Bayesian Multimodal Local False Discovery Rate in Neuroimaging Studies

BY

YUE WANG
B.S., Ocean University of China, Qingdao, China, 1998
Ph.D., University of Nebraska at Lincoln, Lincoln, Nebraska, 2005
M.S., University of Nebraska at Lincoln, Lincoln, Nebraska, 2007

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Public Health Sciences
in the Graduate College of the
University of Illinois at Chicago, 2020

Chicago, Illinois

Defense Committee:

Dulal K. Bhaumik, Chair and Advisor
Robert T. Krafty, University of Pittsburgh
Runa Bhaumik
Olusola A. Ajilore, Psychiatry
Theresa L. Bender Pape, Edward Hines, Jr. VA Hospital

This thesis is dedicated to my family.

ACKNOWLEDGMENTS

My study and research for Ph.D. in Biostatistics at University of Illinois at Chicago School of Public Health has been a challenging, intense but absolutely worthy and rewarding journey, both academically and personally. I would like to take this opportunity to reflect on the people who have made this possible.

First, I would like to thank my thesis adviser, Dr. Dulal Bhaumik, a well-respected and dedicated professor, passionate teacher and talented statistician. I am particularly appreciative for your strong mentorship, structured guidance, insight, and kindly support throughout my studies. You consistently and patiently steered me in the right direction, provided critical advises, review and careful attention to details on the research. You have shared with me your wisdom and experiences, showed me your sharpness, prudence and passion as a researcher in theoretical and applied statistics. What I have learnt most from you is your humbleness, kindness, vision and continuous learning, and I shall never stop being a student, and never stop learning.

I would like to given special thanks to my thesis committee members Drs. Robert Krafty, Runa Bhaumik, Olusola Ajilore and Theresa Pape for dedication of your time to provide a thorough review, insightful and constructive discussion and comments on my thesis. I am very grateful for all kindly support from you, especially during this Covid-19 pandemic.

My sincere thanks also go to the faculty members Drs. Huayun Chen, Hakan Demirtas and Hui Xie for their teaching and instructions to help us build up a solid foundation and deep understanding in various statistics topics, Dr. Li Liu for being my first academic adviser and providing great mentorship, guidance and support during the first two years, and Drs. Sanjib

Basu and Saria Awadalla for their advise and help through my study, the staff members Liliana Aceves, Jacob Wilson, Dan Albert, Jose Alvarez, and Linda Johnson for their assistance over the years, and Nina Sandlin for her thorough format review on my thesis. I also want to thank my graduate colleagues Drs. Rawan Rupnow, Fei Jie, Yuanbo Song, Weihan Zhao, Nairita Ghosal and Rachel Nordgren and for their kindly discussion, help and collaboration.

I would like to thank my ex-advisers Drs. Katheryn Hanford, Stephen Kachman, Lawrence Harshman, and L Dale Van Vleck, at University of Nebraska at Lincoln for their inspiration, enthusiasm, continuous encouragement and support.

I want to express my gratitude to my colleagues Chaohui Wang, Di Li, Narinder Nangia, Martin King, Yili Pritchett, Rita Kristy, Mary Beth Blauwet, Weizhong He, James Myles, Peter Bonate, Kevin Chartier and Susanne Gronen for their advise, encouragement, support and help over the years, which has enabled me to enhance my technical skill and maintain statistical sharpness as a clinical statistician in pharmaceutical industry.

Finally, I would like to thank and dedicate this thesis to my family. To my parents Zeyin Wang and Jiazhen Huang, my husband Cedric and my sister Le for their love, patience, understanding and supporting me in all of my pursuits. My parents have taught us the endeavor to leave the world a little better, showed us the importance and value of hard work and integrity, the power of science and education, always had faith in me. To my daughter Chloé for motivating and inspiring me to follow my dreams. Your love, curious, kind and happy nature, and laughter keep me smiling every day and help me persist through challenges and difficult times. I hope that you will see, understand and eventually live by those essential values I inherited from my parents and learned from these great people with beautiful minds I have had pleasure to know.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF ABBREVIATIONS

AD	Alzheimer's Disease
AN	Affective Network
BA	Brodmann Area
BOLD	Blood-Oxygen-Level Dependence
BUGS	Bayesian Inference Using Gibbs Sampling
CBF	Cerebral Blood Flow
CEN	Central Executive Network
dACC	Dorsal Anterior Cingulate Cortex
dlPFC	Dorsolateral Prefrontal Cortex
DMN	Default Mode Network
DP	Dirichlet Process
DPM	Dirichlet Process Mixture
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders, 4^{th} edition
DTI	Diffusion Tensor Imaging
DWI	Diffusion-Weighted Imaging
EB	Empirical Bayes
EC	Effective Connectivity
EPI	Echo-Planar Imaging
FA	Fractional Anisotropy
FC	Functional Connectivity

LIST OF ABBREVIATIONS (Continued)

FDA	Food and Drug Administration
FDR	False Discovery Rate
FFA	Fusiform Face Area
FG	Fusiform Gyrus
fMRI	Functional Magnetic Resonance Imaging
FOV	Field of View
FPR	False Positive Rate
FWER	Family-Wise Error Rate
GLM	Generalized Liner Model
HC	Healthy Control
HDRS	Hamilton Depression Rating Scale
IPL	Inferior Parietal Lobe
JAGS	Just Another Gibbs Sampler
LASSO	Least Absolute Shrinkage and Selection Operator
Lfdr	Local false discovery rate
LLD	Late-Life Depression
MCMC	Markov Chain Monte Carlo
MD	Mean Diffusivity
MDD	Major Depression Disorder
MERM	Mixed-Effects Regression Model
MML	Maximum Marginal Likelihood
MRI	Magnetic Resonance Imaging
NDR	Non-Discovery Rate

LIST OF ABBREVIATIONS (Continued)

PCC	Posterior Cingulate Cortex
RN	Reward Network
ROI	Region of Interest
rs-fMRI	resting-state Functional Magnetic Resonance Imaging
rTMS	Repetitive Transcranial Magnetic Stimulation
SC	Structural Connectivity
SCID-IV	Structured Clinical Interview for DSM-IV
SENSE	Sensitivity-Encoding
SN	Salience Network
TE	Echo Time
TR	Repetition Time
UIC	University of Illinois at Chicago
VN	Visual Network

SUMMARY

In brain connectivity research one of the primary objectives is to identify differential connectivities that may be potentially associated with the underlying pathophysiology of neurological and psychiatric disorders, and further enhance our knowledge to study the complex brain structure, function and the relationship with human behavior. For exploratory neuroimaging studies to compare disease group and healthy control group, the main purpose is to detect any signal and thus thousands of hypotheses are tested simultaneously, which is known as the large-scale simultaneous hypothesis testing problem. Most of the hypotheses tested are null, that is, nothing but noise, while a very small number of them may contain true signals. Over the last three decades, the false discovery rate (FDR) and local FDR (Lfdr) methods have been developed to address this problem (Benjamini and Hochberg, 1995; Efron et al., 2001; Efron, 2007). Typical neuroimaging studies usually have small sample size due to the high economic cost, leading to low statistical power and high probability of falsely significant findings (Szucs and Ioannidis, 2019). The existing methodologies do not provide a satisfied control of FDR especially for small sample size studies, neither capabilities that allow integration of multimodal neuroimaging data into statistical modeling.

The information provided by multimodal imaging techniques can be complementary to each other and thus integrated multimodal analysis enables us to borrow strength from different modalities. A covariate-modulated Lfdr method has been used in genome-wide association studies to incorporate functional genome annotations as covariates in the model and proved to be efficient by increasing power (Zablocki et al., 2014; Torkamani et al., 2011). We extend this method to multimodal neuroimaging data, aimed to improve FDR control and sensitivity

to detect differential functional connectivity (FC) links between disease group and healthy control group in a cross-sectional, comparative, multimodal neuroimaging study with small sample size. We implement a Bayesian multimodal Lfdr approach, which utilizes a Bayesian mixture model to leverage complementary structural connectivity (SC) statistics and enhance the modeling of the density of FC statistics. The approach is general and can be useful in a broad spectrum of applications. The utility of Bayesian multimodal approach is illustrated with extensive simulation study and a neuroimaging study in late-life depression (LLD) in which both FC (measured using resting-state functional magnetic resonance imaging) and SC (measured using diffusion tensor imaging) data were measured on each of the 23 participants (13 healthy subjects and 10 LLD patients). We demonstrate in simulation study, that Bayesian multimodal Lfdr method performs numerically better in terms of FDR control by comparison with the traditional Lfdr method that solely considers FC especially when sample size is small.

Results from the LLD neuroimaging data analysis using Bayesian multimodal Lfdr method are (i) that the right dorsolateral prefrontal cortex (dlPFC), a major node in central executive network (CEN), is identified as a primary hub region that has significantly increased FC to the other seven regions in LLD patients as compared to healthy controls; (ii) that disrupted FC pattern associated with LLD are found within and across the major large-scale neurocognitive and internally-guided resting-state brain networks including CEN, default mode network (DMN) and salience network (SN). The results suggest that degeneration of specific hubs within the CEN, DMN, SN and fusiform face area for facial cognition may contribute to a deterioration of memory and cognitive functions in LLD.

In addition, we have employed a Bayesian multiple comparison method via a non-parametric Bayesian Dirichlet process mixture (DPM) model directly on FC data in the LLD neuroimag-

ing study, as a comparison with the results from Bayesian multimodal Lfdr method. Bayesian DPM model imposes no model assumptions and can also mimic hypothesis testing framework while preserving FDR due to the model-based clustering properties. The results consistently suggest the dlPFC within the CEN as a primary hub with enhanced FCs to regions in the CEN and DMN for LLD. Besides, increased FCs in affective network and sensorimotor network are identified to be associated with LLD.

CHAPTER 1

INTRODUCTION

Human brain is not only the command center of human body, but it is also the most complex dynamic network and biologic system that integrates structural and functional information (Sporns et al., 2005). Human brain consists of 86 billion neurons (i.e., nerve cells) on average (Azevedo et al., 2009). As the basic structure in the central nervous system, the neurons are specialized to process and transmit intracellular signals (communication within the cell) and intercellular signals (communication between cells). Their communication with each other is by the movement of chemicals or electrical signals across synapses, that are the connections between neurons through which the information flows from one neuron to another. Each neuron has approximately 10,000 synapses, that is, connections with 10,000 other neurons. Hence, the neurons forms a a network (known as neural network) to process all information in and out of the central nervous system. For centuries, many research scientists have put great efforts to study human brain in order to have a detailed understanding on how neurons and neural networks work by reading in information, transforming those information into processes, and producing guidance and control on the way we think, feel and behave. Brain is analogous to a super computer, or more precisely, a symphony orchestra in which the neural networks work synchronously in a specifically complex and coordinated way like musicians to produce rhythms as a whole and a tiny out-of-tune disruption can lead to dysfunction and further disorder.

The advances in modern magnetic resonance imaging (MRI) techniques and statistical methodologies for neuroimaging data analysis have enabled the brain research to progress tremendously during the last two decades. However, to address the exquisitely complicated

relationship between brain structure, function and human intelligence and behavior remains a compelling scientific challenge. In the last ten years, the scientific society worldwide has realized that this sophisticated puzzle cannot be resolved with knowledge and efforts from one academic area and started forming interdisciplinary collaborations among expertise from neuroscience, medicine, psychiatry, psychology, mathematics, statistics, computer sciences, engineering, etc.. In 2009, the National Institute of Health launched the Human Connectome Project with an ambitious goal to build a comprehensive map of neural connections in the brain (known as connectome) that will enhance our knowledge on the structural and functional connectivities in healthy human brain and shed light on the pathophysiological activities in the study of brain, behavior, and clinical disorders (Van Essen et al., 2013). Furthermore, in 2013, the European Union started the Human Brain Project, a large ten-year scientific research project, to build a research infrastructure in neuroscience, statistical computing and medicine (Salles et al., 2019).

One of the key aspects of brain research is to study brain connectivity patterns and networks that are specifically involved in the underlying pathology of neurological and psychiatric disorders (Rowe, 2010). Brain connectivity refers to the interactions between two different regions within the cortical and subcortical networks in the brain. Changes in the brain connectivities can relate to dysfunction of certain cognitive processes leading to neurological disorders. There are three main categories of brain connectivities, which are structural, functional and effective connectivity. By definition, structural connectivity (SC) is the structural (or anatomical) link such as fiber pathways between two discrete brain regions, which typically corresponds to the white matter tracts connecting cortical areas/nuclei. Functional connectivity (FC) is defined as "the temporal correlation (or statistical dependency) between spatially remote neurophysiological events" (Friston et al., 1993; Friston, 1994). That is, FC is a "statistical concept"

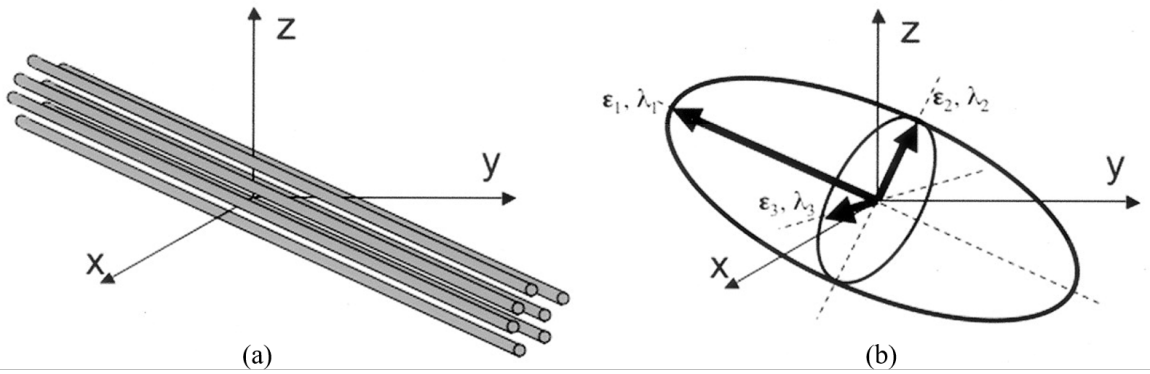
that reflects "the synchronized and correlated patterns of neural activity between two regions that are structurally connected or unconnected" (Friston, 1994). The third category, effective connectivity (EC) refers to "the causal influence of one neural element over another either directly or indirectly" (Friston et al., 1993), that is, a signal or activation in one region directly or indirectly causes change, signal, activation or suppression in another region. Similar to FC, EC can also be found between two structurally unconnected brain regions. Though not clear, the relationships among the three modes of brain connectivity might be described as both segregated and integrated (Tononi et al., 1994). In this thesis, we focus on integration of SC and FC data to help improve false discovery rate (FDR) control to identify disrupted FCs associate with the disorder.

The MRI has been regarded as a milestone breakthrough in medical diagnostics and research. In 2003, Professors Paul Lauterbur and Peter Mansfield were awarded the Nobel Prize in Physiology or Medicine for their discoveries of MRI in 1970's. The MRI technique provides objective and precise measurements of brain activities and it continuously gains popularity, as a result neuroimaging studies have been widely used to investigate brain connectivity within network and between networks in both clinical and cognitive neuroscience research areas since then. Brain connectivity studies are of particular interest to elucidate connectivity patterns associated with pathophysiology of neurodegenerative, psychiatric and neurological disorders such as attention-deficit hyperactivity disorder, autism, dementia, depression, schizophrenia, bipolar, post-traumatic stress disorder, traumatic brain injury, stroke, Alzheimer's disease and epilepsy.

The *in vivo* neuroimaging techniques include structural imaging modalities (for example, diffuse tensor imaging) and functional imaging modalities (for example, functional magnetic

resonance imaging). Structural connectivity can be measured using diffusion weighted imaging (DWI) and the T1-based connectivity. Using DWI, the connectivity matrices can be built based on SC inferred from white matter connections between two brain regions. Physical properties of the fibers like fractional anisotropy and mean diffusivity as well as number of crossing fibers (i.e., percentage of all fibers passing from or ending in two brain regions) can be used. The connectivity matrices based on T1 are built using correlation coefficient of sizes of the two brain regions. Diffusion tensor imaging (DTI) is an anatomical MRI technique that is used to provide detailed information on the microstructure of brain's white matter tracts such as the tracts' position, orientation and anisotropy based on measurement of the diffusion of water molecules along different direction within the brain tissues using a diffusion tensor that elucidates behavior of anisotropic diffusion, in which model diffusion is assumed to follow multivariate normal distribution (Basser et al., 1994).

Approximately 73% of the human brain is composed of water (Mitchell et al., 1945). Diffusion magnetic resonance imaging captures diffusion of water molecules in tissues, here diffusion refers to the random motion driven by thermal energy, or Brownie motion. In an unrestricted environment, the water molecules move randomly and diffusion is same in every direction, which is referred to as isotropy. On the other hand, in a constrained environment, the water molecules move along one axis, which is referred to as anisotropy. In the presence of isotropy, diffusion can be characterized by a single diffusion coefficient D. While in the presence of anisotropy, diffusion cannot be described by a single number, but a three-dimensional diffusion tensor assuming the shape of an ellipsoid as illustrated in Figure 1.



Source: Jellison et al. (2004)

Figure 1. A three-dimensional diffusion tensor assuming the shape of an ellipsoid

(a) Along a fiber tract. (b) In the three-dimensional diffusivity model as an ellipsoid with the degree of diffusivity along the three axes characterized by three eigenvalues (λ_1, λ_2, λ_3) and orientation characterized by corresponding eigenvectors ($\boldsymbol{\epsilon}_1$, $\boldsymbol{\epsilon}_2$, $\boldsymbol{\epsilon}_3$).

The mathematical model to describe the ellipsoid shape for diffuse tensor consists of a 3×3 symmetrical matrix with 6 degrees of freedom (i.e., 6 elements are independent), indicating that constructing a diffusion requires at least 6 directions,

$$\boldsymbol{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}, \tag{1.1}$$

where the three diagonal elements of the tensor matrix denoted by D_{xx}, D_{yy}, and D_{zz} represent the "diffusion coefficients measured in the laboratory frame of reference along each of the principal directions" X, Y and Z, respectively, while the rest six off-diagonal elements denoted by D_{xy}, D_{yz}, D_{xz}, D_{yx}, D_{zy} and D_{zx} indicate "correlations between random motions corresponding to each pair of principal directions" (Elster). The tensor matrix $\boldsymbol{D}$ can be diagonalized into

a diagonal matrix $\boldsymbol{\Lambda}$ containing three positive eigenvalues (λ_1, λ_2, λ_3) (in decreasing order, i.e., $\lambda_1 > \lambda_2 > \lambda_3$) and a matrix $\boldsymbol{E}$ composed of three corresponding linearly independent eigenvectors ($\boldsymbol{\epsilon}_1$, $\boldsymbol{\epsilon}_2$, $\boldsymbol{\epsilon}_3$),

$$\boldsymbol{D} = \boldsymbol{E}^{-1}\boldsymbol{\Lambda}\boldsymbol{E}, \text{ where}$$

$$\boldsymbol{E} = \begin{bmatrix} \boldsymbol{\epsilon}_1 & \boldsymbol{\epsilon}_2 & \boldsymbol{\epsilon}_3 \end{bmatrix} \text{ and } \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}, \tag{1.2}$$

As $\boldsymbol{D}$ is symmetric and $\boldsymbol{D}$ is diagonalized as in Equation (1.2), the eigenvector matrix $\boldsymbol{E}$ is orthogonal, $\boldsymbol{E}^{-1} = \boldsymbol{E}^T$. The three eigenvalues assess the degree of diffusivity along the three axes with orientations defined by the corresponding eigenvectors. The primary eigenvector corresponding to the maximum of the three eigenvalues determines the direction of maximum diffusivity, based on which the orientation of fiber tracts can be traced. These calculated eigenvalues and eigenvectors can be used to create images reflecting various diffusion properties of a tissue.

Mean diffusivity (MD) and fractional anisotropy (FA) are the two main indexes based on the eigenvalues to represent the magnitude of diffusion (Conturo et al., 1996; Pierpaoli and Basser, 1996; Mukherjee et al., 2008). The MD is the average of the three eigenvalues, $MD = (\lambda_1 + \lambda_2 + \lambda_3)/3 = trace/3$. The FA is a scaled measure of the degree of anisotropic diffusion within a voxel (a voxel represents a very small cube of brain tissue composed of approximately a few million neurons), defined in terms of its eigenvalues by,

$$FA = \sqrt{\frac{3}{2}}\sqrt{\frac{(\lambda_1 - MD)^2 + (\lambda_2 - MD)^2 + (\lambda_1 - MD)^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}, \tag{1.3}$$

where the scaling factor $\sqrt{\frac{3}{2}}$ is used to make the value of FA to have a range between 0 and 1. In case of isotropic diffusion, where $\lambda_1 = \lambda_2 = \lambda_3$, "the diffusion ellipsoid is a sphere" and FA = 0. With higher degree of diffusion anisotropy, "the eigenvalues becomes more unequal, the ellipsoid becomes more elongated in one direction" and FA goes to 1 (Elster). Note that FA contains no information about the orientation, so it is invariant to rotation. The FA value is calculated for each voxel, based on which an FA map can be generated.

Functional magnetic resonance imaging (fMRI) measures intrinsic brain activity based on synchronous blood-oxygen-level dependent (BOLD) correlation. It detects changes in cerebral hemodynamic flow associated with neural activity in the cortex of the human brain, which is known as neurovascular coupling, both in the resting-state and task-state. Ogawa et al. (1990) first demonstrated that BOLD signal can measure the inhomogeneities in magnetic field due to changes in the level of oxygen in the blood, and thus suggested the use of MRI as a non-invasive yet *in vivo* (i.e., within living organism) study of brain function. Hemoglobin is an iron-containing protein in the blood that transports oxygen to the tissues. It is made up of "4 polypeptide chains (α_1, α_2, β_1 and β_2)" (Encyclopaedia Britannica, 2020). Each chain is attached to a hemo group, at the center of which is an iron atom (Fe). Each iron ion (Fe^{2+}) can attach an oxygen molecule. Thus hemoglobin has two forms: Oxyhemoglobin is the form of hemoglobin with the bound of oxygen, while deoxyhemoglobin is the form of hemoglobin without the bound of oxygen. The two forms of hemoglobin exhibit different magnetic properties.

Oxyhemoglobin has no unpaired electrons and is weakly diamagnetic (Pauling and Coryell, 1936), it is therefore repelled by magnetic fields. In contrast, deoxyhemoglobin has 4 unpaired electrons at each iron center, and thus exhibits strong paramagnetism and creates local magnetic field distortions (also called susceptibility gradients) in and around the blood vessels. In terms of magnetic susceptibility, which is a quantitative measure of the extent to which a material may be magnetized in an applied magnetic field, paramagnetic deoxyhemoglobin has positive magnetic susceptibility and increases a magnetic field. As the ratio of deoxyhemoglobin to oxyhemoglobin decreases caused a reduction in the magnetic susceptibility, the transverse relaxation time (i.e., the decay of the MRI signal induced by the processing transverse nuclear magnetization, in gradient-echo, referred to as T2*) becomes longer, and hence magnetic resonance (MR) signal intensity increases. The mechanism of BOLD fMRI is depicted in Figure 2.

When a brain region receives a stimulus corresponding to a particular task, it becomes more active and the blood flows at a higher rate that requires more local supply of oxygen and glucose, which leads to an increase of local oxygen and glucose consumption in the brain tissue that is followed by an increase in cerebral blood flow (CBF). The percentage increases in CBF and glucose consumption are comparable and much higher than oxygen consumption (Fox and Raichle, 1986; Fox et al., 1988), resulting in an oversupply of oxygen that forms the basis of BOLD fMRI. Attwell and Iadecola (2002) suggested that the CBF increase associated with the neural activity may not be a result of the increased energy metabolism demands, but controlled by neural signaling mechanisms that require further investigation. Therefore, weakly diamagnetic oxyhemoglobin releases oxygen and becomes strongly paramagnetic deoxyhemoglobin, leading to an increase of the concentration of deoxyhemoglobin. Due to the

paramagnetic property of deoxyhemoglobin, the increase in deoxyhemoglobin concentration can reduce the BOLD signal and the change of BOLD signal can be detected by fMRI.

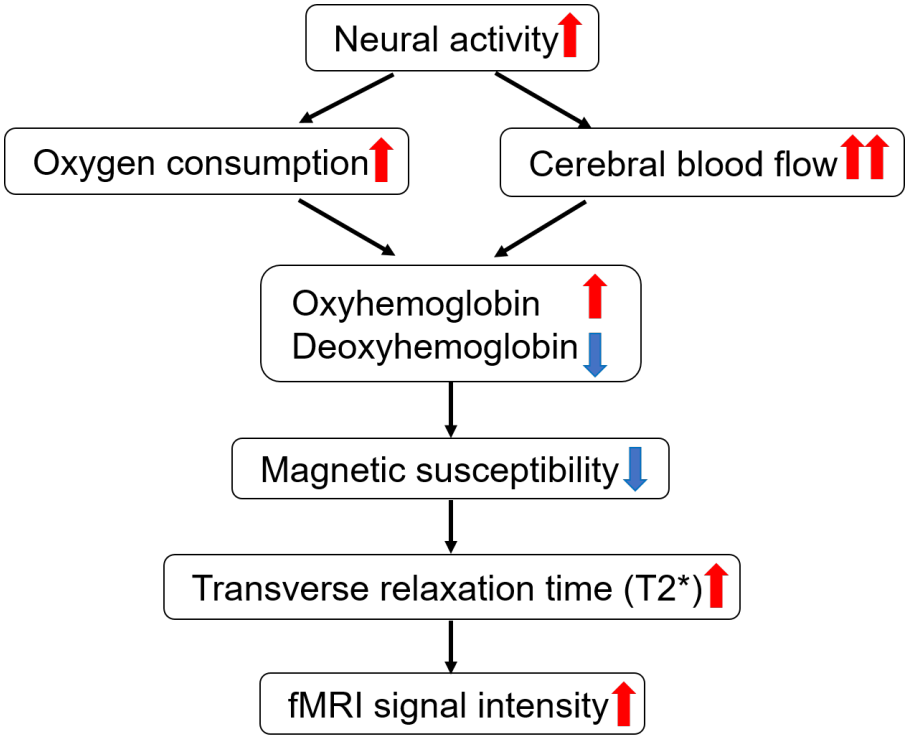
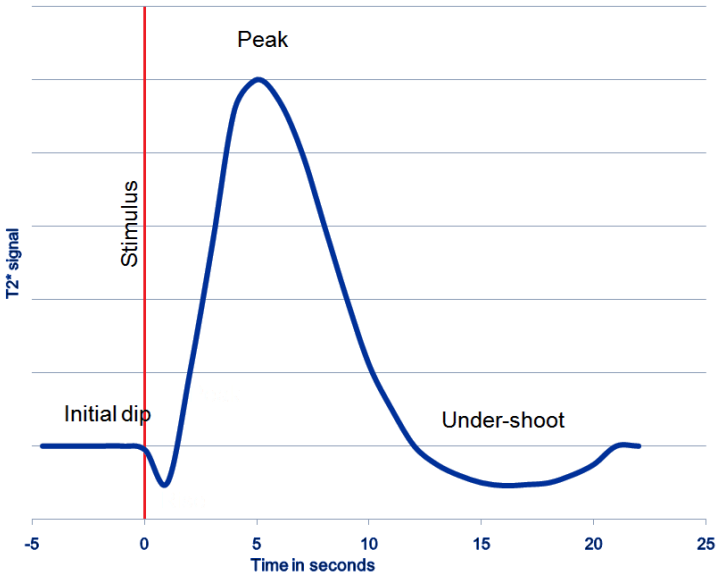


Figure 2. Mechanism of BOLD fMRI

As shown in Figure 3, the time course of a typical BOLD signal in response to neural activity can be characterized by three phases: initial dip, peak and post-stimulus under-shoot. Immediately following neural activity, as a result of increase in local oxygen consumption the concentration of oxyhemoglobin deceases, which causes an initial small and negative dip of BOLD signal. In about 2-3 seconds, the signal drops to its lowest point. To compensate for the increased demand for oxygen, the vascular system increases CBF with increased amount

of oxyhemoglobin relative to the deoxyhemoglobin. Because the deoxyhemoglobin attenuates the MR signal an increase in the relation between oxyhemoglobin and deoxyhemoglobin leads to an increase of the BOLD signal. In about 4-6 seconds, the signal reaches to a peak. This positive BOLD signal phase is the main focus of fMRI, and the underlying reason why more oxygen is delivered to these vessels than is needed is still unclear. If there is no more stimulus, the signal gradually declines to a point that is lower than the baseline and then slowly returns to the baseline level, which is known as post-stimulus under-shoot phase.



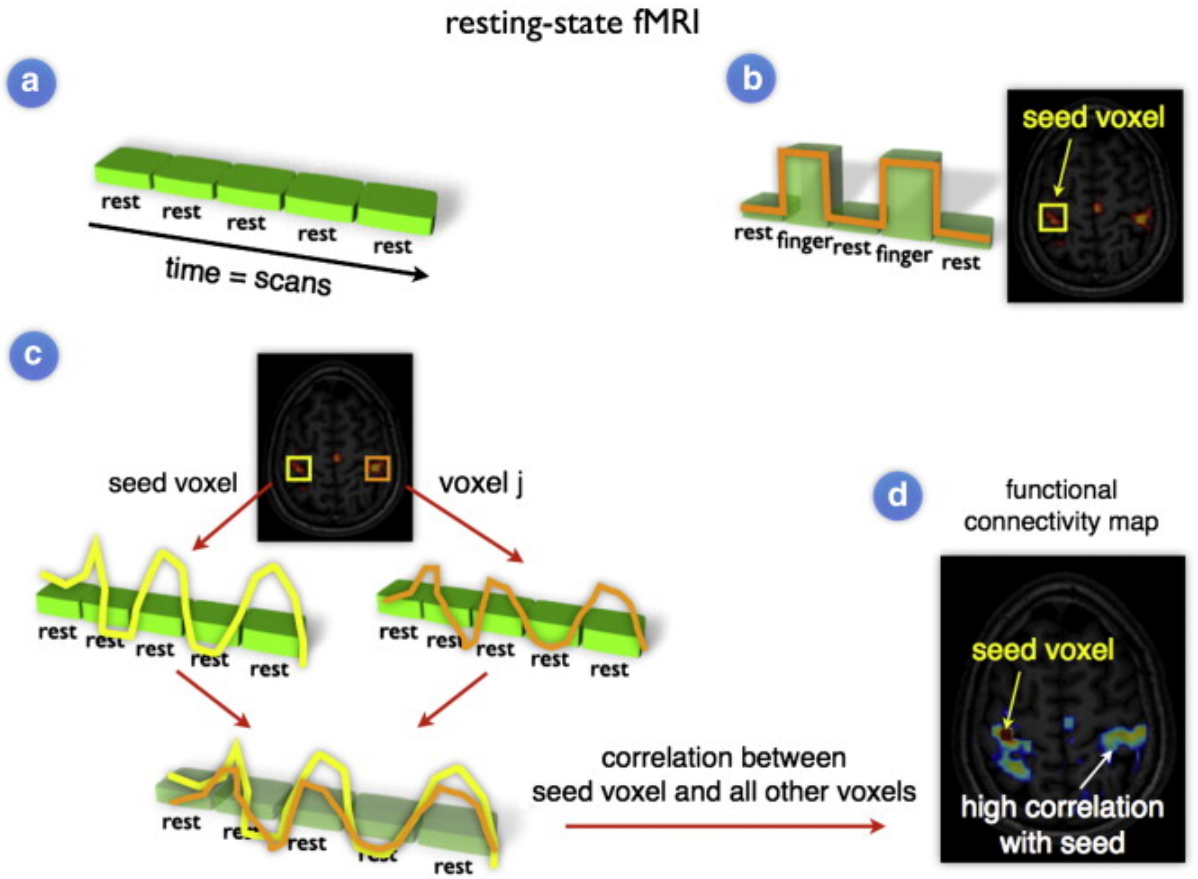
Source: Radiopaedia

Figure 3. Time course of BOLD signal change in response to neural activity

Biswal et al. (1995) first observed high coherent synchronous correlation between the BOLD signals (i.e., FC) from the regions of the somatomotor system in the left and right hemispheres of

the brain in healthy volunteers during resting conditions and demonstrated that BOLD signals reflect neural activities for both resting-state and task-state. Since then the resting-state fMRI (rs-fMRI) has been applied extensively in cognitive neuroscience research to study patients with neurological, mental or psychiatric disorders as well as healthy subjects as control, in order to evaluate behavior-related physiological changes in human brain that occur in a resting state (Uddin and Menon1, 2010), for the rs-fMRI procedure is relatively less demanding in terms of mental tasks and its scan length is short (less than 10 minutes). Studies show that using rs-fMRI can help identify defects in FC in various forms of neurological disorders such as Alzheimer's disease (Binnewijzend et al., 2014; Dai et al., 2015), psychiatric disorders including schizophrenia (Liang et al., 2006; Li et al., 2019), depression (Zhou et al., 2010; Andreescu et al., 2011; Alexopoulos et al., 2012), dementia, attention deficit hyperactivity disorder (Castellanos et al., 2008; Uddin et al., 2009) and autism spectrum disorder (Cherkassky et al., 2006; Uddin et al., 2013; Woodward et al., 2017).

A schematic diagram of construction of FC map in the human brain is illustrated in Figure 4. During the rs-fMRI experiments, the participants are placed in the scanner and instructed to stay quiet with their eyes closed and mind relaxed without thinking of anything particular. The fMRI scan generates a three-dimensional image, that is built up in voxels. The BOLD time-course series from each voxel are measured by MRI scanner as shown in Figure 4a. Next, select seed regions of interest (Figure 4b). Then compute the correlation between the BOLD time-series signal from the two selected seed voxel i and voxel j (Figure 4c) as a quantification of FC with the calculated correlations among all selected seed regions of interest, we can construct a FC map and high correlation indicates a high FC level (Figure 4d).



Source: van den Heuvel and Hulshoff Pol (2010)

Figure 4. Construction of FC map in the human brain based on rs-fMRI

The commonly used statistical methods for calculation of FC based on rs-fMRI can be grouped into model-driven methods and data-driven methods (Li et al., 2009). The model-driven approaches such as cross-correlation analysis (Cao and Worsley, 1999; Hyde and Jesmanowicz, 2012) and coherence analysis (Sun et al., 2004) and statistical parametric mapping employing general linear model (Friston et al., 1994a) are based on selected seed regions of interest. The data-driven decomposition methods include multivariate analyses methods such as principle component analysis (Friston et al., 1993), independent component analysis (Franco et al., 2008) and clustering analysis (Goutte et al., 1999).

In general, cross-correlation analysis is the most commonly used model-based method to measure FC (defined as the statistical dependency between the BOLD time series) due to its simplicity. The FC can be estimated by computing the Pearson correlation coefficients between the two BOLD time courses corresponding to each pair of ROIs collected on the same participant at lag of k, denoted by vectors $\boldsymbol{X}(\boldsymbol{t})$ and $\boldsymbol{Y}(\boldsymbol{t}+k)$, respectively.

$$r_{XY}(k) = \mathrm{Cor}\{\boldsymbol{X}(\boldsymbol{t}), \boldsymbol{Y}(\boldsymbol{t}+k)\} = \frac{\mathrm{Cov}(\boldsymbol{X}(\boldsymbol{t}), \boldsymbol{Y}(\boldsymbol{t}+k))}{\sqrt{\mathrm{Var}(\boldsymbol{X}(\boldsymbol{t}))\mathrm{Var}(\boldsymbol{Y}(\boldsymbol{t}+k))}}, \tag{1.4}$$

The zero-lag correlation has been used in many studies (Ogawa et al., 1990; Bandettini et al., 1993; Friston et al., 1994b; Biswal et al., 1995).

Notwithstanding the major technical advantages of MRI, a few challenges remain. First, high-throughput neuroimaging technologies generate incredibly large amounts of data, so called "big data", which can be very difficult in terms of data analysis and result interpretation. The neuroimaging studies comparing two groups (e.g., disease group and healthy control group) are considered exploratory and the primary objective is to detect any differential connectivities

associated with the disease. As such, thousands of hypotheses are tested simultaneously that is known as the large-scale simultaneous hypothesis testing problem (Efron, 2004). Most of the hypotheses tested simultaneously are null containing nothing but noise, while only a small number of them are true signals. This sparsity issue makes detecting these true signals very challenging (Cai and Sun, 2017). The conventional statistical method used for multiple testing to control family-wise error rate (e.g., Bonferroni correction) tends to be overly conservative in large-scale hypothesis testing, leading to a high chance to miss out a lot of important differences that may be meaningful to the researchers. The false discovery rate (FDR) method has been used to address this multiple testing issue in neuroimaging (Benjamini and Hochberg, 1995; Genovese et al., 2002), where FDR is defined as the expected proportion of false positives among all rejected hypotheses. Efron et al. (2001) further extended the FDR concept from an empirical Bayesian perspective and proposed a test statistics based local false discovery rate (Lfdr), defined as the posterior probability of the null hypothesis is true given the observed test statistics. Efron's Lfdr method allows us to fully explore the distribution of test statistics and hence provides a more precise control of FDR. Zhao (2014) proposed a mixed-effects regression model (MERM) assuming random subject effect and heteroscedastic errors at both group level (e.g., disease group and healthy control group) and connectivity link level to discover disruptive FCs while controlling FDR using rs-fMRI data. Several existing statistical approaches that are commonly used for FDR control in neuroimaging data analysis have been investigated and compared via simulation studies, and Efron's Lfdr method based on densities of test statistics was recommended as the best suited approach for between-group comparisons to detect disrupted FCs (Song, 2016; Bhaumik et al., 2018b).

Second, the current economic cost of the imaging techniques is high and consequently, most of neuroimaging studies have relatively limited number of subjects. Szucs and Ioannidis (2019) conducted a systematic review and evaluation of sample size across the highly cited published MRI study papers between 1990 and 2012 and reported a median sample size of 12.5 per group based on the 107 clinical fMRI studies with more than one group. The high sample variability associated with the small sample size will lead to lower power, and thus higher probability of falsely detecting significant findings. The existing statistical methodologies do not provide a satisfied control of FDR especially for the neuroimaging studies with small sample size (Song, 2016; Bhaumik et al., 2018b).

The third challenge is how to incorporate multimodal neuroimaging data in statistical modeling. Previous neuroimaging studies on FC and SC using quantitative/computational modeling (Koch et al., 2002; Honey et al., 2009) revealed evidence that regions with high SC also exhibit high FC, but high FC can also occur in regions with low SC. These results suggest that strong SC can be a predictor of strong FC, but not for the reverse case (i.e., weak SC cannot be a predictor of weak FC). The information provided by multimodal imaging techniques can be complementary to each other and thus using integrated multimodal analysis enables us to borrow strength from different modalities. In a review paper of studies combining both structural and functional connectivity data Rykhlevskaia et al. (2008) discussed different approaches to integrate FC and SC including analysis of FC informed by SC. Zhao (2014) developed a bivariate linear mixed-effects model with random subject intercepts and heteroscedastic errors to analyze FC and SC data jointly. The model incorporates both the between-modality and within-subject correlations between FC and SC. Xue et al. (2015) introduced a Bayesian multimodal approach directly on FC time series data incorporating SC data into modeling the prior distribution of

FC. Chiang (2016) developed a Bayesian vector autoregressive model that combines multimodal neuroimaging data by integrating structural imaging data into prior information to improve inference of EC. Zhang et al. (2019) introduced a covariate-adaptive method employing a mixture of generalized linear model and Gaussian model to optimize p-value threshold for multiple hypothesis testing, and applied it to the fMRI data with Brodmann area (BA, cerebral cortex regions) label as covariate.

Statistical methods for FDR control such as Efron's Lfdr that solely considers FC data do not account for the potential influence of SC into density estimation of FC test statistics. Thus, my research focuses on development of a statistical method to incorporate the test statistics using MERM from multimodal neuroimaging data in a Bayesian mixture model for FC detection, to improve the FDR, especially for multimodal neuroimaging studies with small to moderate sample sizes.

A covariate-modulated Lfdr method has been used to incorporate functional genome annotations in the model and proved to be efficient by increasing power in genome-wide association studies, where the sample size was sufficiently large (Zablocki et al., 2014; Torkamani et al., 2011). My research focus is on FDR control in a cross-sectional (i.e. , comparative multimodal neuroimaging study with small samples size. In this thesis, we extend the covariate-modulated Lfdr method to multimodal neuroimaging data and implement a Bayesian multimodal Lfdr approach to integrate SC and FC statistics utilizing a Bayesian mixture model. This approach leverages the complementary SC statistic as auxiliary information to enhance the modeling of the FC statistics distribution for the identification of differential FCs between the two groups.

The remainder of the thesis is structured as follows. Chapter 2 introduces a neuroimaging study in late-life depression (LLD), which is the motivating example for this research where

FC and SC data were collected from each participant. Chapter 3 describes a linear MERM with heteroscedastic errors in technical details, which has been used to analyze the FC and SC data in neuroimaging study and provide the test statistics for hypotheses testing associated with each connectivity link. Chapter 4 starts with a brief review on the large-scale simultaneous hypothesis testing problem and the existing statistical methods that have been used to control family-wise error rate and FDR, and then elaborates a Bayesian multimodal local false discovery rate (Lfdr) method utilizing a Bayesian mixture model that integrates FC and SC data. This is followed by an extensive simulation study designed to evaluate the performance of this method by comparison with Efron's Lfdr method that solely considers FC for FDR control in neuroimaging. Chapter 5 describes a semi-parametric Bayesian approach utilizing a non-parametric Dirichlet process mixture (DPM) model, which has been developed by Ghosal (2019) to study differential FC between patients with autism spectrum disorders and controls using rs-fMRI data. As a Bayesian multiple comparison approach, this Bayesian DPM model does not impose any model assumption and also can be used to mimic hypothesis testing framework, while preserving the false discovery rate. Therefore, this method is applied to the motivating study as an alternative approach for FC data analysis for a comparison purpose. Chapter 6 presents results using Bayesian multimodal Lfdr method as well Bayesian DPM model in the motivating multimodal neuroimaging study in LLD for identifying differential FC links. Details on the results focusing on the FC patterns found within and between the major brain networks that may be associated with LLD are discussed. Further, a comprehensive comparison between Efron's Lfdr, Bayesian multimodal Lfdr, and Bayesian DPM model is also provided in this chapter. Chapter 7 concludes the thesis, discusses the limitations of the current framework and provides suggestions and thoughts for future work.

CHAPTER 2

MOTIVATING EXAMPLE: A MULTIMODAL NEUROIMAGING STUDY IN LATE-LIFE DEPRESSION

2.1 Background

What has motivated us to form this research topic was a cross-sectional, comparative neuroimaging study in late-life depression (LLD) to compare LLD group vs. healthy control group investigated by Dr. Ajilore at University of Illinois at Chicago (UIC). In this study, multimodal neuroimaging data including FC (measured using rs-fMRI) and SC (measured using DTI) were collected at a single time point on each participant. The objective of this study is to investigate intrinsic FC patterns of whole brain networks that are impaired during depression, to help prioritize FC patterns for further consideration as potential biomarkers in the future interventional clinical studies.

Particularly, two features of this study are worth to mention. First it has a small sample size, which is typical issue in neuroimaging studies. Second, the study has FC and SC measures from each participant, allowing us to explore ways to incorporate both modality measures to improve efficiency in detecting FCs while controlling FDR.

Patients with LLD are usually over 50 or 60 years old of age and have major depressive symptoms. The most prominent major depressive symptoms are severe and persistent low mood and self-esteem, profound of sadness or a sense of despair, and loss of interest in activities that are usually enjoyable. As the world population of adults aged 60 years and older is expanding rapidly from 900 million in 2015 to 2 billion by 2050 (World Health Organization,

2018), focus on understanding the age-related disorders is becoming prominent and necessary to help accommodate this demographic shift. One of the major disorders in elders is depression associated with socioeconomic, psychiatric and medical factors (Ellison et al., 2012). The LLD is a leading cause of disability in older adults and substantial health care expenses (Snow and Abram, 2016). Major depression disorder (MDD) is a common and complex form of mental disorder in the United States and worldwide. It was reported that approximately 7.1% of U.S. adults have major depression (National Institute of Mental Health, 2017), and more than 264 million people have depression worldwide (World Health Organization, 2019). In U.S., the prevalence of major depression in adults 50 years and older is estimated to be 4.7% (National Institute of Mental Health, 2017). An epidemiological study in five US catchment areas reported that 1.4% to 3.7% were diagnosed of major depression in 5,723 elderly participants (Weissman et al., 1991). The Cache county study conducted in 4,559 adults aged 65 to 100 years old without dementia reported that 4.4% in women and 2.7% in men had major depression in the mid of 1990's (Steffens et al., 2000). According to a national epidemiological survey (Hasin et al., 2005), the estimated prevalence of 12-month and lifetime MDD in adults 65 years and older were 2.7% and 8.2%, respectively. Neuroimaging studies suggest disrupted SCs and FCs in LLD patients and new treatment has been targeting on specific brain networks involved in pathophysiological characterization of brain connectivity associated with LLD (Alexopoulos et al., 2012; Yuen et al., 2014; Alexopoulos, 2019).

2.2 Materials and Methods

2.2.1 Sample Size

This study was a very typical neuroimaging study in terms of small sample size. A total of twenty-three elderly participants (13 healthy subjects and 10 LLD patients) participated in

this study. Their ages ranged from 60 to 82 years, with mean age of 67.8 years and standard deviation of 7.05 years.

2.2.2 Participant Recruitment

The participant recruitment methods were the same as described in Zhang et al. (2012). Individuals of age 55 and older were recruited via "community outreach (e.g., local newspaper, radio, and television advertisements) and relevant outpatient clinics". Initially participants went through a screening procedure via telephone. The major exclusion criteria consisted of "psychotic disorders such as schizophrenia and bipolar disorder, history of any of the following: anxiety disorder outside of major depression episodes, head trauma and substance abuse, MRI contraindications (e.g., pacemaker, metal implants)". The major inclusion criteria for all participants were "over 55 years of age, no history of unstable cardiac or neurological diseases, and medication-naive or anti-depressant free for at least two weeks". After the telephone screen, all eligible participants were scheduled for an assessment with Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, 4^{th} edition (DSM-IV) (SCID-IV) (Spitzer et al., 1992) by a trained research assistant, followed by an evaluation on severity of depression using the 17-item Hamilton Depression Rating Scale (HDRS)) (Hamilton, 1960) by a board certified psychiatrist. The LLD patients must meet "SCID-IV criteria for MDD and a score ≥ 15 on the 17-item HDRS". All study subjects provided written informed consent. The study was approved by the "UIC Institutional Review Board" and performed in compliance with the "Declaration of Helsinki".

2.2.3 Regions of Interest

The FC measured using rs-fMRI and SC measured using DTI were collected from 87 cortical and subcortical gray matter regions of interest (ROI) in each participant for the whole brain

analysis. The 87 ROIs included 43 bilateral regions and brain stem (central), shown in Table I.

2.2.4 Image Acquisition

The rs-fMRI and DTI image acquisition methods were the same as described in Zhang et al. (2012). "The rs-fMRI imaging was performed using a Philips Achieva 3.0T MRI scanner (Philips Medical Systems, The Netherlands) with an 8-channel sensitivity-encoding (SENSE) head coil". During the scan, participants were instructed to "remain still, keep their eyes closed, be relaxed and not to think anything particular, with soft ear plugs and foam pads provided to ensure comfort and reduce the head motion". High resolution three-dimensional T1-weighted image data were obtained with a "magnetization-prepared rapid acquisition gradient-echo sequence" with the following parameters: "field of view (FOV) = 240 mm; repetition time (TR) = 8.4 ms; echo time (TE) = 3.9 ms; flip angle = 8°; voxel size = $1.1 \times 1.1 \times 1.1$ mm; 134 contiguous axial slices". "A single-shot gradient-echo echo-planar imaging (EPI) sequence" was used to collect rs-fMRI imaging data with the following parameters: "EPI factor = 47; FOV = $23 \times 23 \times 15$ cm^3; TR = 2,000 ms; TE = 30 ms; flip angle = 80°; in-plane resolution = 3×3 mm^2; slice thickness/gap = 5/0 mm; slice number = 30; SENSE reduction factor = 1.8; number of repeat scan = 200; total scan time = 6:52".

