shi, Y. Wang, R. Ceschin, X. An, Y. Lao, D. Vanderbilt, M. D. Nelson, P. Thompson, A. Panigrahy, and N. Leporé. A multivariate surface-based analysis of the putamen in premature newborns: Regional differences within the ventral striatum. *PloS One*, 8(7):e66736, 2013.

[142] J. Shi, N. Lepor, B. A. Gutman, P. M. Thompson, L. C. Baxter, R. L. Caselli, Y. Wang, and ADNI. Genetic influence of apolipoprotein e4 genotype on hippocampal morphometry: An n = 725 surface-based alzheimer's disease neuroimaging initiative study. *Human Brain Mapping*, pages 3903–3918, 2014. ISSN 1097-0193. doi: 10.1002/hbm.22447. URL `http://dx.doi.org/10.1002/hbm.22447`.

[143] D. Sinha, M.-H. Chen, and S. K. Ghosh. Bayesian analysis and model selection for interval-censored survival data. *Biometrics*, 55(2):585–590, 1999.

[144] D. Sinha, J. G. Ibrahim, and M. H. Chen. A Bayesian justification of Cox's partial likelihood. *Biometrika*, 90:629–641, 2003.

[145] I. Skoog and D. Gustafson. Update on hypertension and alzheimer's disease. *Neurological research*, 28(6):605–611, 2006.

[146] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

[147] J. D. Storey. The positive false discovery rate: a bayesian interpretation and the q-value. *Annals of statistics*, pages 2013–2035, 2003.

[148] W. J. Strittmatter, A. M. Saunders, D. Schmechel, M. Pericak-Vance, J. Enghild, G. S. Salvesen, and A. D. Roses. Apolipoprotein e: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial alzheimer disease. *Proceedings of the National Academy of Sciences*, 90(5):1977–1981, 1993.

[149] M. Tabert, J. Manly, X. Liu, G. H. Pelton, S. Rosenblum, M. Jacobs, D. Zamora, M. Goodkind, K. Bell, Y. Stern, and D. Devanand. Neuropsychological prediction of conversion to alzheimer disease in patients with mild cognitive impairment. *Archives of General Psychiatry*, 63(8):916–924, 2006.

[150] M. A. Tanner and W. H. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398):528–540, 1987.

[151] P. M. Thompson, K. M. Hayashi, G. I. de Zubicaray, A. L. Janke, S. E. Rose, J. Semple, M. S. Hong, D. H. Herman, D. Gravano, D. M. Doddrell, et al. Mapping hippocampal and ventricular change in alzheimer disease. *Neuroimage*, 22(4):1754–1766, 2004.

[152] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of Royal Statistical Society. Series B.*, 58(1):267–288, 1996. ISSN 0035-9246.

[153] M. P. Van Den Heuvel and H. E. H. Pol. Exploring the brain network: a review on resting-state fmri functional connectivity. *European Neuropsychopharmacology*, 20(8): 519–534, 2010.

[154] P. Vemuri, J. L. Gunter, M. L. Senjem, J. L. Whitwell, K. Kantarci, D. S. Knopman, B. F. Boeve, R. C. Petersen, and C. R. Jack Jr. Alzheimer's disease diagnosis in individual subjects using structural mr images: validation studies. *Neuroimage*, 39(3): 1186–1197, 2008.

[155] S. Walker and B. K. Mallick. A bayesian semiparametric accelerated failure time model. *Biometrics*, 55(2):477–483, 1999.

[156] H. Wang, A. Banerjee, and D. Boley. Common component analysis for multiple covariance matrices. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 956–964. ACM, 2011.

[157] K. Wang, M. Liang, L. Wang, L. Tian, X. Zhang, K. Li, and T. Jiang. Altered functional connectivity in early alzheimer's disease: A resting-state fmri study. *Human brain mapping*, 28(10):967–978, 2007.

[158] L. Wang, Y. Zang, Y. He, M. Liang, X. Zhang, L. Tian, T. Wu, T. Jiang, and K. Li. Changes in hippocampal connectivity in the early stages of alzheimer's disease: evidence from resting state fmri. *Neuroimage*, 31(2):496–504, 2006.

[159] S. Wang, B. Nan, J. Zhu, and D. G. Beer. Doubly penalized buckley–james method for survival data with high-dimensional covariates. *Biometrics*, 64(1):132–140, 2008.

[160] Y. Wang, J. Zhang, B. Gutman, T. F. Chan, J. T. Becker, H. Aizenstein, O. Lopez, R. Tamburo, A. Toga, and P. Thompson. Multivariate tensor-based morphometry on surfaces: application to mapping ventricular abnormalities in hiv/aids. *NeuroImage*, 49(3):2141–2157, 2010.

[161] Y. Wang, Y. Song, P. Rajagopalan, T. An, K. Liu, Y.-Y. Chou, B. Gutman, A. W. Toga, and P. M. Thompson. Surface-based tbm boosts power to detect disease effects on the brain: an n= 804 adni study. *Neuroimage*, 56(4):1993–2010, 2011.

[162] S. T. Weiss, B. A. Raby, and A. Rogers. Asthma genetics and genomics 2009. *Current opinion in genetics & development*, 19(3):279–282, 2009.

[163] T. Wu, Y. Chen, T. Hastie, E. Sobel, and K. Lange. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics*, 25(6):714–721, 2009.

[164] M. Xia, J. Wang, and Y. He. Brainnet viewer: a network visualization tool for human brain connectomics. *PloS one*, 8(7):e68910, 2013.

[165] F. Yao, H. G. Muller, and J. L. Wang. Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33:2873–2903, 2005.

[166] J. Young, M. Modat, M. J. Cardoso, A. Mendelson, D. Cash, and S. Ourselin. Accurate multimodal probabilistic prediction of conversion to alzheimer's disease in patients with mild cognitive impairment. *NeuroImage: Clinical*, 2:735–745, 2013.

[167] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.

[168] A. Zellner. On assessing prior distributions and bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques: Essays in Honor of Bruno De Finetti*, 6:233–243, 1986.

[169] C. Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942, 2010.

[170] D. Zhang, D. Shen, and ADNI. Predicting future clinical changes of mci patients using longitudinal and multimodal biomarkers. *PloS One*, 7(3):e33182, 2012.

[171] H. H. Zhang and W. Lu. Adaptive lasso for cox's proportional hazards model. *Biometrika*, 94(3):691–703, 2007.

[172] X. Zhang, F. Xue, H. Liu, D. Zhu, B. Peng, J. L. Wiemels, and X. Yang. Integrative bayesian variable selection with gene-based informative priors for genome-wide association studies. *BMC genetics*, 15(1):1, 2014.

[173] S. D. Zhao and Y. Li. Principled sure independence screening for cox models with ultra-high-dimensional covariates. *Journal of multivariate analysis*, 105(1):397–411, 2012.