For DTI data acquisition, "a single-shot spin-echo EPI sequence" was used with the following parameters: "FOV = 240 mm; voxel size = $0.83 \times 0.83 \times 2.2$ mm; TR = 6994 ms; TE = 71 ms; flip angel = 90°; SENSE reduction factor = 2.5". "Sixty-seven contiguous axial slices aligned to the anterior commissure-posterior commissure line were collected in 32 gradient directions with b-value = 700 s/mm^2 and one acquisition without diffusion sensitization (referred to as b0 image)".

TABLE I
THE 87 CORTICAL AND SUBCORTICAL GREY MATTER REGIONS OF INTEREST ANALYZED IN LLD NEUROIMAGING STUDY

Region No.	Region Name[a]	Region No.	Region Name[a]
1	L cerebellum cortex	2	L thalamus proper
3	L caudate	4	L putamen
5	L pallidum	6	brain stem (central)
7	L hippocampus	8	L amygdala
9	L accumbens area	10	L ventral diencephalon
11	R cerebellum cortex	12	R thalamus proper
13	R caudate	14	R putamen
15	R pallidum	16	R hippocampus
17	R amygdala	18	R accumbens area
19	R ventral diencephalon	20	L banks of superior temporal sulcus
21	L caudal anterior cingulate	22	L caudal middle frontal
23	L cuneus	24	L entorhinal
25	L fusiform	26	L inferior parietal
27	L inferior temporal	28	L isthmus cingulate
29	L lateral occipital	30	L lateral orbitofrontal
31	L lingual	32	L medial orbitofrontal
33	L middle temporal	34	L parahippocampal
35	L paracentral	36	L pars opercularis
37	L pars orbitalis	38	L pars triangularis
39	L pericalcarine	40	L postcentral
41	L posterior cingulate	42	L precentral
43	L precuneus	44	L rostral anterior cingulate
45	L rostral middle frontal	46	L superior frontal
47	L superior parietal	48	L superior temporal
49	L supramarginal	50	L frontal pole
51	L temporal pole	52	L transverse temporal
53	L insula	54	R banks of superior temporal sulcus
55	R caudal anterior cingulate	56	R caudal middle frontal
57	R cuneus	58	R entorhinal
59	R fusiform	60	R inferior parietal
61	R inferior temporal	62	R isthmus cingulate
63	R lateral occipital	64	R lateral orbitofrontal
65	R lingual	66	R medial orbitofrontal
67	R middle temporal	68	R parahippocampal
69	R paracentral	70	R pars opercularis
71	R pars orbitalis	72	R pars triangularis
73	R pericalcarine	74	R postcentral
75	R posterior cingulate	76	R precentral
77	R precuneus	78	R rostral anterior cingulate
79	R rostral middle frontal	80	R superior frontal
81	R superior parietal	82	R superior temporal
83	R supramarginal	84	R frontal pole
85	R temporal pole	86	R transverse temporal
87	R insula		

[a] L = left; R = right.

2.2.5 Data Processing

The data processing methods are the same as described in Zhao (2014). The individual resting-state FC were measured using the FC toolbox CONN (Whitfield-Gabrieli and Nieto-Castanon, 2012). The CONN performs seed-based correlation analysis by computing the Pearson correlation coefficients between the BOLD time series from a given ROI to all other ROIs in the brain shown in Equation (1.4). The potential confounding factors including "motion artifact, white matter, and cerebrospinal fluid and physiological noise source reduction" were regressed out from the signal prior to data analysis. The SC maps was generated for each of the twenty-three participants using a pipeline that integrates multiple image analysis techniques (GadElkarim et al., 2012). Briefly, diffusion-weighted imaging (DWI) images were first eddy current corrected using the automatic image registration tool in DtiStudio software (Jiang et al., 2006) by registering all DWI images to their corresponding b0 images with a 12-parameter affine transformation, followed by computation of diffusion tensors and deterministic tractography using fiber assignment by a continuous tracking algorithm (Mori et al., 1999). Label maps were generated using T1-weighted images with FreeSurfer (Fischl, 2012). The SC based on the 87 cortical and subcortical gray matter regions were then measured using "an internal Matlab program by counting the number of fiber tracts found by the tractography algorithm connecting each pair of regions". Note that the SC data contains a number of zero values, suggesting no SC between the corresponding brain regions.

Since the FC data is measured using the Pearson's correlation coefficients, we can apply Fisher's Z transformation to stabilize the variance and approximate normal distribution for the transformed data (Bhaumik et al., 2018b,a; Afyouni et al., 2019), $r^*_{XY} = \text{arctanh}(r_{XY}) = \frac{1}{2}\ln\left(\frac{1+r_{XY}}{1-r_{XY}}\right)$. And for SC data, cube-root transformation is applied considering that the DTI

measures are essentially the estimates of fiber counts that can be assumed to follow Poisson distribution (Zhao, 2014).

The adjacency matrices of FC and SC each has $\binom{87}{2} = 87 \times (87-1)/2 = 3741$ unique connectivity measures (or links), denoted as $c_{i,j}$, where $i, j = 1, 2, \cdots, 87$ and $i < j$, shown in Equation (2.1).

$$\begin{bmatrix} - & & & & \\ c_{1,2} & - & & & \\ c_{1,3} & c_{2,3} & - & & \\ \vdots & \vdots & \vdots & \ddots & \\ c_{1,87} & c_{2,87} & \cdots & \cdots & - \end{bmatrix}_{87\times 87} \tag{2.1}$$

In this motivating example, the dataset contains $m = 3741$ FC and SC measurements from each of 23 participants from the two groups including 13 subjects in healthy control (HC) group and 10 patients in LLD group. The primary goal is to detect differential FC links between LLD group and HC group based on rs-fMRI data. For the i^{th} connectivity link, $i = 1, \cdots, 3741$, the corresponding null hypothesis is,

$$H_{0,i} : \text{The } i^{th} \text{ link has no FC difference between LLD group and HC group,}$$

In total, we have 3741 null hypotheses to be tested, $\{H_{0,1}, H_{0,2}, \cdots, H_{0,3741}\}$, with corresponding test statistics $\{t_1, t_2, \cdots, t_{3741}\}$ and p-values $\{p_1, p_2, \cdots, p_{3741}\}$. When the 3741 hypotheses are tested simultaneously, without adjusting for multiplicity, the probability of falsely rejecting at least one null hypothesis will be greatly inflated. This large-scale simultaneous hypothesis problem is discussed in Section 4.1.

CHAPTER 3

MIXED-EFFECTS MODEL FOR NEUROIMAGING DATA

A linear MERM assuming random subject effect and heteroscedastic errors at both group level (e.g., disease group and HC group) and connectivity link level has been applied to data analysis of FC measured using rs-fMRI in neuroimaging studies (Hedeker and Gibbons, 2006; Zhao, 2014; Jie, 2016; Song, 2016; Bhaumik et al., 2018a,b). The simulation studies based on neuroimaging data in LLD and autism studies have demonstrated that the mixed-effects model utilizing an expectation-maximization (EM) algorithm performs well in terms of accuracy and precision in parameter estimation (Zhao, 2014; Bhaumik et al., 2018a,b).

In the mixed-effects model, we assume a random subject effect to account for the within-subject correlation among the connectivity links taken from the same subject, and a random error that is specific to both connectivity link and group to address the different heterogeneity pattern across all connectivity links in each group. The between-group comparisons are based on the connectivity link-specific group information.

The outcome variable y_{ij} is the connectivity measurement for the i^{th} connectivity link in the j^{th} subject, $i = 1, \cdots, m$, $j = 1, 2, \cdots, N$, where m and N refer to the total number of links and the total number of subjects, respectively. The mixed-effects model for y_{ij} with fixed effect parameter β_{0i} and β_{1i}, random subject effect γ_j and random error ϵ_{ij} is then,

$$y_{ij} = \beta_{0i} \times (1 - Grp_j) + \beta_{1i} \times Grp_j + \gamma_j + \epsilon_{ij}, \tag{3.1}$$

where

- Grp_j is the indicator variable for which group the j^{th} subject is from, HC group if $Grp_j = 0$, or disease group if $Grp_j = 1$.
- β_{0i} and β_{1i} are the fixed effects for the i^{th} connectivity link in HC group and disease group, respectively. $\beta_{1i} - \beta_{0i}$ is the between-group difference for the i^{th} connectivity link.
- γ_j is the random subject effect for the j^{th} subject, $\gamma_j \sim N(0, \sigma^2_\gamma)$.
- ϵ_{ij} is the random error for the i^{th} connectivity link in the j^{th} subject, $\epsilon_{ij} \sim N(0, \sigma^2_{Grp_j i})$, that is, $\epsilon_{ij} \sim N(0, \sigma^2_{0i})$ if the subject is from HC group, $\epsilon_{ij} \sim N(0, \sigma^2_{1i})$ if the subject is from disease group.
- γ_j and ϵ_{ij} are independent of each other.

Note that in the LLD neuroimaging study described in Chapter 2 we have $m = 3741$ and $N = 23$.

Using matrix notation, the mixed-effects model in Equation (3.1) for the j^{th} subject can be rewritten as follows:

$$\boldsymbol{y}_j = \boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\gamma_j + \boldsymbol{\epsilon}_j, \tag{3.2}$$

where

- $\boldsymbol{y}_j$ is a $m \times 1$ vector of the connectivity measurement.

$$\boldsymbol{y}_j = \begin{bmatrix} y_{1j} & y_{2j} & \cdots & y_{mj} \end{bmatrix}^T_{1\times m}, \tag{3.3}$$

where ^{T} denotes vector/matrix transpose.

- $\boldsymbol{X}_j$ is a $m \times 2m$ design matrix for the fixed effects $\boldsymbol{\beta}$.

$$\begin{aligned}\boldsymbol{X}_j &= \begin{bmatrix} 1-Grp_j & \cdots & 0 & Grp_j & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1-Grp_j & 0 & \cdots & Grp_j \end{bmatrix}_{m\times 2m} \\ &= \begin{bmatrix} 1-Grp_j & Grp_j \end{bmatrix} \otimes \boldsymbol{I}_m, \end{aligned} \tag{3.4}$$

  where $\boldsymbol{I}_m$ is a m-dimensional identity matrix.

- $\boldsymbol{\beta}$ is a $2m \times 1$ vector of the fixed effects.

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_0 & \boldsymbol{\beta}_1 \end{bmatrix}^T_{1\times 2m} = \begin{bmatrix} \beta_{01} & \cdots & \beta_{0m} & \beta_{11} \cdots & \beta_{1m} \end{bmatrix}^T_{1\times 2m}, \tag{3.5}$$

- $\boldsymbol{Z}_j$ is a $m \times 1$ design matrix for the random subject effects,

$$\boldsymbol{Z}_j = \mathbf{1}_m, \tag{3.6}$$

  where $\mathbf{1}_m$ is a $m \times 1$ vector of all ones.

- γ_j is the random subject effect.

$$\gamma_j \sim N(0, \sigma^2_\gamma), \tag{3.7}$$

- $\boldsymbol{\epsilon}_j$ is a $m \times 1$ vector of random errors.

$$\boldsymbol{\epsilon}_j = \begin{bmatrix} \epsilon_{1j} & \epsilon_{2j} & \ldots & \epsilon_{mj} \end{bmatrix}^T_{1\times m} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{Grp_j}), \tag{3.8}$$

where the variance of $\boldsymbol{\epsilon}_j$ is a $m \times m$ diagonal matrix:

$$\boldsymbol{\Sigma}_{Grp_j} = \begin{bmatrix} \sigma^2_{Grp_j 1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2_{Grp_j m} \end{bmatrix}_{m \times m}, \tag{3.9}$$

To obtain the joint distribution of $\boldsymbol{y}_j$ and γ_j, first we get their expectations and variance-covariance by Equation (3.2),

$$\mathrm{E}(\boldsymbol{y}_j) = \mathrm{E}\big(\boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\gamma_j + \boldsymbol{\epsilon}_j\big) = \boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\mathrm{E}(\gamma_j) + \mathrm{E}(\boldsymbol{\epsilon}_j) = \boldsymbol{X}_j\boldsymbol{\beta} \tag{3.10}$$

$$\mathrm{E}(\gamma_j) = 0 \tag{3.11}$$

$$\mathrm{Var}(\boldsymbol{y}_j) = \mathrm{Var}\big(\boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\gamma_j + \boldsymbol{\epsilon}_j\big) = \mathrm{Var}(\boldsymbol{Z}_j\gamma_j) + \mathrm{Var}(\boldsymbol{\epsilon}_j)$$

$$= \boldsymbol{Z}_j\mathrm{Var}(\gamma_j)\boldsymbol{Z}_j^T + \mathrm{Var}(\boldsymbol{\epsilon}_j) = \boldsymbol{Z}_j\sigma^2_\gamma\boldsymbol{Z}_j^T + \boldsymbol{\Sigma}_{Grp_j} \tag{3.12}$$

$$\mathrm{Var}(\gamma_j) = \sigma^2_\gamma \tag{3.13}$$

$$\mathrm{Cov}(\boldsymbol{y}_j, \gamma_j) = \mathrm{Cov}\big(\boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\gamma_j + \boldsymbol{\epsilon}_j, \gamma_j\big) = \mathrm{Cov}(\boldsymbol{Z}_j\gamma_j, \gamma_j)$$

$$= \boldsymbol{Z}_j\mathrm{Var}(\gamma_j) = \boldsymbol{Z}_j\sigma^2_\gamma \tag{3.14}$$

$$\mathrm{Cov}(\gamma_j, \boldsymbol{y}_j) = \sigma^2_\gamma\boldsymbol{Z}_j^T \tag{3.15}$$

We denote $\mathrm{Var}(\boldsymbol{y}_j)$, $\mathrm{Cov}(\boldsymbol{y}_j, \gamma_j)$ and $\mathrm{Cov}(\gamma_j, \boldsymbol{y}_j)$ by $\boldsymbol{\Sigma}_{\boldsymbol{y}_j}$, $\boldsymbol{\Sigma}_{\boldsymbol{y}_j\gamma_j}$ and $\boldsymbol{\Sigma}_{\gamma_j\boldsymbol{y}_j}$, respectively.

Then the joint distribution of $\boldsymbol{y}_j$ and γ_j follows a multivariate normal distribution,

$$\begin{bmatrix} \boldsymbol{y}_j \\ \gamma_j \end{bmatrix} \sim MVN\left(\begin{bmatrix} \boldsymbol{X}_j\boldsymbol{\beta} \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{Z}_j\sigma^2_\gamma\boldsymbol{Z}_j^T + \boldsymbol{\Sigma}_{Grp_j} & \boldsymbol{Z}_j\sigma^2_\gamma \\ \sigma^2_\gamma\boldsymbol{Z}_j^T & \sigma^2_\gamma \end{bmatrix}\right) \tag{3.16}$$

The quantities we are interested in estimating from this study are the mean connectivity difference between disease group and HC group for each link ($\beta_{1i} - \beta_{0i}, i = 1, \cdots, m$ and the

associated variances, which are then used to calculate the test statistics for the tests. For estimation of model parameters, we employ an EM algorithm (Dempster et al., 1977, 1981; Hedeker and Gibbons, 2006), in which the random subject effect γ_j will be treated as unobserved data. For data with missing values or unobserved latent variables, the iterative procedure in the EM algorithm allows us to compute the maximum likelihood estimates of model parameters. Briefly, in the E-step, the empirical Bayes (EB) estimates of the random subject effect γ_j and the conditional variances of γ_j given data $\boldsymbol{y}_j$ (denoted by $\Sigma_{\gamma_j|\boldsymbol{y}_j}$) are computed based on the initial values of the rest of parameters. In the M-step, maximize the marginal likelihood with respect to the fixed effect vector $\boldsymbol{\beta}$, random error variance $\boldsymbol{\Sigma}_{Grp_j}$ and the variance of the random effects σ^2_γ and obtain the maximum marginal likelihood (MML) estimates for these parameters based on the current estimates of the random effects. The algorithm iterates between E-step and M-step until convergence.

For notational convenience, we denote the parameter vector by $\boldsymbol{\theta} = (\boldsymbol{\beta},\, \sigma^2_\gamma,\, \boldsymbol{\Sigma}_{Grp_j},\, \gamma_j,\, \Sigma_{\gamma_j|\boldsymbol{y}_j})$, and partition $\boldsymbol{\theta}$ into $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, as $\boldsymbol{\theta}_1 = (\boldsymbol{\beta},\, \sigma^2_\gamma,\, \boldsymbol{\Sigma}_{Grp_j})$ and $\boldsymbol{\theta}_2 = (\gamma_j,\, \Sigma_{\gamma_j|\boldsymbol{y}_j})$. The EM algorithm is outlined in the following steps,

1. Give initial values to parameters $\boldsymbol{\theta}_1$, denoted by $\boldsymbol{\theta}_1^{(0)} = (\boldsymbol{\beta}^{(0)},\, \sigma^{2(0)}_\gamma,\, \boldsymbol{\Sigma}^{(0)}_{Grp_j})$.
2. E-step: Based on $\boldsymbol{\theta}_1^{(0)}$, calculate the EB estimates $\boldsymbol{\theta}_2^{(1)} = ({\gamma_j}^{(1)},\, \Sigma^{(1)}_{\gamma_j|\boldsymbol{y}_j})$.
3. M-step: Using $\boldsymbol{\theta}_2^{(1)}$, calculate the MML estimates $\boldsymbol{\theta}_1^{(1)} = (\boldsymbol{\beta}^{(1)},\, \sigma^{2(1)}_\gamma,\, \boldsymbol{\Sigma}^{(1)}_{Grp_j})$.
4. Replace $\boldsymbol{\theta}_1^{(0)}$ by $\boldsymbol{\theta}_1^{(1)}$, carry out the E-step and the M-step again.
5. Repeat the steps 2-4 until the parameter estimates converge.

The detailed methods on EB and MML estimation are described in the following sections 3.1 and 3.2, respectively.

3.1 Empirical Bayes Estimation

Using empirical Bayes approach, the following distributions are specified:

$$\boldsymbol{y}_j \mid \gamma_j; \boldsymbol{\beta}, \boldsymbol{\Sigma}_{Grp_j} \sim N\big(\boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z_j}\gamma_j,\, \boldsymbol{\Sigma}_{Grp_j}\big), \tag{3.17}$$

$$\gamma_j; \sigma^2_\gamma \sim N(0, \sigma^2_\gamma). \tag{3.18}$$

By application of Bayes' theorem, the posterior distribution of the random subject effect γ_j given the data $\boldsymbol{y}_j$ is

$$f(\gamma_j \mid \boldsymbol{y}_j) = \frac{f(\boldsymbol{y}_j \mid \gamma_j; \boldsymbol{\beta}, \boldsymbol{\Sigma}_{Grp_j}) f(\gamma_j)}{f(\boldsymbol{y}_j; \sigma^2_\gamma)}, \tag{3.19}$$

where $f(\boldsymbol{y}_j)$ is the marginal distribution of $\boldsymbol{y}_j$.

$$f(\boldsymbol{y}_j) = \int f(\boldsymbol{y}_j \mid \gamma_j; \boldsymbol{\beta}, \boldsymbol{\Sigma}_{Grp_j}) f(\gamma_j; \sigma^2_\gamma) d\gamma_j. \tag{3.20}$$

Since $f(\gamma_j \mid \boldsymbol{y}_j)$ does not depend on γ_j, maximizing $f(\gamma_j \mid \boldsymbol{y}_j)$ is equivalent to maximizing the numerator in the Bayes theorem Equation (3.19).

We need to find the posterior mean and posterior variance of γ_j conditional on observed data vector $\boldsymbol{y}_j$. Based on the joint multivariate distribution of $\boldsymbol{y}_j$ and γ_j in Equation (3.16), we can obtain the expectation and variance of γ_j conditional on $\boldsymbol{y}_j$:

$$\begin{aligned} \mathrm{E}(\gamma_j \mid \boldsymbol{y}_j) &= \mathrm{E}(\gamma_j) + \boldsymbol{\Sigma}_{\gamma_j \boldsymbol{y}_j} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{y}_j} \big[\boldsymbol{y}_j - \mathrm{E}(\boldsymbol{y}_j)\big] \\ &= \sigma^2_\gamma \boldsymbol{Z}^T_j (\boldsymbol{Z}_j \sigma^2_\gamma \boldsymbol{Z}^T_j + \boldsymbol{\Sigma}_{Grp_j})^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta}). \end{aligned} \tag{3.21}$$

$$\begin{aligned}\mathrm{Var}(\gamma_j \mid \boldsymbol{y}_j) = \boldsymbol{\Sigma}_{\gamma_j|\boldsymbol{y}_j} &= \sigma^2_\gamma - \boldsymbol{\Sigma}_{\gamma_j \boldsymbol{y}_j} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{y}_j} \boldsymbol{\Sigma}_{\boldsymbol{y}_j \gamma_j} \\ &= \sigma^2_\gamma - \sigma^2_\gamma \boldsymbol{Z}_j^T (\boldsymbol{Z}_j \sigma^2_\gamma \boldsymbol{Z}_j^T + \boldsymbol{\Sigma}_{Grp_j})^{-1} \boldsymbol{Z}_j \sigma^2_\gamma. \end{aligned} \tag{3.22}$$

Then the conditional distribution of $\gamma_j \mid \boldsymbol{y}_j$ is

$$\begin{aligned}\gamma_j \mid \boldsymbol{y}_j \sim N\big(&\sigma^2_\gamma \boldsymbol{Z}_j^T (\boldsymbol{Z}_j \sigma^2_\gamma \boldsymbol{Z}_j^T + \boldsymbol{\Sigma}_{Grp_j})^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta}), \\ &\sigma^2_\gamma - \sigma^2_\gamma \boldsymbol{Z}_j^T (\boldsymbol{Z}_j \sigma^2_\gamma \boldsymbol{Z}_j^T + \boldsymbol{\Sigma}_{Grp_j})^{-1} \boldsymbol{Z}_j \sigma^2_\gamma\big). \end{aligned} \tag{3.23}$$

Hence the EB estimates of γ_j and $\Sigma_{\gamma_j|\boldsymbol{y}_j}$ (denoted by $\tilde{\gamma}_j$ and $\tilde{\Sigma}_{\gamma_j|\boldsymbol{y}_j}$, respectively) are equivalent to the expectation and variance of γ_j conditional on $\boldsymbol{y}_j$ in Equations (3.21) and (3.22):

$$\tilde{\gamma}_j = \mathrm{E}(\gamma_j \mid \boldsymbol{y}_j) = \sigma^2_\gamma \boldsymbol{Z}_j^T (\boldsymbol{Z}_j \sigma^2_\gamma \boldsymbol{Z}_j^T + \boldsymbol{\Sigma}_{Grp_j})^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta}). \tag{3.24}$$

$$\tilde{\Sigma}_{\gamma_j|\boldsymbol{y}_j} = \mathrm{Var}(\gamma_j \mid \boldsymbol{y}_j) = \sigma^2_\gamma - \sigma^2_\gamma \boldsymbol{Z}_j^T (\boldsymbol{Z}_j \sigma^2_\gamma \boldsymbol{Z}_j^T + \boldsymbol{\Sigma}_{Grp_j})^{-1} \boldsymbol{Z}_j \sigma^2_\gamma. \tag{3.25}$$

Both Equations (3.24) and (3.25) can be computationally challenging due to the inversion of $m \times m$ matrix involved. Therefore, it is necessary to apply the following matrix inversion lemma,

$$(\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B} + \boldsymbol{D})^{-1} = \boldsymbol{D}^{-1} - \boldsymbol{D}^{-1}\boldsymbol{C}(\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C} + \boldsymbol{A})^{-1}\boldsymbol{B}\boldsymbol{D}^{-1}, \tag{3.26}$$

$$(\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B} + \boldsymbol{D})^{-1} = \boldsymbol{D}^{-1} - \boldsymbol{D}^{-1}\big[(\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} + \boldsymbol{D}^{-1}\big]^{-1}\boldsymbol{D}^{-1}. \tag{3.27}$$

By Equations (3.26) and (3.27), we have

$$\boldsymbol{C}(\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}+\boldsymbol{A})^{-1}\boldsymbol{B}=\left[(\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}+\boldsymbol{D}^{-1}\right]^{-1}. \tag{3.28}$$

Let $\boldsymbol{A}=\boldsymbol{\Sigma}_{Grp_j}$, $\boldsymbol{B}=\boldsymbol{Z}_j$, $\boldsymbol{C}=\boldsymbol{Z}_j^T$, and $\boldsymbol{D}=\sigma_\gamma^{-2}$, by Equations (3.26) and (3.27) we have

$$\begin{aligned}
&\sigma_\gamma^2-\sigma_\gamma^2\boldsymbol{Z}_j^T(\boldsymbol{Z}_j\sigma_\gamma^2\boldsymbol{Z}_j^T+\boldsymbol{\Sigma}_{Grp_j})^{-1}\boldsymbol{Z}_j\sigma_\gamma^2\\
=&(\boldsymbol{Z}_j^T\boldsymbol{\Sigma}_{Grp_j}^{-1}\boldsymbol{Z}_j+\sigma_\gamma^{-2})^{-1} &(3.29)\\
=&\sigma_\gamma^2-\sigma_\gamma^2\left[(\boldsymbol{Z}_j^T\boldsymbol{\Sigma}_{Grp_j}^{-1}\boldsymbol{Z}_j)^{-1}+\sigma_\gamma^2\right]^{-1}\sigma_\gamma^2. &(3.30)
\end{aligned}$$

By Equation (3.28), we have

$$\begin{aligned}
&\boldsymbol{Z}_j^T(\boldsymbol{Z}_j\sigma_\gamma^2\boldsymbol{Z}_j^T+\boldsymbol{\Sigma}_{Grp_j})^{-1}\boldsymbol{Z}_j=\left[(\boldsymbol{Z}_j^T\boldsymbol{\Sigma}_{Grp_j}^{-1}\boldsymbol{Z}_j)^{-1}+\sigma_\gamma^2\right]^{-1}\\
\Longrightarrow\ &\boldsymbol{Z}_j^T(\boldsymbol{Z}_j\sigma_\gamma^2\boldsymbol{Z}_j^T+\boldsymbol{\Sigma}_{Grp_j})^{-1}\boldsymbol{Z}_j^T\boldsymbol{Z}_j^{-1}=\left[(\boldsymbol{Z}_j^T\boldsymbol{\Sigma}_{Grp_j}^{-1}\boldsymbol{Z}_j)^{-1}+\sigma_\gamma^2\right]^{-1}\boldsymbol{Z}_j^{-1}\\
\Longrightarrow\ &\boldsymbol{Z}_j^T(\boldsymbol{Z}_j\sigma_\gamma^2\boldsymbol{Z}_j^T+\boldsymbol{\Sigma}_{Grp_j})^{-1}=\left[(\boldsymbol{Z}_j^T\boldsymbol{\Sigma}_{Grp_j}^{-1}\boldsymbol{Z}_j)^{-1}+\sigma_\gamma^2\right]^{-1}\boldsymbol{Z}_j^{-1}. &(3.31)
\end{aligned}$$

Then

$$\begin{aligned}
\boldsymbol{Z}_j^T(\boldsymbol{Z}_j\sigma_\gamma^2\boldsymbol{Z}_j^T+\boldsymbol{\Sigma}_{Grp_j})^{-1}&=\left[(\boldsymbol{Z}_j^T\boldsymbol{\Sigma}_{Grp_j}^{-1}\boldsymbol{Z}_j)^{-1}+\sigma_\gamma^2\right]^{-1}\boldsymbol{Z}_j^{-1}\\
&=\left[(\boldsymbol{Z}_{jk}^T\boldsymbol{\Sigma}_{Grp_j}^{-1}\boldsymbol{Z}_j)^{-1}+\sigma_\gamma^2\right]^{-1}(\boldsymbol{Z}_j^T\boldsymbol{Z}_j)^{-1}(\boldsymbol{Z}_{jk}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j^{-1}\\
&=\left[(\boldsymbol{Z}_{jk}^T\boldsymbol{\Sigma}_{Grp_j}^{-1}\boldsymbol{Z}_j)^{-1}+\sigma_\gamma^2\right]^{-1}(\boldsymbol{Z}_j^T\boldsymbol{Z}_j)^{-1}\boldsymbol{Z}_j^T. &(3.32)
\end{aligned}$$

By Equation (3.32), $\tilde{\gamma}_j$ in Equation (3.24) can be simplified as:

$$\begin{aligned}\tilde{\gamma}_j &= \sigma_\gamma^2 \boldsymbol{Z}_j^T (\boldsymbol{Z}_j \sigma_\gamma^2 \boldsymbol{Z}_j^T + \boldsymbol{\Sigma}_{Grp_j})^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta}) \\ &= \sigma_\gamma^2 \big[(\boldsymbol{Z}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} \boldsymbol{Z}_j)^{-1} + \sigma_\gamma^2 \big]^{-1} (\boldsymbol{Z}_j^T \boldsymbol{Z}_j)^{-1} \boldsymbol{Z}_j^T (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta}) \\ &= R (\boldsymbol{Z}_j^T \boldsymbol{Z}_j)^{-1} \boldsymbol{Z}_j^T (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta}). \end{aligned} \tag{3.33}$$

where $R = \sigma_\gamma^2 \big[(\boldsymbol{Z}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} \boldsymbol{Z}_j)^{-1} + \sigma_\gamma^2 \big]^{-1}$. And by Equations (3.29) and (3.30), $\tilde{\Sigma}_{\gamma_j | \boldsymbol{y}_j}$ in Equation (3.25) can be simplified as:

$$\begin{aligned}\tilde{\Sigma}_{\gamma_j | \boldsymbol{y}_j} &= \sigma_\gamma^2 - \sigma_\gamma^2 \boldsymbol{Z}_j^T (\boldsymbol{Z}_j \sigma_\gamma^2 \boldsymbol{Z}_j^T + \boldsymbol{\Sigma}_{Grp_j})^{-1} \boldsymbol{Z}_j \sigma_\gamma^2 \\ &= (\boldsymbol{Z}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} \boldsymbol{Z}_j + \sigma_\gamma^{-2})^{-1} \\ &= \sigma_\gamma^2 - \sigma_\gamma^2 \big[(\boldsymbol{Z}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} \boldsymbol{Z}_j)^{-1} + \sigma_\gamma^2 \big]^{-1} \sigma_\gamma^2 \\ &= \sigma_\gamma^2 - R \sigma_\gamma^2 \\ &= (1 - R) \sigma_\gamma^2. \end{aligned} \tag{3.34}$$

As compared to Equations (3.24) and (3.25) which involves inversion of $m \times m$ matrix, Equations (3.33) and (3.34) only require inversion of a scalar and are much more efficient computationally.

3.2 Maximum Marginal Likelihood Estimation

In the mixed-effects model, the maximum likelihood estimation for the fixed effects $\boldsymbol{\beta}$, the variance of random errors $\boldsymbol{\Sigma}_{Grp_j}$, and the variance of random subject effect σ_γ^2 are based

on the marginal distribution of y_j for the j^{th} subject that is obtained by integrating over the distribution of the random effects $\boldsymbol{\gamma}_j$ (Hedeker and Gibbons, 2006),

$$f(\boldsymbol{y}_j) = \int f(\boldsymbol{y}_j \mid \gamma_j; \boldsymbol{\beta}, \boldsymbol{\Sigma}_{Grp_j}) f(\gamma_j; \sigma^2_\gamma) d\gamma_j, \tag{3.35}$$

where

$$f(\boldsymbol{y}_j \mid \gamma_j; \boldsymbol{\beta}, \boldsymbol{\Sigma}_{Grp_j}) = (2\pi)^{-\frac{m}{2}} |\boldsymbol{\Sigma}_{Grp_j}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\gamma_j)^T \boldsymbol{\Sigma}^{-1}_{Grp_j}(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\gamma_j)\right], \tag{3.36}$$

$$f(\gamma_j; \sigma^2_\gamma) = (2\pi\sigma^2_\gamma)^{-1/2} \exp\left(-\frac{1}{2\sigma^2_\gamma}\gamma_j^2\right), \tag{3.37}$$

Then the marginal log-likelihood function is the sum of the log-likelihood in Equation (3.35) over all N subjects,

$$\log L = \sum_{j=1}^{N} \log\big[f(\boldsymbol{y}_j)\big]. \tag{3.38}$$

The detailed derivations on maximum marginal likelihood estimates of parameters are provided in Appendix A.

3.3 Hypothesis Testing

For group comparisons of the mean connectivity measures between the HC group and the diseased group at the link level, a total of m hypotheses are tested simultaneously. For each of

the m connectivity links, the null hypothesis $H_{0,i}$ is that the i^{th} connectivity link has no FC difference between the two groups, and the hypothesis testing is:

$$H_{0,i} : \beta_{0i} = \beta_{1i} \text{ vs. } H_{1,i} : \beta_{0i} \neq \beta_{1i}, i = 1, \cdots, m, \tag{3.39}$$

which is the same as using contrast $\boldsymbol{h} = \begin{bmatrix} 1 & -1 \end{bmatrix}_{1\times 2}$:

$$H_{0,i} : \boldsymbol{h}\boldsymbol{\beta}_i = 0 \text{ vs. } H_{0,1} : \boldsymbol{h}\boldsymbol{\beta}_i \neq 0, i = 1, \cdots, m, \tag{3.40}$$

To obtain the maximum likelihood estimates of the fixed effects for each connectivity link in Equation (3.1) $\boldsymbol{\beta}_i = \begin{bmatrix} \beta_{0i} & \beta_{1i} \end{bmatrix}^T$, $i = 1, \cdots, m$, and the associated variances, we can rewrite Equation (3.1) as:

$$y_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta}_i + \gamma_j + \epsilon_{ij}, \tag{3.41}$$

where

- $\boldsymbol{\beta}_i = \begin{bmatrix} \beta_{0i} & \beta_{1i} \end{bmatrix}^T$ is a vector of the fixed effects for the i^{th} connectivity link in HC group (β_{0i}) and disease group (β_{1i}).
- $\boldsymbol{X}_{ij}$ is a 1×2 design matrix for the fixed effects,

$$\boldsymbol{X}_{ij} = \begin{bmatrix} 1 - Grp_j & Grp_j \end{bmatrix}, \tag{3.42}$$

- γ_j and ϵ_{ij} are defined the same as for Equation (3.1).

Similarly as in Equation (3.35), the marginal distribution of y_{ij} is

$$f(y_{ij}) = \int f(y_{ij} \mid \gamma_j; \boldsymbol{\beta}_i, \sigma^2_{Grp_j i}) f(\gamma_j; \sigma^2_\gamma) d\gamma_j, \tag{3.43}$$

where

$$f(y_{ij} \mid \gamma_j; \boldsymbol{\beta}_i, \sigma^2_{Grp_j i}) = \left(\frac{1}{2\pi\sigma^2_{Grp_j i}} \right)^{\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2_{Grp_j i}} (y_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta}_i - \gamma_j)^2 \right], \tag{3.44}$$

and $f(\gamma_j; \sigma^2_\gamma)$ is defined in Equation (3.37).

The marginal log-likelihood function is then the sum of the log-likelihood in Equation (3.43) over all N subjects for the i^{th} connectivity link,

$$\log L_i = \sum_{j=1}^{N} \log\big[f(y_{ij}) \big]. \tag{3.45}$$

The first derivative of the marginal log-likelihood function in Equation (3.45) with respect to $\boldsymbol{\beta}_i$ is,

$$\begin{aligned}
\frac{\partial \log L_i}{\partial \boldsymbol{\beta}_i} &= \frac{\partial}{\partial \boldsymbol{\beta}_i} \sum_{j=1}^{N} \log\big[f(y_{ij}) \big] \\
&= \sum_{j=1}^{N} \frac{\partial \log f(y_{ij})}{\partial \boldsymbol{\beta}_i} \\
&= \sum_{j=1}^{N} \frac{1}{f(y_{ij})} \frac{\partial f(y_{ij})}{\partial \boldsymbol{\beta}_i} \\
&= \sum_{j=1}^{N} \frac{1}{f(y_{ij})} \frac{\partial \big[\int f(y_{ij} \mid \gamma_j) f(\gamma_j) d\gamma_j \big]}{\partial \boldsymbol{\beta}_i} \\
&= \sum_{j=1}^{N} \frac{1}{f(y_{ij})} \int \frac{\partial f(y_{ij} \mid \gamma_j)}{\partial \boldsymbol{\beta}_i} f(\gamma_j) d\gamma_j
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{N} \frac{1}{f(y_{ij})} \int f(y_{ij} \mid \gamma_j) f(\gamma_j) \frac{\partial \log f(y_{ij} \mid \gamma_j)}{\partial \boldsymbol{\beta}_i} d\gamma_j \\
&= \sum_{j=1}^{N} \int \frac{f(y_{ij} \mid \gamma_j) f(\gamma_j)}{f(y_{ij})} \boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} (y_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta}_i - \gamma_j) d\gamma_j \\
&= \sum_{j=1}^{N} \int f(\gamma_j \mid y_{ij}) \boldsymbol{X}_j^T \sigma_{Grp_j i}^{-2} (y_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta}_i - \gamma_j) d\gamma_j \\
&= \sum_{j=1}^{N} \int \left[\boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} (y_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta}_i) f(\gamma_j \mid y_{ij}) - \boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} f(\gamma_j \mid y_{ij}) \gamma_j \right] d\gamma_j \\
&= \sum_{j=1}^{N} \left[\int \boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} (y_j - \boldsymbol{X}_{ij}\boldsymbol{\beta}_i) f(\gamma_j \mid y_{ij}) d\gamma_j - \int \boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} f(\gamma_j \mid y_{ij}) \gamma_j d\gamma_j \right] \\
&= \sum_{j=1}^{N} \left[\boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} (y_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta}_i) \int f(\gamma_j \mid y_{ij}) d\gamma_j - \boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} \int f(\gamma_j \mid y_{ij}) \gamma_j d\gamma_j \right] \\
&= \sum_{j=1}^{N} \left[\boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} (y_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta}_i) - \boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} \tilde{\gamma}_j \right] \\
&= \sum_{j=1}^{N} \boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} (y_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta}_i - \tilde{\gamma}_j), \qquad (3.46)
\end{aligned}
$$

as $\int f(\gamma_j \mid y_{ij}) d\gamma_j = 1$ and $\int f(\gamma_j \mid y_{ij}) \gamma_j d\gamma_j = \tilde{\gamma}_j$.

By equating Equation (3.46) to zero we have

$$
\begin{aligned}
&\frac{\partial \log L_i}{\partial \boldsymbol{\beta}_i} = \sum_{j=1}^{N} \boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} (y_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta}_i - \tilde{\gamma}_j) = \mathbf{0} \\
\implies &\sum_{j=1}^{N} \boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} \boldsymbol{X}_{ij} \boldsymbol{\beta}_i = \sum_{j=1}^{N} \boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} (y_{ij} - \tilde{\gamma}_j) \\
\implies &\hat{\boldsymbol{\beta}}_i = \left(\sum_{j=1}^{N} \boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} \boldsymbol{X}_{ij} \right)^{-1} \left[\sum_{j=1}^{N} \boldsymbol{X}_{ij}^T \sigma_{Grp_j i}^{-2} (y_{ij} - \tilde{\gamma}_j) \right]. \qquad (3.47)
\end{aligned}
$$

By Equation (A.11), the variance of $\hat{\boldsymbol{\beta}}_i$ is

$$
\mathrm{Var}(\hat{\boldsymbol{\beta}}_i) = \left[\boldsymbol{X_i}^T \left(\boldsymbol{Z}_i \sigma_\gamma^2 \boldsymbol{Z}_i^T + \boldsymbol{\Sigma}_{\epsilon i} \right)^{-1} \boldsymbol{X} \right]^{-1}, \qquad (3.48)
$$

where

- $\boldsymbol{X}_i$ is a $N \times 2$ design matrix for the fixed effects $\boldsymbol{\beta}_i$.

$$\boldsymbol{X}_i = \begin{bmatrix} 1 - Grp_1 & Grp_1 \\ 1 - Grp_2 & Grp_2 \\ \vdots & \vdots \\ 1 - Grp_N & Grp_N \end{bmatrix}_{N \times 2} = \mathbf{1}_N \otimes \begin{bmatrix} 1 - Grp_j & Grp_j \end{bmatrix}, \tag{3.49}$$

- $\boldsymbol{Z}$ is a $N \times N$ identity design matrix for the random subject effects,

$$\boldsymbol{Z} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{N \times N} = \boldsymbol{I}_N, \tag{3.50}$$

- $\boldsymbol{\Sigma}_{\epsilon i}$ is the variance of $\boldsymbol{\epsilon}_i$ is a $N \times N$ diagonal matrix:

$$\boldsymbol{\Sigma}_{\epsilon i} = \begin{bmatrix} \sigma_{Grp_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{Grp_N} \end{bmatrix}_{N \times N}, \tag{3.51}$$

Based on Equations (3.47) and (3.48), we have the asymtotic distribution of $\hat{\boldsymbol{\beta}}_i$ as:

$$\hat{\boldsymbol{\beta}}_i \sim N\left(\boldsymbol{\beta}_i, \left[\boldsymbol{X_i}^T \left(\boldsymbol{Z}_i \sigma_\gamma^2 \boldsymbol{Z}_i^T + \boldsymbol{\Sigma}_{\epsilon i}\right)^{-1} \boldsymbol{X}\right]^{-1}\right), \tag{3.52}$$

Then the asymptotic distribution of $\boldsymbol{h}\hat{\boldsymbol{\beta}}_i$ is:

$$\boldsymbol{h}\hat{\boldsymbol{\beta}}_i \sim N\left(\boldsymbol{h}\boldsymbol{\beta}_i, \boldsymbol{h}\left[\boldsymbol{X_i}^T \left(\boldsymbol{Z}_i \sigma_\gamma^2 \boldsymbol{Z}_i^T + \boldsymbol{\Sigma}_{\epsilon i}\right)^{-1} \boldsymbol{X}\right]^{-1} \boldsymbol{h}^T\right), \tag{3.53}$$

Under $H_{0,i}$,

$$\boldsymbol{h}\hat{\boldsymbol{\beta}}_i \sim N\left(\mathbf{0},\, \boldsymbol{h}\left[\boldsymbol{X_i}^T\left(\boldsymbol{Z}_i\sigma_\gamma^2\boldsymbol{Z}_i^T + \boldsymbol{\Sigma}_{\epsilon i}\right)^{-1}\boldsymbol{X}\right]^{-1}\boldsymbol{h}^T\right), \tag{3.54}$$

The Wald test statistic to test the null hypothesis in Equation (3.40) is:

$$c_i = (\boldsymbol{h}\hat{\boldsymbol{\beta}}_i)\left(\boldsymbol{h}\left[\boldsymbol{X_i}^T\left(\boldsymbol{Z}_i\hat{\sigma}_\gamma^2\boldsymbol{Z}_i^T + \hat{\boldsymbol{\Sigma}}_{\epsilon i}\right)^{-1}\boldsymbol{X}\right]^{-1}\boldsymbol{h}^T\right)^{-1}(\boldsymbol{h}\hat{\boldsymbol{\beta}}_i)^T, \tag{3.55}$$

which follows an asymptotic χ^2-distribution with 1 degree of freedom under $H_{0,i}$. $H_{0,i}$ is rejected if $c_i \geq \chi^2_{1-\alpha}(1)$. The square root of c_i (denoted by t_i) follows an asymptotic standard normal distribution,

$$t_i = \sqrt{c_i} = (\boldsymbol{h}\hat{\boldsymbol{\beta}}_i)\left(\boldsymbol{h}\left[\boldsymbol{X_i}^T\left(\boldsymbol{Z}_i\hat{\sigma}_\gamma^2\boldsymbol{Z}_i^T + \hat{\boldsymbol{\Sigma}}_{\epsilon i}\right)^{-1}\boldsymbol{X}\right]^{-1}\boldsymbol{h}^T\right)^{-\frac{1}{2}} \sim N(0,1). \tag{3.56}$$

And thus equivalently we can reject $H_{0,i}$ if $t_i \geq z_{1-\alpha}$, where $z_{1-\alpha}$ denotes the critical value at level $1-\alpha$ from the standard normal distribution. The test statistics $t_i, i = 1, \cdots, m$, are then used for the further analysis described in the next chapter.

CHAPTER 4

A BAYESIAN APPROACH FOR CONTROLLING THE FALSE DISCOVERY RATE

4.1 Large-Scale Simultaneous Hypothesis Testing Problem

The objective of a hypothesis testing is to decide which hypothesis, the null hypothesis (denoted by H_0) or the alternative hypothesis (denoted by H_1), is true given observed test statistics t based on a sample taken from a population. In practice, the null hypothesis assumes no effect, for example, no difference between the means of the two groups. We can treat the decision on H_0 as a binary variable, in which $H_0 = 0$ when H_0 is true, or $H_0 = 1$ otherwise. The probability of the null hypothesis being true given observed test statistics is

$$P(H_0 = 0 \mid T = t), \tag{4.1}$$

The probability of the alternative hypothesis being true (or the null hypothesis is false) given observed t is

$$P(H_0 = 1 \mid T = t) = 1 - P(H_0 = 0 \mid T = t), \tag{4.2}$$

The larger $P(H_0 = 0 \mid T = t)$ is, the greater the likelihood that the H_0 is true, and vice versa.

However, the p-value used in classic hypothesis testing does not provide a statistical measure of the probability that the null hypothesis is true based on the observed data (Nuzzo, 2014; Wasserstein and Lazar, 2016). The p-value relies on the tail regions of the distribution under null hypothesis and measures the probability that the test statistic is equal to or more extreme

(i.e, greater for a right-sided test, or lower for a left-sided test) than the observed value, under the assumption that H_0 is true,

$$p\text{-value} = \begin{cases} P(T \geq t \mid H_0 = 0) = \int_t^{\infty} f_0(t)dt, & \text{for a right-sided test,} \\ P(T \leq t \mid H_0 = 0) = \int_{-\infty}^{t} f_0(t)dt, & \text{for a left-sided test,} \\ P(T \geq |t| \mid H_0 = 0) = \int_{-\infty}^{-|t|} f(t)dt + \int_{|t|}^{\infty} f(t)dt, & \text{for a two-sided test.} \end{cases} \tag{4.3}$$

where $f_0(t)$ is the probability density function of test statistics under null hypothesis.

There are two main differences between the concept of p-value, using the one for a right-sided test in Equation (4.3) as an example, the probability that the null hypothesis is true given the observed data in Equation (4.1). First, $T = t$ is substituted by $T \geq t$ (i.e., the tail area under the null hypothesis) since for continuous variables $P(T = t \mid H_0 = 0) = 0$, yet it is still possible that the observed test statistics is exactly equal to t. Second, the two have a reversed order in terms of the conditional probability, Equation (4.1) is the conditional probability of H_0 being true given t, while Equation (4.3) is the conditional probability of $T \geq t$ given H_0 is true. Based on the statements on "statistical significance and p-values" made by the American Statistical Association (Wasserstein and Lazar, 2016), p-value measures the "statistical incompatibility between the observed data and the null hypothesis". That is, assuming the null hypothesis is true a smaller p-value would suggest a greater incompatibility between the data and the null hypothesis. It is important to note that the p-value cannot be used to calculate the probability that the null hypothesis is true or not backwards.

For a single hypothesis testing, the probability of falsely rejecting a true null hypothesis H_0, also known as type I error rate denoted by α $(0 < \alpha < 1)$ is controlled by choosing a rejection region. For multiple testing that involves the simultaneous testing of more than one hypothesis,

when there are m null hypotheses to be tested simultaneously, the overall null hypothesis is the intersection of the m null hypotheses,

$$H_0 : \bigcap_{i=1}^{m} H_{0,i}, \tag{4.4}$$

That is, the intersection null means that all of the individual null hypotheses are true. The four possible outcomes from m hypotheses tested are shown in Table II, where only number of null hypotheses tested (denoted by m) and number of rejected hypotheses (denoted by D) are known.

TABLE II
CLASSIFICATION OF ALL NULL HYPOTHESES TESTED

		Testing status		Total
		Accept H_0	Reject H_0	
True status	True H_0	TND[a]	FD[b]	m_0[c]
	False H_0	FND[d]	TD[e]	m_1[f]
	Total	$m - D$	D[g]	m[h]

[a] TND = number of true non-discoveries (also called true negatives).
[b] FD = number of false discoveries (also called false positives, type I errors).
[c] m_0 = number of true null hypotheses.
[d] FND = number of false non-discoveries (also called false negatives, type II errors).
[e] TD = number of true discoveries (also called true positives).
[f] m_1 = number of false null hypotheses (also called true alternative hypotheses), $m_1 = m - m_0$.
[g] D = number of discoveries (also called positives).
[h] m = number of null hypotheses tested.

The probability of at least one type I error (i.e., falsely rejecting at least one $H_{0,i}$), is referred to as family-wise error rate (FWER),

$$FWER = P(FD \geq 1) = 1 - P(FD = 0), \tag{4.5}$$

where FD is number of false discoveries / type I errors. Without adjusting for multiplicity, FWER can be calculated as $1 - (1 - \alpha)^m$, assuming all the m hypotheses are independent of each other. It is very close to 1 when m is large. One commonly used method to control FWER for multiple testing is the Bonferroni correction (Dunn, 1961), in which the significance level for each individual test is set at $\frac{\alpha}{m}$. The Bonferroni correction rejects $H_{0,i}$ if $p_i \leq \frac{\alpha}{m}$, thus guaranteeing the FWER $\leq \alpha$ by the Bonferroni inequality as shown below,

$$P\left\{\bigcup_{i=1}^{m}(p_i \leq \frac{\alpha}{m})\right\} \leq \alpha, \tag{4.6}$$

Though the Bonferroni method is simple and assumption-free, it tends to be extremely conservative especially when the number of comparisons m is large. The very small significance level $\frac{\alpha}{m}$ makes it very difficult to reject any individual null hypothesis, yielding to a very high probability of false negative error and thus a very low power (Nakagawa, 2004). We should note that when one hypothesis is rejected at significance level of $\frac{\alpha}{m}$, the number of remaining tests is reduced to $m - 1$. Using the same significance level of $\frac{\alpha}{m}$ for the rest of the tests actually ignores such fact. To take into account the sequential rejecting with adjusted significance level, Holm (1979) proposed a sequential Bonferroni correction procedure to control FWER while maintaining adequate power simultaneously irrespective of the independence of the test statistics. Holm's step-down procedure employs stepwise adjustments to the significance level

based on the rank order of the unadjusted p-values of the multiple tests, which is illustrated as follows:

1. First order the m unadjusted p-values in ascending order: $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$, where $p_{(1)}$ and $p_{(m)}$ refer to the smallest and the largest p-values, respectively. Accordingly, $H_{0,(i)}$ is the null hypothesis corresponding to $p_{(i)}$, the i^{th} ordered p-value.
2. Start from the smallest p-value $p_{(1)}$: If $p_{(1)} \leq \frac{\alpha}{m}$, reject $H_{0,(1)}$ and then move to $p_{(2)}$; if not, accept all $H_{0,(1)}, \cdots, H_{0,(m)}$ and stop.
3. For $i = 2, \cdots, m-1$: If $p_{(i)} \leq \frac{\alpha}{m-i+1}$, reject $H_{0,(i)}$ and move to $p_{(i+1)}$; if not, accept $H_{0,(i)}, \cdots, H_{0,(m)}$ and stop.
4. For the largest p-value $p_{(m)}$: If $p_{(m)} \leq \alpha$, reject $H_{0,(m)}$; if not, accept $H_{0,(m)}$.

Holm's procedure provides a strong control of FWER and is more powerful than Bonferroni. However, when m is large, the power is still low. Hochberg (1988) developed a similar but more powerful procedure to control FWER as compared to Holm's method. Hochberg's step-up procedure starts from the largest p-value $p_{(m)}$ and continue until find the first i that $p_{(i)} \leq \frac{\alpha}{m-i+1}$ at which points all smaller p-values are considered significant. That is, for any $i = m, \cdots, 1$, if $p_{(i)} \leq \frac{\alpha}{m-i+1}$, reject all $H_{(j)}$, $j = 1, \cdots, i$. Meanwhile, Hommel (1988) gave another modified Bonferroni procedure that is in general more powerful than Hochberg's procedure (Hommel, 1989). Using Hommel's method, reject all hypotheses whose p-values are $\leq \frac{\alpha}{k}$, where k is the largest integer for which $p_{(m-k+j)} > \frac{j}{k}\alpha$, for all $j = 1, \cdots, k$. If k does not exist, reject all $H_{0,i}$, $i = 1, \cdots, m$. One needs to be careful that Hochberg's method requires justification, that is, it only applies when the tests are independent of each other. Holm's method does not require such independence assumption and hence Holm's procedure should be used if there are concerns

about potential dependencies among the tests, otherwise using Hochberg or Hommel's method instead.

4.2 False Discovery Rate

The FWER control is fit for the purpose when we wish to avoid any false positive. It can be useful when there are a fairly small number of multiple tests and few of them might be significant. However, for large-scale simultaneous hypothesis testing such as genomics and neuroimaging studies with a large number of multiple tests, the majority of the hypotheses tested are pure noise and only a very small proportion of them contains signals, known as sparsity issue. In these cases, FWER control methods are so stringent that some important differences that may be meaningful to the researchers have little chance to be detected. Given this sparsity issue, we are more concerned if all potentially true alternatives are detected and thus, we need an approach that allows for a certain number of false positives but controls the proportion of false positives among all positive results under a certain desired significance level.

In order to identify as many significant findings as possible while still maintaining a low false positive rate, Benjamini and Hochberg (1995) established the key concept of false discovery rate (FDR) as a measure of the expected proportion of falsely rejected hypotheses (also called false discoveries, false positives, or type I errors) among all rejected hypotheses (also called discoveries or positives) when a large number of null hypotheses are tested simultaneously. Using the same notation as in Table II, FDR is defined as the expected proportion of false positives among all rejected hypotheses,

$$FDR = \mathrm{E}\left(\frac{FD}{FD+TD}\right) = \mathrm{E}\left(\frac{FD}{D}\right), \tag{4.7}$$

where FD, TD and D refer to number of false discoveries, number of true discoveries and number of discoveries, respectively. Generally FDR method provides a more powerful approach than FWER in multiple testing (Benjamini and Hochberg, 1995). We can show that FDR is less than FWER in general by,

$$\begin{aligned} FDR = \mathrm{E}\left(\frac{FD}{D}\right) &= \mathrm{E}\left(\frac{FD}{D} \mid FD = 0\right) P(FD = 0) + \mathrm{E}\left(\frac{FD}{D} \mid FD \geq 1\right) P(FD \geq 1) \\ &= 0 \times P(FD = 0) + \mathrm{E}\left(\frac{FD}{D} \mid FD \geq 1\right) P(FD \geq 1) \\ &< FWER, \end{aligned} \tag{4.8}$$

As a contrast to FDR, the false positive rate (FPR) is the expected proportion of true null hypotheses that are incorrectly called significant, $\mathrm{E}\left(\frac{FD}{m_0}\right)$. Note that FDR and FPR share the same numerator, the number of false discoveries, FD; what distinguish between the two terms is the denominator, FDR has the number of discoveries (D) as FDR focuses on all discoveries, while FPR uses the number of true null hypothesis (m_0). Storey and Tibshirani (2003) shows that the FDR measure balances between sensitivity (defined as true positive rate $= \frac{TD}{m_1}$) and specificity (defined as true negative rate $= \frac{TND}{m_0}$), by expressing the FDR in terms of sensitivity and specificity as follows,

$$\begin{aligned} FDR = \mathrm{E}\left(\frac{FD}{D}\right) &= \mathrm{E}\left(\frac{FD}{FD + TD}\right) \\ &= \mathrm{E}\left(\frac{m_0 - TND}{(m_0 - TND) + TD}\right) \\ &= \mathrm{E}\left[\frac{m_0 \times (1 - \frac{TND}{m_0})}{m_0 \times (1 - \frac{TND}{m_0}) + m_1 \times \frac{TD}{m_1}}\right] \\ &= \mathrm{E}\left[\frac{m_0 \times (1 - \text{specificity})}{m_0 \times (1 - \text{specificity}) + m_1 \times \text{sensitivity}}\right]. \end{aligned} \tag{4.9}$$

Equivalently, FDR can be defined as the expected proportion of false discoveries among all rejected hypotheses given at least one rejection is made times the probability of making at least one rejection,

$$FDR = \mathrm{E}\left(\frac{FD}{D} \mid D > 0\right) P(D > 0), \tag{4.10}$$

Storey (2002, 2003) defined a positive FDR denoted by pFDR,

$$pFDR = \mathrm{E}\left(\frac{FD}{D} \mid D > 0\right), \tag{4.11}$$

here "positive" is used to indicate that the expectation is conditional on occurrence of rejecting at least one hypothesis. Note that $FDR = pFDR \times P(D > 0) \leq pFDR$. In most genomics and neuroimaging studies with large-scale simultaneous hypothesis testing, we can guarantee that $P(D > 1) \approx 1$ so FDR and pFDR are very close. Storey (2002, 2003) proved that pFDR can be interpreted using a Bayesian mixture model as the posterior probability of a null hypothesis being true given that it is rejected,

$$\begin{aligned} pFDR &= P(H_0 = 0 \mid p\text{-value} \leq \gamma) \\ &= \frac{p_0 P(p\text{-value} \leq \gamma \mid H_0 = 0)}{P(p\text{-value} \leq \gamma)} \\ &= \frac{p_0 P(p\text{-value} \leq \gamma \mid H_0 = 0)}{p_0 P(p\text{-value} \leq \gamma \mid H_0 = 0) + (1 - p_0) P(p\text{-value} \leq \gamma \mid H_0 = 1)}, \end{aligned} \tag{4.12}$$

where p_0 is the prior probability that the null hypothesis is true, $p_0 = P(H_0 = 0)$, and γ is the fixed threshold at which level the test is called significant (i.e., the hypothesis is rejected).

Another FDR definition is marginal FDR denoted by mFDR (Genovese and Wasserman, 2002a; Storey, 2003; Sun and Cai, 2007),

$$mFDR = \frac{\mathrm{E}(FD)}{\mathrm{E}(D)}, \tag{4.13}$$

Storey (2003) showed that pFDR and mFDR are equivalent under the assumption that the test statistics are random variables with a mixture of null and alternative distributions. Genovese and Wasserman (2002a) proved that under independence assumption, mFDR and FDR have the same asymptotic properties when number of hypotheses m is large,

$$mFDR = FDR + O(m^{-1/2}), \tag{4.14}$$

To control FDR, Benjamini and Hochberg (1995) employed a step-up procedure (referred to as Benjamini and Hochberg procedure) described as follows,

1. Sort unadjusted p-values in ascending order by $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$ and the corresponding null hypotheses are $H_{0,(1)}, H_{0,(2)}, ..., H_{0,(m)}$,
2. For a given q, $0 < q < 1$, let $k = \arg\max_i \left\{ i : p_{(i)} \leq \frac{i}{m} q \right\}$,
3. Then reject all $H_{0,(i)}, i = 1, \cdots, k$,

Benjamini and Hochberg (1995) proved that for independent tests, using the above procedure can guarantee $FDR \leq \frac{m_0}{m} q \leq q$, and thus the FDR is controlled at q.

Since Genovese et al. (2002) proposed the Benjamini and Hochberg procedure for controlling the FDR in fMRI data analysis, the method has been widely applied in neuroimging data analysis. Like single hypothesis testing, classical Benjamini and Hochberg FDR theory is based

on the well-known p-values that rely on the tail areas of null hypothesis. Thus, it is also named as global FDR. The FDR provides an overall accuracy for a set of significant tests and requires less assumptions, but it gives weaker results in general and does not provide a direct measure of the probability of each individual test is a false positive.

Benjamini and Hochberg (2000) proposed an adaptive procedure to incorporate the information of unknown proportion of null hypotheses for FDR control (referred to as adaptive Benjamini and Hochberg procedure). The adaptive procedure has similar setting as the Benjamini and Hochberg procedure described above, in which we determine $k = \arg\max_i \left\{ i : p_{(i)} \leq \frac{i}{\hat{p}_0 m} q \right\}$, where $\hat{p}_0$ is a conservative estimate of p_0. As compared to Benjamini and Hochberg procedure, it has also been shown that the adaptive procedure gained more power.

Storey (2002) introduced a q-value approach as a measure of the minimum FDR that can be reached if the test is considered significant, as well as an algorithm to estimate the q-value based on the estimated proportion of p-values under null hypothesis from a range of threshold at which level the test is called significant from 0 to 1. Conceptually the q-value is similar to the p-value, and it can be interpreted as the "Bayesian posterior p-value" (Storey, 2003; Storey and Tibshirani, 2003). For a given test, the p-value of the test tells the minimum FPR that is incurred if the test is declared significant, while the q-value of the test is the minimum of FDR estimates among all thresholds at which the test is declared significant.

4.3 Local False Discovery Rate

Efron et al. (2001); Efron (2007) proposed an empirical Bayes local false discovery rate (Lfdr) method (henceforth described as Efron's Lfdr method) based on the densities of test statistics, as an alternative of the tail area global FDR for FDR control in large-scale simultaneous hypothesis testing. Efron's Lfdr method has enabled us to directly compute the

probability of the null hypothesis being true given the observed test statistic, which is thought of as a "local" FDR (Efron, 2004), and hence controls the FDR more precisely as compared to the p-value based Benjamini and Hochberg procedures (Cai and Sun, 2017) that rely on the tail areas of null hypothesis. It should be also noted that the local FDR method provides a direct measure of the probability of the null hypothesis being true given the observed t, which is in line with the objective of hypothesis testing in Equation (4.1).

In Lfdr theory, the truth of hypothesis $H_{0,i}$ is assumed to be i.i.d. Bernoulli random variables. Let $H_{0,i} = 0$ when $H_{0,i}$ is true, and $H_{0,i} = 1$ otherwise, and t_i be the corresponding test statistic, we have,

$$\begin{aligned} H_{0,i} &\overset{iid}{\sim} Bernoulli(p_1), \\ t_i &\sim \begin{cases} f_0, & \text{if } H_{0,i} = 0 \\ f_1, & \text{if } H_{0,i} = 1 \end{cases}, \end{aligned}$$

where $p_0 = P(H_{0,i} = 0)$ and $p_1 = P(H_{0,i} = 1) = 1 - p_0$ are the prior probabilities, f_0 and f_1 are the probability density functions of test statistic under null hypothesis and under alternative hypothesis, respectively.

Thus, t_i follows a mixture distribution marginally,

$$f(t_i) = p_0 f_0(t_i) + p_1 f_1(t_i).$$

Then, in the Bayesian framework, $Lfdr_i(z_i)$ is defined as the posterior probability that $H_{0,i}$ being true at a given observation t_i. By Bayes' rule we have,

$$Lfdr_i(t_i) = P(H_{0,i} = 0 \mid T = t_i) = \frac{p_0 f_0(t_i)}{f(t_i)}, \tag{4.15}$$

The Lfdr has exactly the same form as the objective of hypothesis testing as shown in Equation (4.1).

Efron (2004, 2007) provided details in estimation of p_0, $f_0(t)$ and $f(t)$ that are required to calculate Lfdr in Equation (4.15). Assuming that the mixture density of $f(t_i)$ for all test statistics under null and alternative hypothesis is smooth and has a $(p+1)$-parameter exponential family form as,

$$f(t) = \exp\left(\sum_{j=0}^{p} \beta_j t^j\right), \tag{4.16}$$

Lindsey's method utilizing a smooth and flexible parametric model based on Poisson regression is used to estimate $f(t_i)$ (Lindsey, 1974a,b; Efron and Tibshirani, 1996). First, discretize the likelihood by partitioning the entire sample space $\mathcal{B} = \{t_i\}_{i=1}^{m}$ into K bins of equal width Δt, denoted by $\{\mathcal{B}_1, \cdots, \mathcal{B}_K\}$. Let m_k be the number of t_i's in B_k, $k = 1, \cdots, K$,

$$m_k = \sum_{i=1}^{m} I\{t_i \in \mathcal{B}_k\},\ k = 1, \cdots, K,$$

where $I\{\cdot\}$ is an indicator function.

$(m_1, \cdots, m_K)$ has a multinomial distribution with m (the total number of t_i's) and probabilities $(\pi_1(\boldsymbol{\beta}), \cdots, \pi_K(\boldsymbol{\beta}))$, where $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_p)$,

$$(m_1, \cdots, m_K) \sim Multinomial_K\left(m, \left(\pi_1(\boldsymbol{\beta}), \cdots, \pi_K(\boldsymbol{\beta})\right)\right), \tag{4.17}$$

Denote the middle point of the interval in $\mathcal{B}_k$ by v_k, the probability π_k can be approximated using,

$$\pi_k = P(t \in B_k) = \int_{B_k} f(t)dt = \int_{v_k-\Delta t/2}^{v_k+\Delta t/2} f(t)dt \approx f(v_k)\Delta t. \tag{4.18}$$

Then, the expected value of m_k is,

$$\mathrm{E}(m_k) = mf(v_k)\Delta t. \tag{4.19}$$

For a sufficiently large K, Poisson regression trick can be applied to model m_k's as independent random Poisson variables,

$$m_k \overset{ind}{\sim} Poisson(\lambda_k),\ k = 1, \cdots, K, \tag{4.20}$$

where λ_k is the expected value of m_k. Note that in Equation (4.17) m_k's are dependent as $\sum_{k=1}^{K} m_k = m$, while Equation (4.20) assumes m_k's are independent. By Equation (4.19), we have

$$\lambda_k = \mathrm{E}(m_k) = m\pi_k = m\Delta t f(v_k) \propto f(v_k), \tag{4.21}$$

Thus, $\log(\lambda_k)$ can be modeled as a $(p-1)^{th}$ degree polynomial function of v_k using a generalized linear model (GLM) with Poisson link function,

$$\log(\lambda_k) = \sum_{j=0}^{p} \beta_j v_k^j, \tag{4.22}$$

The maximum likelihood estimates $(\hat{\beta}_1, \cdots, \hat{\beta}_p)$ can be obtained from a standard GLM Poisson regression. Then calculate $\hat{\beta}_0$, the normalizing constant, by making the $f(t)$ integrate to 1,

$$\hat{\beta}_0 = \log \left\{ \int_{\mathcal{B}} \exp \left(\sum_{j=1}^{p} \hat{\beta}_p t^j \right) \right\}^{-1}, \tag{4.23}$$

Efron (2004) has discussed that a 6^{th}-order polynomial model would provide a sufficiently good of fit.

Regarding estimation of $f_0(t)$ and p_0, Efron (2007) proposed a "central matching" method based on "zero assumption" that the central peak of the test statistics distribution is comprised mostly of the null hypotheses. As in large-scale hypothesis testing problem for genomics and neuroimaging studies, the majority of the hypotheses are null, which leads to a common and reasonable assumption that p_0 is close to 1. The underlying distribution of the test statistics under null hypothesis is assumed to be standard normal $f_0(t) \sim N(0, 1)$ known as "theoretical null", or $f_0(t) \sim N(\mu_0, \sigma_0^2)$ known as"empirical null" (Efron, 2004, 2007). Assuming the empirical null, $p_0 f_0$ can be estimated by fitting a quadratic curve to approximate the histogram around zero area through Poisson regression as described above,

$$\log \big(p_0 f_0(t) \big) = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2. \tag{4.24}$$

Given $f_0(t) \sim N(\mu_0, \sigma_0^2)$,

$$\begin{aligned}
\log(p_0 f_0(t)) &= \log\left\{ p_0 \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2\sigma_0^2}(t-\mu_0)^2 \right) \right\} \\
&= \log p_0 - \frac{1}{2}\log(2\pi\sigma_0^2) - \frac{1}{2\sigma_0^2}(t-\mu_0)^2 \\
&= \log p_0 - \frac{1}{2}\log(2\pi\sigma_0^2) - \frac{\mu_0^2}{2\sigma_0^2} + \frac{\mu_0}{\sigma_0^2}t - \frac{1}{2\sigma_0^2}t^2. \qquad (4.25)
\end{aligned}$$

By setting Equations (4.24) and (4.25) equal,

$$\begin{aligned}
\hat{\beta}_0 &= \log \hat{p}_0 - \frac{1}{2}\log(2\pi\hat{\sigma}_0^2) - \frac{\hat{\mu}_0^2}{2\hat{\sigma}_0^2}, \\
\hat{\beta}_1 &= \frac{\hat{\mu}_0}{\hat{\sigma}_0^2}, \\
\hat{\beta}_2 &= -\frac{1}{2\hat{\sigma}_0^2}, \qquad (4.26)
\end{aligned}$$

Now $\hat{\mu}_0$, $\hat{\sigma}_0^2$ and $\hat{p}_0$ can be calculated as,

$$\begin{aligned}
\hat{\sigma}_0^2 &= -\frac{1}{2\hat{\beta}_2}, \\
\hat{\mu}_0 &= -\frac{\hat{\beta}_1}{2\hat{\beta}_2}, \\
\hat{p}_0 &= \exp\left\{ \hat{\beta}_0 + \frac{1}{2}\log\left(-\frac{\pi}{\hat{\beta}_2} \right) - \frac{\hat{\beta}_1^2}{4\hat{\beta}_2} \right\}, \qquad (4.27)
\end{aligned}$$

As long as the "zero assumption" holds when p_0 is close to 1, the estimates by "central matching" method are nearly unbiased.

Efron and Tibshirani (2002) showed the close relationship between the test statistics based Lfdr and the p-value based Benjamini and Hochberg FDR. Rewrite FDR in Bayesian form by replacing the probability density functions f_0 and f_1 in Equation (4.15) with the corresponding

cumulative distribution functions F_0 and F_1, respectively, and denote the Bayesian FDR by $Fdr_i(T \leq t_i)$, the posterior probability that $H_{0,i}$ is true given $T \leq t_i$ is,

$$Fdr_i(T \leq t_i) = P(H_{0,i} = 0 \mid T \leq t_i) = \frac{p_0 F_0(t_i)}{F(t_i)}, \tag{4.28}$$

Then the relationship between Fdr_i and $Lfdr_i$ is given by (Efron and Tibshirani, 2002):

$$Fdr_i(T \leq t_i) = \frac{\int_{-\infty}^{t_i} Lfdr(t) f(t) dz}{\int_{-\infty}^{t_i} f(t) dz} = \mathrm{E}\big[Lfdr(t) \mid T \leq t_i\big], \tag{4.29}$$

So $Fdr_i(T \leq t_i)$ is the conditional expectation of $Lfdr_i(t)$ given $T \leq t_i$ and can be estimated as an average of $Lfdr_i(t)$ over all $T \leq t_i$. When $|t|$ increases $Lfdr_i(t)$ decreases, so it is expected that $Fdr_i(T \leq t_i)$ is always smaller than $Lfdr_i(t)$.

Sun and Cai (2007) introduced an oracle procedure based on test statistics for FDR control by regarding multiple testing as a compound decision problem, intended to minimize false non-discovery rate (i.e., maximizing power, see Section 4.5 for detailed review) that is subject to a constraint on FDR $\leq q$. In the oracle and adaptive procedure,

1. Sort Lfdr estimates in ascending order by $\widehat{Lfdr}_{(1)} \leq \widehat{Lfdr}_{(2)} \leq \cdots \leq \widehat{Lfdr}_{(m)}$ and the corresponding hypotheses are $H_{(1)}, H_{(2)}, ..., H_{(m)}$,
2. For a given q, , where $q = 0.2 - 0.3$ is reasonable, let $k = max_i \left\{ i : \frac{1}{i} \sum_{j=1}^{i} \widehat{Lfdr}_{(j)} \leq q \right\}$,
3. Then reject all $H_{(i)}, i = 1, ..., k$.

Cai (2008) has commented on the inefficiency problem with the p-value based FDR control procedures and suggested that Lfdr be the fundamental quantity that can be used for optimal FDR control in large-scale multiple testing. Jin and Cai (2007) proposed an empirical

characteristic function approach using normal mixtures and Fourier analysis to estimate null distribution parameters and p_0.

4.4 Considerations for FDR Level q

The desired FDR level q is the maximum proportion of false discoveries among rejected hypotheses that can be tolerated on average. For a single hypothesis testing, generally we fix the type I error rate α at 5%, and then try to achieve a power at 80%. For multiple hypothesis testing, similarly we need to determine the value of q at which level the FDR is controlled. For example, choosing a q of 0.10 means that on average 10% of the significant tests we find are actually false positives. Genovese et al. (2002) suggested a range of FDR levels between 0.1 and 0.2 be acceptable for neuroimaging studies. Bhaumik et al. (2018b) further discussed the considerations to determine the FDR level q in terms of the proportion of null hypothesis (p_0) and adjustment for type I error rate (α) and power (also called average power for multiple hypothesis testing).

Let m_0 and m_1 denote the number of truly null hypotheses and truly alternative hypotheses, respectivey. Based on mFDR defined in Equation (4.14), in which E(FD) and E(D) can be approximated as, E(FD) $\approx \alpha m_0$ and E(D) $\approx \alpha m_0 + \text{power} \times m_1$, the expected FDR is calculated as,

$$mFDR \approx \frac{\text{E}(FD)}{\text{E}(D)} = \frac{\alpha m_0}{\alpha m_0 + \text{power} \times m_1} = \frac{\alpha p_0}{\alpha p_0 + \text{power} \times (1 - p_0)}, \tag{4.30}$$

To illustrate the relationship between q, p_0, α and power, we use two multiple testing examples shown below.

In the first example, a total of 2000 hypotheses are tested, in which 1000 hypotheses are true null (i.e., the proportion of true nulls is $p_0 = \frac{1000}{2000} = 0.50$). Given the overall type I

error rate $\alpha = 5\%$, the expected number of rejections among the 1000 truly null hypothesis is $1000 \times 5\% = 50$. With 80% power, the expected number of rejections among the 1000 false null hypotheses is $1000 \times 80\% = 800$. Then by Equation (4.30) the expected FDR level can be calculated as: $q = \frac{50}{50+800} \approx 0.059$.

In the second example, out of the 1000 hypotheses, 900 are true null (i.e., $p_0 = 0.90$) and 100 are false null. Similarly, given $\alpha = 5\%$ and power of 80%, by Equation (4.30) we can calculate the expected FDR as: $q = \frac{900\times5\%}{900\times5\%+100\times80\%} = \frac{45}{45+80} = 0.36$. Note that the power in neuroimaging and genomics studies is usually much less than 80% as a result of insufficient sample size (that will be discussed briefly in the next section), thus the expected FDR level is even greater than the calculated.

As this procedure continues to a very large value of p_0 such as $p_0 = 0.98$, the value of q will be close to 1, which is not acceptable due to the high probability of false positives. Therefore, Bhaumik et al. (2018b) recommended that in a multiple testing problem the FDR level of q should be adjusted for the value of p_0, the type I error rate and power, such that FDR does not exceed a desired threshold (e.g., 0.20 or 0.30). Also considering the assumption of the large proportion of null hypotheses typically found in neuroimaging and genomics studies in which $p_0 \geq 95\%$, it is not reasonable to have a small value of q (e.g., $q = 0.05$ or 0.1).

In addition, we explore the impact of average effect size on the expected FDR level. For multiple testing in neuroimaging, the effect sizes are not fixed as we normally assume for alternative hypothesis in case of single hypothesis testing, instead they are varying due to the heteroscedastic errors at both group level and link level. Song (2016) assessed the 95% confidence interval of the varying effect sizes via simulation study. For example, the 95% confidence interval of the effect sizes is $(0.50, 1.50)$ for a true mean between-group difference

in FC of 0.15, assuming the total sample size of 40 and $\alpha = 5\%$, yielding an average power of 80.2%. Figure 5 shows the expected FDR level and average effect size for different total sample sizes ranging from 40 to 100 assuming $p_0 = 0.90$. It is expected that given a fixed sample size, smaller average effect size will lead to a higher expected FDR level, suggesting that the more stricter FDR level control (i.e., smaller q) requires sufficiently large sample size and average effect size.

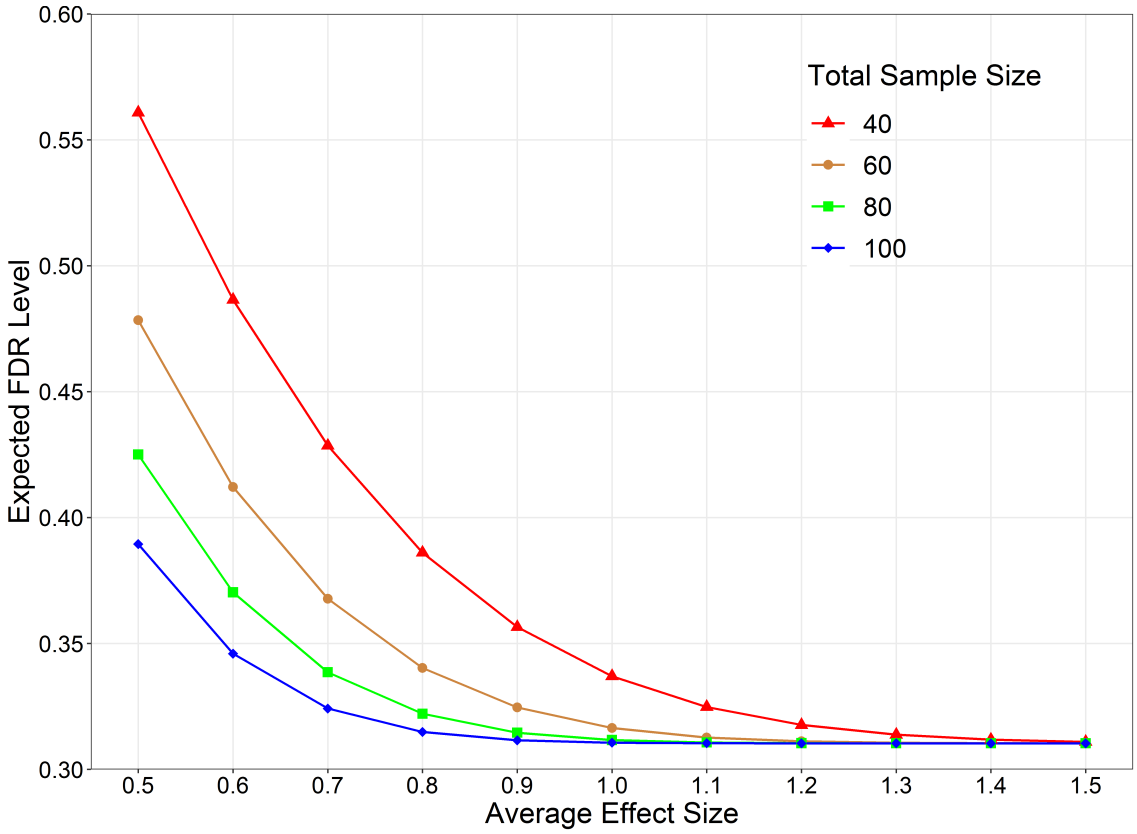


Figure 5. The expected FDR as a function of the average effect size for different total sample size assuming true null proportion $p_0 = 0.90$

4.5 FDR, Non-Discovery Rate, Power and Sample Size

Though the emphasis of this thesis is on FDR, another key component related to multiple hypothesis testing is power, which can be quite challenging especially for neuroimaging studies with small to moderate sample size. In this section we provide a brief overview on the relationship between non-discovery rate (NDR), power and sample size for multiple hypothesis testing in neuroimaging study.

In terms of a single hypothesis testing, power is known as the probability of rejecting a null hypothesis when the null hypothesis is false. In the context of multiple testing, power can have different forms (Dudoit et al., 2003) that are (1) the probability of rejecting at least one false null hypothesis, $P(TD > 1)$; (2) the probability of rejecting all m_1 false null hypotheses $P(TD = m_1)$; and (3) the average probability of rejecting false null hypotheses, i.e., average power, $\mathrm{E}(TD)/m_1$. In this review we only consider the third definition of power, that is, the expected proportion of correctly rejected hypotheses among all false null hypotheses, in the same form as for sensitivity,

$$\text{power} = \frac{\mathrm{E}(TD)}{m_1} = \text{sensitivity},\ m_1 > 0, \tag{4.31}$$

Define NDR as the expected proportion of accepted false null hypotheses among all false null hypotheses (Craiu and Sun, 2008),

$$NDR = \frac{\mathrm{E}(FND)}{m_1} = 1 - \frac{\mathrm{E}(TD)}{m_1} = 1 - \text{power},\ m_1 > 0, \tag{4.32}$$

Thus NDR can be interpreted as the average type II error rate over all false null hypotheses tested,

$$NDR = 1 - \text{power} = \frac{\sum_{i=1}^{m_1} P(p_i > \gamma \mid H_{0,i} = 1)}{m_1} = \bar{\beta}, \tag{4.33}$$

where γ is the fixed threshold at which level the individual hypothesis is rejected. We can rewrite NDR in Equation (4.32) as an expression of the true null proportion p_0 and FDR level q,

$$NDR = \frac{\mathrm{E}(FND)}{m_1} = \frac{m_1 - \mathrm{E}(TD)}{m_1} = 1 - \frac{(1-q)\mathrm{E}(D)}{(1-p_0)m}, \tag{4.34}$$

By replacing p_0 and $\mathrm{E}(D)$ with the estimated $\hat{p}_0$ in Equation (4.27) by "central matching" method (Efron, 2007) and the observed number of rejected hypotheses D, respectively, we can obtain an estimate of NDR as,

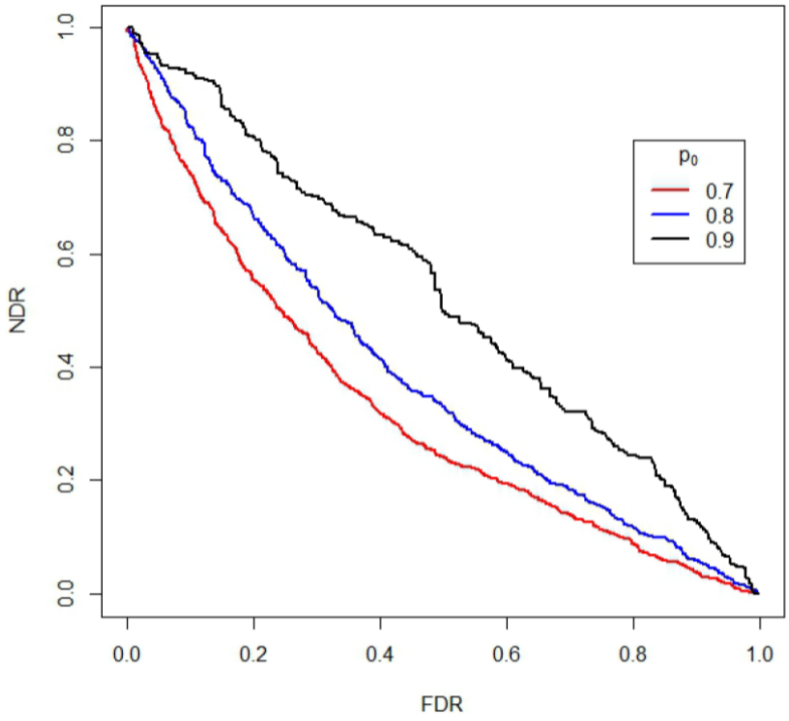
$$\widehat{NDR} = 1 - \frac{(1-q)D}{(1-\hat{p}_0)m}, \tag{4.35}$$

Recall that mFDR can be expressed as a function of power and p_0 in Equation (4.30), and by replacing power with 1-NDR, we have

$$mFDR \approx \frac{\mathrm{E}(FD)}{\mathrm{E}(D)} = \frac{\alpha m_0}{\alpha m_0 + \text{power} \times m_1} = \frac{\alpha p_0}{\alpha p_0 + (1 - NDR) \times (1 - p_0)}, \tag{4.36}$$

Song (2016) has investigated the relationship between FDR level q, NDR, true null proportion p_0, power and sample size in neuroimaging studies, where the results are summarized in Figures 6, 7and 8. Those findings align with the results of a previous study on FDR, sensitivity

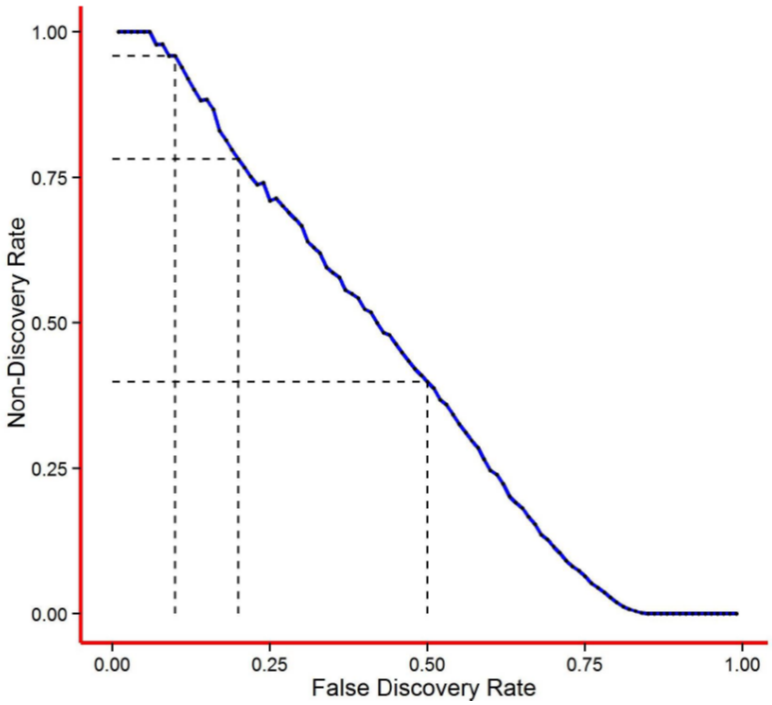
(i.e., power) and sample size for microarray studies (Pawitan et al., 2005). Figure 6 shows the trade-off relationship between FDR level q and NDR for various p_0 ranging from 0.7 to 0.9 via simulation study with similar data structure as the FC data in the neuroimaging study for LLD (Song, 2016). As can be seen, as FDR level q increases, the NDR level decreases regardless of p_0 value. However, p_0 value controls the steepness of NDR decrease, where lower p_0 is, faster NDR decreases. For example, at $q = 0.2$, NDR levels are approximately 0.55, 0.67 and 0.80 for p_0 equal to 0.7, 0.8, and 0.9, respectively. This is expected as lower p_0 means a larger number of false null hypotheses, which yields to a higher power and thus lower NDR.



Source: Figure 11 in Song (2016)

Figure 6. Trade-off relationship between FDR and NDR for different null proportion p_0 via simulation study

Red, blue and black curves indicate $p_0 = 0.7$, 0.8 and 0.9, respectively.



Source: Figure 12 in Song (2016)

Figure 7. Trade-off relationship between FDR and NDR using fMRI data in LLD neuroimaging study with estimated null proportion $\hat{p}_0 = 0.98$

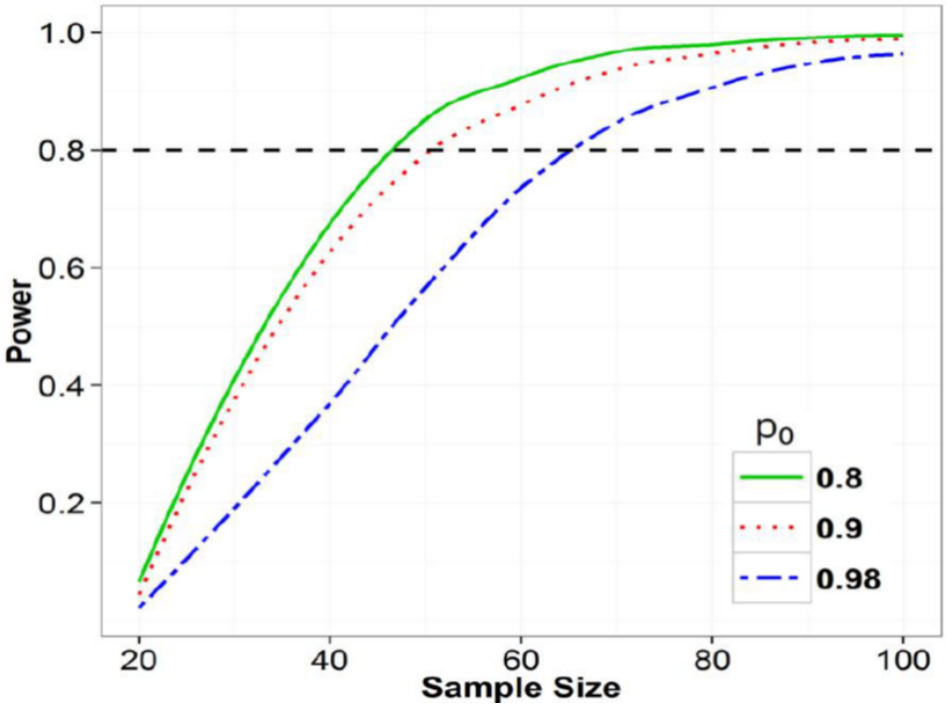
The three dashed lines indicate NDR levels of 0.94, 0.79 and 0.41 at FDR level $q = 0.1$, 0.2 and 0.5, respectively.

Figure 7 depicts the relationship between FDR level q and NDR using fMRI data in the LLD neuroimaging data described in Chapter 2 with the null proportion estimated as $\hat{p}_0 = 0.98$ based on Efron's "central matching" method illustrated in Section 4.3 (Song, 2016). There is almost a linear relationship between NDR and FDR level q within the range of (0.10, 0.75), where the slope = -1.358. At $q = 0.1$, NDR is substantially high at 0.94 that corresponds to a very low power of 0.06, which may not be acceptable. When q increases to 0.2, NDR decreases to 0.79 while the power increases to 0.21, though still considerably low. NDR level drops to 0.41 resulting in a reasonable power of 0.59 at $q = 0.5$, which is considered too high to be

accepted. Therefore, in this study we consider q in an acceptable range between 0.2 and 0.3, which correspond to NDR level between 0.79 and 0.67 and power between 0.21 and 0.33. This lower power is expected as a result of the small sample size and high null proportion in this study.

There are two important notes we can take from this trade-off relationship between FDR and NDR: (1) low FDR level q leads to high NDR and low power, that is, one can decrease FDR at the cost of raising NDR (reducing power), or vice versa. This provides a strong argument why use of low FDR level can be problematic; (2) control for FDR is also a control of NDR/power, based on which sample size determination for multiple testing can be simplified.

As aforementioned, typical neuroimaging studies usually have small to moderate sample size and thus it is of particular importance for researchers to understand the consequences especially with regards to the power. Figure 8 is a plot of sample size vs. power for different null proportion at FDR level $q = 0.2$ via simulation study assuming the true mean between-group difference in FC is 0.20 for false null hypotheses (Song, 2016). We can see that greater the null proportion, larger the sample size is required to achieve the same power. To achieve an average power of 0.8, the required sample size is approximately 46, 50 and 65 for p_0 equal to 0.8, 0.9 and 0.98, respectively. Again, we need to be mindful that for a study with small sample size of 20, the power we can expect is as low as 0.1 that makes detecting differential connectivity links very difficult. Taking all the above discussions in this section as well as the previous section into considerations, we therefore use $q = 0.2$ or 0.3 as desired FDR level for simulation study and $q = 0.2$ for data analysis in this thesis.



Source: Figure 13 in Song (2016)

Figure 8. Sample size and power for different null proportion p_0 at FDR level $q = 0.2$ via simulation study assuming the true mean between-group difference in FC = 0.20 for false null hypotheses

The green solid, red dotted and blue dash-dotted curves indicate $p_0 = 0.8$, 0.9 and 0.98, respectively.

4.6 Bayesian Multimodal Local False Discovery Rate Method

Efron's Lfdr method has gained increased popularity in genomics and neuroimaging studies over the years and been regarded as the benchmark method for FDR control for large-scale simultaneous hypothesis testing. However, in cases when sample size is small, Efron's Lfdr method cannot provide a sufficient control of FDR due to the high sampling variability. Table III presents the comparison of the performance of the four existing statistical approaches for FDR control in detecting disrupted FCs via simulation study with similar data structure as the FC data in the neuroimaging study for LLD (Song, 2016; Bhaumik et al., 2018b). We see that under small sample size $N = 23$ (10 LLD and 13 HC, same as the LLD study), all the four methods failed to control FDR at the pre-specified q level including $q = 0.2$ and 0.3. With the sample size increasing to $N = 100$ (50 in each group), the performance of the four methods all improved to certain degree, and Efron's approach is the best by yielding the closet FDR to the pre-specified q level for neuroimaging comparisons.

Furthermore, Efron's Lfdr method does not have the capability to account for the potential influence from other information into density estimations of the test statistics for the measurement of interest. In the multimodal neuroimaging study, the disrupted SC may be a predictor of the disrupted FC. Borrowing strength from the SC statistics into the modeling for FC statistics would increase the sensitivity of testing and help improve efficiency in both FDR control and detection of differential FC links. Zablocki et al. (2014) proposed a covariate-modulated Lfdr method using a parametric Bayesian two-group mixture model to incorporate functional genome annotations as covariates in genome-wide association studies, which has been proved to be efficient by increasing power substantially. We extend this method to the multimodal neuroimaging data aimed to better control FDR and detect differential FC links between the

TABLE III
COMPARISON OF PERFORMANCES OF DIFFERENT FDR CONTROL PROCEDURES[a]

N[b]	q[c]	Adaptive BH[d]	Theoretical Null	Empirical Null (Efron)[e]	Empirical Null (Sun & Cai)[f]
23	0.30	0.848	0.942	0.592	0.603
	0.20	0.804	0.916	0.497	0.508
	0.10	0.791	0.860	0.385	0.382
	0.05	0.752	0.777	0.291	0.286
50	0.30	0.674	0.900	0.373	0.506
	0.20	0.560	0.841	0.261	0.361
	0.10	0.402	0.700	0.150	0.191
	0.05	0.322	0.535	0.088	0.098
100	0.30	0.541	0.830	0.355	0.450
	0.20	0.399	0.714	0.235	0.320
	0.10	0.221	0.472	0.110	0.181
	0.05	0.122	0.265	0.052	0.092

Source: Table VIII in Song (2016)

[a] Results are based on the simulation study with similar data structure as the FC data in the LLD neuroimaging study (Song, 2016; Bhaumik et al., 2018b).
[b] N = Total sample size.
[c] q = desired FDR level.
[d] Benjamini and Hochberg method (Benjamini and Hochberg, 1995).
[e] Efron's Lfdr method (Efron et al., 2001; Efron, 2007).
[f] Sun and Cai method (Sun and Cai, 2007).

two groups in cross-sectional, multimodal neuroimaging studies with small to moderate sample size. We implement a Bayesian multimodal Lfdr method, which utilizes a Bayesian mixture model to leverage complementary SC statistics and enhance the modeling of the density of FC statistics. The approach is general and can be useful in a broad spectrum of applications, such as genomics, neuroimaging, and microbiome data.

This approach has two distinguishing features. First, for each individual FC link the prior probability $p_0 = P(H_{0,i} = 0)$ (also known as proportion of null hypotheses) is modeled using a logistic regression model with link-specific SC statistics as covariate, as opposed to assuming the same p_0 estimate for all FC links in Efron's Lfdr method that considers solely FCs. Second, the densities of FC test statistics under both null hypothesis and alternative hypothesis are estimated using parametric models, where the posterior sampling for the model parameters are obtained using Gibbs sampling. The SC test statistics serve as auxiliary information in the mixture model aiding identification of differential FCs between the disease group and control group. Figure 9 presents the evolution of FDR methods in neuroimaging from Benjamini and Hochberg FDR, Efron's Lfdr, and to Bayesian multimodal Lfdr in neuroimaging studies.

Several computing algorithms are implemented to deal with intensive computations in posterior sampling that naturally arise in this approach. The utility of the method is illustrated with an extensive simulation study and an application to the motivating LLD study, in which FC and SC data were measured on each participant and the objective is to identify differential FC links between LLD and HC groups.

4.6.1 Bayesian Mixture Models

In this approach, the densities of FC test statistics under both null hypothesis and alternative hypothesis are estimated using parametric models. Given zero assumption, most of

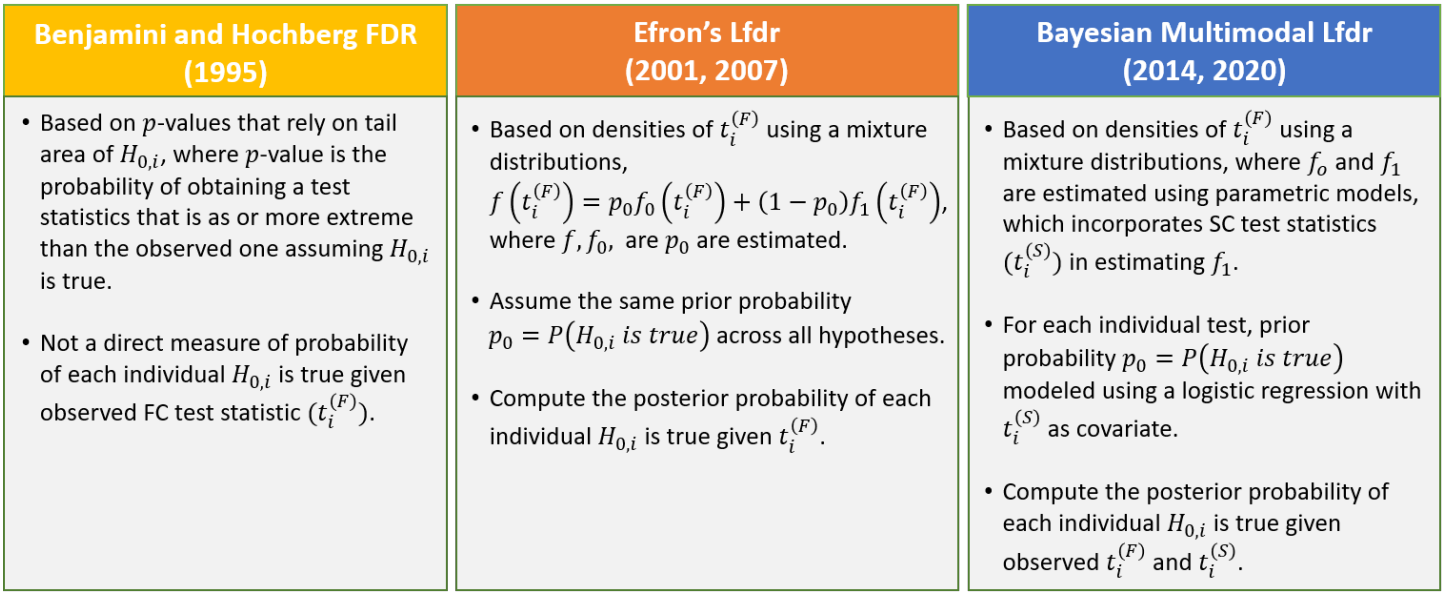


Figure 9. The evolution of FDR methods from Benjamini and Hochberg FDR, Efron's Lfdr, and to Bayesian multimodal Lfdr in neuroimaging studies

Note: In Bayesian multimodal Lfdr approach, absolute values of test statistics without directions are used for modeling, considering that test statistics located in the lower and upper tail areas are equally important to researchers and combining both tail areas will provide more information to estimate alternative density using a parametric model.

the hypotheses tested are null, so there may not be sufficient number of test statistics under alternative hypotheses in either lower or upper tail area that can be used for parametric density estimation for the two directions. The test statistics located in the lower and upper tail areas are considered equally important to researchers and thus combining both tail areas will provide more information to help estimate alternative density using a parametric model. Therefore, we use the absolute values of test statistics without directions (signs) for modeling, as suggested by the genome-wide association studies (Zablocki et al., 2014).

From both theoretical and empirical perspectives it is assumed that the underlying distribution of the test statistics under null hypothesis is normal as described in the previous section (Efron, 2004, 2007). Thus, the absolute values of test statistics under null hypothesis follows a folded normal distribution (Leone et al., 1961). Figure 10 depicts the probability density function of a folded normal random variable $|x|$, where x follows a standard normal distribution with mean zero and the standard deviation one (same as the theoretical null). We see that the probability mass values on the left half of the normal distribution are folded over the right half, as "folded" literally means.

Let $t_i^{(F)}$ and $t_i^{(S)}$, $i = 1, \cdots, m$, denote the absolute value of test statistics to compare mean connectivity measures between disease group and control group for the i^{th} connectivity link, for FC and SC, respectively. The $t_i^{(F)}$ and $t_i^{(S)}$ statistics are obtained by fitting the FC and SC data using a linear MERM with heteroscedastic errors described in Chapter 3, separately. And $\boldsymbol{t}^{(F)} = \begin{bmatrix} t_1^{(F)} & t_2^{(F)} & \cdots & t_m^{(F)} \end{bmatrix}^T$ and $\boldsymbol{t}^{(S)} = \begin{bmatrix} t_1^{(S)} & t_2^{(S)} & \cdots & t_m^{(S)} \end{bmatrix}^T$ are the vector of the absolute values of test statistics for FC and SC, respectively.

A Bayesian two-group mixture model (Efron et al., 2001; Storey, 2002; Efron, 2008; Zablocki et al., 2014) is utilized to model the density of FC test statistics, with auxiliary information

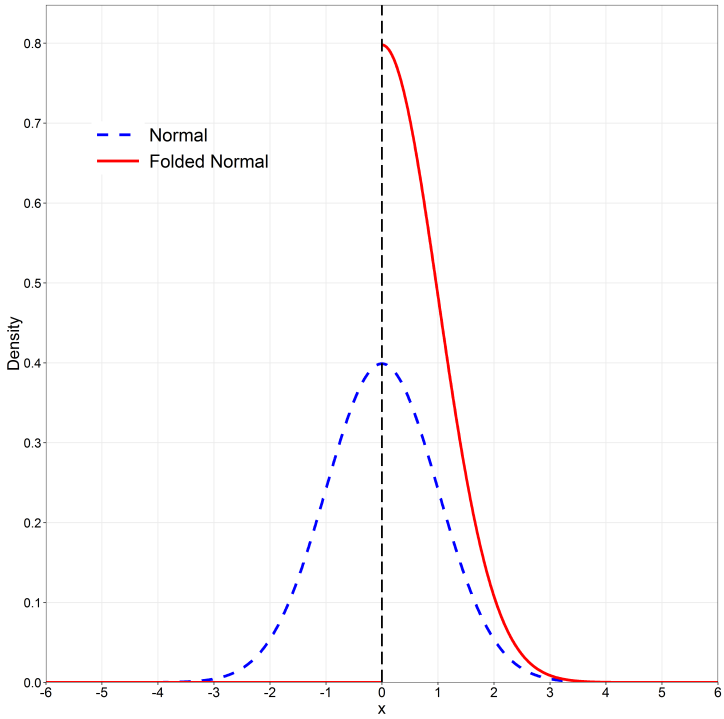


Figure 10. Densities of standard normal and folded normal

from SC test statistics. We formulate the Bayesian two-group mixture model according to Zablocki et al. (2014). Assume that for the connectivity links under the null hypothesis of no FC difference between disease group and control group (referred to as null connectivity links), their $t_i^{(F)}$ statistics follow a folded normal distribution, denoted by f_0, with zero location parameter ($\mu_0 = 0$) and unknown scale parameter $\sigma_0^2 > 0$. For a two group comparison study,

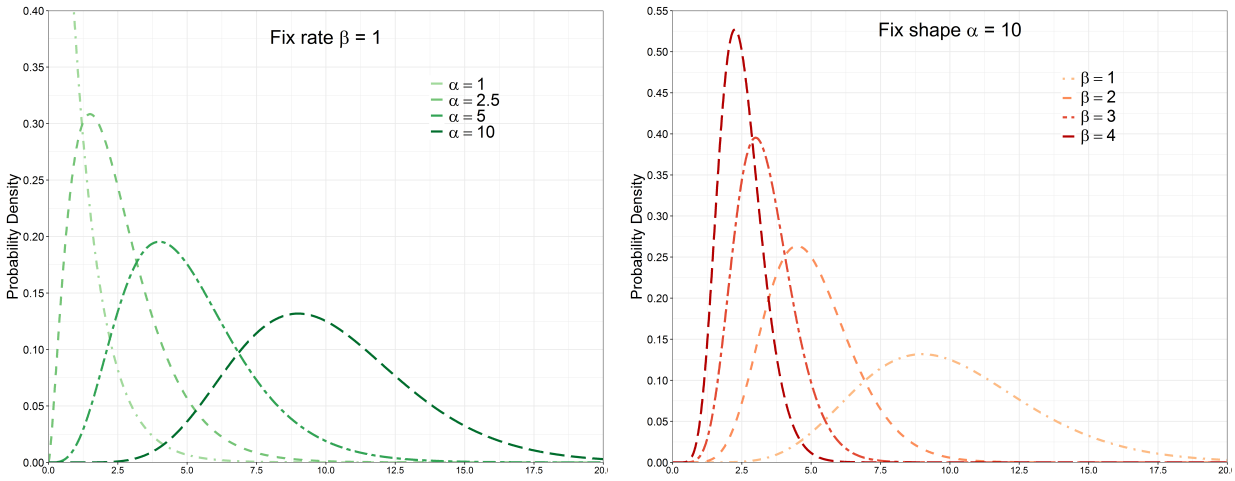
the zero mean assumption under the null hypothesis is justified, and σ_0^2 assumption provides a better flexibility to explain variability.

$$\begin{aligned} f_0\big(t_i^{(F)} \mid \mu_0 = 0, \sigma_0^2\big) &= \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{(t_i^{(F)} - \mu_0)^2}{2\sigma_0^2}\right] + \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{(t_i^{(F)} + \mu_0)^2}{2\sigma_0^2}\right] \\ &= 2\frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{\big(t_i^{(F)}\big)^2}{2\sigma_0^2}\right], \\ &= \sqrt{\frac{2}{\pi\sigma_0^2}} \exp\left[-\frac{\big(t_i^{(F)}\big)^2}{2\sigma_0^2}\right], t_i^{(F)} \geq 0, \end{aligned} \tag{4.37}$$

Note that the SC test statistic $t_i^{(S)}$ is not involved in Equation (4.37), under the assumption that SC statistic has no influence on density of FC statistic for the null connectivity links (Zablocki et al., 2014; Zhang et al., 2019).

Next, we assume that for the connectivity links under the alternative hypothesis of differential (or significant) FC difference between disease group and control group (referred to as alternative connectivity links), their $t_i^{(F)}$ statistic follows a gamma distribution. The gamma distribution is considered because it offers flexibility to accommodate different shapes of the density function using various parameter settings and good fit for right-skewed data in general. Figure 11 presents the shape of gamma distribution using various specifications of shape and rate parameters (denoted by α and β, respectively), where Figure 11a shows densities for different shape parameter values with fixed rate $\beta = 1$, while Figure 11b shows densities for different rate parameter values with fixed shape $\alpha = 10$. As seen in Figure 11a, the shape parameter α controls skewness of the density given skewness $= \frac{2}{\sqrt{\alpha}}$, which is reduced as α increases. When $\alpha = 1$, Gamma(1, β) is an exponential distribution; when $\alpha \to \infty$, the gamma distribution can be approximated by the normal distribution with mean $\frac{\alpha}{\beta}$ and variance $\frac{\alpha}{\beta^2}$. In Figure 11b where the fixed shape parameter is fairly large $\alpha = 10$, we can see that the densities are close

to normal distribution and the rate parameter β have the effect to stretch (when β is small, e.g., $\beta = 1$) or compress (when β is large, e.g., $\beta = 4$) the range of the distribution as variance $= \frac{\alpha}{\beta^2}$.



(a) (b)

Figure 11. Density of Gamma(α, β), where α and β are the shape and rate parameters, respectively

(a) Densities for various shape parameter values from 1 to 10 with fixed rate at $\beta = 1$.
(b) Densities for various rate parameter values from 1 to 4 with fixed shape at $\alpha = 10$.

Denote the gamma distribution for alternative connectivity links by f_1, with a fixed location parameter μ_1, unknown shape parameter $\alpha(t_i^{(S)} \mid \boldsymbol{\alpha}) > 0$ and unknown rate parameter $\beta > 0$,

$$f_1\left(t_i^{(F)} \mid \mu_1, \alpha(t_i^{(S)} \mid \boldsymbol{\alpha}), \beta\right) = \frac{\beta^{\alpha\left(t_i^{(S)}|\boldsymbol{\alpha}\right)}}{\Gamma\left(\alpha\left(t_i^{(S)} \mid \boldsymbol{\alpha}\right)\right)}\left(t_i^{(F)} - \mu_1\right)^{\alpha\left(t_i^{(F)}|\boldsymbol{\alpha}\right)-1} \exp\left[-\beta\left(t_i^{(F)} - \mu_1\right)\right],$$
$$t_i^{(F)} \geq \mu_1, \tag{4.38}$$

Further we assume that $\alpha\big(t_i^{(S)} \mid \boldsymbol{\alpha}\big)$ is a log-linear function of SC statistic $t_i^{(S)}$ with unknown parameters α_0 and α_1, that is,

$$\alpha(t_i^{(S)} \mid \boldsymbol{\alpha}) = \exp(\alpha_0 + \alpha_1 t_i^{(S)}), \tag{4.39}$$

where $\boldsymbol{\alpha} = \begin{bmatrix}\alpha_0 & \alpha_1\end{bmatrix}^T$ is a vector of the unknown parameters.

In order to avoid the nonconvergence issue and gain some efficiency, we assume β does not depend on $t_i^{(S)}$. The location parameter μ_1 in Equation (4.38) is purposely fixed at 0.674, the median of the standard folded normal distribution $f_0(0,1)$ (Zablocki et al., 2014). Based on the zero assumption (Efron, 2007; Storey, 2002; Storey and Tibshirani, 2003; Zablocki et al., 2014) that the most of the test statistics around zero are null hypotheses, it is reasonable to assume $t_i^{(F)} \leq 0.674$ is null *a priori* so the alternative density is bound away from zero. Also this restriction on μ_1 in f_1 is necessary to help avoid any identifiable model issue when strong assumptions about parameters are not available.

To complete the mixture model, we now define a latent variable w_i, $i = 1, \cdots, m$, to indicate which component of the mixture model each $t_i^{(F)}$ follows, null distribution if $w_i = 0$, or alternative distribution if $w_i = 1$,

$$t_i^{(F)}, \quad i = 1, ..., m, \sim \begin{cases} f_0\big(t_i^{(F)} \mid \sigma_0^2\big), & \text{if } w_i = 0, \\ f_1\big(t_i^{(F)} \mid \alpha\big(t_i^{(S)} \mid \boldsymbol{\alpha}\big), \beta\big), & \text{if } w_i = 1. \end{cases} \tag{4.40}$$

The prior probability of $w_i = 1$ depends upon the SC statistic $t_i^{(S)}$, denoted by $\pi_1\big(t_i^{(S)} \mid \boldsymbol{\gamma}\big)$, and is modeled using a logistic regression with $t_i^{(S)}$ as covariate,

$$\pi_1\big(t_i^{(S)} \mid \boldsymbol{\gamma}\big) = P\big(w_i = 1 \mid t_i^{(S)}\big) = \frac{\exp\big(\gamma_0 + \gamma_1 t_i^{(S)}\big)}{1 + \exp\big(\gamma_0 + \gamma_1 t_i^{(S)}\big)}, \tag{4.41}$$

where $\boldsymbol{\gamma} = \begin{bmatrix}\gamma_0 & \gamma_1\end{bmatrix}^T$ is a vector of unknown parameters. The prior probability of $w_i = 0$ given $t_i^{(S)}$, denoted by $\pi_0\big(t_i^{(S)} \mid \boldsymbol{\gamma}\big)$, is

$$\pi_0\big(t_i^{(S)} \mid \boldsymbol{\gamma}\big) = P\big(w_i = 0 \mid t_i^{(S)}\big) = 1 - \pi_1\big(t_i^{(S)} \mid \boldsymbol{\gamma}\big) = \frac{1}{1 + \exp\big(\gamma_0 + \gamma_1 t_i^{(S)}\big)}, \tag{4.42}$$

Then the density of the mixture model is

$$f\big(t_i^{(F)} \mid t_i^{(S)}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \beta, \sigma_0^2\big) = \pi_0\big(t_i^{(S)} \mid \boldsymbol{\gamma}\big) f_0\big(t_i^{(F)} \mid \sigma_0^2\big) + \pi_1\big(t_i^{(S)} \mid \boldsymbol{\gamma}\big) f_1\big(t_i^{(F)} \mid \boldsymbol{\alpha}, \beta, t_i^{(S)}\big), \tag{4.43}$$

And the augmented likelihood function is

$$\begin{aligned}
&\boldsymbol{L}\big(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \beta, \sigma_0^2 \mid \boldsymbol{w}, \boldsymbol{t}^{(F)}, \boldsymbol{t}^{(S)}\big) \\
&= \prod_{i=1}^{m} \left\{ \Big[\pi_0\big(t_i^{(S)} \mid \boldsymbol{\gamma}\big) f_0\big(t_i^{(F)} \mid \sigma_0^2\big)\Big]^{(1-w_i)} \Big[\pi_1\big(t_i^{(S)} \mid \boldsymbol{\gamma}\big) f_1\big(t_i^{(F)} | \boldsymbol{\alpha}, \beta, t_i^{(S)}\big)\Big]^{w_i} \right\},
\end{aligned} \tag{4.44}$$

where $\boldsymbol{w} = \begin{bmatrix} w_1 & w_2 & \cdots & w_m \end{bmatrix}^T$ is a vector of w_i's.

By Bayes' theorem, $BLfdr_i\big(t_i^{(F)}, t_i^{(S)}\big)$, defined as the posterior probability that the i^{th} connectivity link is null (i.e., no differential FC between disease group and control group) given $t_i^{(F)}$ and $t_i^{(S)}$ statistics, and the posterior estimates of the parameters $\{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}}, \hat{\beta}, \hat{\sigma}_0^2\}$ is

$$\begin{aligned}
BLfdr_i\big(t_i^{(F)}, t_i^{(S)}\big) &= P\big(H_{0,i} = 0 \mid t_i^{(F)}, t_i^{(S)}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}}, \hat{\beta}, \hat{\sigma}_0^2\big) \\
&= \frac{\pi_0\big(t_i^{(S)} \mid \hat{\boldsymbol{\gamma}}\big) f_0\big(t_i^{(F)} \mid \hat{\sigma}_0^2\big)}{f\big(t_i^{(F)} \mid t_i^{(S)} \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}}, \hat{\beta}, \hat{\sigma}_0^2\big)} \\
&= \frac{\pi_0\big(t_i^{(S)} \mid \hat{\boldsymbol{\gamma}}\big) f_0\big(t_i^{(F)} \mid \hat{\sigma}_0^2\big)}{\pi_0\big(t_i^{(S)} \mid \hat{\boldsymbol{\gamma}}\big) f_0\big(t_i^{(F)} \mid \hat{\sigma_0}^2\big) + \pi_1\big(t_i^{(S)} \mid \hat{\boldsymbol{\gamma}}\big) f_1\big(t_i^{(F)} \mid \hat{\boldsymbol{\alpha}}, \beta, t_i^{(S)}\big)}.
\end{aligned} \tag{4.45}$$

4.6.2 Prior Distributions

For the unknown parameters $\{\boldsymbol{\alpha}, \boldsymbol{\gamma}, \beta, \sigma_0^2\}$ in the mixture model (4.43), we assume generic weakly informative prior distributions. Details are illustrated as follows.

For $\boldsymbol{\alpha} = \begin{bmatrix}\alpha_0 & \alpha_1\end{bmatrix}^T$ in Equation (4.39) and $\boldsymbol{\gamma} = \begin{bmatrix}\gamma_0 & \gamma_1\end{bmatrix}^T$ in Equation (4.41), we assume that α_0 and α_1 are *a priori* independent, γ_0 and γ_1 are *a priori* independent, and their values are concentrated between -1 and 1, so we have $\boldsymbol{\alpha} \sim N_2(\mathbf{0}, \Sigma_{\boldsymbol{\alpha}})$ and $\boldsymbol{\gamma} \sim N_2(\mathbf{0}, \Sigma_{\boldsymbol{\gamma}})$ given by

$$P(\boldsymbol{\alpha}) = \frac{1}{2\pi|\Sigma_{\boldsymbol{\alpha}}|^{1/2}} \exp\left\{-\frac{1}{2}\boldsymbol{\alpha}^T\Sigma_{\boldsymbol{\alpha}}^{-1}\boldsymbol{\alpha}\right\}, \tag{4.46}$$

$$P(\boldsymbol{\gamma}) = \frac{1}{2\pi|\Sigma_{\boldsymbol{\gamma}}|^{1/2}} \exp\left\{-\frac{1}{2}\boldsymbol{\gamma}^T\Sigma_{\boldsymbol{\gamma}}^{-1}\boldsymbol{\gamma}\right\}, \tag{4.47}$$

where $\Sigma_{\boldsymbol{\alpha}} = \Sigma_{\boldsymbol{\gamma}} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$.

For the rate parameter β in Equation (4.38), we choose a gamma distribution as the prior for mathematical convenience since it is the conjugate prior for the gamma distribution, $\beta \sim \text{Gamma}(a_\beta, b_\beta)$, given by

$$P(\beta \mid a_\beta, b_\beta) = \frac{\beta^{(a_\beta - 1)}}{{b_\beta}^{a_\beta}\Gamma(a_\beta)} \exp\left(-\frac{\beta}{b_\beta}\right), \beta > 0, \tag{4.48}$$

where $a_\beta > 0$ and $b_\beta > 0$ are the shape and rate hyperparameters, respectively. The prior mean $= a_\beta b_\beta$ and variance $= a_\beta {b_\beta}^2$.

For the scale parameter σ_0^2 in Equation (4.37), similarly we choose the conditionally conjugate prior distribution $\sigma_0^2 \sim \text{Inverse-Gamma}(a_{\sigma_0}, b_{\sigma_0})$ given by

$$P(\sigma_0^2 \mid a_{\sigma_0}, b_{\sigma_0}) = \frac{{b_{\sigma_0}}^{a_{\sigma_0}}}{\Gamma(a_{\sigma_0})}(\sigma_0^2)^{-(a_{\sigma_0}+1)} \exp\left(-\frac{b_{\sigma_0}}{\sigma_0^2}\right), \sigma_0^2 > 0, \tag{4.49}$$

where $a_{\sigma_0} > 0$ and $b_{\sigma_0} > 0$ are the shape and rate hyperparameters, respectively. The prior mean $= \frac{b_{\sigma_0}}{a_{\sigma_0}-1}$ and variance $= \frac{{b_{\sigma_0}}^2}{(a_{\sigma_0}-1)^2(a_{\sigma_0}-2)}$.

Note that in the form of gamma distribution (4.48) and the inverse gamma distribution (4.49), when both shape and rate hyperparameters have very small values (e.g., 0.001), the prior distributions will approach the Jeffreys' prior as:

$$p(\beta) \propto \beta^{a_\beta - 1} \exp\left(-\beta b_\beta\right), \lim_{a_\beta \to 0, b_\beta \to 0} p(\beta) = \frac{1}{\beta}, \tag{4.50}$$

$$p(\sigma_0^2) \propto (\sigma_0^2)^{-(a_{\sigma_0}+1)} \exp\left(-\frac{b_{\sigma_0}}{\sigma_0^2}\right), \lim_{a_{\sigma_0} \to 0, b_{\sigma_0} \to 0} p(\sigma_0^2) = \frac{1}{\sigma_0^2}. \tag{4.51}$$

Jeffreys' prior refers to the non-informative prior distribution for a parameter that is invariant under smooth monotone transformation of the parameter based on the Fisher information for the parameter (Jeffreys, 1946). For a random variable X distributed as $p(x|\theta)$ with parameter θ and $\theta \in \Theta$, the Fisher information that measures information contained in X contains about the parameter θ is defined as,

$$I_1(\theta) = \mathrm{E}_\theta \left[\left(\frac{d \log p(x|\theta)}{d\theta} \right)^2 \Big| \theta \right] = -\mathrm{E}_\theta \left[\frac{d^2 \log p(x|\theta)}{d\theta^2} \Big| \theta \right], \tag{4.52}$$

For a data with n observations $(X_1, X_2, \cdots, X_n)$ that are independent and identically distributed random variables, $X_i \overset{i.i.d.}{\sim} p(x_i|\theta)$, $i = 1, \cdots, n$, we have $p(x_1, x_2, \cdots, x_n|\theta) = \prod_{i=1}^n p(x_i|\theta)$ and then

$$I(\theta) = \sum_{i=1}^n \left\{ -\mathrm{E}_\theta \left[\frac{d^2 \log p(x_i|\theta)}{d\theta^2} \Big| \theta \right] \right\} = n I_1(\theta), \tag{4.53}$$

Jeffreys' prior is proportional to $\sqrt{I(\theta)}$,

$$p(\theta) \propto \sqrt{I(\theta)} = \text{const.} \times \sqrt{I(\theta)}, \tag{4.54}$$

If the integral of $\sqrt{I(\theta)}$ over the parameter space Θ, $\int_\Theta \sqrt{I(\theta)}d\theta$, is a finite number, then the constant can be taken as $\frac{1}{\int_\Theta \sqrt{I(\theta)}d\theta}$ so that

$$\int_\Theta p(\theta)d\theta = \int_\Theta \frac{1}{\int_\Theta \sqrt{I(\theta)}d\theta}\sqrt{I(\theta)}d\theta = \frac{1}{\int_\Theta \sqrt{I(\theta)}d\theta}\int_\Theta \sqrt{I(\theta)}d\theta = 1,$$

indicating that the prior $p(\theta)$ is a proper probability density. If the integral is infinite, then the constant cannot be specified and the prior $p(\theta)$ is an improper probability density.

To see how Jeffreys' prior is invariant under reparameterization. For $\phi = g(\theta)$, where $g()$ is a smooth monotone function, the Fisher information for ϕ is,

$$\begin{aligned}
I(\phi) &= \sum_{i=1}^{n}\left\{\mathrm{E}_\phi\left[\left(\frac{d\log p(x_i|\phi)}{d\phi}\right)^2\Big|\phi\right]\right\} \\
&= \sum_{i=1}^{n}\left\{\mathrm{E}_\theta\left[\left(\frac{d\log p(x_i|\theta)}{d\theta}\frac{d\theta}{d\phi}\right)^2\Big|\theta\right]\right\} \\
&= \sum_{i=1}^{n}\left\{\mathrm{E}_\theta\left[\left(\frac{d\log p(x_i|\theta)}{d\theta}\right)^2\left(\frac{d\theta}{d\phi}\right)^2\Big|\theta\right]\right\} \\
&= \sum_{i=1}^{n}\left\{\mathrm{E}_\theta\left[\left(\frac{d\log p(x_i|\theta)}{d\theta}\right)^2\Big|\theta\right]\right\}\left(\frac{d\theta}{d\phi}\right)^2 \\
&= I(\theta)\left(\frac{d\theta}{d\phi}\right)^2,
\end{aligned} \tag{4.55}$$

Thus,

$$\sqrt{I(\phi)} = \sqrt{I(\theta)}\left|\frac{d\theta}{d\phi}\right|, \tag{4.56}$$

Then the prior density on ϕ is

$$p(\phi) = \text{const.} \times \sqrt{\mathrm{I}(\phi)} = \text{const.} \times \sqrt{\mathrm{I}(\theta)}\left|\frac{d\theta}{d\phi}\right| = p(\theta)\left|\frac{d\theta}{d\phi}\right|, \tag{4.57}$$

Gamma(0.001, 0.001) and Inverse-Gamma(0.001, 0.001) have been commonly used as non-informative prior in the examples in Bayesian Inference Using Gibbs Sampling (BUGS) (Lunn et al., 2009). However, there are major concerns being raised and discussed with regards to the use of these non-informative priors. First, the priors are not distributions because they are unbounded and the integrals over the entire parameter space are infinite,

$$\int_0^\infty f(\theta)d\theta = \int_0^\infty \frac{1}{\theta}d\theta = \ln\theta\Big|_0^\infty = \infty - (-\infty) = \infty, \tag{4.58}$$

Hence, the priors are improper and may also yield an improper posterior density (Gelman, 2006; Gelman et al., 2013). Both prior distributions have the mean of 1 but extremely large variance, with high probability density very close to 0 and a very long tail (as shown in dashed red lines in Figures 12a and 12b). The high prior probability density near 0 has a strong influence on the posterior density so the posterior tends to behave in a similar way as the prior with majority mass near 0 and a long and heavy tail. This will lead to an unregularized posterior distribution that diffuse to extreme large values. In this sense, the priors are not non-informative but actually quite influential. In addition, Robert and Casella (1999) pointed out that in complex cases the improper prior often results in convergence issue in Gibbs sampler that draws samples directly from the conditional distributions.

Given the undesirable properties of the non-informative priors alluded above, it has been suggested and recommended to use weakly informative prior, which contains little but sufficient

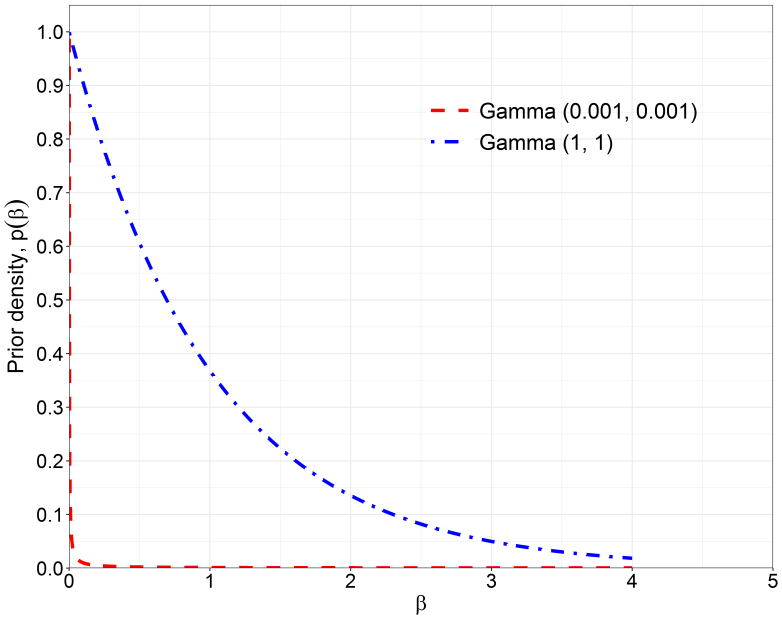
information to provide some regularization to ensure that the posterior be bounded within a reasonable range (Gelman, 2006; Gelman et al., 2008, 2013). As such, we set those hyperparameters using the method of moments estimation (Kim et al., 2009), choose Gamma($a_\beta = 1, b_\beta = 1$) and Inverse-Gamma ($a_{\sigma_0} = 3, b_{\sigma_0} = 2$) both with mean = 1 and variance = 1 as the weakly informative prior for β and σ_0^2, respectively (as shown in dash-dotted blue lines in Figures 12a and 12b). These weakly informative priors offer more stable inferences than would be obtained from non-informative priors through regularization in parameter estimation, as well as enough vagueness to ensure that posteriors are dominated by the data.

4.6.3 Posterior Sampling

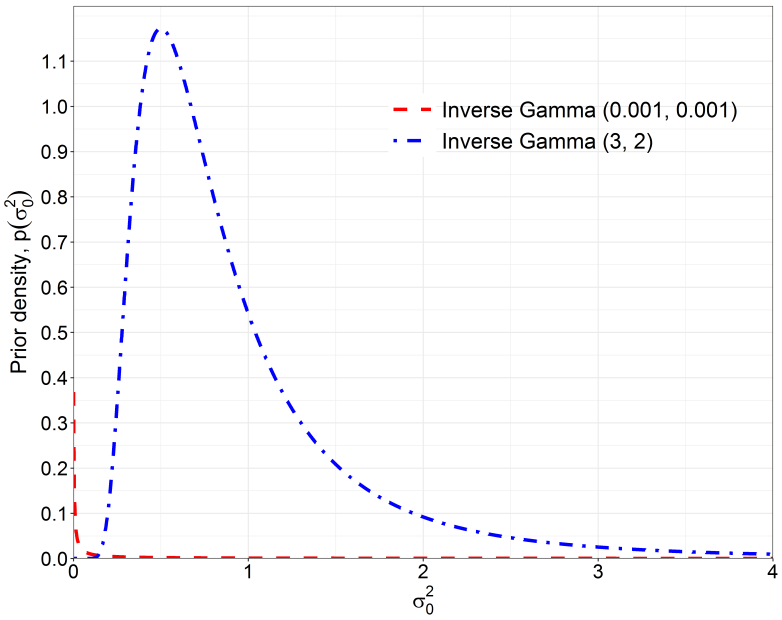
We obtain the posterior sampling distributions for the unknown parameters using Gibbs Sampler (Geman and Geman, 1984; Casella and George, 1992; Robert and Casella, 1999). Gibbs sampler is a special Markov Chain Monte Carlo (MCMC) algorithm (Hastings, 1970; Gelfand and Smith, 1990), which generates samples from the conditional distributions given all other variables without requiring the marginal distributions. The full conditional distributions of the parameters $\{\boldsymbol{\alpha}, \boldsymbol{\gamma}, \beta, \sigma_0^2\}$ in the mixture model (4.43) are described below.

By Equations (4.38) and (4.46), the conditional distribution of $\boldsymbol{\alpha}$ given β, $\boldsymbol{w}$, $\boldsymbol{t}_x$ and $\boldsymbol{t}_z$ is

$$
\begin{aligned}
& f(\boldsymbol{\alpha} \mid \beta, \boldsymbol{w}, \boldsymbol{t}^{(F)}, \boldsymbol{t}^{(S)}) \\
& = P(\boldsymbol{\alpha}) \prod_{\{i|w_i=1\}} f_1\big(t_i^{(F)} \mid \boldsymbol{\alpha}, \beta, t_i^{(S)}\big) \\
& = \frac{1}{2\pi|\Sigma_{\boldsymbol{\alpha}}|^{1/2}} \exp\left\{-\frac{1}{2}\boldsymbol{\alpha}^T\Sigma_{\boldsymbol{\alpha}}^{-1}\boldsymbol{\alpha}\right\} \prod_{\{i|w_i=1\}} \frac{\beta^{\alpha\left(t_i^{(S)}|\boldsymbol{\alpha}\right)}}{\Gamma\big(\alpha\big(t_i^{(S)} \mid \boldsymbol{\alpha}\big)\big)} \big(t_i^{(F)} - \mu_1\big)^{\alpha\left(t_i^{(S)}|\boldsymbol{\alpha}\right)-1} \exp\Big[-\beta\big(t_i^{(F)} - \mu_1\big)\Big] \\
& \propto \exp\left\{-\frac{1}{2}\boldsymbol{\alpha}^T\Sigma_{\boldsymbol{\alpha}}^{-1}\boldsymbol{\alpha}\right\} \prod_{\{i|w_i=1\}} \frac{\beta^{\alpha\left(t_i^{(S)}|\boldsymbol{\alpha}\right)}}{\Gamma\big(\alpha\big(t_i^{(S)} \mid \boldsymbol{\alpha}\big)\big)} \big(t_i^{(F)} - \mu_1\big)^{\alpha\left(t_i^{(S)}|\boldsymbol{\alpha}\right)}, \qquad (4.59)
\end{aligned}
$$



(a)



(b)

Figure 12. Prior distributions

(a) Gamma for β. (b) Inverse-Gamma for σ_0^2. The dashed red lines are the non-informative priors, and the dash-dotted blue lines are the weakly informative priors.

where $\alpha\big(t_i^{(S)} \mid \boldsymbol{\alpha}\big) = \exp\big(\alpha_0 + \alpha_1 t_i^{(S)}\big)$, $\mu_1 = 0.674$.

By Equations (4.38), (4.41) and (4.42), the conditional distribution of $\boldsymbol{\gamma}$ given $\boldsymbol{w}$ and $\boldsymbol{t}_z$ is

$$
\begin{aligned}
& f\big(\boldsymbol{\gamma} \mid \boldsymbol{w}, \boldsymbol{t}^{(F)}\big) \\
& = P(\boldsymbol{\gamma}) \prod_{i=1}^{m} \Big\{ \pi_0\big(t_i^{(S)} \mid \boldsymbol{\gamma}\big)^{1-w_i} \pi_1\big(t_i^{(S)} \mid \boldsymbol{\gamma}\big)^{w_i} \Big\} \\
& = \frac{1}{2\pi|\Sigma_{\boldsymbol{\gamma}}|^{1/2}} \exp\left\{ -\frac{1}{2} \boldsymbol{\gamma}^T \Sigma_{\boldsymbol{\gamma}}^{-1} \boldsymbol{\gamma} \right\} \prod_{i=1}^{m} \left\{ \left[\frac{1}{1 + \exp(\gamma_0 + \gamma_1 t_i^{(S)})} \right]^{1-w_i} \left[\frac{\exp(\gamma_0 + \gamma_1 t_i^{(S)})}{1 + \exp(\gamma_0 + \gamma_1 t_i^{(S)})} \right]^{w_i} \right\} \\
& \propto \exp\left\{ -\frac{1}{2} \boldsymbol{\gamma}^T \Sigma_{\boldsymbol{\gamma}}^{-1} \boldsymbol{\gamma} \right\} \prod_{i=1}^{m} \frac{\exp\big(\gamma_0 + \gamma_1 t_i^{(S)}\big)^{w_i}}{1 + \exp\big(\gamma_0 + \gamma_1 t_i^{(S)}\big)}, \qquad (4.60)
\end{aligned}
$$

By Equations (4.38) and (4.48), the conditional distribution of β given $\boldsymbol{\alpha}$, $\boldsymbol{w}$, $\boldsymbol{t}^{(F)}$ and $\boldsymbol{t}^{(S)}$ is

$$
\begin{aligned}
& f\big(\beta \mid \boldsymbol{\alpha}, \boldsymbol{w}, \boldsymbol{t}^{(F)}, \boldsymbol{t}^{(S)}\big) \\
& = P(\beta \mid a_\beta, b_\beta) \prod_{\{i|w_i=1\}} f_1\big(t_i^{(F)} \mid \boldsymbol{\alpha}, \beta, t_i^{(S)}\big) \\
& = \frac{\beta^{(a_\beta - 1)}}{b_\beta{}^{a_\beta} \Gamma(a_\beta)} \exp\left(-\frac{\beta}{b_\beta}\right) \prod_{\{i|w_i=1\}} \frac{\beta^{\alpha\left(t_i^{(S)}|\boldsymbol{\alpha}\right)}}{\Gamma\big(\alpha\big(t_i^{(S)} \mid \boldsymbol{\alpha}\big)\big)} \big(t_i^{(F)} - \mu_1\big)^{\alpha\left(t_i^{(S)}|\boldsymbol{\alpha}\right)-1} \exp\Big[-\beta\big(t_i^{(F)} - \mu_1\big)\Big] \\
& \propto \beta^{(a_\beta - 1)} \exp\left(-\frac{\beta}{b_\beta}\right) \prod_{\{i|w_i=1\}} \beta^{\alpha\left(t_i^{(S)}|\boldsymbol{\alpha}\right)} \exp\Big[-\beta\big(t_i^{(F)} - \mu_1\big)\Big] \\
& \propto \beta^{(a_\beta - 1)} \exp\left(-\frac{\beta}{b_\beta}\right) \beta^{\sum\limits_{\{i|w_i=1\}} \alpha\left(t_i^{(S)}|\boldsymbol{\alpha}\right)} \exp\left[-\beta \sum_{\{i|w_i=1\}} \big(t_i^{(F)} - \mu_1\big)\right] \\
& \propto \beta^{a_\beta + \sum\limits_{\{i|w_i=1\}} \alpha\left(t_i^{(S)}|\boldsymbol{\alpha}\right) - 1} \exp\left\{ -\beta \left[b_\beta + \sum_{\{i|w_i=1\}} \big(t_i^{(F)} - \mu_1\big) \right] \right\} \\
& \sim \text{Gamma}\left(a_\beta + \sum_{\{i|w_i=1\}} \alpha\big(t_i^{(S)} \mid \boldsymbol{\alpha}\big),\ b_\beta + \sum_{\{i|w_i=1\}} \big(t_i^{(F)} - \mu_1\big) \right), \qquad (4.61)
\end{aligned}
$$

where $\alpha\big(t_i^{(S)} \mid \boldsymbol{\alpha}\big) = \exp\big(\alpha_0 + \alpha_1 t_i^{(S)}\big)$, $\mu_1 = 0.674$, $a_\beta = 1$ and $b_\beta = 1$.

By Equations (4.37) and (4.49), the conditional distribution of σ_0^2 given $\boldsymbol{w}$ and $\boldsymbol{t}^{(F)}$ is

$$
\begin{aligned}
f\big(\sigma_0^2 \mid \boldsymbol{w}, \boldsymbol{t}^{(F)}\big) &= P(\sigma_0^2 \mid a_{\sigma_0}, b_{\sigma_0}) \prod_{\{i|w_i=0\}} f_0\big(t_i^{(F)} \mid \mu_0 = 0, \sigma_0^2\big) \\
&= \frac{{b_{\sigma_0}}^{a_{\sigma_0}}}{\Gamma(a_{\sigma_0})} (\sigma_0^2)^{-(a_{\sigma_0}+1)} \exp\left(-\frac{b_{\sigma_0}}{\sigma_0^2}\right) \prod_{\{i|w_i=0\}} 2\frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{\big(t_i^{(F)}\big)^2}{2\sigma_0^2}\right) \\
&\propto (\sigma_0^2)^{-(a_{\sigma_0}+1)} \exp\left(-\frac{b_{\sigma_0}}{\sigma_0^2}\right) (\sigma_0^2)^{\left(-\frac{1}{2}\sum_{i=1}^{m} I_{\{w_i=0\}}\right)} \exp\left(-\frac{1}{2\sigma_0^2} \sum_{\{i|w_i=0\}} \big(t_i^{(F)}\big)^2\right) \\
&\propto (\sigma_0^2)^{-\left(a_{\sigma_0}+\frac{1}{2}\sum_{i=1}^{m} I_{\{w_i=0\}}+1\right)} \exp\left\{-\frac{1}{\sigma_0^2}\left(b_{\sigma_0} + \frac{1}{2} \sum_{\{i|w_i=0\}} \big(t_i^{(F)}\big)^2\right)\right\} \\
&\sim \text{Inverse-Gamma}\left(a_{\sigma_0} + \frac{1}{2}\sum_{i=1}^{m} I_{\{w_i=0\}},\ b_{\sigma_0} + \frac{1}{2} \sum_{\{i|w_i=0\}} \big(t_i^{(F)}\big)^2\right). \qquad (4.62)
\end{aligned}
$$

where $I_{\{w_i=0\}} = 1$ if $w_i = 0$ and 0 otherwise, $a_{\sigma_0} = 3$ and $b_{\sigma_0} = 2$.

Based on MCMC method using Gibbs sampler and Metropolis-Hasting algorithm (Metropolis et al., 1953; Hastings, 1970; Geman and Geman, 1984; Robert and Casella, 1999), the posterior sampling for β and σ_0^2, are directly from the full conditional distributions (4.61) and (4.62), respectively. For $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, the posterior sampling are drawn from the proposal multivariate t distribution with a small number of degree of freedom ν such as $\nu = 4$, which provides three finite moments for the approximating density (Gelman et al., 2013). Specifically, a multiple-try Metropolis algorithm is employed to increase the step size and acceptance rate (Liu et al., 2000) with details shown below.

For each iteration k, $k = 1, ..., K$, where K is the total number of MCMC iterations:

1. Draw J candidate samples from the proposal multivariate t function $q()$ based on $\boldsymbol{\theta}^{(k-1)}$:

$$\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \cdots, \boldsymbol{\theta}_J^* \sim q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k-1)}),$$

2. Let $\pi(\boldsymbol{\theta})$ be the likelihood function, which is multivariate normal distribution in this case. Given $q()$ is symmetric, we can calculate the importance weights as

$$w(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}^{(k-1)}) = \pi(\boldsymbol{\theta}_j^*), j = 1, \cdots, J,$$

3. Select one sample $\boldsymbol{\theta}^* \in \{\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \cdots, \boldsymbol{\theta}_J^*\}$ by multinomial sampling, here the weights are the normalized importance weights:

$$\hat{w}(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}^{(k-1)}) = \frac{w(\boldsymbol{\theta}_j^*)}{\sum_{j=1}^J w(\boldsymbol{\theta}_j^*)},$$

4. Draw $J-1$ auxiliary samples based on selected $\boldsymbol{\theta}^*$

$$\boldsymbol{\theta}_1^{**}, \boldsymbol{\theta}_2^{**}, \cdots, \boldsymbol{\theta}_{J-1}^{**} \sim q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*),$$

Set $\boldsymbol{\theta}_J^{**} = \boldsymbol{\theta}^{(k-1)}$.

5. Calculate the importance weights for the auxiliary samples as

$$w(\boldsymbol{\theta}_j^{**} \mid \boldsymbol{\theta}^*) = \pi(\boldsymbol{\theta}_j^{**}), j = 1, \cdots, J,$$

6. Let $\rho = \frac{\sum_{j=1}^J w(\boldsymbol{\theta}_j^* | \boldsymbol{\theta}^{(k-1)})}{\sum_{j=1}^J w(\boldsymbol{\theta}_j^{**} | \boldsymbol{\theta}^*)}$.

7. Generate a uniform random number $u \sim Unif(0,1)$.

8. If $u \leq min(1, \rho)$, then accept and set $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^*$. Otherwise, reject and set $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^{(k-1)}$.

For the Markov chains, the initial values are set as $\boldsymbol{\alpha}^{(0)} = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$, $\beta^{(0)} = 0.1$, $\sigma_0^{2(0)} = 1$ and $\pi_0^{(0)} = 0.94$. Denote the threshold that corresponds to $\pi_0^{(0)} = 0.94$ by $\lambda^{(0)}$, $w_i^{(0)} = 1$ if

$t_i^{(F)} \geq \lambda^{(0)}\}$ and 0 otherwise. Then we can get the estimates of the coefficients based on the logistic regression model described in Equation (4.41) on $w_i^{(0)}$ given $t_i^{(S)}$, $i = 1, \cdots, m$, as the initial values of $\boldsymbol{\gamma}^{(0)} = \begin{bmatrix} \gamma_0^{(0)} & \gamma_1^{(0)} \end{bmatrix}^T$.

At the k^{th} MCMC iteration, $k = 1, ..., K$, when $\{\boldsymbol{\alpha}^{(k)}, \boldsymbol{\gamma}^{(k)}, \beta^{(k)}, \sigma_0^{2(k)}\}$ are drawn, the probability of the latent variable $w_i^{(k)} = 1$ given $t_i^{(F)}$, $t_i^{(S)}$ and $\{\boldsymbol{\alpha}^{(k)}, \boldsymbol{\gamma}^{(k)}, \beta^{(k)}, \sigma_0^{2(k)}\}$ denoted by $p_i^{(k)}$, $i = 1, ...m$, is updated using

$$\begin{aligned} p_i^{(k)} &= P\big(w_i = 1 \mid t_i^{(F)}, t_i^{(S)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\gamma}^{(k)}, \beta^{(k)}, \sigma_0^{2(k)}\big) \\ &= \frac{\pi_1(t_i^{(S)} \mid \boldsymbol{\gamma}^{(k)}) f_1\big(t_i^{(F)} \mid t_i^{(S)}, \boldsymbol{\alpha}^{(k)}, \beta^{(k)}\big)}{\pi_0\big(t_i^{(S)} \mid \boldsymbol{\gamma}^{(k)}\big) f_0\big(t_i^{(F)} \mid \sigma_0^{2(k)}\big) + \pi_1(t_i^{(S)} \mid \boldsymbol{\gamma}^{(k)}) f_1\big(t_i^{(F)} \mid t_i^{(S)}, \boldsymbol{\alpha}^{(k)}, \beta^{(k)}\big)}, \end{aligned} \tag{4.63}$$

Then $w_i^{(k)}$ is sampled from the full conditional Bernoulli distribution with probability $p_i^{(k)}$,

$$w_i^{(k)} \overset{iid}{\sim} Bernoulli\,(p_i^{(k)}), \tag{4.64}$$

After obtaining the samples for each parameter from their posterior distributions, we can take the posterior medians as the posterior estimates of the parameters $\{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}}, \hat{\beta}, \hat{\sigma}_0^2\}$. By Equation (4.45), the posterior probability of the i^{th} connectivity link is null given $t_i^{(F)}$ and $t_i^{(S)}$, denoted by $\widehat{BLfdr}_i(t_i^{(F)}, t_i^{(S)})$, is calculated as

$$\begin{aligned} \widehat{BLfdr}_i\big(t_i^{(F)}, t_i^{(S)}\big) &= P\big(H_{0,i} = 0 \mid t_i^{(F)}, t_i^{(S)}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}}, \hat{\beta}, \hat{\sigma}_0^2\big) \\ &= \frac{\pi_0\big(t_i^{(S)} \mid \hat{\boldsymbol{\gamma}}\big) f_0\big(t_i^{(F)} \mid \hat{\sigma}_0^2\big)}{\pi_0\big(t_i^{(S)} \mid \hat{\boldsymbol{\gamma}}\big) f_0\big(t_i^{(F)} \mid \hat{\sigma}_0^2\big) + \pi_1\big(t_i^{(S)} \mid \hat{\boldsymbol{\gamma}}\big) f_1\big(t_i^{(F)} \mid \hat{\boldsymbol{\alpha}}, \hat{\beta}, t_i^{(S)}\big)}, \end{aligned} \tag{4.65}$$

where $f_0\big(t_i^{(F)} \mid \hat{\sigma}_0^2\big)$, $f_1\big(t_i^{(F)} \mid t_i^{(S)}, \hat{\boldsymbol{\alpha}}, \hat{\beta}\big)$, $\pi_0\big(t_i^{(S)} \mid \hat{\boldsymbol{\gamma}}\big)$ and $\pi_1\big(t_i^{(S)} \mid \hat{\boldsymbol{\gamma}}\big)$ are calculated using Equations (4.37), (4.38), (4.41) and (4.42), respectively.

Then we apply the oracle procedure introduced by Sun and Cai (2007) for multiple testing:

1. Calculate $\widehat{BLfdr}_i$, for $i = 1, \cdots, m$, using Equation (4.65) and then sort them in ascending order denoted by $\widehat{BLfdr}_{(1)} \leq \widehat{BLfdr}_{(2)} \leq \cdots \widehat{BLfdr}_{(m)}$.
2. For a given q $(0 < q < 1)$, find k for which $k = \arg\max_i \left\{ i : \frac{1}{i} \sum_{j=1}^{i} \widehat{BLfdr}_{(j)} \leq q \right\}$.
3. Then reject all $H_{(i)}$, $i = 1, \cdots, k$.

Using this procedure the average FDR is controlled at a pre-specified level of q, where $q = 0.2 - 0.3$ is considered reasonable as discussed in previous sections 4.4 and 4.5 (Genovese et al., 2002; Song, 2016; Bhaumik et al., 2018b).

4.7 Simulation Study

The purpose of the simulation study is to evaluate the performance of Bayesian multimodal Lfdr method and compare it with Efron's Lfdr method that considers solely FC data (Efron et al., 2001; Efron and Tibshirani, 2002) in terms of FDR control to identify differential FCs between disease group and control group in a cross-sectional, multimodal neuroimaging study with small to moderate sample size, in which both FC and SC are measured.

4.7.1 Simulation Algorithm

The simulation study is designed to have a similar data structure as the motivating study as described in Chapter 2. To account for the potential correlation between FC and SC observations that may exist, we use a bivariate mixed-effects model with heteroscedastic errors (Zhao, 2014) to simulate FC and SC data jointly and simultaneously based on the parameters estimated from the mixed-effects model described in Chapter 3 using the LLD neuroimaging data.

By mimicking the sample size of the LLD study with 10 in LLD group and 13 in HC group, in this simulation we consider small to moderate sample sizes and assume equal sample sizes for the LLD group and HC group: $n = 15, 25, 35$ and 45 per group, with FC and SC measurement taken from 87 regions for each participant. The total number of connectivity links measured is $m = 87 \times 86/2 = 3741$.

Let $H_{0,i}^{(F)}$ and $H_{0,i}^{(S)}$ be the null hypothesis that the i^{th} link has no difference between LLD group and HC group in FC and SC, respectively. For simplicity, we refer to the connectivity links under the null hypothesis as null connectivity links, and the connectivity links under the alternative hypothesis as alternative connectivity links. The 3741 connectivity links can be classified into four link classes based on $H_{0,i}^{(F)}$ and $H_{0,i}^{(S)}$ (see Table IV).

TABLE IV
CLASSIFICATION OF CONNECTIVITY LINKS IN SIMULATION STUDY

$H_{0,i}^{(F)}$	$H_{0,i}^{(S)}$	Link Class Description
False	False	alternative in both FC and SC
False	True	alternative in FC, null in SC
True	False	null in FC, alternative in SC
True	True	null in both FC and SC

Following the LLD neuroimaging dataset, we assume that

1. The proportion of connectivity links that are alternative in both FC and SC is 1%,
2. The proportion of connectivity links that are alternative in FC and null in SC is 1%,
3. The proportion of connectivity links that are null in FC and alternative in SC is 1%,

4. The proportion of connectivity links that are null in both FC and SC is 97%.

So the total proportion of alternative links in FC is 2%, consistent with the assumption used in simulation study based on the same LLD neuroimaging data (Song, 2016; Bhaumik et al., 2018b). Next we assume, without a loss of generality, that in each simulated dataset the 3741 connectivity links are set in the following order:

1. The first 37 (1%) connectivity links are alternative in both FC and SC,
2. The following 37 (1%) connectivity links are alternative in FC and null in SC,
3. The next 37 (1%) connectivity links are null in FC and alternative in SC,
4. The rest 3630 (97%) connectivity links are null in both FC and SC.

Let $\delta^{(F)}$ and $\delta^{(S)}$ denote the true differences between LLD group and HC group for the alternative connectivity links in FC and SC, respectively. We assume $\delta^{(F)} = 0.075$ and $\delta^{(S)} = 0.25$ that are considered moderate sizes. Let $y_{ij}^{(F)}$ and $y_{ij}^{(S)}$ be the FC and SC observation, respectively, for the i^{th} connectivity link ($i = 1, \cdots, m$) of the j^{th} subject ($j = 1, \cdots, N$, where N is the total number of subjects and $N = 2n$), $y_{ij}^{(F)}$ and $y_{ij}^{(S)}$ are generated using the bivariate mixed-effects model with the following settings,

$$\boldsymbol{y}_{ij} = \begin{bmatrix} y_{ij}^{(F)} & y_{ij}^{(S)} \end{bmatrix}^T, \tag{4.66}$$

In the bivariate mixed-effects model, $\boldsymbol{y}_{ij}$ can be partitioned into fixed effects $\boldsymbol{\beta}_{0i}$ and $\boldsymbol{\beta}_{1i}$, random subject effects $\boldsymbol{\gamma}_j$ and random errors $\boldsymbol{\epsilon}_{ij}$,

$$\boldsymbol{y}_{ij} = \boldsymbol{\beta}_{0i} \times (1 - Grp_j) + \boldsymbol{\beta}_{1i} \times Grp_j + \boldsymbol{\gamma}_j + \boldsymbol{\epsilon}_{ij}, \tag{4.67}$$

where

- Grp_j is the indicator variable for which group the subject is from, HC group if $Grp_j = 0$, or LLD group if $Grp_j = 1$.
- $\boldsymbol{\beta}_{0i}$ and $\boldsymbol{\beta}_{1i}$ are the vectors of fixed FC and SC effects for the i^{th} connectivity link in HC group and LLD group, respectively.
- $\boldsymbol{\gamma}_j = \begin{bmatrix} \gamma_j^{(F)} & \gamma_j^{(S)} \end{bmatrix}^T$ is the vector of random subject effects for FC and SC in the j^{th} subject.
- $\boldsymbol{\epsilon}_{ij} = \begin{bmatrix} \epsilon_{ij}^{(F)} & \epsilon_{ij}^{(S)} \end{bmatrix}^T$ is the vector of random errors for FC and SC of the i^{th} connectivity link in the j^{th} subject.
- $\boldsymbol{\gamma}_j$ and $\boldsymbol{\epsilon}_{ij}$ are independent of each other.

For fixed Effects $\boldsymbol{\beta}_{0i}$ and $\boldsymbol{\beta}_{1i}$, the vectors of fixed FC and SC effects for the i^{th} connectivity link in HC group and LLD group are,

$$\boldsymbol{\beta}_{0i} = \begin{bmatrix} \beta_{0i}^{(F)} & \beta_{0i}^{(S)} \end{bmatrix}^T,$$
$$\boldsymbol{\beta}_{1i} = \begin{bmatrix} \beta_{1i}^{(F)} & \beta_{1i}^{(S)} \end{bmatrix}^T,$$

Let $\bar{\hat{\beta}}_{0+}^{(F)}$ and $\bar{\hat{\beta}}_{0+}^{(S)}$ be the average of the fixed effect estimates over all m connectivity links in HC group for FC and SC from mixed-effects model using the LLD neuroimaging data, respectively, $\bar{\hat{\beta}}_{0+}^{(F)} = \frac{1}{3741}\sum_{i=1}^{3741}\hat{\beta}_{0i}^{(F)} = 0.138$, $\bar{\hat{\beta}}_{0+}^{(S)} = \frac{1}{3741}\sum_{i=1}^{3741}\hat{\beta}_{0i}^{(S)} = 1.099$. We assume,

- For HC group: $\beta_{0i}^{(F)} = \bar{\hat{\beta}}_{0+}^{(F)} = 0.138,\ \beta_{0i}^{(S)} = \bar{\hat{\beta}}_{0+}^{(S)} = 1.099,\ i = 1, ..., 3741$. That is,

$$\boldsymbol{\beta}_0^{(F)} = (\underbrace{0.138, \cdots, 0.138}_{3741})^T,$$
$$\boldsymbol{\beta}_0^{(S)} = (\underbrace{1.099, \cdots, 1.099}_{3741})^T,$$

- For LLD group:

$$\beta_{1i}^{(F)} = \begin{cases} \bar{\hat{\beta}}_{0+}^{(F)} = 0.138 & \text{for null links in FC ,} \\ \bar{\hat{\beta}}_{0+}^{(F)} + \delta^{(F)} = 0.213 & \text{for alternative links in FC,} \end{cases}$$

$$\beta_{1i}^{(S)} = \begin{cases} \bar{\hat{\beta}}_{0+}^{(S)} = 1.099 & \text{for null links in SC,} \\ \bar{\hat{\beta}}_{0+}^{(S)} + \delta^{(S)} = 1.349 & \text{for alternative links in SC,} \end{cases}$$

That is,

$$\boldsymbol{\beta}_1^{(F)} = (\underbrace{0.213, \cdots, 0.213}_{74}, \underbrace{0.138, \cdots, 0.138}_{3667})^T,$$
$$\boldsymbol{\beta}_1^{(S)} = (\underbrace{1.349, \cdots, 1.349}_{37}, \underbrace{1.099, \cdots, 1.099}_{37}, \underbrace{1.349, \cdots, 1.349}_{37}, \underbrace{1.099, \cdots, 1.099}_{3630})^T,$$

For random Subject Effects $\boldsymbol{\gamma}_i$, the vector of random subject effects for FC and SC in the j^{th} subject is $\boldsymbol{\gamma}_j = \begin{bmatrix} \gamma_j^{(F)} & \gamma_j^{(S)} \end{bmatrix}^T$. Assume $\gamma_j^{(F)}$ and $\gamma_j^{(F)}$ are independent of each other,

$$\boldsymbol{\gamma_j} = \begin{bmatrix} \gamma_j^{(F)} & \gamma_j^{(S)} \end{bmatrix}^T \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\gamma),$$
$$\boldsymbol{\Sigma}_\gamma = \begin{bmatrix} \sigma_\gamma^{2(F)} & 0 \\ 0 & \sigma_\gamma^{2(S)} \end{bmatrix},$$

where $\sigma^{2(F)}_{\gamma} = 0.004$ and $\sigma^{2(S)}_{\gamma} = 0.108$, estimated from mixed-effects model using the LLD neuroimaging data.

For random Errors $\boldsymbol{\epsilon_{ij}}$, the vector of random errors for FC and SC of the i^{th} connectivity link in the j^{th} subject is $\boldsymbol{\epsilon_{ij}} = \begin{bmatrix} \epsilon^{(F)}_{ij} & \epsilon^{(S)}_{ij} \end{bmatrix}^T$ and we assume,

- For HC group:

$$
\begin{aligned}
&\boldsymbol{\epsilon_{ij}} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_{\epsilon,0}\right), \\
&\boldsymbol{\Sigma}_{\epsilon,0} = \begin{bmatrix} \sigma^{2(F)}_{\{\epsilon,0i\}} & \rho_{\epsilon,0}\sigma^{(F)}_{\{\epsilon,0i\}}\sigma^{(S)}_{\{\epsilon,0i\}} \\ \rho_{\epsilon,0}\sigma^{(F)}_{\{\epsilon,0i\}}\sigma^{(S)}_{\{\epsilon,0i\}} & \sigma^{2(S)}_{\{\epsilon,0i\}} \end{bmatrix}, \\
&\sigma^{2(F)}_{\{\epsilon,0i\}} \sim \Gamma\left(\alpha^{(F)}_0 = 4.34, \beta^{(F)}_0 = 120.67\right), \\
&\sigma^{2(S)}_{\{\epsilon,0i\}} \sim \Gamma\left(\alpha^{(S)}_0 = 0.71, \beta^{(S)}_0 = 0.64\right),
\end{aligned}
$$

  where the shape and rate parameters for the variances are estimated by fitting a gamma distribution to the variance estimates of all 3741 links in HC group for FC and SC separately, $\rho_{\epsilon,0}$ is the correlation between the random FC and SC errors in subjects from HC group and we set $\rho_{\epsilon,0} = 0.1$, estimated from the correlation between the variance estimates of all links in HC group for FC and SC, from the mixed-effects model in the lLD neuroimaging study.

- For LLD group:

$$
\begin{aligned}
&\boldsymbol{\epsilon_{ij}} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_{\epsilon,1}\right), \\
&\boldsymbol{\Sigma}_{\epsilon,1} = \begin{bmatrix} \sigma^{2(F)}_{\{\epsilon,1i\}} & \rho_{\epsilon,1}\sigma^{(F)}_{\{\epsilon,1i\}}\sigma^{(S)}_{\{\epsilon,1i\}} \\ \rho_{\epsilon,1}\sigma^{(F)}_{\{\epsilon,1i\}}\sigma^{(S)}_{\{\epsilon,1i\}} & \sigma^{2(S)}_{\{\epsilon,1i\}} \end{bmatrix},
\end{aligned}
$$

$$\sigma^{2(F)}_{\{\epsilon,1i\}} \sim \Gamma\left(\alpha_1^{(F)} = 3.49, \beta_1^{(F)} = 89.47\right),$$
$$\sigma^{2(S)}_{\{\epsilon,1i\}} \sim \Gamma\left(\alpha_1^{(S)} = 0.66, \beta_1^{(S)} = 0.69\right),$$

where the shape and rate parameters for the variances are estimated by fitting a gamma distribution to the variance estimates of all 3741 links in LLD group for FC and SC separately, from the mixed-effects model using the lLD neuroimaging study.

With respect to the correlation coefficient between the random FC and SC errors in subjects from LLD group (denoted by $\rho_{\epsilon,1}$), we assume that (i) weak correlation (i.e., $\rho_{\epsilon,1} = 0.1$) for null connectivity FC links; (ii) Varying scenarios in terms of $\rho_{\epsilon,1}$ for alternative connectivity FC links including weak, mild, moderate and strong correlations between random FC and SC errors in LLD patients as described in Table V, aimed to investigate the influence of $\rho_{\epsilon,1}$ specifically for alternative FC links on FDR control.

TABLE V
SCENARIOS CONSIDERING VARYING CORRELATION COEFFICIENTS BETWEEN THE RANDOM FC AND SC ERRORS IN LLD PATIENTS ($\rho_{\epsilon,1}$) FROM WEAK TO STRONG BY LINK CLASS

Scenario	Link Class Description[a]			
	both alter	FC alter, SC null	FC null, SC alter	both null
1. Weak correlation	0.1	0.1	0.1	0.1
2. Mild correlation	0.4	0.1	0.1	0.1
3. Moderate correlation	0.4	0.8	0.1	0.1
4. Strong correlation	0.9	0.1	0.1	0.1

[a] alter = alternative links; null = null links. Refer to Table IV for the link class description.

To illustrate the data structure in matrix notation, we can rewrite the bivariate linear MERM in Equation (4.67) for the j^{th} subject using matrix notation as,

$$\boldsymbol{y}_j = \boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\boldsymbol{\gamma}_j + \boldsymbol{\epsilon}_j \tag{4.68}$$

where

- $\boldsymbol{y}_j$ is a $2m \times 1$ vector with the first m elements for FC measurements and the last m elements for SC measurements,

$$\boldsymbol{y}_j = \begin{bmatrix} y_{1j}^{(F)} & y_{2j}^{(F)} & \dots & y_{mj}^{(F)} & y_{z,1j}^{(S)} & y_{2j}^{(S)} & \dots & y_{mj}^{(S)} \end{bmatrix}^T_{1\times 2m}, \tag{4.69}$$

- $\boldsymbol{X}_j$ is a $2m \times 4m$ design matrix for the fixed effects $\boldsymbol{\beta}_k$,

$$\begin{aligned} \boldsymbol{X} &= \begin{bmatrix} 1-Grp_j & \cdots & 0 & Grp_j & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1-Grp_j & 0 & \cdots & Grp_j \end{bmatrix}_{2m\times 4m} \\ &= \begin{bmatrix} 1-Grp_j & Grp_j \end{bmatrix} \otimes \boldsymbol{I}_{2m}, \end{aligned} \tag{4.70}$$

  where $\boldsymbol{I}_{2m}$ is a $2m$-dimensional identity matrix.

- $\boldsymbol{\beta}$ is a $4m \times 1$ fixed effect vector,

$$\begin{aligned} \boldsymbol{\beta} &= \begin{bmatrix} \boldsymbol{\beta}_0^{(F)T} & \boldsymbol{\beta}_0^{(S)T} & \boldsymbol{\beta}_1^{(F)T} & \boldsymbol{\beta}_1^{(S)T} \end{bmatrix}^T_{1\times 4m} \\ &= \begin{bmatrix} \beta_{01}^{(F)} & \dots & \beta_{0m}^{(F)} & \beta_{01}^{(S)} & \dots & \beta_{0m}^{(S)} & \beta_{11}^{(F)} & \dots & \beta_{1m}^{(F)} & \beta_{11}^{(S)} & \dots & \beta_{1m}^{(S)} \end{bmatrix}^T_{1\times 4m}, \end{aligned} \tag{4.71}$$

- $\boldsymbol{Z}_j$ is a $2m \times 2$ design matrix for the random subject effects $\boldsymbol{\gamma}_j$,

$$\boldsymbol{Z}_j = \begin{bmatrix} \mathbf{1}_m & \mathbf{0}_m \\ \mathbf{0}_m & \mathbf{1}_m \end{bmatrix}_{2m \times 2}, \tag{4.72}$$

  where $\mathbf{1}_m$ is a $m \times 1$ vector of all ones and $\mathbf{0}_m$ is a $m \times 1$ vector of all zeros.

- $\boldsymbol{\gamma}_j$ is a 2×1 vector of the random subject effects for FC and SC measurements,

$$\boldsymbol{\gamma}_j = \begin{bmatrix} \gamma_j^{(F)} \\ \gamma_j^{(S)} \end{bmatrix}, \tag{4.73}$$

- $\boldsymbol{\epsilon}_j$ is a $2m \times 1$ vector with the first m elements for the random errors of FC and the rest m elements for the random errors of SC,

$$\boldsymbol{\epsilon}_j = \begin{bmatrix} \epsilon_{1j}^{(F)} & \epsilon_{2j}^{(F)} & \dots & \epsilon_{mj}^{(F)} & \epsilon_{1j}^{(S)} & \epsilon_{2j}^{(S)} & \dots & \epsilon_{mj}^{(S)} \end{bmatrix}^T_{1 \times 2m}. \tag{4.74}$$

For each scenario of correlation coefficients under each sample size of $n = 15$, 25, 35 and 45 subjects per group, 1,000 datasets are simulated. The FC and SC data in each simulation are analyzed separately using mixed-effects model to compute the FC and SC test statistics for all the m connectivity links, based on which the Lfdr using Efron's Lfdr and BLfdr using Bayesian multimodal Lfdr for each of the m connectivity links are calculated. The oracle procedure described above is then applied to control FDR and determine which connectivity links are differential between the two groups.

For Bayesian multimodal Lfdr method, we run one MCMC chain for 20,000 iterations with the first 10,000 iterations discarded as burn-in and a thinning interval of 10 (i.e., retaining only every 10^{th} iteration), to produce a total of 1,000 random posterior draws for each parameter.

4.7.2 Simulation Results

Table VI, Figures 13 and 14 summarize the simulation results in terms of FDR at pre-specified level of $q = 0.2$ and 0.3 by Efron's Lfdr method and Bayesian multimodal Lfdr method for different scenarios assuming the true FC and SC differences between LLD group and HC group for the alternative connectivity links are $\delta^{(F)} = 0.075$ and $\delta^{(S)} = 0.25$. Bayesian multimodal Lfdr method outperforms Efron's Lfdr method in terms of controling FDR consistently for all scenarios across different sample sizes, orrelation between the random FC and SC errors in true alternative connectivity links for both FC and SC measurements in LLD group, and q values.

Sample size has a major impact on controlling FDR at the desired q level. For small sample sizes, Bayesian multimodal Lfdr method yields considerably lower FDR than Efron's Lfdr at the same q level. For example, with $n = 15$ subjects per group the FDR at $q = 0.2$ ranges between 0.429 to 0.549 by Bayesian multimodal Lfdr method, as compared to a range between 0.675 to 0.686 by Efron's Lfdr method. As discussed in Chapter 1, neuroimaging studies usually have a small number of subjects, and thus improving FDR control in the presence of small sample sizes is particularly helpful for researchers. As sample size increases, FDR by both methods are reduced greatly. With $n = 45$ subjects per group, FDRs from the two methods are very close and both converging to the pre-specified level q, suggesting that the asymptotic properties are similar between the two methods.

It is also noted that FDR by Bayesian multimodal Lfdr decreases when correlation between the random FC and SC errors in true alternative connectivity links for both FC and SC measurements in LLD group (denoted by $\rho_{\epsilon,1}$) is larger, for small sample sizes. For example, for $n = 15$ subjects per group when $\rho_{\epsilon,1}$ increases from 0.1 to 0.9, FDR using Bayesian mul-

timodal Lfdr method is reduced from 0.549 to 0.429 at $q = 0.2$, and from 0.654 to 0.555 at $q = 0.3$. As aforementioned, Koch et al. (2002) and Honey et al. (2009) both demonstrated the direct inference of strong FC from strong SC, suggesting the high functional-structural correlation ($\rho_{\epsilon,1} = 0.9$) in true alternative connectivity links for both FC and SC that is assumed in Scenario 4. The effect of the $\rho_{\epsilon,1}$ on FDR control seems to be attenuated as the sample size increases. For example, for $n = 45$ subjects per group FDR at $q = 0.2$ using Bayesian multimodal Lfdr method is 0.219 and 0.226 for $\rho_{\epsilon,1} = 0.1$ and 0.9, respectively. In contrast, FDRs based on Efron's Lfdr method are invariant to different values of $\rho_{\epsilon,1}$, which totally makes sense since Efron's method uses information from FC data only and hence the functional-structural correlation should not have any influence on FDR.

In addition, we have observed that larger q value yields a larger FDR as FDR controlled at $q = 0.3$ are larger as compared to those controlled at $q = 0.2$. FDR by both Efron and Bayesian multimodal Lfdr methods reaches to the pre-specified q values for $n = 45$ subjects per group.

TABLE VI
FDR AT PRE-SPECIFIED LEVELS OF 0.2 AND 0.3 BY EFRON'S LFDR METHOD AND BAYESIAN MULTIMODAL LFDR METHOD USING SIMULATION STUDIES[a]

Scenario[b]	n[c]	$q = 0.2$[d]		$q = 0.3$[d]	
		Efron's Lfdr	Bayesian multimodal Lfdr	Efron's Lfdr	Bayesian multimodal Lfdr
1. Weak correlation					
	15	0.675	0.549	0.744	0.654
	25	0.461	0.405	0.551	0.494
	35	0.308	0.245	0.409	0.345
	45	0.249	0.219	0.354	0.310
2. Mild correlation					
	15	0.676	0.496	0.750	0.575
	25	0.451	0.362	0.535	0.463
	35	0.298	0.295	0.413	0.378
	45	0.248	0.242	0.365	0.336
3. Moderate correlation					
	15	0.678	0.522	0.758	0.602
	25	0.467	0.384	0.537	0.463
	35	0.328	0.317	0.436	0.392
	45	0.258	0.217	0.361	0.320
4. Strong correlation					
	15	0.686	0.429	0.741	0.555
	25	0.473	0.400	0.563	0.514
	35	0.321	0.291	0.419	0.387
	45	0.247	0.226	0.352	0.323

[a] Results are based on 1000 simulated datasets for each scenario, assuming the true FC and SC differences between LLD group and HC group for the alternative connectivity links $\delta^{(F)} = 0.075$ and $\delta^{(S)} = 0.25$. Each simulated data is fit with the mixed-effects model to obtain the test statistics for each of the $m = 3741$ connectivity links for FC data and SC data, separately.
[b] Including varying correlation coefficients between the random FC and SC errors in LLD patients from weak to strong. Refer to Table V for the details of each scenario.
[c] n = sample size per group.
[d] q = desired FDR level.

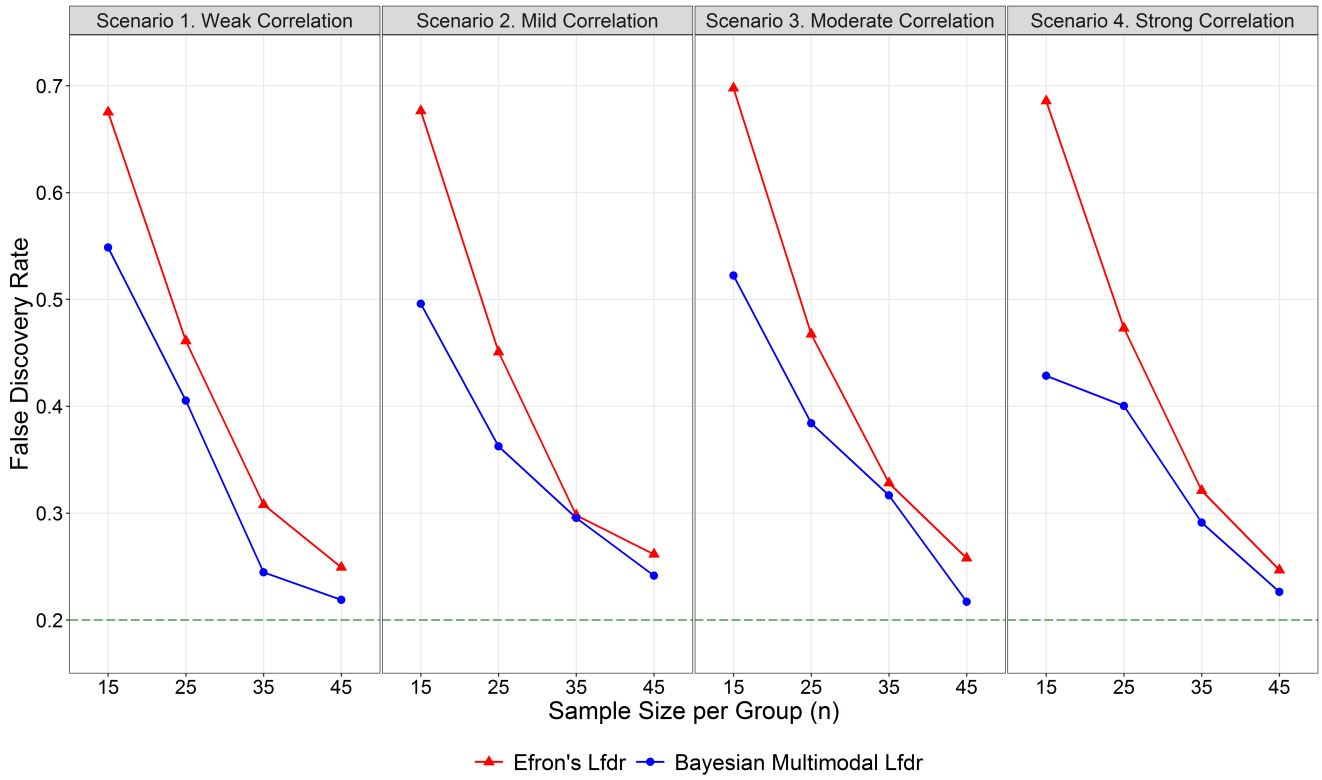


Figure 13. FDR control at pre-specified level of $q = 0.2$ by Efron's Lfdr method and Bayesian multimodal Lfdr method using simulation study

Refer to Table V for the details of each scenario.

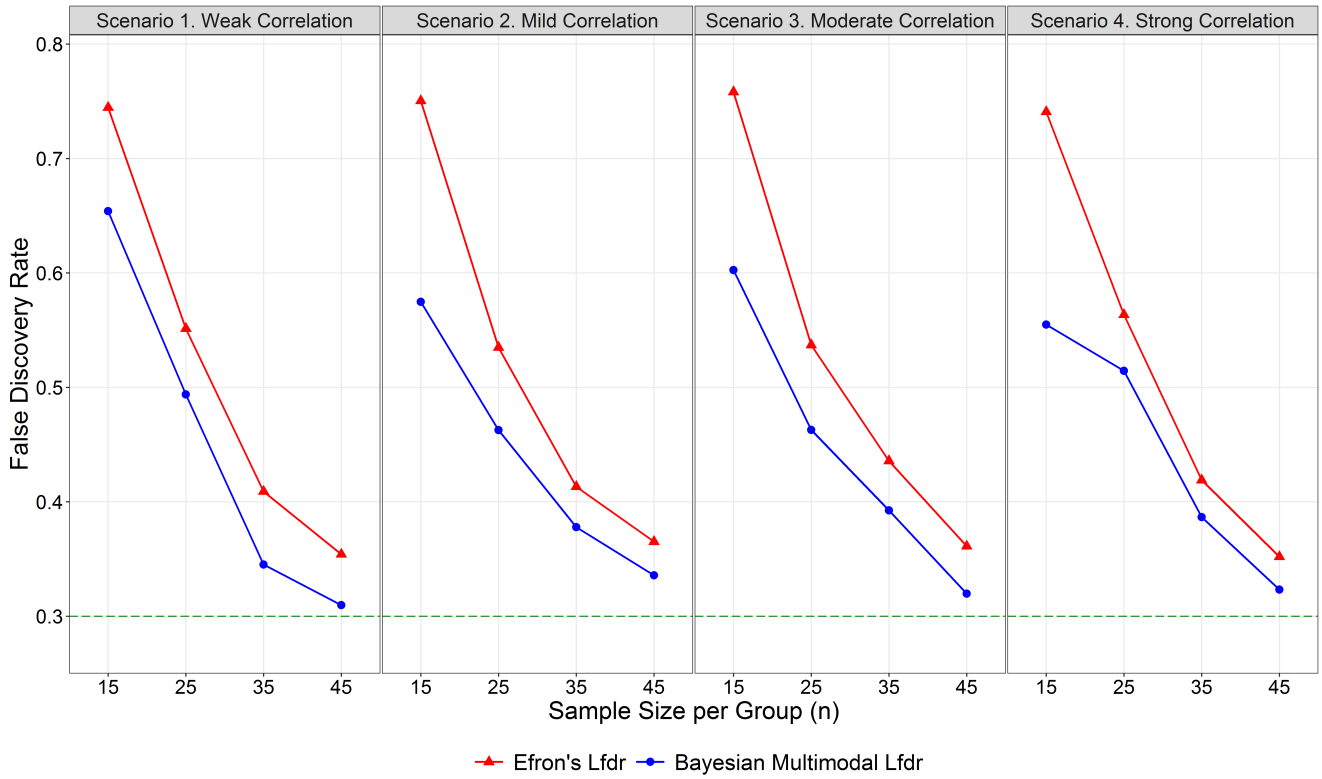


Figure 14. FDR control at pre-specified level of $q = 0.3$ by Efron's Lfdr method and Bayesian multimodal Lfdr method using simulation study

Refer to Table V for the details of each scenario.

CHAPTER 5

DIRICHLET PROCESS FOR DISCOVERING SIGNIFICANT FUNCTIONAL CONNECTIVITIES

The previous chapter described Bayesian multimodal Lfdr method for FDR control based on the test statistics for group comparison between disease group and control group for FC and SC obtained from the mixed-effects model presented in Chapter 3. In this chapter, the FC data are directly modeled via a semi-parametric Bayesian approach utilizing a non-parametric Dirichlet process mixture model. This Bayesian method has several advantages. It does not impose any model assumption and thus allows flexibility in distributional assumptions in support of exploring the space of all possible and measurable distributions. It also can reduce dimension and induce multiplicity adjustment through the inherent clustering property of the Dirichlet process. Therefore, this method can be used to mimic hypothesis testing framework, while preserving the false discovery rate.

Gopalan and Berry (1998) introduced Dirichlet process priors for distribution of the parameters of interest in multiple comparison. Since then, the Bayesian multiple comparison has been explored with a focus on genomics research. Dahl and Newton (2007) proposed a Dirichlet process mixture model for multiple testing of correlated hypothesis in microarray genomics data and showed in simulation that model-based clustering of treatment effects naturally incorporates information among all tests and thus induces multiplicity adjustment, leading to improved efficiency in FDR control. Kim et al. (2009) also presented a non-parametric Bayesian model utilizing a spiked Dirichlet process prior for distribution of random effects for multiple testing

in genomics, and demonstrated in simulation study that the method is effective in reducing FDR.

Recently, Ghosal (2019) developed a semi-parametric Bayesian approach exploiting a non-parametric Dirichlet process mixture model for discoveries of differential FCs in neuroimaging study, and has demonstrated in simulation study, that the model yields higher true positive rate (power) while controlling false positive rate (type I error) using receiver operating characteristics curves as compared to the parametric model and neighboring model. We therefore apply this method to analyze the FC data from the LLD neuroimaging study, as a comparison with the results based on Bayesian multimodal Lfdr method and mixed-effects model.

5.1 Dirichlet Process

5.1.1 Definition

Dirichlet process (DP) is a stochastic process to draw probability distributions over a set of distributions and it is often used in Bayesian non-parametric models as a prior over probability distributions (Ferguson, 1973; Teh et al., 2006; Teh, 2010). The definition of DP was provided by Ferguson (1973). Given a measurable space $(\Theta, \mathcal{A})$ where Θ be a set and $\mathcal{A}$ is a σ-algebra of subsets of Θ, let G_0 be a base probability measure on $(\Theta, \mathcal{A})$, which is often assumed to be continuous (or smooth), and $\alpha > 0$ be the concentration (or scale) parameter, for all pairwise disjoint partitions $\{\mathcal{A}_1, \cdots, \mathcal{A}_k\}$ of space Θ where $\bigcup_{i=1}^{k} A_i = \Theta$, if there exists a random probability distribution $G(\cdot)$ that yields

$$\{G(\mathcal{A}_1), \cdots, G(\mathcal{A}_k)\} \sim Dirichlet\left(\alpha G_0(\mathcal{A}_1), \cdots, \alpha G_0(\mathcal{A}_k)\right),\ \alpha > 0 \tag{5.1}$$

Then $G(\cdot)$ is a DP on $(\Theta, \mathcal{A})$,

$$G(\cdot) \sim DP\big(\alpha, G_0(\cdot)\big),\ \alpha > 0 \tag{5.2}$$

For any measurable set $\mathcal{A}_j \subset \mathcal{A}$, the marginal distributions of $G(\mathcal{A}_j)$ is a Beta distribution (Kotz et al., 2000):

$$G(\mathcal{A}_j) \sim Beta\big(\alpha G_0(\mathcal{A}_j), \alpha(1 - G_0(\mathcal{A}_j))\big),\ j = 1, \cdots, k \tag{5.3}$$

Thus, the prior mean and variance of the random probability measure $G(\mathcal{A}_j)$ are:

$$\mathrm{E}\big(G(\mathcal{A}_j)\big) = G_0(\mathcal{A}_j), \tag{5.4}$$

$$\mathrm{Var}\big(G(\mathcal{A}_j)\big) = \frac{G_0(\mathcal{A}_j)(1 - G_0(\mathcal{A}_j))}{\alpha + 1}. \tag{5.5}$$

We see that the base probability measure $G_0(\mathcal{A}_j)$ is the prior mean of the DP, and the concentration parameter α controls the prior variance of the DP: larger α is, smaller the variance (i.e., more concentration of the mass around the mean).

5.1.2 Posterior Distribution

Let $G(\cdot) \sim DP\big(\alpha, G_0(\cdot)\big)$, a sample of size n can be drawn from $G(\cdot)$, $\theta_1, \cdots, \theta_n \overset{iid}{\sim} G(\cdot)$. As $G(\cdot)$ is a probability distribution over Θ, the values of θ_i, $i = 1, \cdots, n$ are in Θ. Since the

Dirichlet distribution is a conjugate prior for the multinomial distribution, by Equation (5.1) the posterior distribution of $G(\cdot)$ given the observed values of $\theta_1, \cdots, \theta_n$ is also a DP on $(\Theta, \mathcal{A})$,

$$\begin{aligned}
&\{G(\mathcal{A}_1), \cdots, G(\mathcal{A}_k)\} \mid \theta_1, \cdots, \theta_n \\
&\sim Dirichlet\left(\alpha G_0(\mathcal{A}_1) + \sum_{i=1}^{n} 1_{\{\theta_j \in \mathcal{A}_1\}}, \cdots, \alpha G_0(\mathcal{A}_k) + \sum_{i=1}^{n} 1_{\{\theta_j \in \mathcal{A}_k\}}\right), \qquad (5.6)
\end{aligned}$$

where $\sum_{i=1}^{n} 1_{\{\theta_i \in \mathcal{A}_j\}}$ denotes the number of observed values in the set $\mathcal{A}_j$, $j = 1, \cdots, k$. The sum of $\sum_{i=1}^{n} 1_{\{\theta_i \in \mathcal{A}_j\}}$ is $\sum_{j=1}^{k} \sum_{i=1}^{n} 1_{\{\theta_i \in \mathcal{A}_j\}} = n$, so the sum of the parameters in Equation (5.6) is $\sum_{j=1}^{k} \left[\alpha G_0(\mathcal{A}_j) + \sum_{i=1}^{n} 1_{\{\theta_j \in \mathcal{A}_j\}}\right] = \alpha + n$. Thus we can rewrite $\alpha G_0(\mathcal{A}_j) + \sum_{i=1}^{n} 1_{\{\theta_i \in \mathcal{A}_j\}}$ in Equation (5.6) as:

$$\begin{aligned}
\alpha G_0(\mathcal{A}_j) + \sum_{i=1}^{n} 1_{\{\theta_i \in \mathcal{A}_j\}} &= (\alpha + n)\left(\frac{\alpha}{\alpha + n} G_0(\mathcal{A}_j) + \frac{1}{\alpha + n} \sum_{i=1}^{n} 1_{\{\theta_i \in \mathcal{A}_j\}}\right) \\
&= (\alpha + n)\left(\frac{\alpha}{\alpha + n} G_0(\mathcal{A}_j) + \frac{n}{\alpha + n} \frac{1}{n} \sum_{i=1}^{n} 1_{\{\theta_i \in \mathcal{A}_j\}}\right) \\
&= (\alpha + n)\left[\left(\frac{\alpha}{\alpha + n}\right) G_0(\mathcal{A}_j) + \left(\frac{n}{\alpha + n}\right) \frac{1}{n} \sum_{i=1}^{n} \delta_{\theta_i}\right], \qquad (5.7)
\end{aligned}$$

where $\delta_{\theta_i}(\mathcal{A}_j) = 1_{\{\theta_i \in \mathcal{A}_j\}}$ is the point mass (or degenerated distribution) at observed $\theta_i \in \mathcal{A}_j$.

Hence the concentration parameter α' and the base probability measure $G_0(\mathcal{A}_j)'$ in the posterior distribution of the DP are updated as:

$$\alpha' = \alpha + n, \qquad (5.8)$$

$$G_0(\mathcal{A}_j)' = \left(\frac{\alpha}{\alpha + n}\right) G_0(\mathcal{A}_j) + \left(\frac{n}{\alpha + n}\right) \frac{1}{n} \sum_{i=1}^{n} \delta_{\theta_i}, \qquad (5.9)$$

In Equation (5.8), as n is the sample size of observed data, in this sense α can be interpreted as the prior sample size.

Now Equation (5.6) can be expressed in terms of $\alpha^{'}$ and $G_0(\mathcal{A}_j)^{'}$,

$$\{G(\mathcal{A}_1), \cdots, G(\mathcal{A}_k)\} \mid \theta_1, \cdots, \theta_n \sim Dirichlet\left(\alpha^{'} G_0(\mathcal{A}_1)^{'}, \cdots, \alpha^{'} G_0(\mathcal{A}_k)^{'}\right), \tag{5.10}$$

By Equation (5.3), the posterior mean of $G(\mathcal{A}_j)$ is:

$$\mathrm{E}\big(G(\mathcal{A}_j) \mid \theta_1, \cdots, \theta_n\big) = G_0(\mathcal{A}_j)^{'} = \left(\frac{\alpha}{\alpha + n}\right) G_0(\mathcal{A}_j) + \left(\frac{n}{\alpha + n}\right) \frac{1}{n} \sum_{i=1}^{n} \delta_{\theta_i}(\mathcal{A}_j), \tag{5.11}$$

It is noted that the posterior mean is a weighted average of the prior base probability measure $G_0(\mathcal{A}_j)$ with a weight proportional to α and the empirical measure $\frac{1}{n}\sum_{i=1}^{n}\delta_{\theta_i}$ with weight proportional to the sample size of observed data n. In cases when $\alpha \to 0$ or the sample size n increases, the posterior mean will be dominated by the empirical measure. In addition, as $n \to \infty$, the posterior mean will converge to the empirical distribution.

5.1.3 Predictive Probability

Let $\tilde{\theta}_{n+1}$ be the yet to be observed data, the predictive probability of $\tilde{\theta}_{n+1} \in \mathcal{A}_j$ where $\mathcal{A}_j$ is a measurable set, given observed data points $\theta_1, \cdots, \theta_n$, can be obtained by marginalizing G out:

$$\begin{aligned} P(\tilde{\theta}_{n+1} \in \mathcal{A}_j \mid \theta_1, \cdots, \theta_n) &= \int_G P(\tilde{\theta}_{n+1}, G(\mathcal{A}_j) \mid \theta_1, \cdots, \theta_n) dG(\mathcal{A}_j) \\ &= \int_G P(\tilde{\theta}_{n+1} \mid G(\mathcal{A}_j), \theta_1, \cdots, \theta_n) P(G(\mathcal{A}_j) \mid \theta_1, \cdots, \theta_n) dG(\mathcal{A}_j), \end{aligned} \tag{5.12}$$

Assuming conditional independence of $\tilde{\theta}_{n+1}$ and $\theta_1, \cdots, \theta_n$ given $G(\mathcal{A}_j)$, Equation (5.12) becomes:

$$\begin{aligned}
P(\tilde{\theta}_{n+1} \in \mathcal{A}_j \mid \theta_1, \cdots, \theta_n) &= \int_G P(\tilde{\theta}_{n+1} \mid G(\mathcal{A}_j), \theta_1, \cdots, \theta_n) P(G(\mathcal{A}_j) \mid \theta_1, \cdots, \theta_n) dG(\mathcal{A}_j) \\
&= \int_G P(\tilde{\theta}_{n+1} \mid G(\mathcal{A}_j)) P(G(\mathcal{A}_j) \mid \theta_1, \cdots, \theta_n) dG(\mathcal{A}_j) \\
&= \int_G G(\mathcal{A}_j) P(G(\mathcal{A}_j) \mid \theta_1, \cdots, \theta_n) dG(\mathcal{A}_j) \\
&= \mathrm{E}\big(G(\mathcal{A}_j) \mid \theta_1, \cdots, \theta_n\big) \\
&= \left(\frac{\alpha}{\alpha + n}\right) G_0(\mathcal{A}_j) + \left(\frac{n}{\alpha + n}\right) \frac{1}{n} \sum_{i=1}^{n} \delta_{\theta_i}(\mathcal{A}_j) \\
&= G_0(\mathcal{A}_j)^{'}, \qquad (5.13)
\end{aligned}$$

We see that the predictive probability of $\tilde{\theta}_{n+1}$ is also the posterior base probability distribution $G_0(\mathcal{A}_j)^{'}$, which is a mixture of prior base measure G_0 and the point masses at the previously observed data points $\theta_1, \cdots, \theta_n$. The probability of $\tilde{\theta}_{n+1}$ taking one of the previously observed data points $\theta_1, \cdots, \theta_n$ is $\frac{n}{\alpha+n}$, which approaches to 1 as $n \to \infty$ or $\alpha \to 0$. That means, it is very likely that the yet to be observed data repeats the previously observed data. This suggests the discreteness property of samples drawn from a DP, which leads to the use of DP for clustering. Imagine each measurable set $\mathcal{A}_j$ is a cluster, larger a cluster is, higher the probability the next data point will be sampled from the cluster, so the probability the cluster is getting bigger is also higher, which is known as rich-gets-richer property of DP. Nevertheless, there is still a probability of $\frac{\alpha}{\alpha+n}$ to pick a data point from a new cluster, implying that there is always some chance to start a new distribution.

Next we describe two common metaphors of DP, Blackwell-MacQueen Urn scheme and Chinese restaurant process.

5.1.4 Blackwell-MacQueen Urn Scheme

Blackwell and MacQueen (1973) introduced Blackwell-MacQueen urn scheme that satisfies properties of DP. As a generalized form of Pólya urn scheme, it can be used to justify the definition of DP as well as to interpret the predictive distribution of DP in Equation (5.13). Imagine there is a non-transparent urn with α black (non-color) balls. Let Θ be the color set with non-black colors, $\mathcal{A}_j \subset \Theta$, $j = 1, ..., k$, be the unique non-black color groups (for example, $\mathcal{A}_1 = \{\text{red}\}$), and G_0 the base probability measure on $(\Theta, \mathcal{A})$. The scheme is as follows: If we draw a black ball from the urn, we put the ball back along with a new non-black ball randomly drawn from G_0; If we draw a non-black ball from the urn, then put the ball back along with another ball with the same color. Let θ_i be a latent variable indicating which color the i^{th} ball is, $i = 1, \cdots, n$. We are interested in the probability of drawing a ball of color $\mathcal{A}_j$ in the next draw.

For the 1^{st} draw, since the urn contains α black balls initially, we always get black balls, so we draw a new ball with non-black color $\theta_1 \sim G_0$ and put it back to the urn along with the black ball.

For the 2^{nd} draw, the urn contains $(\alpha + 1)$ balls including α black balls and one ball with non-black color θ_1. With probability $\frac{\alpha}{\alpha+1}$, we draw a black ball and draw a new non-black ball with color $\theta_2 \sim G_0$ and put it back into the urn along with the black ball, or with probability $\frac{1}{\alpha+1}$ we draw a ball of color $\theta_2 = \theta_1$, pick a new ball with the same color and put both back into the urn.

Continuing in this way to the $(n+1)^{th}$ draw, the urn has a total of $(\alpha+n)$ balls including α black balls and n non-black balls. With probability $\frac{\alpha}{\alpha+n}$, we draw a black ball and a new ball with color $\theta_{n+1} \sim G_0$ and put both back into the urn, or with probability $\frac{n}{\alpha+n}$ we draw a ball from the empirical distribution $\frac{1}{n}\sum_{i=1}^{n}\delta_{\theta_i}(\mathcal{A}_j)$, pick a new ball with the same color and put both back into the urn.

In this manner, we can draw each θ_i given $\theta_1, \cdots, \theta_{i-1}$ until the urn has a total of n balls. Based on the conditional distribution, we have the joint distribution of $\theta_1, \cdots, \theta_n$ as,

$$\begin{aligned} P(\theta_1, \cdots, \theta_n) &= P(\theta_1)P(\theta_2 \mid \theta_1)\cdots P(\theta_n \mid \theta_1, \theta_2, \cdots, \theta_{n-1}) \\ &= P(\theta_1)\prod_{i=2}^{n} P(\theta_i \mid \theta_1, \cdots, \theta_{i-1}), \end{aligned} \tag{5.14}$$

Note that the joint distribution in Equation (5.14) is invariant to any finite permutation of the sequence $\{\theta_1, \theta_2, \cdots, \theta_n\}$ and hence the sequence itself is exchangeable. By de Finetti's theorem, for an exchangeable sequence there exists a random probability measure G where all elements in the sequence are independently and identically distributed from,

$$P(\theta_1, \cdots, \theta_n) = \int (G(\theta_1)\cdots G(\theta_n))\, dP(G) = \int \left(\prod_{i=1}^{n} G(\theta_i)\right) dP(G), \tag{5.15}$$

This prior distribution over distributions $P(G)$ is Dirichlet process $DP(\alpha, G_0)$.

5.1.5 Chinese Restaurant Process and Clustering Property

As aforementioned in predictive distribution, the discreteness property of DP can be further used to partition observations into groups (i.e., clustering). We will use Chinese restaurant process to illustrate this clustering property of DP. Chinese restaurant process is another popular representation of DP, and in this metaphor suppose there is a very large Chinese restau-

rant with infinite round tables and each table has no limit of seats. Let Θ be the table set, $\mathcal{A}_j \subset \Theta$, $j = 1, ..., k$, represent each table. Each customer entering the restaurant and sitting at a table is an observation, and can choose to sit at any of the occupied tables, or a currently unoccupied table (say new table). Assume that a total of n customers already sit at a total of m tables in the restaurant, with number of customers at the j^{th} table denoted by n_j, $j = 1, \cdots, m$, and $\sum_{j=1}^{k} n_j = n$. Let θ_i be a latent variable indicating which table the i^{th} customer is sitting at with values taking from $\{1, \cdots, m\}$, $i = 1, \cdots, n$.

For the next $(n+1)^{th}$ customer entering the restaurant, the predictive probability that the customer will sit at the j^{th} table is proportional to the number of customers already sit at that table (n_j),

$$P(\tilde{\theta}_{n+1} = j \mid \theta_1, \cdots, \theta_n) = \frac{n}{\alpha + n} \frac{1}{n} \sum_{i=1}^{n} \delta_{\theta_i}(\mathcal{A}_j) = \frac{n}{\alpha + n} \frac{1}{n} n_j = \frac{n_j}{\alpha + n}, \tag{5.16}$$

More people sits at the table, higher the probability the size of that table will grow. This is also the underlying reason for the rich-gets-richer characteristic of DP. When summing Equation (5.16) over all m currently occupied tables, we have

$$\sum_{j=1}^{m} P(\tilde{\theta}_{n+1} = j \mid \theta_1, \cdots, \theta_n) = \sum_{j=1}^{m} \frac{n_j}{\alpha + n} = \frac{\sum_{j=1}^{m} n_j}{\alpha + n} = \frac{n}{\alpha + n}, \tag{5.17}$$

With Equation (5.17), we see that there is non-zero probability that the customer will sit at a new table labeled as $m + 1$,

$$P(\tilde{\theta}_{n+1} = m + 1 \mid \theta_1, \cdots, \theta_n) = \frac{\alpha}{\alpha + n}, \tag{5.18}$$

Note that conceptually, Chinese restaurant process and Blackwell-MacQueen urn scheme are quite similar where tables and ball colors are both corresponding to clusters. The only subtle difference is that in the Blackwell-MacQueen urn scheme, along with a black ball drawn a new ball will be drawn randomly from the base distribution G_0, while the Chinese restaurant process does not have G_0 in place. This can be addressed by adding a dish variable associated with each table denoted by d_j where $d_j \sim G_0$.

Each occupied table $j = 1, \cdots, m$ can be considered as a cluster. As the number of existing tables m is always equal to less than the number of customers n, we wonder how fast the number of clusters increase as number of observations goes up. By Equation (5.18), the probability of adding more new table (cluster) when the i^{th} customer enters the restaurant given there are $i-1$ customers in the restaurant is $\frac{\alpha}{\alpha+i-1}$, based on which we can get expected m given n,

$$\mathrm{E}(m \mid n) = \sum_{i=1}^{n} \frac{\alpha}{\alpha + i - 1}, \tag{5.19}$$

Equation (5.19) can be approximated as $\alpha \log(1 + \frac{n}{\alpha})$ (Teh, 2010). We see that the expected number of clusters grows slowly as a logarithmic function of n, which is consistent with the rich-gets-richer property. Also, α can be interpreted as the prior number of clusters.

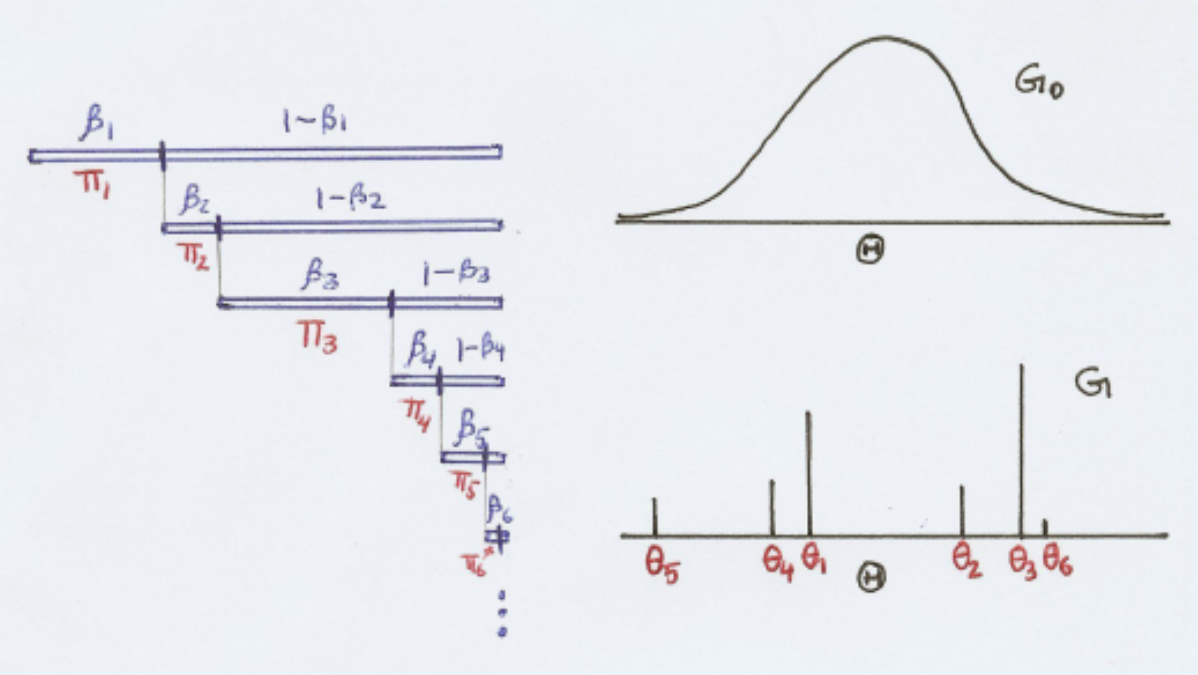
5.1.6 Stick-Breaking Construction of Dirichlet Process

Sethuraman (1994) described stick-breaking method in a simple and straightforward way to construct a DP. In this metaphor, suppose there is a stick of length 1 unit. Let $\beta_k \sim Beta(1, \alpha)$ be the fraction being taken away from the reminder of the stick, and π_k the length (or weight) of the part being taken away, at the k^{th} iteration, $k = 1, \cdots$. At the 1^{st} iteration, take β_1 away, $\pi_1 = \beta_1$. At the 2^{nd} iteration, take β_2 away from the reminder $1-\beta_1$, $\pi_2 = \beta_2(1-\beta_1)$. Continue

to the K^{th} iteration, take β_K away from the reminder $\prod_{k=1}^{K-1}(1-\beta_k)$, $\pi_K = \beta_K \prod_{k=1}^{K-1}(1-\beta_k)$. When the number of iterations becomes very large, we have $\sum_{k=1}^{\infty} \pi_k \to 1$ almost surely. Given $\theta_k \overset{iid}{\sim} G_0(\cdot)$, we can construct samples of G from a DP as:

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k} = \sum_{k=1}^{\infty} \left[\beta_k \prod_{j=1}^{k-1} (1-\beta_j) \right] \delta_{\theta_k}, \tag{5.20}$$

where $\delta_{\theta_k} = 1_{\{\theta_k\}}$ is the point mass (or degenerated distribution) located at θ_k.



Source: Ghahramani (2005)

Figure 15. Stick-breaking representation of Dirichlet process

Using stick-breaking method, the procedure to obtain $G \sim DP(\alpha, G_0)$ is as follows:

- Generate $\beta_1, \cdots, \beta_K \overset{iid}{\sim} Beta(1, \alpha)$,
- Calculate the corresponding weights as $\pi_k = \beta_k \prod_{j=1}^{k-1}(1-\beta_j)$, $k = 1, \cdots, K$,
- Generate $\theta_k \overset{iid}{\sim} G_0$, $k = 1, \cdots, K$,
- Define $G = \sum_{k=1}^{K} \pi_k \delta_{\theta_k} \sim DP(\alpha, G_0)$.

Figure 15 shows a graphic depiction of stick-breaking process (Ghahramani, 2005). Note that as the stick-breaking process continues, the weights will decrease to very small numbers that can be neglected. This suggests truncation of stick-breaking as an approximation of DP using a finite number of clusters with details provided in next section.

5.2 Dirichlet Process Mixture Model Specification

5.2.1 Dirichlet Process Mixture Model

One most important and profound application of Dirichlet process is for clustering analysis using mixture model (Antoniak, 1974; Neal, 1992; Rasmussen, 2000). Let $\boldsymbol{y} = \{y_1, \cdots, y_n\}$ be a set of independently observed data points, a general Dirichlet process mixture (DPM) model is given by

$$\begin{aligned} & y_i \mid \theta_i, \boldsymbol{\phi} \sim f(y_i \mid \theta_i, \boldsymbol{\phi}),\ i = 1, \cdots, n; \\ & \theta_1, \cdots, \theta_n \mid G \overset{iid}{\sim} G, \\ & G \mid \alpha, G_0 \sim DP(\alpha, G_0(\cdot \mid \boldsymbol{\kappa})), \\ & (\boldsymbol{\phi}, \boldsymbol{\kappa}, \alpha) \sim \mathcal{P} \end{aligned} \tag{5.21}$$

where $\boldsymbol{\phi}$ and $\boldsymbol{\kappa}$ denote the vectors of unknown parameters associated with the distribution of y_i and base probability measure G_0, respectively, and $\theta_1, \cdots, \theta_n$ are the latent parameters that

are conditionally independent given G. Note that as G is a discrete distribution, different θ_i's may have the same values. Then we complete the Bayesian mixture model by assuming that the model parameter vector $\boldsymbol{\phi}$, hyperparameter vector $\boldsymbol{\kappa}$ and the concentration parameter α follow a parametric distribution $\mathcal{P}$.

As aforementioned in stick-Breaking construction of Dirichlet process, truncated stick-breaking process can be an approximation of DP. Let K be the total number of clusters, $y_1, \cdots, y_n$ a set of observed data points, $s_1, \cdots, s_n$ the latent variable associated with each y_i indicating which cluster y_i is assigned to, a finite mixture model using truncated stick-breaking is given by

$$
\begin{aligned}
&\beta_k \sim Beta(1, \alpha), \text{ for } k = 1, \cdots, K-1, \\
&\pi_k = \beta_k \prod_{j=1}^{k-1} (1 - \beta_j), \text{ for } k = 1, \cdots, K-1, \\
&\pi_K = 1 - \sum_{k=1}^{K-1} \pi_k, \\
&\theta_k \sim G_0, \text{ for } k = 1, \cdots, K, \\
&G = \sum_{k=1}^{K} \pi_k \delta_{\theta_k}
\end{aligned}
\tag{5.22}
$$

where G has an approximate DP.

Figures 16a and 16b present the graphic models for a DPM model and a finite DPM model using stick-breaking truncation, respectively.

5.2.2 Blocked Gibbs Sampler Using Stick-Breaking Truncation

Introduced by Ishwaran and James (2001, 2002), the blocked Gibbs sampling method using stick-breaking truncation as an approximation to DP provides a straightforward MCMC

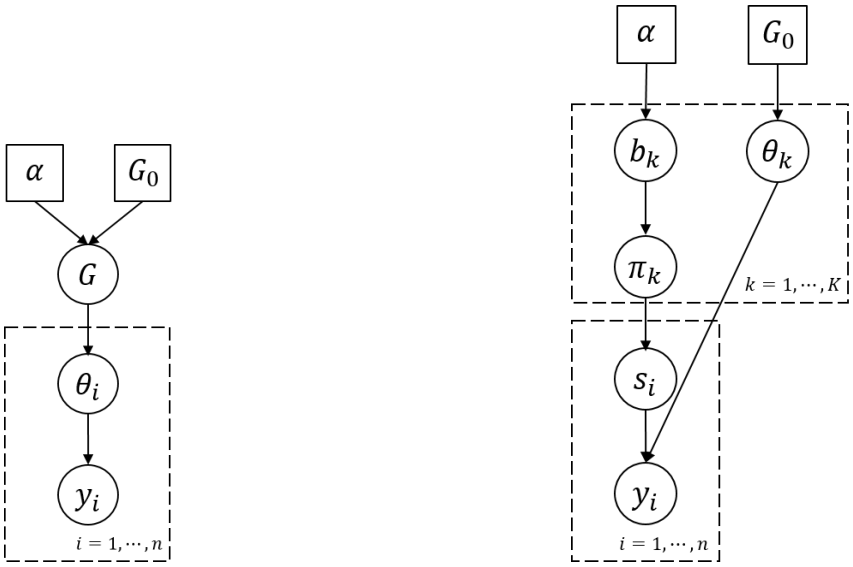


Figure 16. Graphic illustrations of Dirichlet process mixture (DPM) models

(a) A general DPM model. (b) A finite DPM model using truncated stick-breaking process. Rectangles represent replications of the model, with number of replicates shown in the bottom right corner.

algorithm to obtain the non-parametric posterior sampling distributions. The blocked Gibbs sampling approach does not require marginalization over the prior distribution of the random probability measure G. The finite mixture model with K clusters is given by

$$
\begin{aligned}
& y_i \mid \boldsymbol{\theta}, s_i, \boldsymbol{\phi} \overset{ind}{\sim} f(y_i \mid \theta_{s_i}, \boldsymbol{\phi}),\ i = 1, \cdots, n; \\
& \boldsymbol{\theta} = \{\theta_1, \cdots, \theta_K\} \overset{iid}{\sim} G_0(\cdot \mid \boldsymbol{\kappa}); \\
& s_i \mid \boldsymbol{\pi} \overset{iid}{\sim} \sum_{k=1}^{K} \pi_k \delta_{\theta_k} \equiv Multinomial(K, \boldsymbol{\pi}),
\end{aligned} \tag{5.23}
$$

where s_i is the latent indicator for which mixture component (cluster) y_i is assigned to, $s_i \in \{1, \cdots, K\}$, and $\boldsymbol{\pi} = \{\pi_1, \cdots, \pi_K\}$ are generated by the stick-breaking process in Equation (5.22). Note that $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ are independent.

Using the blocked Gibbs sampler, $\boldsymbol{\theta}$, $\boldsymbol{s}$ and $\boldsymbol{\pi}$ are updated in multivariate blocks. The full conditional distributions of $(\boldsymbol{\theta} \mid \boldsymbol{s}, \boldsymbol{\phi}, \boldsymbol{\kappa}, \boldsymbol{y})$ and $(\boldsymbol{s} \mid \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\kappa}, \boldsymbol{y})$ and $(\boldsymbol{\pi} \mid \boldsymbol{s}, \alpha)$ are given in Ishwaran and James (2001) Section 5.2. Let $\{s_1^*, \cdots, s_m^*\}$ denote the m current unique values of $\boldsymbol{s}$, the blocked Gibbs sampler algorithm draws samples in the following order

1. Draw $\theta_k \overset{iid}{\sim} G_0(\cdot \mid \boldsymbol{\kappa})$, for $k \notin \{s_1^*, \cdots, s_m^*\}$. Also draw $\theta_{s_j^*}$ from

$$\theta_{s_j^*} \overset{iid}{\sim} P(\theta_{s_j^*} \mid \boldsymbol{s}, \boldsymbol{\theta}, \boldsymbol{y}) \propto G_0(\cdot \mid \boldsymbol{\kappa}) \prod_{i:s_i=s_j^*} P(y_i \mid \theta_{s_j^*}, \cdot),\ j = 1, \cdots, m,$$

2. Draw each s_i independently from

$$s_i \mid \cdot \overset{ind}{\sim} \sum_{k=1}^{K} \pi_{k,i} \delta_{\theta_k},\ i = 1, \cdots, n,$$

   where $\pi_{k,i} \propto \pi_k P(y_i \mid \theta_k)$, $k = 1, \cdots, K$.

3. Update $\boldsymbol{\pi}$ by $\pi_k = b_k \prod_{j=1}^{k-1}(1 - b_j)$, $k = 1, \cdots, K-1$, and $\pi_K = 1 - \sum_{k=1}^{K-1} \pi_k$, where $b_k \sim Beta(1 + n_k, \alpha + \sum_{j=k+1}^{K} n_j)$, $c = 1, \cdots, K-1$, where n_k is the number of s_i values that equal to k, that is, the sizes of each cluster in $s_1, \cdots, s_n$.

5.2.3 Model Specification

The Bayesian Dirichlet process mixture model used to analyze the FC data (using rs-fMRI) in LLD neuroimaging study is specified as follows.

Let $y_{ij}^{(F)}$ denote the FC measurement for the i^{th} link ($i = 1, \cdots, m$) of the j^{th} subject ($j = 1, \cdots, N$), the Bayesian hierarchical model using DPM is

$$\begin{aligned} y_{ij}^{(F)} &\sim N\big(\mu_{ij}, \sigma^2\big), \\ \mu_{ij} &= \beta_{1i} + \beta_{2i} Grp_j + \gamma_j, \\ \sigma^2 &\sim \text{Inverse-Gamma}(0.01, 0.01), \\ \gamma_j &\sim N(0, 0.004), \end{aligned} \tag{5.24}$$

where Grp_j is the indicator variable for which group the j^{th} subject is from, HC group if $Grp_j = 0$, or LLD group if $Grp_j = 1$, γ_j is the random subject effect for the j^{th} subject.

The parameter β_{1i} denotes the intercept term for the i^{th} link,

$$\begin{aligned} \beta_{1i} &\sim N(\mu_{\beta_1}, \sigma_{\beta_1}^2), \\ \mu_{\beta_1} &\sim N(0.138, 0.0308), \\ \sigma_{\beta_1}^2 &\sim \text{Inverse-Gamma}(4.1, 0.1), \end{aligned} \tag{5.25}$$

For subjects in HC group:

$$\mu_{ij} = \beta_{1i} + \gamma_j, \tag{5.26}$$

The parameter β_{2i} denotes the differential FC effect between LLD group and HC group for the i^{th} link. For subjects in LLD group:

$$\mu_{ij} = \beta_{1i} + \beta_{2i} + \gamma_j, \tag{5.27}$$

Under the null hypothesis of no difference in FC between LLD group and HC group for the i^{th} link, $\beta_{2i} = 0$. And we assume the majority of the links are under null hypothesis (i.e., null links). To mimic the hypothesis testing framework, $H_{0,i}$: $\beta_{2,i} = 0$ vs. $H_{1,i}$: $\beta_{2,i} \neq 0$, we use a spike-and-slab mixture prior (Mitchell and Beauchamp, 1988) to model β_{2i}.

$$\beta_{2i} \overset{iid}{\sim} w\delta_0(\cdot) + (1-w)G(\cdot), \tag{5.28}$$

where $\delta_0(\cdot)$ denotes a degenerate point mass at zero (the "spike component") indicating $H_{0,i}$ is true (i.e., $\beta_{2i} = 0$) and w is the prior probability that $H_{0,i}$ is true. We fix $w = 0.9$ as a reflection on "zero assumption" that most of the links are null cases. The "slab component" of the mixture $G(\cdot)$ represents the distribution of non-zero β_{2i} for a small proportion of links with differential FC between LLD group and HC group (i.e., alternative links). Spike-and-slab mixture model is a useful technique in Bayesian variable selection. As the "spike component" has a point mass at zero, variable selection depends on the posterior probability that the corresponding coefficients are selected to the "slab component". This prior helps reduce β_{2i} coefficients to zero for null links in which $\beta_{2i} = 0$ by making the posterior mean values small, and hence allows for a Bayesian stochastic search on the sparse but large dimensional model space of differential FCs for alternative links. The motivation of modeling $G(\cdot)$ using a non-parametric DP prior is considering that one normal model for β_{2i}'s might be inaccurate and unrealistic, specifying a

DP prior for $G(\cdot)$ will provide flexibility for all possible and measurable distributions of β_{2i} that are driven by data.

To approximate $G(\cdot)$ we use the truncated stick-breaking process based on finite mixture model with the following specifications:

$$b_k \sim Beta(1, \alpha), \text{ for } k = 1, \cdots, K-1, \tag{5.29}$$

$$\pi_k = b_k \prod_{j=1}^{k-1}(1 - b_j), \text{ for } k = 1, \cdots, K-1, \text{ and } \pi_K = 1 - \sum_{k=1}^{K-1} \pi_k \left(\text{so} \sum_{k=1}^{K} \pi_k = 1 \right), \tag{5.30}$$

$$\theta_k \overset{iid}{\sim} G_0 \equiv N(0, \sigma_\theta^2), \text{ for } k = 1, \cdots, K, \text{ where } \sigma_\theta^2 \sim \text{Inverse-Gamma}(3.2, 0.1), \tag{5.31}$$

$$G(\cdot) = \sum_{k=1}^{K} \pi_k \delta_{\theta_k} \sim DP(\alpha, G_0), \tag{5.32}$$

where K is the number of clusters. As Gelman et al. (2013) suggested that $K \leq 15$ can be sufficient number of clusters, we specify K in a range from 5 to 15. For this DP prior, we use the concentration parameter $\alpha = 1$, which implies that the prior sample size (or prior number of clusters) is 1.

To determine which cluster from $\{\theta_1, \cdots, \theta_K\}$ in Equation (5.31) β_{2i} is assigned to, we define a latent variable s_i associated with each β_{2i}, $s_i \in \{1, \cdots, K\}$. Given $\boldsymbol{\pi} = \{\pi_1, \cdots, \pi_K\}$ in Equation (5.30), we can draw s_i from $G(\cdot)$ by Equation (5.32)

$$s_i \mid \boldsymbol{\pi} \overset{ind}{\sim} G(\cdot) \equiv \sum_{k=1}^{K} \pi_k \delta_{\theta_k} \equiv Multinomial(K, \boldsymbol{\pi}),$$

$$\beta_{2i} = \theta_{s_i}. \tag{5.33}$$

CHAPTER 6

APPLICATION TO NEUROIMAGING STUDY IN LATE-LIFE DEPRESSION

In this chapter, we apply Bayesian multimodal Lfdr method as described in Chapter 4 as well as the semi-parametric Bayesian approach utilizing a non-parametric Dirichlet process mixture model as reviewed in Chapter 5 to the motivating multimodal neuroimaging study in LLD for identifying differential FC links and altered networks that may be potentially involved in pathology of LLD.

6.1 Bayesian Multimodal Local False Discovery Rate Method

6.1.1 Results

We performed the mixed-effects model analysis for FC (rs-fMRI) data in the LLD neuroimaging study with 10 LLD patients and 13 HC subjects described in Chapter 2. Figure 17 reveals the histogram of p-values from t-tests to compare mean FC measures between LLD group and HC group obtained from the mixed-effects model for all 3741 links. Out of the total of 3741 hypothesis tests, 281 (7.5%) tests have two-sided p-value ≤ 0.025. As mentioned in Section 4.1, determination of significance solely based on p-values will result in a substantial inflation in family-wise error rate. Under the null hypothesis of no FC difference between LLD group and HC group, p-values are expected be uniformly distributed in the range between 0 and 1. By contrast, the p-values under the alternative hypotheses will tend to be very close to 0. In Figure 17, we can see that the p-values in the range between 0.025 and 1 seem to have a more even (or uniform) distribution, and the p-values close to 0 such as in the range

between 0 and 0.025 have a relatively higher density compared to the rest. This suggests that there may be a small proportion of alternative connectivity links whose FC measurement differ significantly between LLD group and HC group, together with some potential false positives.

Figure 18 presents the histogram of the 3741 absolute test statistics for FC. The folded normal distribution with mean = 0 and standard deviation = 1.01 estimated using the 3584 absolute test statistics with p-value > 0.025 (95.8%) seems to be a reasonable fit for the connectivity links under null hypothesis, while the tail of the histogram for the potential alternative connectivity links can be fitted using a gamma distribution described in Section 4.6.1. Then we can obtain a conservative but unbiased estimate of proportion of null cases (Storey and Tibshirani, 2003), denoted by $\hat{\pi}_0$, as,

$$\hat{\pi}_0 = \frac{\sum_{i=1}^{m} I\{p\text{-value}_i > 0.025\}}{m(1 - 0.025)} = \frac{3584}{3741 \times 0.975} = 0.983. \tag{6.1}$$

In order to have a comprehensive understanding of the impact on results from different perspectives with regards to q-value in controlling FDR, we applied q-value approach (Storey, 2002) using R qvalue package and the results are illustrated in Figure 19. Figure 19a is a plot of the estimated overall proportion of true null hypotheses (denoted by $\hat{\pi}_0$) as a function of the tuning parameter λ, where the natural cubic spline function with degree of freedom 3 fits the data well and $\hat{\pi}_0 = 0.935$ suggesting the majority of the hypotheses are null that aligns with the "zero assumption". Figure 19b shows the number of significant links identified for each q-value. We notice that there are fewer significant links when q-value < 0.3, indicating that the q-value = 0.2 would be a very conservative choice. Figure 19c displays the relationship between q-values and the corresponding p-values, from which we can see the expected proportion of

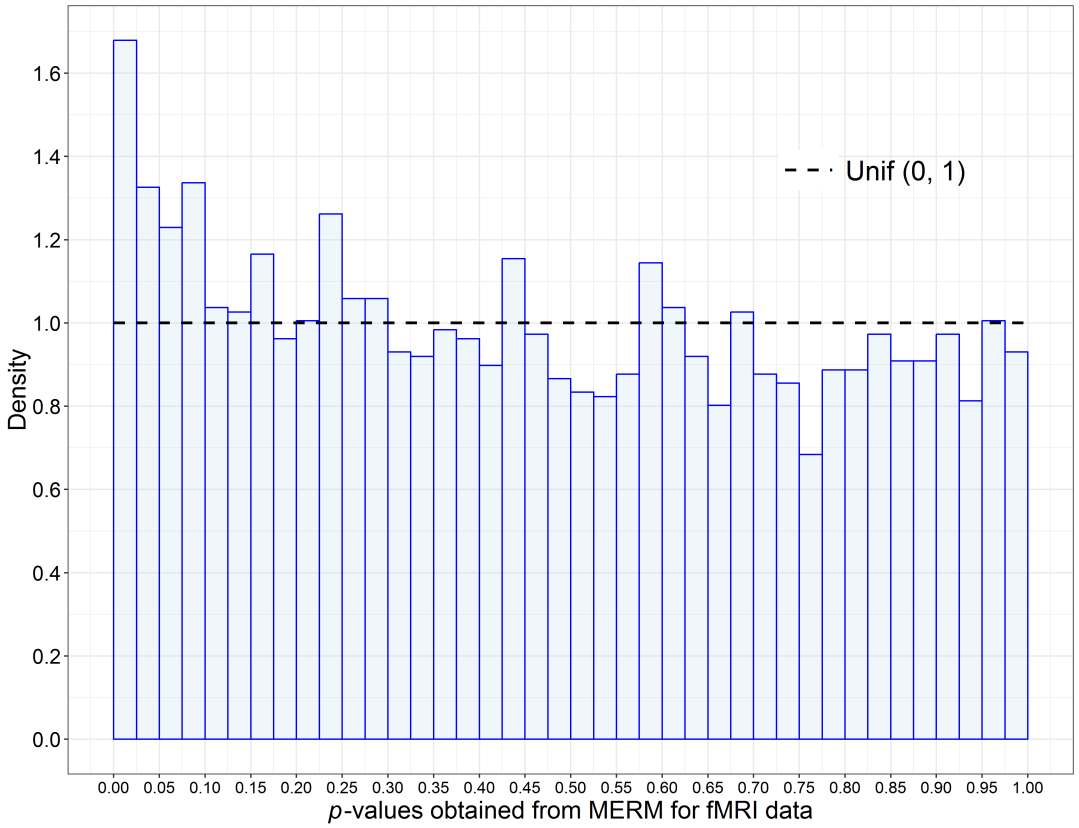


Figure 17. Histogram of the 3741 p-values for between-group comparison for FC (rs-fMRI) using the mixed-effects model in LLD neuroimaging study

The dashed line is the density of uniform distribution if all connectivity links were under null hypothesis (i.e., no differential FC).

false positives for different p-value cutoffs. Figure 19d shows the linear relationship between expected false positives and positives. Using q-value $= 0.2$ yields a small number of significant links (≈ 10), among which approximately 5 ($\approx \frac{5}{10} = 50\%$) are expected to be false positives.

Prior to applying Bayesian multimodal Lfdr method to control FDR for large-scale multiple testing, we first conducted the mixed-effects model analysis on the SC data measured by DTI and obtained the absolute test statistics comparing mean SC measures between LLD group and HC group for all 3741 connectivity links, which are then incorporated into the Bayesian

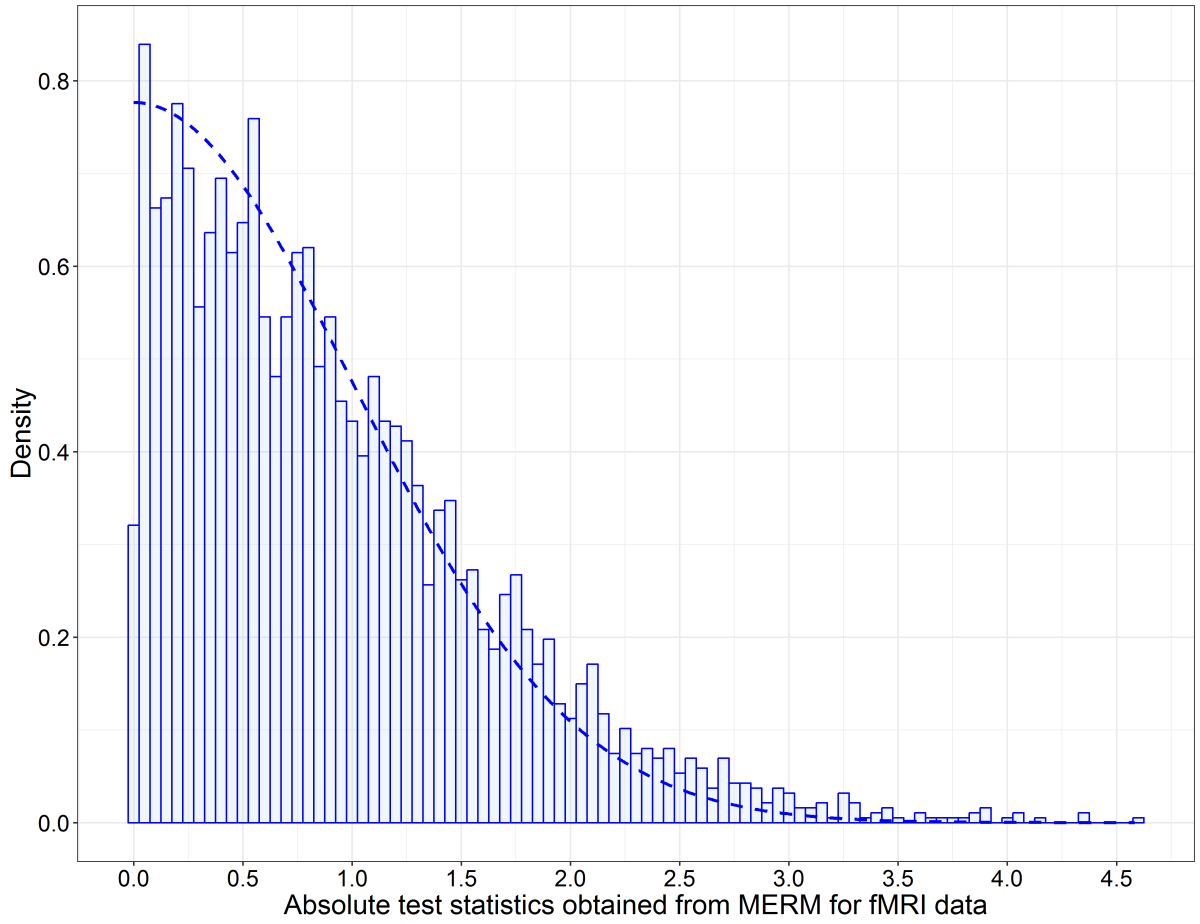


Figure 18. Histogram of the 3741 absolute test statistics from the mixed-effects model for FC (rs-fMRI) using the mixed-effects model in LLD neuroimaging study

The dashed blue line is the weighted null density $\hat{\pi}_0 \hat{f}_0\left(t_i^{(F)}\right)$, where $\hat{\pi}_0 = 0.983$ and $f_0\left(t_i^{(F)}\right)$ is the estimated density of folded normal distribution with mean = 0 and standard deviation = 1.01. Results are approximated using the 3584 absolute test statistics with p-value > 0.025 (95.8%) assuming these FC links were under null hypothesis.

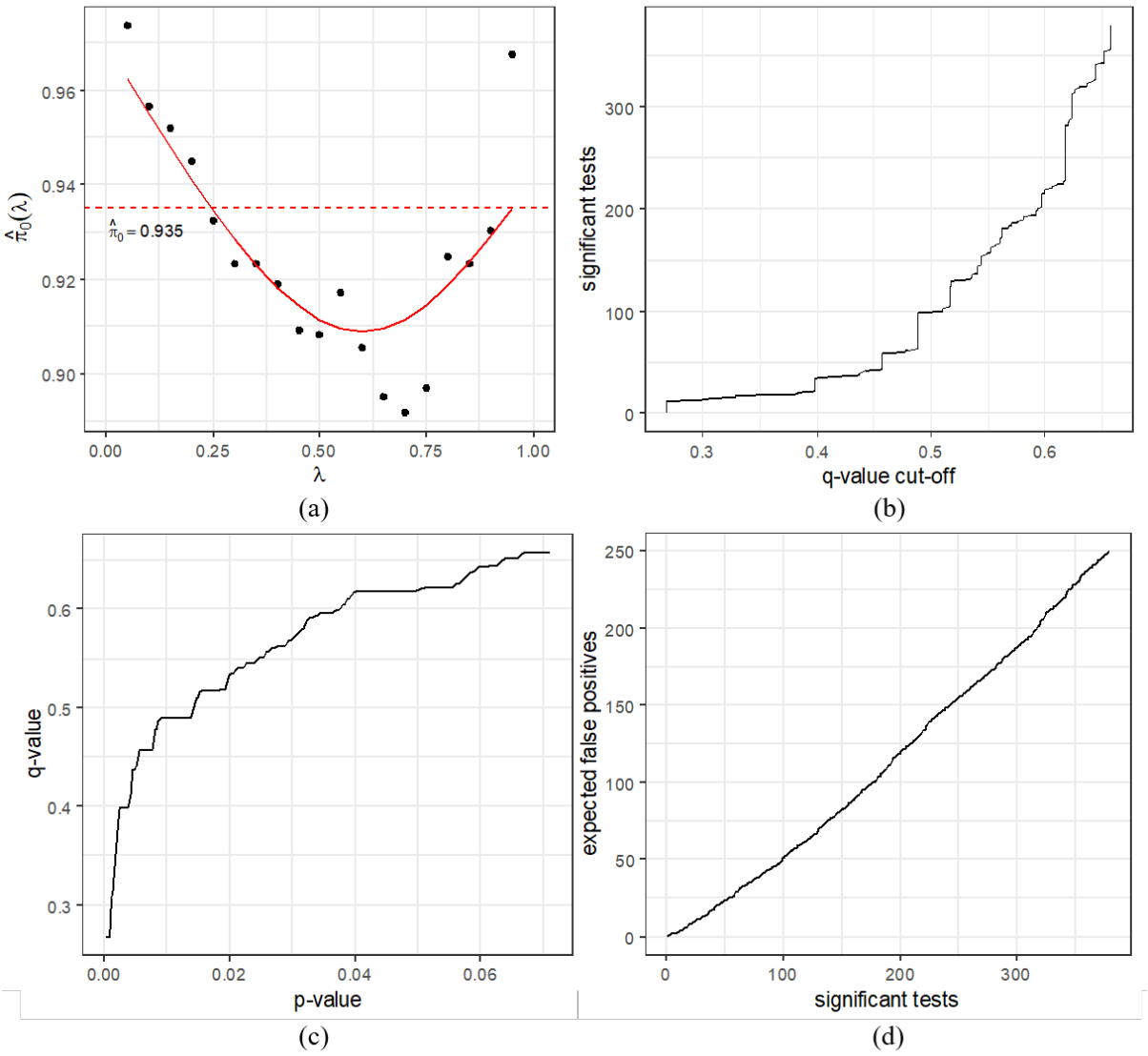


Figure 19. Results from the FC data analysis in LLD neuroimaging study using q-value approach (Storey, 2002)

(a) The estimated overall proportion of true null hypothesis (denoted by $\hat{\pi}_0$) as a function of the tuning parameter λ, where the red solid line indicates a natural cubic spline function with degree of freedom set to 3.

(b) The number of significant links as a function of q-value cutoff.

(c) The q-values vs. the corresponding p-values.

(d) The expected false positives as a linear function of positives.

mixture model. Figure 20 shows the scatter plot of the absolute test statistics for the between-group difference in SC vs. for FC from the mixed-effects model for all 3741 connectivity links. Using critical value $t_{1-0.025,\nu}$, where $\nu = 23 - 2 = 21$ is the degree of freedom, as the cutoff to determine significance in a naive way, out of 3741 connectivity links, 3334 (89.1%) are considered non-significant for both FC and SC (shown in blue), 271 (7.2%) are significant only for FC measures (shown in orange), 126 (3.4%) are significant only for SC measures (shown in green), and 10 (0.3%) are considered significant for both FC and SC connectivity measures (shown in red). We do not observe any simple correlation pattern between the absolute test statistics for between-group differences in FC and SC measures.

Using the prior distribution for model parameters specified in Section 4.7.2, we run three Markov chains in parallel where each chain has 325,000 iterations with the first 25,000 iterations of each chain discarded as burn-in and a thinning interval of 100 after burn-in, so the total number of iterations is 3,000 for each chain. We have tried different thinning intervals of 10, 50 and 100, and found that by taking every 100^{th} iteration the autocorrelation of the posterior samples for each model parameter is reduced to an acceptable level. The acceptance rates for $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ using the multiple-try Metropolis within Gibbs algorithm are approximately 38% and 30%, respectively, for all three chains. As a rule of thumb, the desired acceptance rate range is between 25% and 45%. Thus it suggests that the algorithm is efficient.

We perform model convergence diagnostics by checking mixing and stationarity (Gelman et al., 2013). To assess the mixing of the three Markov chains, the difference between the chains are analyzed based on the Gelman-Rubin convergence diagnostics (Gelman and Rubin, 1992; Brooks and Gelman, 1997). Potential scale reduction factor, denoted by $\hat{R}$, is used to compares the between-chain and within-chain variance estimates for each parameter. As shown

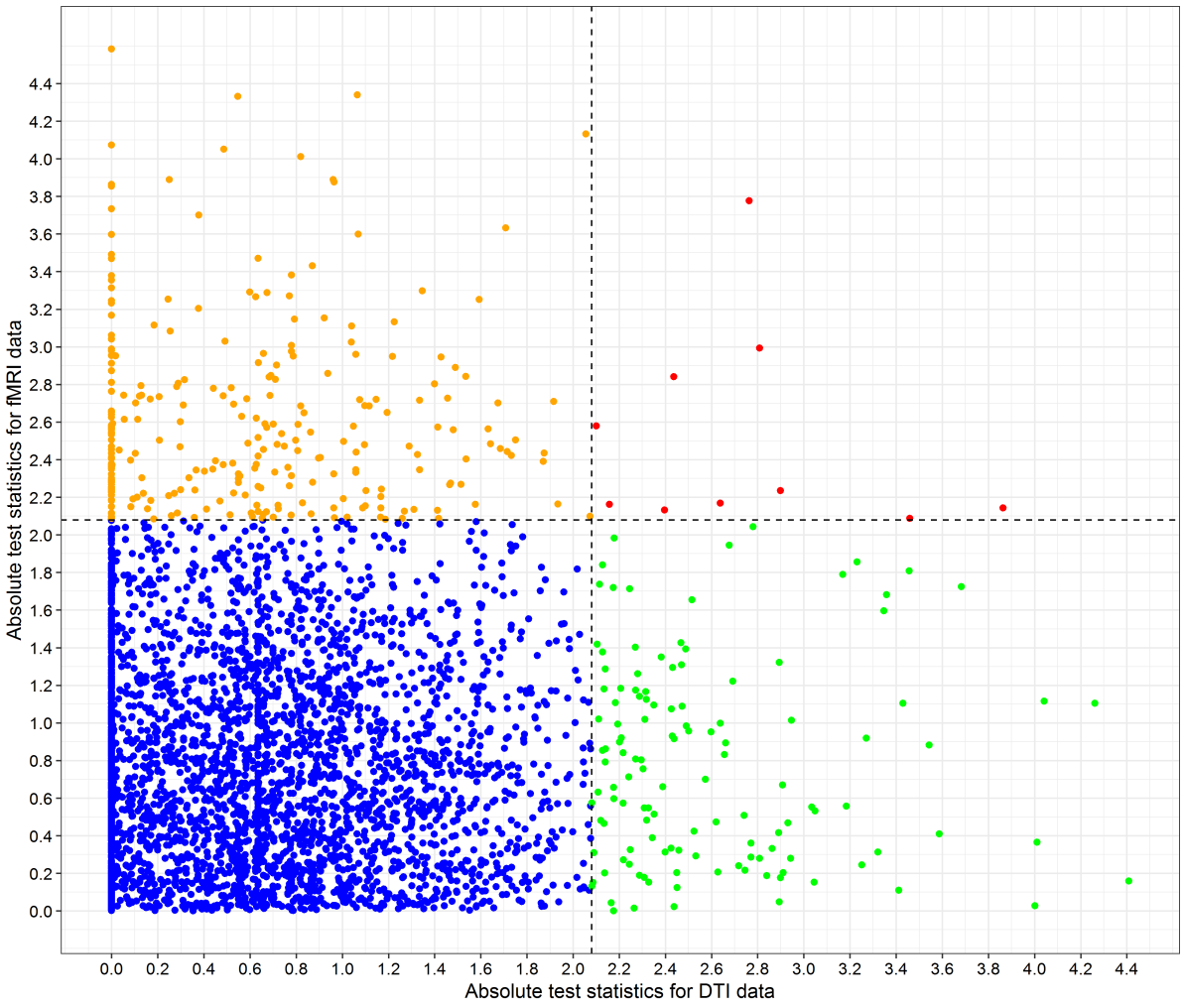


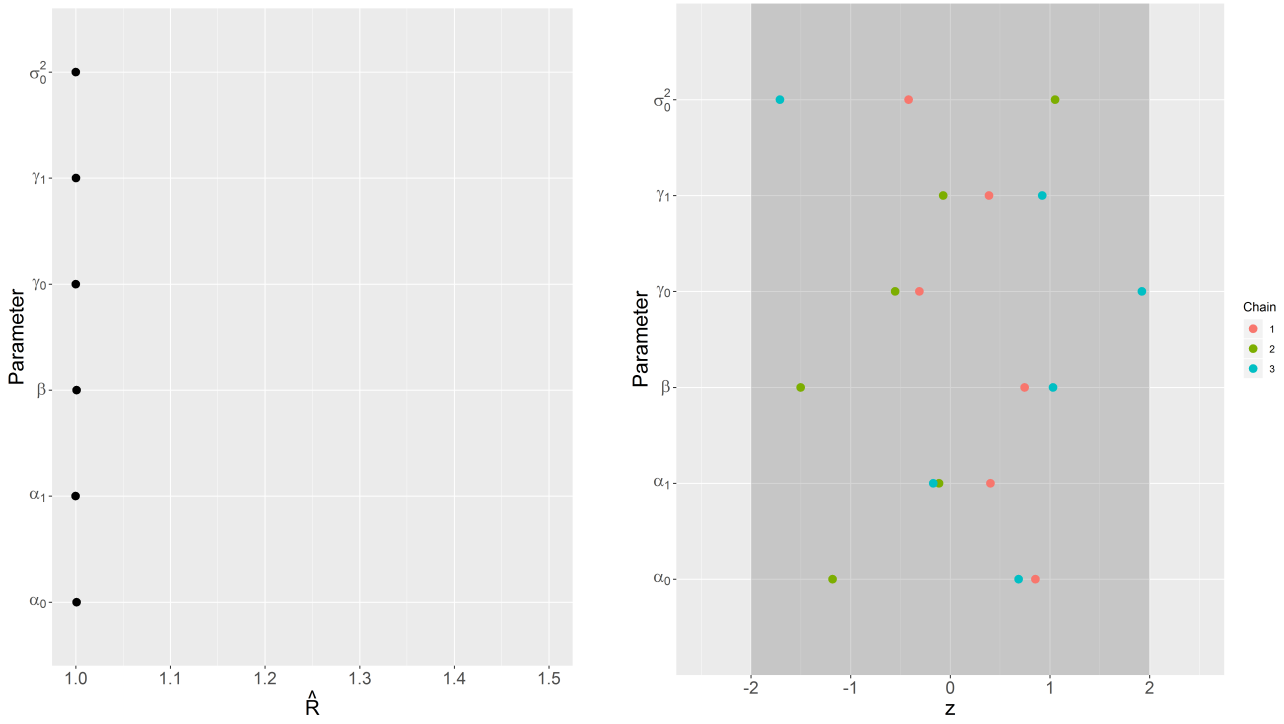
Figure 20. Scatter plot of the absolute test statistics for the between-group difference in SC (DTI) vs. for FC (rs-fMRI) from the mixed-effects model for all 3741 connectivity links in LLD neuroimaging study

The dashed lines indicate the critical value $t_{1-0.025,\nu}$, where $\nu = 23 - 2 = 21$ is the degree of freedom.

in Figure 21a, $\hat{R}$ for all model parameters are very close to 1.0 indicating that the three chains have converged to the target posterior distribution. Geweke z-score diagnostics (Geweke, 1992) provides a formal check of stationarity by comparing the means between the first 10% and the last 50% of the chain using a modified z-score. In Figure 21b, z-scores for all model parameters by each chain are within the range of $[-2, 2]$, suggesting that the three chains have reached stationarity.

Figures 22, 23 and 24 display the overlapped trace, marginal density and autocorrelation plots of posterior draws of each parameter with different colors by chain, respectively. As illustrated in Figure 22, three chains have mixed well and each individual chain has achieved stationarity. Figure 23 shows that estimation of the posterior densities from these three chains are consistent and each chain has converged to a similar sample space. From Figure 24, lag one autocorrelations are equal to or slightly larger than 0.5 for α_0, β, and σ_0^2, and small for α_1, γ_0 and γ_1. We see the autocorrelations decrease quickly with respect to the lag and are near zero with lag ≥ 8.

By combining the three chains into one long chain with 9,000 iterations in total, we take the posterior medians as the posterior estimates of each parameter. The posterior inferences including medians and 95% credible intervals are presented in Table VII. The estimated variance of the null density ($\hat{\sigma}_0^2$) is 1.100. The negative coefficient estimate for SC test statistics $\hat{\gamma}_1 = -0.692$ indicates that the lower alternative probability (i.e., higher null probability) is associated with higher value of $t_i^{(S)}$. Based on the posterior estimates of the model parameters $\{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}}, \hat{\beta}, \hat{\sigma}_0^2\}$, we calculate the posterior probability of each individual link is null given the observed FC and SC test statistics, $\widehat{BLfdr}_i$, $i = 1, ..., 3741$, by Equation (4.65), and then determine which connectivity links to be rejected using the oracle procedure for FDR control (Sun and Cai,



(a) (b)

Figure 21. Plots for model convergence diagnostics

(a) The estimated potential scale reduction $\hat{R}$. (b) Geweke z-scores.

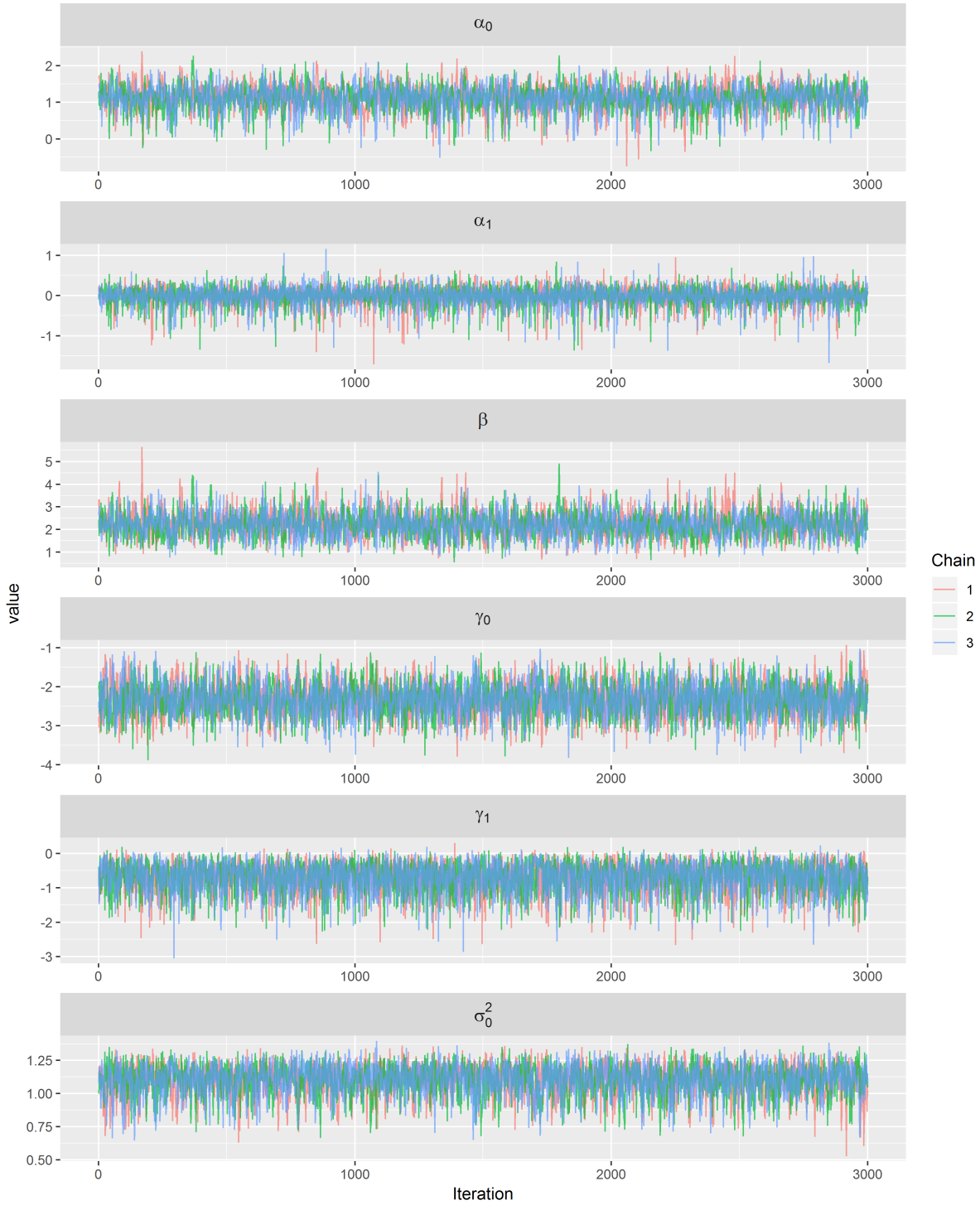


Figure 22. Trace plots of posterior draws of each parameter by chain

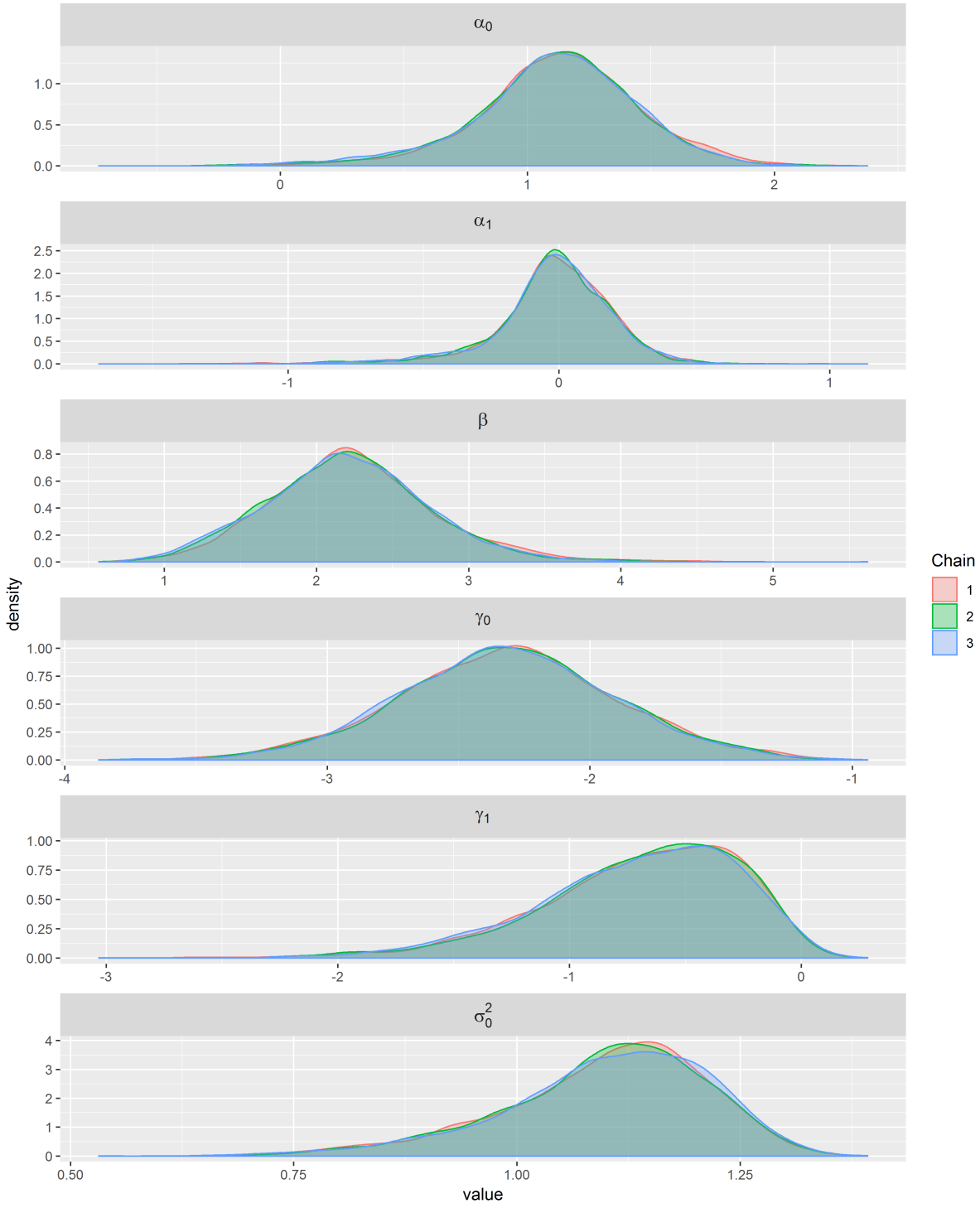


Figure 23. Marginal density plots of posterior draws of each parameter by chain

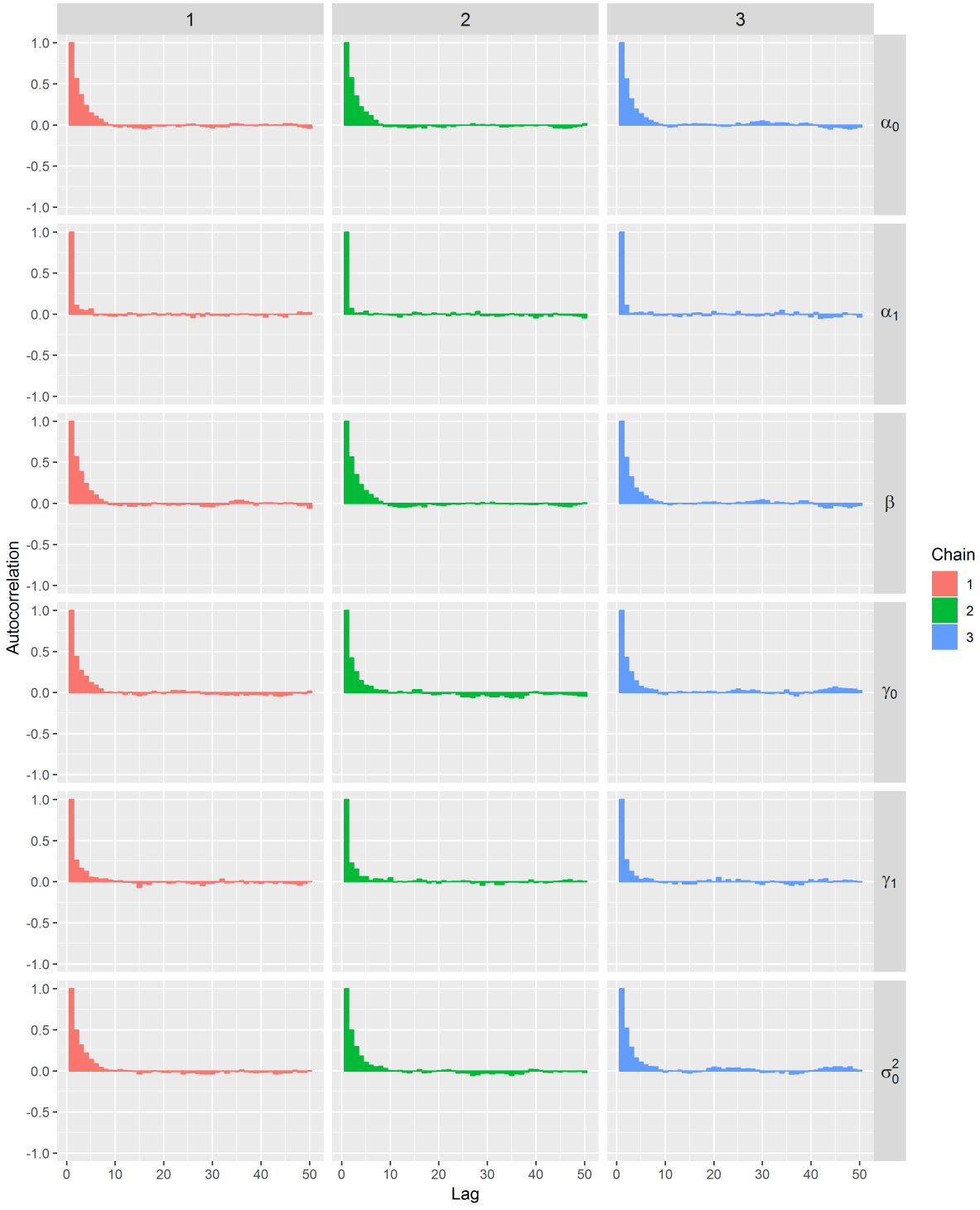


Figure 24. Autocorrelation plots of posterior draws of each parameter by chain

2007). For each $t_i^{(F)}$, $i = 1, \cdots, 3741$, the estimated weighted null and alternative densities as well as the mixture density based on the posterior estimates are presented in Figure 25.

TABLE VII
POSTERIOR INFERENCES FOR MODEL PARAMETERS

Parameter	Median (95% Credible Interval)[a]
α_0	1.120 (0.362, 1.739)
α_1	-0.018 (-0.548, 0.353)
β	2.214 (1.204, 3.359)
γ_0	-2.313 (-3.138, -1.484)
γ_1	-0.692 (-1.680, -0.055)
σ_0^2	1.100 (0.829, 1.283)

[a] Based on 9,000 posterior samples combined from three parallel Markov chains, where each chain has 325,000 iterations with the first 25,000 as burn-in and a thinning interval of 100 after burn-in.

Bayesian multimodal Lfdr method detected 21 connectivity links showing significant FC difference between LLD group ($n_1 = 10$) and HC group ($n_2 = 13$) at $q = 0.2$ presented in Table VIII, where 15 connectivity links have increased FC significantly (hyperconnectivities) and 6 connectivity links have decreased FC significantly (hypoconnectivities).

To assess the consistency of results between Bayesian multimodal Lfdr and Efron's Lfdr method, we compared the significant FC links identified using the two methods. Using Efron's Lfdr method by R locfdr package (Efron et al., 2015), the central matching estimates of the null distribution is $f_0 = N(\hat{\mu}_0 = 0.038, \hat{\sigma}_0^2 = 1.250)$ and the null proportion is $\hat{p} = 0.988$. The estimated null density has a variance of 1.250, which is larger as compared to 1.100 using Bayesian multimodal Lfdr method, implying that by Bayesian multimodal Lfdr, the estimated

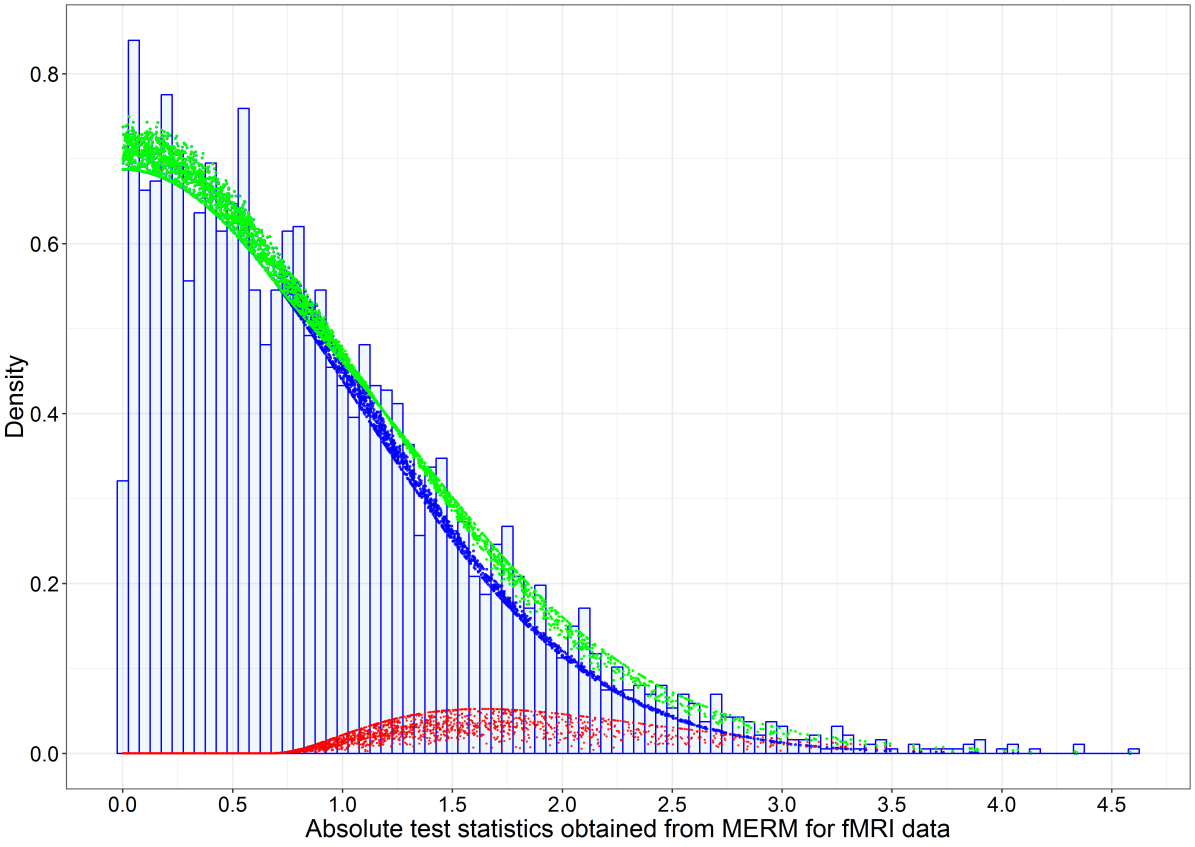


Figure 25. The estimated weighted null, weighted alternative and mixture densities based on the posterior estimates of model parameters for the 3741 absolute test statistics obtained from the mixed-effects model on FC (fMRI) data in LLD neuroimaging study

The blue dots are the weighted null density $\pi_0(t_i^{(S)}|\hat{\boldsymbol{\gamma}})f_0\big(t_i^{(F)} \mid \hat{\sigma}_0^2\big)$, the red dots are the weighted alternative density $\pi_1\big(t_i^{(S)} \mid \hat{\boldsymbol{\gamma}}\big)f_1\big(t_i^{(F)} \mid \hat{\boldsymbol{\alpha}}, \hat{\beta}, t_i^{(S)}\big)$, and green dots are the mixture density $\pi_0\big(t_i^{(S)} \mid \hat{\boldsymbol{\gamma}}\big)f_0\big(t_i^{(F)} \mid \hat{\sigma}_0^2\big)+\pi_1\big(t_i^{(S)} \mid \hat{\boldsymbol{\gamma}}\big)f_1\big(t_i^{(F)} \mid \hat{\boldsymbol{\alpha}}, \hat{\beta}, t_i^{(S)}\big)$, where $\{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}}, \hat{\beta}, \hat{\sigma}_0^2\}$ are the posterior estimates shown in Table VII.

alternative density with SC statistics incorporated has heavier tail. Efron's Lfdr method detects 12 significant FC links at $q = 0.2$, among which 11 links are overlapping with those identified by Bayesian multimodal Lfdr method. This is expected because the two method are based on the distributions of test statistics using Bayesian mixture model. The results suggest that Bayesian multimodal Lfdr method is more powerful as compared to Efron's Lfdr method as it yields more findings, which is also expected for it combines and borrows strength across multimodal neuroimaging data.

In addition, we compared the results with the bivariate linear mixed-effects model to fit FC data and SC data jointly based on the same LLD neuroimaging study (Zhao, 2014), which has identified 40 significant links at $q = 0.2$ using p-value based Benjamini and Hochberg FDR method. There are only 2 links (the left thalamus proper - the left posterior cingulate, and the left ventral diencephalon - the right caudal anterior cingulate) overlapping with those found by Bayesian multimodal method. The two methods have different purposes, since Bayesian multimodal Lfdr is intended to detect significant FC links using a model for FC statistics from a univariate mixed-effects model with leverage of complementary SC statistics, while the bivariate approach in Zhao (2014) utilized a joint modeling to identify connectivity links that are significant in both FC and SC. Therefore we would anticipate that the results from the two approaches are different with little overlap. Also we would like to point out two main limitations of the bivariate approach. First, it may be impractical given the inconsistent overall structure-function relationship (Koch et al., 2002; Honey et al., 2009). And second, it removed the excess zero values of SC data resulting in a substantial reduction in the total number of connectivity links included in the analysis from 3741 to 1433.

TABLE VIII
TWENTY-ONE DIFFERENTIAL FC LINKS AND THE CORRESPONDING TEST STATISTICS BY BAYESIAN MULTIMODAL LFDR METHOD AT FDR LEVEL OF 0.2 FOR LLD NEUROIMAGING STUDY

Region 1[a]	Region 2[a]	Test statistic[b]
Hyperconnectivities:		
R thalamus proper (RTP)	R caudal middle frontal (RCMF)	4.340 (E)
R pallidum (RP)	L inferior parietal (LIP)	4.011 (E)
R pallidum (RP)	R caudal middle frontal (RCMF)	4.051 (E)
R accumbens area (RAA)	L isthmus cingulate (LIC)	3.890 (E)
R ventral diencephalon (RVD)	R fusiform (RF)	4.132 (E)
L isthmus cingulate (LIC)	R caudal middle frontal (RCMF)	4.583 (E)
L posterior cingulate (LPC)	R rostral middle frontal (RRMF)	3.877 (E)
L posterior cingulate (LPC)	R supramarginal (RS)	3.378
L rostral middle frontal (LRMF)	R caudal middle frontal (RCMF)	4.074 (E)
L superior parietal (LSP)	R pars opercularis (RPO)	3.355
R caudal anterior cingulate (RCAC)	R caudal middle frontal (RCMF)	3.855 (E)
R caudal middle frontal (RCMF)	R isthmus cingulate (RIC)	3.733 (E)
R caudal middle frontal (RCMF)	R posterior cingulate (RPC)	3.598
R pars triangularis (RPT)	R rostral anterior cingulate (RRAC)	3.469
R posterior cingulate (RPC)	R supramarginal (RS)	3.491
Hypoconnectivities:		
L thalamus proper (LTP)	L posterior cingulate (LPC)	-3.701
L ventral diencephalon (LVD)	R caudal anterior cingulate (RCAC)	-3.889
R caudate (RCau)	L cuneus (LCun)	-3.245
L entorhinal (LE)	L supramarginal (LS)	-3.314
L fusiform (LF)	L pars triangularis (LPT)	-3.864
L fusiform (LF)	L supramarginal (LS)	-4.332 (E)

[a] L = left; R = right.
[b] Based on between-group comparison in FC, positive value indicates hyperconnectivity, negative value indicates hypoconnectivity. The 11 FC links also identified by Efron's Lfdr method at $q = 0.2$ are indicated using E in parentheses.

6.1.2 Discussion

Figure 26 displays the network analysis of the 21 connectivity links showing significant FC difference between LLD group and HC group by Bayesian multimodal Lfdr method at $q = 0.2$ based on cortical and subcortical gray matter regions by left and right hemisphere of the brain.

There are two main findings we have learned from the network analysis using Bayesian multimodal Lfdr method.

First, we identify a primary hub region, the right caudal middle frontal (also known as the right dorsolateral prefrontal cortex [dlPFC]) that has significantly increased FC to seven other regions including the left isthmus cingulate, the left rostral middle frontal, the right thalamus proper, the right pallidum, the right caudal anterior cingulate, the right isthmus cingulate and the right posterior cingulate in LLD patients as compared to the healthy controls.

The dlPFC (BA 9 and 46) is long known for its critical role within central executive network (CEN) for cognitive and complex mental processes including emotion modulation, selective attention and working memory. The dlPFC area has been implicated as the key neural substrates for MDD from the literature (Grimm et al., 2008; Koenigs and Grafman, 2009; Hamilton et al., 2012). During working memory task, healthy subjects show bilateral activation in dlPFC and anterior cingulate cortex, while depressed patients exhibit asymmetric activity in dlPFC where the left dlPFC shows increased activation as reported in a number of neuroimaging studies (Matsuo et al., 2007; Grimm et al., 2008; Sheline et al., 2009; Perrin et al., 2012). Alexopoulos et al. (2012) found lower FC in the left dlPFC area in LLD patients relative to healthy subjects in a rs-fMRI study. The left dlPFC has been the target site of repetitive transcranial magnetic stimulation (rTMS), a noninvasive procedure using magnetic field pulses to stimulate nerve cells that was approved by US Food and Drug Administration (FDA) in

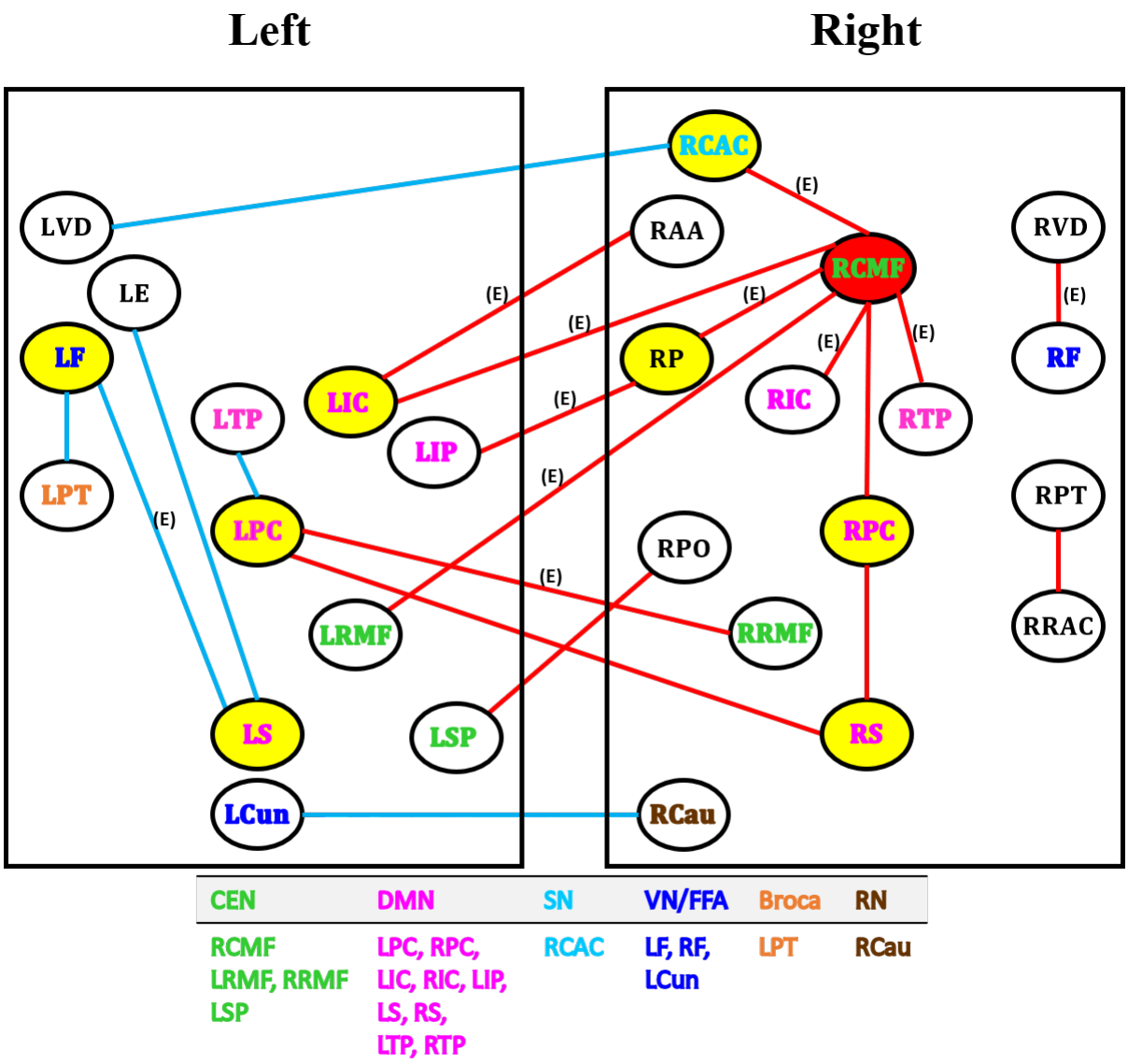


Figure 26. Network analysis of the 21 FC links by Bayesian multimodal Lfdr method at FDR level of 0.2 based on cortical and subcortical gray matter regions by left and right hemisphere of the brain

Red line indicates an increase in FC (hyperconnectivity) and blue line indicates a decrease in FC (hypoconnectivity). Red circle denotes the primary hub, and yellow circle denotes the secondary hub. The 11 FC links also identified by Efron's Lfdr method are indicated using E in parentheses.

Abbreviations: CEN = central executive network; DMN = default mode network; SN = salience network; VN/FFA = visual network / fusiform face area; Broca = Broca's area; RN = reward network; LCun = left cuneus; LE = left entorhinal; LF/RF = left/right fusiform; LIC/RIC = left/right isthmus cingulate; LIP = left inferior parietal; LPC/RPC = left/right posterior cingulate; LPT/RPT = left/right pars triangularis; LRMF/RRMF = left/right rostral middle frontal; LS/RS = left/right supramarginal; LSP = left superior parietal; LTP/RTP = left/right thalamus proper; LVD/RVD = left/right ventral diencephalon; RAA = right accumbens area; RCAC = right caudal anterior cingulate; RCMF = right caudal middle frontal; RCau = right caudate; RP = right pallidum; RPO = right pars opercularis; RRAC = right rostral anterior cingulate.

2008 as a treatment for medication resistant MDD (refer to FDA approval K061053). The clinical effectiveness of rTMS on the left dlPFC was established in randomized clinical trials for depression (George et al., 2010; Blumberger et al., 2018). By contrast, our analysis reveals evidence of the laterization of the right dlPFC with significantly increased FC and therefore suggests a specific and distinctive FC activation path via the right dlPFC in LLD patients. The right dlPFC was also found as a key hub of the altered FCs in chronic net pain patients who are at high risk to experience depression (Elbinoune et al., 2016; Ihara et al., 2019), and for failed back surgery syndrome patients (Kornelsen et al., 2013). Chronic pain is one of the most common co-occurring (i.e., comorbid) condition with depression in LLD patients (Aziz and Steffens, 2013) and thus, the right dlPFC with increase FCs is the key region involved in both cognition impairment and pain related to LLD.

We noticed an increased FC between the left superior parietal and the right pars opercularis in LLD. The left superior parietal is also part of CEN, and the right pars opercularis (BA44) is located within the right interior frontal gyrus. It is not clear about the function of the right pars opercularis. From the limited literature we have learnt that that the disrupted right pars opercularis may be related to speech inhibition in a case reported by Herbet et al. (2015), and hyperconnectivity of the right pars opercularis was observed in adolescents with MDD (Tang et al., 2018). There is evidence showing that the right interior frontal gyrus especially the right pars opercularis may be specialized in music neurocognition for (Cheung et al., 2018). It has been long known that music and emotion are connected, and music therapy has been applied to treat various psychiatric disorders including depression, autism, etc. (Trimble and Hesdorffer, 2017). Our finding consistently suggests that the right pars opercularis has enhanced FCs in LLD, and music therapy may be a target treatment to improve the depression symptoms.

The second main finding is that as compared to healthy subjects, LLD patients exhibit increased FC pattern comprising the right dlPFC and the right rostral middle frontal within the CEN, the right caudal anterior cingulate within salience network, bilateral isthmus cingulate, bilateral posterior cingulate, the right thalamus proper and the right supramarginal within the default mode network (DMN), and decreased FC between the left posterior cingulate and the left thalamus proper within the DMN.

The rostral middle frontal (or rostral middle frontal gyrus) is part of dlPFC. The right middle frontal gyrus is believed to be a converging site of the dorsal and ventral attention networks by playing a role in reorientation of attention (Japee et al., 2015). The posterior cingulate, or posterior cingulate cortex (PCC, BA 23 and 31), is considered as the posterior hub of the DMN and may play a role in emotion, cognition, awareness, arousal and regulatory modulation (Buckner et al., 2008; Leech and Sharp, 2014; Wang et al., 2019). The DMN is a group of brain regions with default mode of brain functions for internally directed self-referential cognitions that are activated at rest and deactivated during cognitive and mental tasks requiring attention (Raichle et al., 2001) and responsible. The PCC exhibit higher activities when the brain is at rest (Raichle, 2015). A functional neuroimaging study in healthy subjects has observed task-related FC decrease in the PCC region during cognitive processing than during resting states (Greicius et al., 2003). Our results revealed a decreased FC between the left PCC and the left thalamus proper within the DMN in the left hemisphere, and increased FC in bilateral PCC regions. Our findings are in line with the previous research with convergent data implicating disruptions of FC in the DMN in neural mechanisms for psychiatric disorders. A meta-analysis of rs-fMRI studies in MDD (Kaiser et al., 2015) reported increased FCs within the DMN associated with MDD. Review on a number of rs-fMRI studies (Leech and Sharp, 2014) showed

reduced FC in the PCC region in healthy aging subjects, patients with Alzheimer's disease (AD) (Binnewijzend et al., 2014), schizophrenia (Li et al., 2019), autism and attention deficit hyperactivity disorder. A meta-analysis of AD based on 34 selected rs-fMRI studies Badhwar et al. (2017) found consistent FC decrease in the DMN in AD patients, and both FC increase and decrease within the DMN in patients with AD, dementia and mild cognitive impairment. Also increased FC has been reported in the posterior DMN area in major depression (Zhou et al., 2010) and late-life anxious depression (Andreescu et al., 2011), in PCC region within the DMN in LLD (Alexopoulos et al., 2012), in the left PCC region of DMN in MDD (Zhao et al., 2019), as well as in the DMN in autistic children (Uddin et al., 2013).

The supramarginal gyrus (BA 40) is part of the inferior parietal lobe (IPL), a brain region within posterior DMN that functions in sensorimotor and cognitive domains. We observed an increased FC between the right supramarginal with the bilateral PCC region in the DMN. In a recent randomized and controlled clinical trial in LLD that compared multidomain intervention to supportive therapy to study both the change in depressive symptoms as primary clinical outcome and the change in FC using rs-fMRI data, the intervention group exhibited statistically significantly more reduction in the depressive symptoms and decreased FC between the PCC and the left IPL within the DMN as opposed to the supportive group as control (Roh et al., 2020). Also increased FC in the posterior DMN regions including the bilateral PCC and the left IPL was found in unhappy people as compared to happy people in an rs-fMRI study with healthy young adults (Luo et al., 2016). Together with our results of increased FCs in bilateral PCC, the left IPL and the left supramarginal gyrus DMN regions in LLD patients, these studies provide strong evidence suggesting that the intervention to help decrease FC within the DMN may have beneficial clinical effect as potential treatment for LLD.

The thalamus proper (or dorsal thalamus) is known as a hub for relaying sensory and motor signals from the senses to the cerebral cortex and receiving feedback from the cortex and plays a mediator role in-between cortico-cortical communication processing (Sherman and Guillery, 2002). Greicius et al. (2007) has found increased FC in thalamus with DMN during resting state in patients with MDD. A recent study based on a large sample of individuals using the Autism Brain Imaging Data Exchange also reported that patients with autism spectrum disorder exhibited hyperconnectivity between prefrontal cortex and thalamus as compared to healthy controls (Woodward et al., 2017).

The caudal anterior cingulate (also known as dorsal anterior cingulate cortex [dACC], BA 32) and anterior insula are the two major cortical structures of the salience network (SN) (Seeley et al., 2007; Menon, 2015). The SN is a large-scale brain network involved in detecting, filtering and integrating external salient stimuli with internal states to orchestrate brain network dynamics in the service of goal-directed behaviors (Uddin, 2015; Menon, 2015) and motivated behaviors (Yuen et al., 2014), as well as playing a mediating role in switching between activation and deactivation of internally directed cognitions of DMN and externally directed cognitions of CEN (Sridharan et al., 2008). The SN has been identified as a potential locus of dysfunction in autism spectrum disorder (Uddin and Menon, 2009), and hyperconnectivity within SN was observed in children with autism (Uddin et al., 2013). Perrin et al. (2012) had found that electroconvulsive therapy reduces FC between the right anterior cingulate cortex and the right dlPFC in patients with severe depression. Yuen et al. (2014) investigated the FC pattern of the SN in a rs-fMRI study to compare LLD with low and high apathy to healthy subjects, while apathy is a common symptom in major depression. Their results showed that relative to healthy controls, LLD patients with high apathy had increased FC pattern within SN between the right

anterior insular cortex and dACC, and increased FC of the right anterior insular cortex to the right dlPFC within the CEN, and to the right posterior cingulate (BA 31) within the DMN. Our results show a consistent increased FC between the right dlPFC within the CEN, the right dACC within the SN and the right PCC within the DMN in LLD patients. Also increased FC between the right dlPFC and the right anterior insular context was found in chronic neck pain patients (Ihara et al., 2019). These findings suggest that increased FC of the right dlPFC to the right regions of the SN may be associated with the neural mechanisms of depression in LLD and chronic neck pain patients.

Our results show increased FC pattern associated with LLD within and across the major large-scale neurocognitive brain networks including CEN, DMN and SN, which align with the previous findings of aberrant FCs among large-scale brain networks in MDD (Kaiser et al., 2015). Further, Zhao et al. (2019) found disrupted functions in regions of CEN and DMN in MDD patients. Recent research in AD including autosomal dominant AD, late-onset AD, AD, dementia and mild cognitive impairment have identified disrupted FCs within DMN, SN, CEN, dorsal attention network and sensorimotor network, and connectivities between SN and CEN (Thomas et al., 2014; Dai et al., 2015; Badhwar et al., 2017). Taken together, the results provide convergent evidence of degeneration of specific hubs within the internally-guided resting-state networks DMN, SN and CEN (Lee and Franggou, 2017) that may contribute to a deterioration of memory and cognitive functions in elderly patients with LLD or AD.

The bilateral fusiform (or fusiform gyrus [FG], BA 37) has been reported to be involved in visual cognition especially facial cognition known as fusiform face area (FFA) (Kanwisher et al., 1997; Mangun et al., 1998; Kanwisher and Yovel, 2006). Disrupted connectivity of bilateral FG to other regions may cause dysfunction in face recognition in patients with mild cognitive

impairment, along with the development of neurodegenerative disease such as AD (Yetkin et al., 2006; Cai et al., 2015). The right ventral diencephalon appears to be an important region in prediction of mild cognitive impairment and dementia in LLD patients (Lebedeva et al., 2017). We found increased FC between the right FG and the right ventral diencephalon, suggesting LLD patients show some degree of cognitive impairment especially in facial cognition.

The left pars triangularis (BA 45) and the left pars opercularis (BA 44) constitute Broca's area, which has been known for its importance in speech process by integrating and coordinating information within brain network (Stuhrmann et al., 2015). In an fMRI study with healthy subjects (Friederici et al., 2000), the left pars triangularis has been observed to activate in semantic processing. Further, another fMRI study of simultaneous language translation in healthy interpreters has reported higher activity in the left pars triangularis during backward translation and hence suggested pars triangularis be considered as a "hub" of the language-control network (Elmer, 2016). Our results show reduced FC between the left pars triangularis, the left FG, the left supramarginal and the left entorhinal in LLD group, suggesting potential functional semantic and facial recognition inhibition in the left hemisphere of brain may be associated with LLD. A meta-analysis of previous studies on antidepressant (Ma, 2015) found that the treatments have improved the neural response to positive emotion in the right dlPFC and the left FG.

In addition, we find eight secondary hubs, which are defined as regions showing at least two significantly increased or decreased FCs to other regions. These regions are the left fusiform (FFA), the right pallidum, the right dACC (SN), the left isthmus cingulate (DMN), bilateral posterior cingulate cortex (DMN) and bilateral supramarginal (DMN). Moreover, significant FC activities are observed in eight bilateral regions: rostral middle frontal gyrus (CEN), dorsal

thalamus, ventral diencephalon, fusiform (FFA), pars triangularis, isthmus cingulate (DMN), posterior cingulate cortex (DMN) and supramarginal (DMN).

We also observe decreased FC between the right caudate and the left cuneous in LLD group. The right caudate is involved in reward network (RN), in which hypoconnectivity was detected in depression (Keren et al., 2018). The left cuneous belongs to visual network (VN). Our present finding aligns with the results of a previous clinical study that the right caudate has been identified a key region with enhanced neural response to positive emotions in patients with treatment-resistant MDD following treatment of antidepressant ketamine (Murrough et al., 2015).

It should also be noted that our results also suggest that LLD patients exhibit decreased FCs within the left hemisphere (the logical/language side), while increased FCs within the right hemisphere. This asymmetric FC pattern may be worthy of further investigation.

6.2 Bayesian Dirichlet Process Mixture Model

6.2.1 Results

We analyze the FC data measured by rs-fMRI in the LLD neuroimaging study using the Bayesian Dirichlet process mixture model specified in Section 5.2.3. The posterior computation is carried out by the blocked Gibbs sampler using stick-breaking truncation described in Section 5.2.2 that is implemented using R runjags package via Just Another Gibbs Sampler (JAGS) (Denwood, 2016). The number of significant differential connectivity links are consistently between 18 and 19 across the different numbers of clusters K ranged from 5 to 12. Hence, we choose $K = 5$. Based on the results from 25,000 posterior iterations, taken from one MCMC chain, which has 1,275,000 iterations with the first 25,000 MCMC iterations discarded as burn-

in and a thinning interval of 50, we compute the posterior probability of $\beta_{2i} \neq 0$ given data for each connectivity link, denoted by r_i,

$$r_i = P(\beta_{2i} \neq 0 \mid \cdot), \tag{6.2}$$

If the posterior probability $r_i \geq c$, where c is the threshold, we reject $H_{0,i}$ and claim the link is significantly differential. Kim et al. (2009) has provided the average of Bayesian FDR and the associated optimal threshold, which is defined as $\arg\max_c\{c : \widehat{FDR}(c) \leq q\}$, computed from simulated datasets, where Bayesian FDR (Genovese and Wasserman, 2002b; Newton et al., 2004) can be calculated as :

$$\widehat{FDR}(c) = \frac{\sum_{i=1}^{m} I\{r_i \geq c\}(1 - r_i)}{\sum_{i=1}^{m} I\{r_i \geq c\}}, \tag{6.3}$$

Based on the average of Bayesian FDR and the associated optimal threshold presented in Figure 4 in Kim et al. (2009), we can extrapolate the optimal thresholds that are approximately 0.45 and 0.43 for FDR level (referred to as "Realized FDR" on the plot) and Bayesian FDR level both at 0.2 on average, respectively, based on 30 simulated datasets (shown in Figure 27).

Nineteen differential connectivity links are identified by Bayesian DPM model with number of clusters $K = 5$ at the threshold 0.43 to control Bayesian FDR level at 0.2, as presented in Table IX.

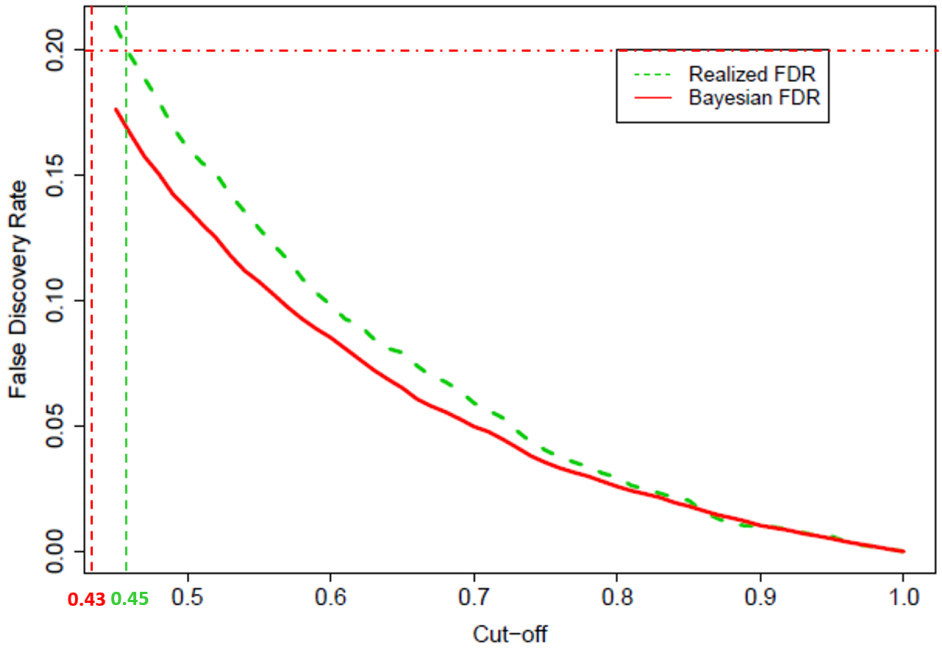
6.2.2 Discussion

The network analysis of the 19 FC links by Bayesian DPM model is illustrated in Figure 28, based on cortical and subcortical gray matter regions by left and right hemisphere of the brain. The main finding aligns with the results by Bayesian multimodal Lfdr method that the

TABLE IX
NINETEEN DIFFERENTIAL FC LINKS BY BAYESIAN DPM MODEL AT BAYESIAN FDR LEVEL OF 0.2[a] FOR LLD NEUROIMAGING STUDY

Region 1[b]	Region 2[b]	$\hat{\beta}_{2i}$[c]	r_i[d]
Hyperconnectivities:			
R thalamus proper (RTP)	R caudal middle frontal (RCMF)	0.060	0.583 (B, E)
R ventral diencephalon (RVD)	R fusiform (RF)	0.044	0.453 (B, E)
L isthmus cingulate (LIC)	R caudal middle frontal (RCMF)	0.052	0.523 (B, E)
L lateral orbitofrontal (LLO)	L medial orbitofrontal (LMO)	0.054	0.531
L lateral orbitofrontal (LLO)	R medial orbitofrontal (RMO)	0.043	0.445
L postcentral (LPoC)	L precentral (LPrC)	0.062	0.606
L posterior cingulate (LPC)	R supramarginal (RS)	0.063	0.611 (B)
L precentral (LPrC)	R postcentral (RPoC)	0.059	0.576
L precentral (LPrC)	R precentral (RPrC)	0.063	0.606
L rostral middle frontal (LRMF)	R caudal middle frontal (RCMF)	0.048	0.483 (B, E)
L superior parietal (LSP)	R pars opercularis (RPO)	0.043	0.444 (B)
L superior parietal (LSP)	R superior parietal (RSP)	0.055	0.552
L superior parietal (LSP)	R supramarginal (RS)	0.059	0.579
R caudal middle frontal (RCMF)	R posterior cingulate (RPC)	0.049	0.497 (B)
R caudal middle frontal (RCMF)	R superior frontal (RSF)	0.057	0.560
R inferior parietal (LIP)	R posterior cingulate (RPC)	0.050	0.501
R posterior cingulate (RPC)	R supramarginal (RS)	0.047	0.476 (B)
Hypoconnectivities:			
R caudate (RCau)	R cuneus (RCun)	-0.058	0.458
L fusiform (LF)	L supramarginal (LS)	-0.092	0.635 (B, E)

[a] Based on Bayesian DPM model with number of clusters $K = 5$ at Bayesian FDR level of 0.2 with the threshold $c = 0.43$ (refer to Figure 27).
[b] L = left; R = right.
[c] Denotes the estimate of differential FC effect between LLD group and HC group for the link from the 25,000 posterior draws, positive value indicating hyperconnectivity while negative value indicating hypoconnectivity.
[d] Denotes the posterior probability of $\beta_{2i} \neq 0$ given data. The 9 FC links also identified by Bayesian multimodal Lfdr method at $q = 0.2$ are indicated using B in parentheses, while the 5 FC links also identified by Efron's Lfdr method at $q = 0.2$ are indicated using E in parentheses.



Source: Figure 4 in Kim et al. (2009)

Figure 27. The optimal thresholds (i.e., cut-off values) corresponding to FDR and Bayesian FDR levels

right caudal middle frontal cortex (dlPFC) within the CEN is identified as the primary hub with enhanced FCs to another CEN region the left rostral middle frontal, as well as four DMN regions including the right superior frontal, the right posterior cingulate, the right thalamus proper and the left isthmus cingulate.

Besides, there are two major distinctive findings. One is LLD patients exhibit increased FCs the left lateral orbitofrontal cortex to the bilateral medial orbitofrontal cortex regions, which is consistent with the previous findings of increased FC of the lateral orbitofrontal cortex in depression (Cheng et al., 2016; Li et al., 2018; Wang et al., 2018; Rolls, 2019). The orbitofrontal cortex has been known for its function in affective network (AN), which is as-

sociated with emotion processing and reward/non-reward system in decision making (Bechara et al., 2000). Rolls (2016) proposed a "non-reward attractor theory of depression", in which the lateral orbitofrontal cortex plays a key role in non-reward and the medial orbitofronal cortex in reward system.

The other unique finding is that increased FCs of the left precentral gyrus to the right precentral gyrus, and the bilateral postcentral gyrus within the sensorimotor network (SMN) are related to LLD. The SMN is constituted by the precentral gyrus for somatosensory functions and postcentral gyrus or motor functions. Research show that dysfunctions in the SMN may be linked to depressive and manic phases of bipolar disorder (Martino et al., 2016).

It is worth noting that the FC alterations involved in the resting-state networks for cognition (CEN), affection (AN) and sensory functions (SMN) associated with LLD are also observed in chronic pain (Kornelsen et al., 2013; Kolesar et al., 2017). Considering that chronic pain is one common ailment in the elderly, the results are anticipated and suggest that the disrupted FCs be related to the underlying pathology of the comorbid depression and chronic pain as well as the cause of LLD.

6.3 Results Comparison

Among all 3741 connectivities analyzed in this study, a total of 32 differential FC links are identified using Efron's Lfdr, Bayesian multimodal Lfdr or DPM model at $q = 0.2$, including 25 hyperconnectivity links and 7 hypoconnectivity links, as shown in Table X. The rest 3709 (99.1%) links are considered non-differential by the three methods.

Figure 29 presents a Venn diagram showing the overlap of the differential FCs detected using the three methods. In comparison with the results by Efron's Lfdr and Bayesian multimodal Lfdr methods, there is an overlap of 9 connectivities between Bayesian multimodal Lfdr

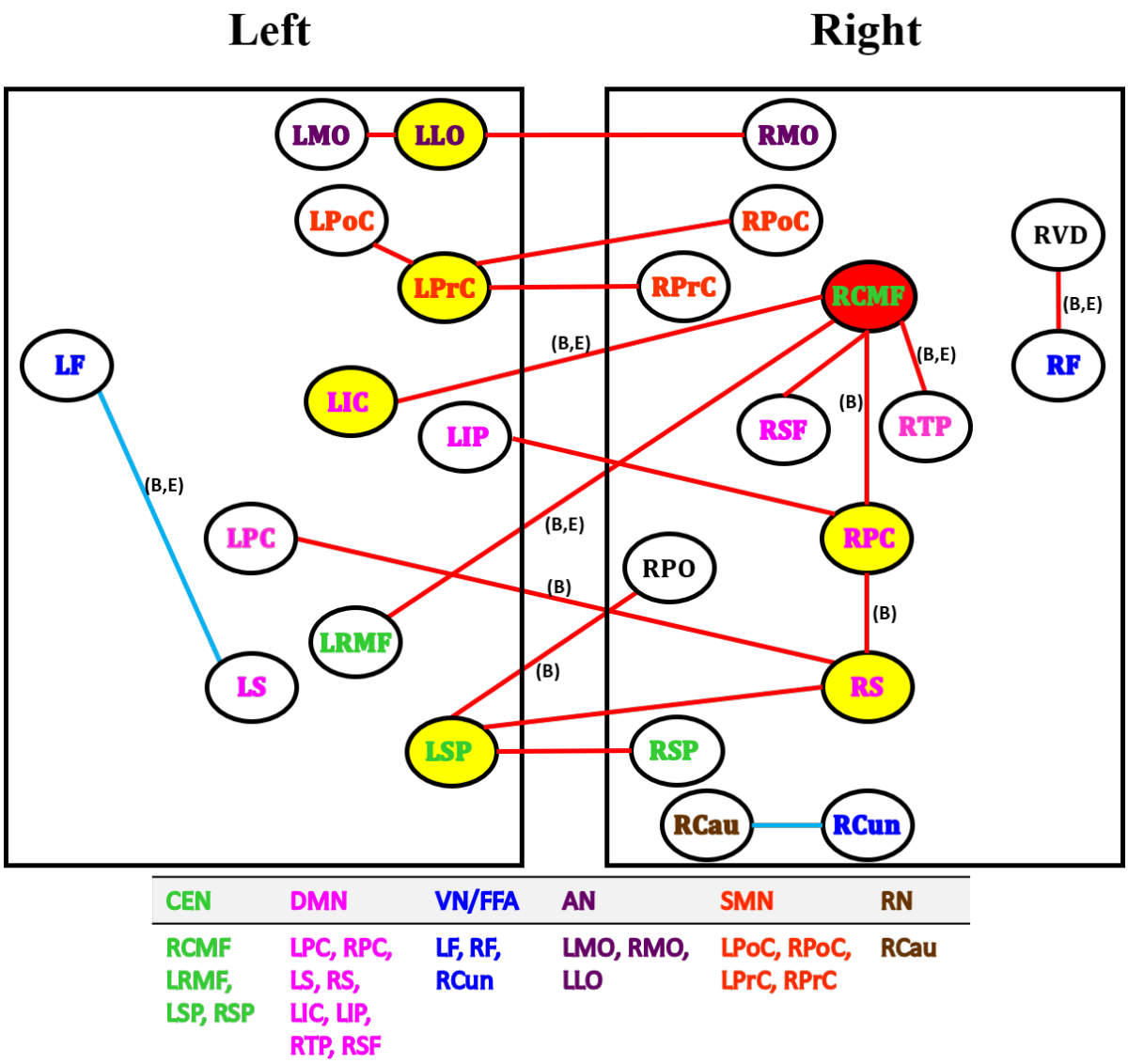


Figure 28. Network analysis of the 19 differential FC links by Bayesian DPM model at Bayesian FDR level of 0.2 based on cortical and subcortical gray matter regions by left and right hemisphere of the brain

Red line indicates an increase in FC (hyperconnectivity) and blue line indicates a decrease in FC (hypoconnectivity). Red circle denotes the primary hub, and yellow circle denotes the secondary hub.
The 9 FC links also identified by Bayesian multimodal Lfdr method are indicated using B in parentheses.
The 5 FC links also identified by Efron's Lfdr method are indicated using E in parentheses.
Abbreviations: CEN = central executive network; DMN = default mode network; VN/FFA = visual network / fusiform face area; AN = affective network; SMN = sensorimotor network; RN = reward network;
LIC = left isthmus cingulate; LIP = left inferior parietal; LLO = left lateral orbitofrontal;
LRMF = left rostral middle frontal; LF/RF = left/right fusiform;
LMO/RMO = left/right medial orbitofrontal; LPC/RPC = left/right posterior cingulate;
LPoC/RPoC = left/right postcentral; LPrC/RPrC = left/right precentral;
LS/RS = left/right supramarginal; LSP/RSP = left/right superior parietal;
RCMF = right caudal middle frontal; RCau = right caudate; RCun = right cuneus;
RPO = right pars opercularis; RSF = right superior frontal;
RTP = right thalamus proper; RVD = right ventral diencephalon.

and DPM, 5 between Efron's Lfdr and DPM, as well as 5 among the three methods. Figure 30 displays the network analysis of the 9 differential FC links identified by both Bayesian multimodal Lfdr method and DPM model at $q = 0.2$ based on cortical and subcortical gray matter regions by left and right hemisphere of the brain. Among the 9 FC links, 8 are hyperconnectivities and 1 hypoconnectivity linking left fusiform in FFA and left supramargional in DMN. The right dlPFC (i.e., the right caudal middle frontal) region in CEN remains a primary hub with increased connectivity to 4 other regions in DMN and CEN directly, and 2 regions in DMN indirectly. Three bilateral regions including fusiform in FFA, supramargional and posterior cingulate in DMN, and the major networks including CEN, DMN and FFA are involved.

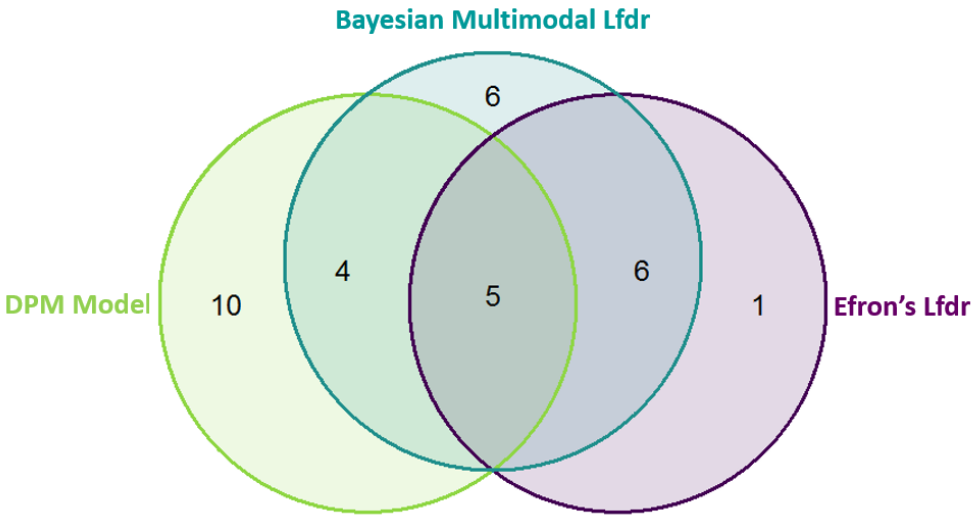


Figure 29. Venn diagram showing overlap between differential FC links identified using Efron's Lfdr, Bayesian multimodal Lfdr and DPM model at FDR or Bayesian FDR level of 0.2 for LLD neuroimaging study

TABLE X
THIRTY-TWO DIFFERENTIAL FC LINKS IDENTIFIED USING EFRON'S LFDR, BAYESIAN MULTIMODAL LFDR OR DPM MODEL AT FDR LEVEL OF 0.2 FOR LLD NEUROIMAGING STUDY

Region 1[a]	Region 2[a]	Efron's Lfdr[b]	Bayesian multimodal Lfdr[b]	DPM Model[b]
Hypoconnectivities:				
R thalamus proper (RTP)	R caudal middle frontal (RCMF)	X	X	X
R pallidum (RP)	L inferior parietal (LIP)	X	X	
R pallidum (RP)	R caudal middle frontal (RCMF)	X	X	
R accumbens area (RAA)	L isthmus cingulate (LIC)	X	X	
R ventral diencephalon (RVD)	R fusiform (RF)	X	X	X
L isthmus cingulate (LIC)	R caudal middle frontal (RCMF)	X	X	X
L lateral orbitofrontal (LLO)	L medial orbitofrontal (LMO)			X
L lateral orbitofrontal (LLO)	R medial orbitofrontal (RMO)			X
L postcentral (LPoC)	L precentral (LPrC)			X
L posterior cingulate (LPC)	R rostral middle frontal (RRMF)	X	X	
L posterior cingulate (LPC)	R supramarginal (RS)		X	X
L precentral (LPrC)	R postcentral (RPoC)			X
L precentral (LPrC)	R precentral (RPrC)			X
L rostral middle frontal (LRMF)	R caudal middle frontal (RCMF)	X	X	X
L superior frontal (LSF)	R isthmus cingulate (RIC)	X		
L superior parietal (LSP)	R pars opercularis (RPO)		X	X
L superior parietal (LSP)	R superior parietal (RSP)			X
L superior parietal (LSP)	R supramarginal (RS)			X
R caudal anterior cingulate (RCAC)	R caudal middle frontal (RCMF)	X	X	
R caudal middle frontal (RCMF)	R isthmus cingulate (RIC)	X	X	
R caudal middle frontal (RCMF)	R posterior cingulate (RPC)		X	X
R caudal middle frontal (RCMF)	R superior frontal (RSF)			X
R inferior parietal (LIP)	R posterior cingulate (RPC)			X
R pars triangularis (RPT)	R rostral anterior cingulate (RRAC)		X	
R posterior cingulate (RPC)	R supramarginal (RS)		X	X
Hypoconnectivities:				
L thalamus proper (LTP)	L posterior cingulate (LPC)		X	
L ventral diencephalon (LVD)	R caudal anterior cingulate (RCAC)		X	
R caudate (RCau)	L cuneus (LCun)		X	
R caudate (RCau)	R cuneus (RCun)			X
L entorhinal (LE)	L supramarginal (LS)		X	
L fusiform (LF)	L pars triangularis (LPT)		X	
L fusiform (LF)	L supramarginal (LS)	X	X	X

[a] L = left; R = right.
[b] The connectivity links identified by each method are indicated using X.

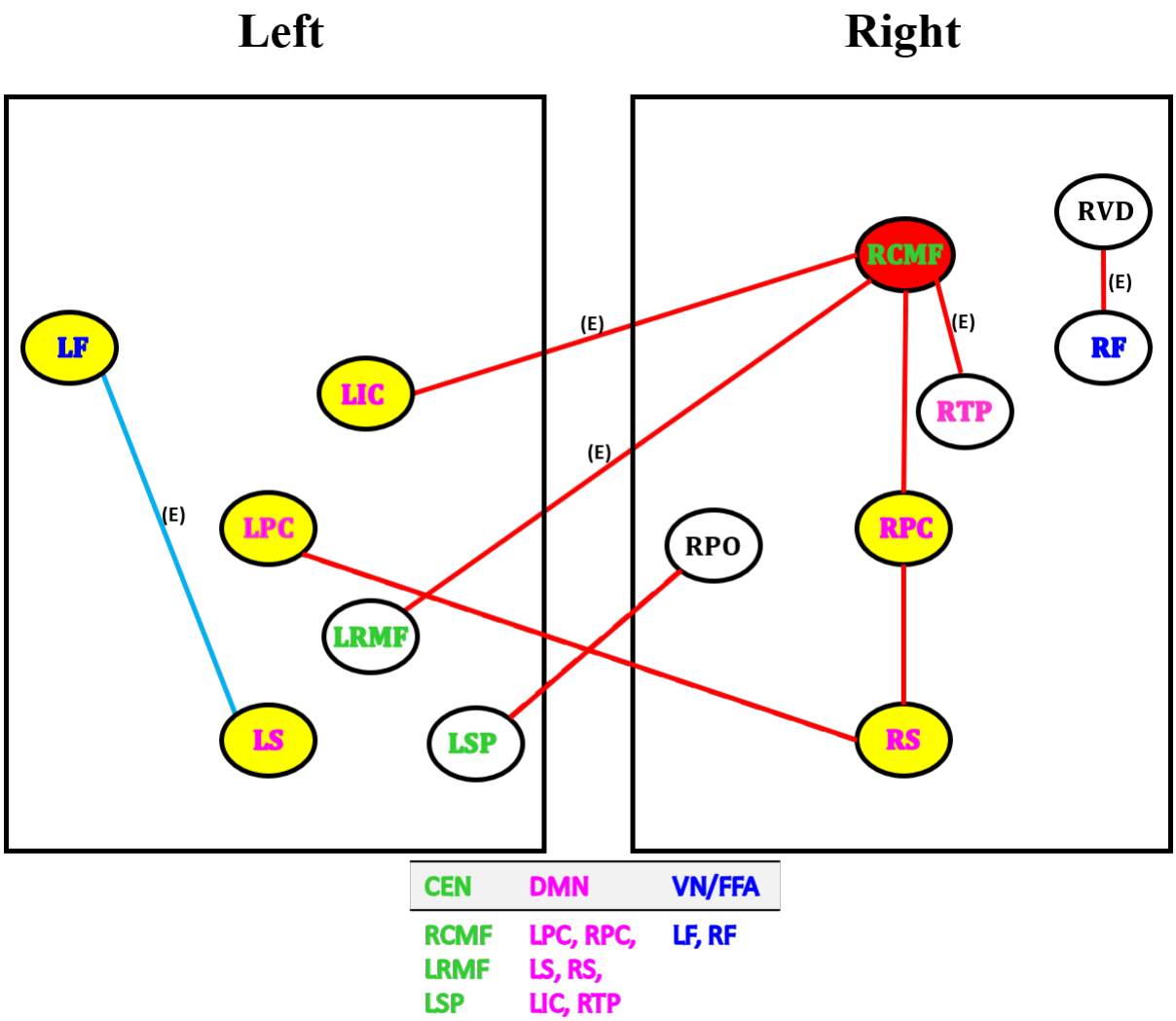


Figure 30. Network analysis of the 9 differential FC links identified by both Bayesian multimodal Lfdr method and DPM model at FDR level of 0.2 based on cortical and subcortical gray matter regions by left and right hemisphere of the brain

Red line indicates an increase in FC (hyperconnectivity) and blue line indicate a decrease in FC (hypoconnectivity). The 5 FC links also identified by Efron's Lfdr method are indicated using E in parentheses.
Abbreviations: CEN = central executive network; DMN = default mode network; VN/FFA = visual network / fusiform face area;
LF/RF = left/right fusiform;
LIC = left isthmus cingulate;
LPC/RPC = left/right posterior cingulate;
RPO = right pars opercularis;
LRMF = left rostral middle frontal;
LS/RS = left/right supramarginal;
LSP = left superior parietal;
RTP = right thalamus proper;
RVD = right ventral diencephalon;
RCMF = right caudal middle frontal.

CHAPTER 7

CONCLUSION AND FUTURE WORK

The main focus of my thesis is on FDR control in a cross-sectional, comparative multimodal neuroimaging study aimed to discover aberrant FCs in patients with psychiatric or neurological disorder as compared to healthy controls, where the sample size is usually small. Based on the test statistics from a linear mixed-effects model with heteroscedastic errors at both group level and connectivity link level, Bayesian multimodal Lfdr method combines and borrows strength across multimodal neuroimaging data in a Bayesian mixture model framework and hence increases the sensitivity of testing and helps improve efficiency in FDR control as well as detection of disrupted FCs. We illustrate Bayesian multimodal Lfdr method with the motivating multimodal neuroimaging study in LLD with a total of 23 participants (13 healthy subjects and 10 LLD patients) and conduct a rigorous comparison of its performance with Efron's Lfdr method and a non-parametric Bayesian DPM model that consider only FC data.

Bayesian multimodal Lfdr method gives more findings than the other two methods, by identifying 21 differential FC links, as compared to 12 and 19 by Efron's Lfdr method and DPM model, respectively, at the same FDR level of $q = 0.2$. Out of the findings from the three methods, a total of 32 (0.9%) differential FC links are detected and the rest 3709 (99.1%) links are considered non-differential. Though the quantity of differential FC links does not appear to be substantial, these findings suggest a profound disruption of functional neurocognitive brain networks including CEN, DMN, SN, SMN, AN, VN/fusiform face area, Broca's area and RN is implicated in LLD, providing crucial evidence for researchers to further investigate their association with clinical outcome in future studies.

Of the 12 FC links by Efron's Lfdr method, the vast majority - 11 out of 12 - are overlapping with Bayesian multimodal Lfdr method, which is expected because the two methods are based on the distributions of test statistics using Bayesian mixture model. The DPM model, as a Bayesian multiple comparison method directly on the FC data, has an overlap of 9 with Bayesian multimodal Lfdr method. These results confirm the robustness and consistency of the main findings by Bayesian multimodal Lfdr method alone and suggest that it is more powerful than Efron's Lfdr method.

We observe disrupted FCs within and between the major neurocogntive brain networks including CEN, DMN, and fusiform face area that are associated with LLD, with the right dlPFC region in the CEN identified as a primary bub showing increased FCs to regions in CEN and DMN. Those findings provide more detailed information on the disrupted FC regions and patterns involved in the underlying pathology of LLD, and further insights on potential neuroimaging biomarkers in the future clinical development to treat and prevent LLD.

Bayesian multimodal Lfdr method has its limitations, based on the fact that the current framework is specifically for a cross-sectional, comparative neuroimaging study that compares the disease group with the control group using connectivity data measured at one single time point. This places certain research topics beyond the reach of the methods. First, the mixed-effects model cannot be applied to neuroimaging data collected at multiple time points. Second, the methods do not yet have the capability to incorporate association analysis between connectivity and disease outcomes based on clinical measurements. Over the past years, the prospective longitudinal and interventional neuroimaging studies in patients with psychiatric or neurological disorder have gained popularity. For these studies, the primary objective is to investigate the association between the connectivity changes and clinical measurement (e.g.,

neurobehavioral measurement) changes within and between the intervention group and the control group and evaluate the effect of intervention group vs. control group. The multimodal neuroimaging data and clinical data are collected at the baseline and scheduled post-baseline visits during treatment period on each participant in both intervention and control groups.

Therefore, we have the following suggestions for the future work. To handle the more complex study design and data structure in a prospective and interventional neuroimaging study, we will need to make substantial extensions to the current framework. The basic ideas include using a mixed-effects model repeated measure method to analyze the longitudinal connectivity data and compare the intervention group vs. control group. The subject-level demographics and baseline characteristics (e.g., age, baseline disease activity such as disease duration) can be included in the model as covariate(s). The test statistics are then modeled using Bayesian multimodal Lfdr with other available modality data incorporated. For association between connectivity data and clinical measurement data, one can compute Pearson correlation coefficients in order to reveal neural connectivity changes from baseline associated with the change from baseline in clinical measurement related to the intervention group and the control group at each scheduled post-baseline visits to identify specific connectivities that are correlated with disease outcome (Pape et al., 2020).

Bayesian multimodal Lfdr method can be expanded to include more resources of data besides SC statistics (e.g., test statistics from analysis of FA measures of SC and/or positron emission tomography, if applicable) in modeling alternative density shown in Equation (4.39) and prior probability shown in Equation (4.42). In addition, Bayesian multimodal Lfdr method can be improved in the following ways. The assumption that all hypotheses tested are independent may not hold for neuroimaging data, as the correlations among connectivities within

and/or between the major brain networks often exist. To better account for the dependencies among the hypotheses, one can consider using Ising model, a binary Markov random field, for prior probability of being null (or alternative) to incorporate SC statistics as well as spatial interaction between the regions, which has been implemented in genome-wide association studies (Zablocki, 2016). Also, the parametric model using gamma distribution assumed for the density of alternative hypotheses in the tail regions could be changed when it does not provide a sufficient goodness of fit for the data in some cases, which may introduce bias as a result of lack of fit. Zablocki et al. (2017) extended the covariate-modulated Lfdr method by proposing a semi-parametric approach to offer more flexibility for a genome-wide association study in Schizophrenia. In this approach, the estimation for null density and prior probability of being null remain the same, while the alternative density is estimated using a non-parametric method that utilizes a mixture of B-spline densities with the weights for each B-spline density modeled using a multinomial logistic regression with functional annotations as covariates. In general, the parametric model has more power if the model fit is sufficient. Thus, the advantage and disadvantage of this semi-parametric approach will need to be assessed and compared with the parametric approach for multimodal neuroimaging studies.

For Bayesian DPM model to analyze the FC data in the motivating LLD neuroimaging study, we also have a couple thoughts for further extension. First, the SC data can be incorporated into the logit of prior probability for the spike-and-slab mixture prior shown in Equation (5.28). Second, it is noted that Bayesian DPM model using MCMC algorithm to obtain the posterior sampling of the model parameters is computationally demanding, which can be a problem in case of simulation study. Variational Bayesian approach circumvents this problem by approximating posterior density using optimization, as an alternative to the direct MCMC

sampling (Blei and Jordan, 2006; Blei et al., 2017). Particularly, we may consider applying the spike-and-slab LASSO (least absolute shrinkage and selection operator) method proposed by Ročková and George (2018) to the spike-and-slab prior for β_{2i} in Equation (5.28) to efficiently detect the posterior mode of each β_{2i} via coordinate ascent optimization algorithm.

Last but not least, machine learning and classification techniques is another useful tool for dimensionality reduction in the high dimensional neuroimaging data where there are thousands of connectivities but only a few subjects (Chén, 2019). Predictive-modeling based variable selection methods can be applied to neuroimaging data to select connectivities that may differentiate the disease group from the control group. The regularization approaches, such as LASSO (Tibshirani, 1996), ridge (Friedman et al., 2010) and elastic net regression (Zou and Hastie, 2005), intended to reduce variance at the cost of introducing some bias. These regression methods shrink the regression coefficients of irrelevant variables by augmenting a penalty to the ordinary least squares function and then discard those variables with coefficients close to zero. Briefly, the LASSO regression penalizes the sum of absolute value of the regression coefficients $\sum_{j=1}^{p} |\beta_j|$ (known as l_1-norm penalty), while the ridge regression penalizes the sum of squared regression coefficients $\sum_{j=1}^{p} \beta_j^2$ (known as l_2-norm penalty), where β_j is the regression coefficient for the j^{th} predictor variable. While performing variable selections simultaneously with highly correlated variables, LASSO regression tends to do a sparse selection by taking one with the large coefficient and discarding the rest with almost zero coefficient, while Ridge regression keeps all as it shrinks coefficients of the correlated variables toward the same.

In presence of correlated variables, the elastic net regression method might be a better choice, as it provides a compromise to balance between LASSO l_1-norm penalty and ridge l_2-norm penalty in the mixture form of $\sum_{j=1}^{p} \left[\alpha|\beta_j| + (1-\alpha)\beta_j^2\right]$, where $0 \leq \alpha \leq 1$ controls the

mixture proportion of each penalty. Recent rs-fMRI neuroimaging studies have utilized elastic net logistic regression as a regularization method (Teipel et al., 2017; Bhaumik et al., 2018a). Algamal and Lee (2015) proposed an adaptive elastic net regularized logistic regression model in a high dimensional genomics study, which can also be applied to neuroimaging data. Typically, this is a two-step procedure: First step, divide the dataset into a training dataset (about 2/3) and a testing dataset (about 1/3); Second step, perform the 10-fold cross-validation using the training dataset and then evaluate the classification accuracy using the test dataset. Normally sufficient sample size is required in order to partition the data into two sets one for training and the other for testing. In the case of the LLD neuroimaging data with small sample size of 23, it will be impractical to split the already small data into even smaller training and testing data. One solution is to develop data pooling strategies to combine the data with other LLD neuroimaging study datasets using similar study design (Costafreda, 2009) and then perform mega-analysis of the pooled data using elastic net method.

APPENDIX

MAXIMUM MARGINAL LIKELIHOOD PARAMETER ESTIMATES

Below we illustrate steps in details how to find solutions that maximize the marginal log-likelihood in Equation (3.38) by differentiating it with respect to the parameters in the linear mixed-effects model in Chapter 3 based on Hedeker and Gibbons (2006); Zhao (2014); Song (2016).

The first derivative of the marginal log-likelihood function in Equation (3.38) with respect to $\boldsymbol{\beta}$ is,

$$
\begin{aligned}
\frac{\partial \log L}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}} \sum_{j=1}^{N} \log\big[f(\boldsymbol{y}_j)\big] \\
&= \sum_{j=1}^{N} \frac{\partial \log f(\boldsymbol{y}_j)}{\partial \boldsymbol{\beta}} \\
&= \sum_{j=1}^{N} \frac{1}{f(\boldsymbol{y}_j)} \frac{\partial f(\boldsymbol{y}_j)}{\partial \boldsymbol{\beta}} \\
&= \sum_{j=1}^{N} \frac{1}{f(\boldsymbol{y}_j)} \frac{\partial \big[\int f(\boldsymbol{y}_j \mid \gamma_j) f(\gamma_j) d\gamma_j\big]}{\partial \boldsymbol{\beta}} \\
&= \sum_{j=1}^{N} \frac{1}{f(\boldsymbol{y}_j)} \int \frac{\partial f(\boldsymbol{y}_j \mid \gamma_j)}{\partial \boldsymbol{\beta}} f(\gamma_j) d\gamma_j \\
&= \sum_{j=1}^{N} \frac{1}{f(\boldsymbol{y}_j)} \int f(\boldsymbol{y}_j \mid \gamma_j) f(\gamma_j) \frac{\partial \log f(\boldsymbol{y}_j \mid \gamma_j)}{\partial \boldsymbol{\beta}} d\gamma_j \\
&= \sum_{j=1}^{N} \int \frac{f(\boldsymbol{y}_j \mid \gamma_j) f(\gamma_j)}{f(\boldsymbol{y}_j)} \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta} - \boldsymbol{Z}_j \gamma_j) d\gamma_j \\
&= \sum_{j=1}^{N} \int f(\gamma_j \mid \boldsymbol{y}_j) \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta} - \boldsymbol{Z}_j \gamma_j) d\gamma_j \\
&= \sum_{j=1}^{N} \int \Big[\boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta}) f(\gamma_j \mid \boldsymbol{y}_j) - \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} \boldsymbol{Z}_j f(\gamma_j \mid \boldsymbol{y}_j) \gamma_j\Big] d\gamma_j
\end{aligned}
$$

APPENDIX (Continued)

$$
\begin{aligned}
&= \sum_{j=1}^{N} \left[\int \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta}) f(\gamma_j \mid \boldsymbol{y}_j) d\gamma_j - \int \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} \boldsymbol{Z}_j f(\gamma_j \mid \boldsymbol{y}_j) \gamma_j d\gamma_j \right] \\
&= \sum_{j=1}^{N} \left[\boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta}) \int f(\gamma_j \mid \boldsymbol{y}_j) d\gamma_j - \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} \boldsymbol{Z}_j \int f(\gamma_j \mid \boldsymbol{y}_j) \gamma_j d\gamma_j \right] \\
&= \sum_{j=1}^{N} \left[\boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta}) - \boldsymbol{Z}_j^T \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} \tilde{\gamma}_j \right] \\
&= \sum_{j=1}^{N} \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta} - \boldsymbol{Z}_j \tilde{\gamma}_j), \qquad \text{(A.1)}
\end{aligned}
$$

as $\int f(\gamma_j \mid \boldsymbol{y}_j) d\gamma_j = 1$ and $\int f(\gamma_j \mid \boldsymbol{y}_j) \gamma_j d\gamma_j = \tilde{\gamma}_j$.

Setting Equation (A.1) equal to zero yields

$$
\begin{aligned}
&\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{N} \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta} - \boldsymbol{Z}_j \tilde{\gamma}_j) = \boldsymbol{0} \\
\implies &\sum_{j=1}^{N} \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} \boldsymbol{X}_j \boldsymbol{\beta} = \sum_{j=1}^{N} \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} (\boldsymbol{y}_j - \boldsymbol{Z}_j \tilde{\gamma}_j) \\
\implies &\hat{\boldsymbol{\beta}} = \left(\sum_{j=1}^{N} \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} \boldsymbol{X}_j \right)^{-1} \left[\sum_{j=1}^{N} \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} (\boldsymbol{y}_j - \boldsymbol{Z}_j \tilde{\gamma}_j) \right]. \qquad \text{(A.2)}
\end{aligned}
$$

Note that as $\tilde{\gamma}_j \to 0$, the random subject effects are negligible and $\hat{\boldsymbol{\beta}}$ is approaching the ordinary least square estimates $\left(\sum_{j=1}^{N} \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} \boldsymbol{X}_j \right)^{-1} \left(\sum_{j=1}^{N} \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{Grp_j}^{-1} \boldsymbol{y}_j \right)$.

To obtain the variance of $\hat{\boldsymbol{\beta}}$, first we rewrite Equation (A.2) as

$$
\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T \boldsymbol{\Sigma}_{\epsilon}^{-1} \boldsymbol{X} \right)^{-1} \left[\boldsymbol{X}^T \boldsymbol{\Sigma}_{\epsilon}^{-1} (\boldsymbol{y} - \boldsymbol{Z} \tilde{\boldsymbol{\gamma}}) \right], \qquad \text{(A.3)}
$$

where

APPENDIX (Continued)

- $\boldsymbol{y}$ is a $mN \times 1$ vector of the connectivity measurement.

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 & \boldsymbol{y}_2 & \cdots & \boldsymbol{y}_N \end{bmatrix}^T_{1 \times mN} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{A.4}$$
$$\boldsymbol{y} \sim N\left(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{Z}\boldsymbol{\Sigma}_\gamma \boldsymbol{Z}^T + \boldsymbol{\Sigma}_\epsilon\right),$$

  where $\boldsymbol{y}_j$, $\boldsymbol{X}$, $\boldsymbol{Z}$, $\boldsymbol{\Sigma}_\gamma$, $\boldsymbol{\Sigma}_\epsilon$ are defined in Equations (3.3), (A.5), (A.6), (A.8) and (A.10), respectively.

- $\boldsymbol{X}$ is a $mN \times 2m$ design matrix for the fixed effects.

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix}_{mN \times 2m}, \tag{A.5}$$

  where $\boldsymbol{X}_j$ is defined in Equation (3.4).

- $\boldsymbol{\beta}$ is a $2m \times 1$ vector of the fixed effects, as defined in Equation (3.5).

- $\boldsymbol{Z}$ is a $mN \times N$ design matrix for the random subject effects,

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{Z}_N \end{bmatrix}_{mN \times N}, \tag{A.6}$$

  where $\boldsymbol{Z}_j$ is defined in Equation (3.6).

- $\boldsymbol{\gamma}$ is a $N \times 1$ vector of random subject effects.

$$\boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_N \end{bmatrix}^T_{1 \times N} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\gamma), \tag{A.7}$$

APPENDIX (Continued)

where γ_j, $j = 1, \cdots, N$, is defined in Equation (3.7) and the variance of $\boldsymbol{\gamma}$ $\boldsymbol{\Sigma}_\gamma$ is a $N \times N$ diagnonal matrix:

$$\boldsymbol{\Sigma}_\gamma = \begin{bmatrix} \sigma_\gamma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_\gamma^2 \end{bmatrix}_{N \times N} = \sigma_\gamma^2 \boldsymbol{I}_N, \tag{A.8}$$

- $\boldsymbol{\epsilon}$ is a $mN \times 1$ vector of random errors.

$$\boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 & \boldsymbol{\epsilon}_2 & \cdots & \boldsymbol{\epsilon}_N \end{bmatrix}^T_{1 \times mN} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon), \tag{A.9}$$

where the variance of $\boldsymbol{\epsilon}$ $\boldsymbol{\Sigma}_\epsilon$ is a $mN \times mN$ diagonal matrix:

$$\boldsymbol{\Sigma}_\epsilon = \begin{bmatrix} \boldsymbol{\Sigma}_{Grp_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\Sigma}_{Grp_N} \end{bmatrix}_{mN \times mN}, \tag{A.10}$$

where $\boldsymbol{\Sigma}_{Grp_j}, j = 1, \cdots, N$, is defined in Equation (3.9).

Then the variance of $\hat{\boldsymbol{\beta}}$ is:

$$\begin{aligned}
\mathrm{Var}(\hat{\boldsymbol{\beta}}) &= \mathrm{Var}\left\{ \left(\boldsymbol{X}^T \boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{X}\right)^{-1} \left[\boldsymbol{X}^T \boldsymbol{\Sigma}_\epsilon^{-1} (\boldsymbol{y} - \boldsymbol{Z}\tilde{\boldsymbol{\gamma}})\right] \right\} \\
&= \left(\boldsymbol{X}^T \boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{X}\right)^{-1} \left[\boldsymbol{X}^T \boldsymbol{\Sigma}_\epsilon^{-1} \mathrm{Var}(\boldsymbol{y}) \boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{X}\right] \left[\left(\boldsymbol{X}^T \boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{X}\right)^{-1}\right]^T \\
&= \boldsymbol{X}^{-1} \boldsymbol{\Sigma}_\epsilon (\boldsymbol{X}^T)^{-1} \left[\boldsymbol{X}^T \boldsymbol{\Sigma}_\epsilon^{-1} \left(\boldsymbol{Z} \boldsymbol{\Sigma}_\gamma \boldsymbol{Z}^T + \boldsymbol{\Sigma}_\epsilon\right) \boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{X}\right] \boldsymbol{X}^{-1} \boldsymbol{\Sigma}_\epsilon (\boldsymbol{X}^{-1})^T \\
&= \boldsymbol{X}^{-1} \boldsymbol{\Sigma}_\epsilon (\boldsymbol{X}^T)^{-1} \boldsymbol{X}^T \boldsymbol{\Sigma}_\epsilon^{-1} \left(\boldsymbol{Z} \boldsymbol{\Sigma}_\gamma \boldsymbol{Z}^T + \boldsymbol{\Sigma}_\epsilon\right) \boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{X} \boldsymbol{X}^{-1} \boldsymbol{\Sigma}_\epsilon (\boldsymbol{X}^{-1})^T \\
&= \boldsymbol{X}^{-1} \left(\boldsymbol{Z} \boldsymbol{\Sigma}_\gamma \boldsymbol{Z}^T + \boldsymbol{\Sigma}_\epsilon\right) (\boldsymbol{X}^{-1})^T \\
&= \boldsymbol{X}^{-1} \left(\boldsymbol{Z} \boldsymbol{\Sigma}_\gamma \boldsymbol{Z}^T + \boldsymbol{\Sigma}_\epsilon\right) (\boldsymbol{X}^T)^{-1}
\end{aligned}$$

APPENDIX (Continued)

$$
\begin{aligned}
&= \left[\boldsymbol{X}^T \left(\boldsymbol{Z}\boldsymbol{\Sigma}_\gamma \boldsymbol{Z}^T + \boldsymbol{\Sigma}_\epsilon\right)^{-1} \boldsymbol{X}\right]^{-1} \\
&= \left[\boldsymbol{X}^T \left(\boldsymbol{Z}\sigma_\gamma^2 \boldsymbol{I}\boldsymbol{Z}^T + \boldsymbol{\Sigma}_\epsilon\right)^{-1} \boldsymbol{X}\right]^{-1} \\
&= \left[\boldsymbol{X}^T \left(\boldsymbol{Z}\sigma_\gamma^2 \boldsymbol{Z}^T + \boldsymbol{\Sigma}_\epsilon\right)^{-1} \boldsymbol{X}\right]^{-1}. \qquad \text{(A.11)}
\end{aligned}
$$

The first derivative of the marginal log-likelihood function in Equation (3.38) with respect to the inverse of $\boldsymbol{\Sigma}_{Grp_j}$ is,

$$
\begin{aligned}
\frac{\partial \log L}{\partial \boldsymbol{\Sigma}_{Grp_j}^{-1}} &= \frac{\partial}{\partial \boldsymbol{\Sigma}_{Grp_j}^{-1}} \sum_{j=1}^{N} \log\big[f(\boldsymbol{y}_j)\big] \\
&= \sum_{j=1}^{N} \frac{\partial \log f(\boldsymbol{y}_j)}{\partial \boldsymbol{\Sigma}_{Grp_j}^{-1}} \\
&= \sum_{j=1}^{N} \frac{1}{f(\boldsymbol{y}_j)} \frac{\partial f(\boldsymbol{y}_j)}{\partial \boldsymbol{\Sigma}_{Grp_j}^{-1}} \\
&= \sum_{j=1}^{N} \frac{1}{f(\boldsymbol{y}_j)} \frac{\partial \big[\int f(\boldsymbol{y}_j \mid \gamma_j) f(\gamma_j) d\gamma_j\big]}{\partial \boldsymbol{\Sigma}_{Grp_j}^{-1}} \\
&= \sum_{j=1}^{N} \frac{1}{f(\boldsymbol{y}_j)} \int \frac{\partial f(\boldsymbol{y}_j \mid \gamma_j)}{\partial \boldsymbol{\Sigma}_{Grp_j}^{-1}} f(\gamma_j) d\gamma_j \\
&= \sum_{j=1}^{N} \int \frac{f(\boldsymbol{y}_j \mid \gamma_j) f(\gamma_j)}{f(\boldsymbol{y}_j)} \frac{\partial \log f(\boldsymbol{y}_j \mid \gamma_j)}{\partial \boldsymbol{\Sigma}_{Grp_j}^{-1}} d\gamma_j \\
&= \sum_{j=1}^{N} \int f(\gamma_j \mid \boldsymbol{y}_j) \left\{ \frac{1}{2} \left[\frac{\partial \log |\boldsymbol{\Sigma}_{Grp_j}|^{-1}}{\partial \boldsymbol{\Sigma}_{Grp_j}^{-1}} - \frac{\partial (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\gamma_j)^T \boldsymbol{\Sigma}_{Grp_j}^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\gamma_j)}{\partial \boldsymbol{\Sigma}_{Grp_j}^{-1}} \right] \right\} d\gamma_j,
\end{aligned}
\qquad \text{(A.12)}
$$

For computational convenience, we apply the following trace trick for quadratic forms and the fact that the inverse of the determinant of a matrix is the determinant of the inverse of the matrix to Equation (A.12):

APPENDIX (Continued)

$$\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X} = \mathrm{tr}[\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}] = \mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{A}], \tag{A.13}$$

$$|\boldsymbol{A}|^{-1} = |\boldsymbol{A}^{-1}|. \tag{A.14}$$

Now Equation (A.12) can be simplified to

$$\begin{aligned}
&\sum_{j=1}^{N}\int f(\gamma_j \mid \boldsymbol{y}_j)\left\{\frac{1}{2}\left[\frac{\partial \log|\boldsymbol{\Sigma}_{Grp_j}^{-1}|}{\partial \boldsymbol{\Sigma}_{Grp_j}^{-1}} - \frac{\partial \,\mathrm{tr}\big[(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\gamma_j)(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\gamma_j)^T\boldsymbol{\Sigma}_{Grp_j}^{-1}\big]}{\partial \boldsymbol{\Sigma}_{Grp_j}^{-1}}\right]\right\}d\gamma_j \\
&= \frac{1}{2}\sum_{j=1}^{N}\int f(\gamma_j \mid \boldsymbol{y}_j)\Big[\boldsymbol{\Sigma}_{Grp_j} - (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\gamma_j)(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\gamma_j)^T\Big]d\gamma_j, \\
&= \frac{1}{2}\sum_{j=1}^{N}\left[\boldsymbol{\Sigma}_{Grp_j}\int f(\gamma_j \mid \boldsymbol{y}_j)d\gamma_j - \int(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\gamma_j)(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\gamma_j)^T f(\gamma_j \mid \boldsymbol{y}_j)d\gamma_j\right] \\
&= \frac{1}{2}\sum_{j=1}^{N}\Big[\boldsymbol{\Sigma}_{Grp_j} - \boldsymbol{\epsilon}_j\boldsymbol{\epsilon}_j^T - \mathrm{tr}\big(\boldsymbol{Z}_j\tilde{\boldsymbol{\Sigma}}_{\gamma_j|\boldsymbol{y}_j}\boldsymbol{Z}_j^T\big)\Big],
\end{aligned} \tag{A.15}$$

as

$$\begin{aligned}
&\int(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\gamma_j)(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\gamma_j)^T f(\gamma_j \mid \boldsymbol{y}_j)d\gamma_j \\
=&\int\big[(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\tilde{\gamma}_j) - \boldsymbol{Z}_j(\gamma_j - \tilde{\gamma}_j)\big]\big[(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\tilde{\gamma}_j) - \boldsymbol{Z}_j(\gamma_j - \tilde{\gamma}_j)\big]^T f(\gamma_j \mid \boldsymbol{y}_j)d\gamma_j \\
=&(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\tilde{\gamma}_j)(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\tilde{\gamma}_j)^T + \int\big[\boldsymbol{Z}_j(\gamma_j - \tilde{\gamma}_j)\big]\big[\boldsymbol{Z}_j(\gamma_j - \tilde{\gamma}_j)\big]^T f(\gamma_j \mid \boldsymbol{y}_j)d\gamma_j \\
=&\boldsymbol{\epsilon}_j\boldsymbol{\epsilon}_j^T + \boldsymbol{Z}_j\tilde{\boldsymbol{\Sigma}}_{\gamma_j|\boldsymbol{y}_j}\boldsymbol{Z}_j^T \\
=&\boldsymbol{\epsilon}_j\boldsymbol{\epsilon}_j^T + \mathrm{tr}\big(\boldsymbol{Z}_j\tilde{\boldsymbol{\Sigma}}_{\gamma_j|\boldsymbol{y}_j}\boldsymbol{Z}_j^T\big),
\end{aligned} \tag{A.16}$$

where $\boldsymbol{\epsilon}_j = \boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{Z}_j\tilde{\gamma}_j$.

APPENDIX (Continued)

Setting Equation (A.15) equal to zero yields

$$\begin{aligned}
\frac{\partial \log L}{\partial \boldsymbol{\Sigma}_{Grp_j}^{-1}} &= \frac{1}{2}\sum_{j=1}^{N}\left[\boldsymbol{\Sigma}_{Grp_j} - \boldsymbol{\epsilon}_j\boldsymbol{\epsilon}_j^T - \mathrm{tr}(\boldsymbol{Z}_j\tilde{\boldsymbol{\Sigma}}_{\gamma_j|\boldsymbol{y}_j}\boldsymbol{Z}_j^T)\right] = \mathbf{0} \\
\implies \sum_{j=1}^{N}\boldsymbol{\Sigma}_{Grp_j} &= \sum_{j=1}^{N}\left[\boldsymbol{\epsilon}_j\boldsymbol{\epsilon}_j^T + \mathrm{tr}(\boldsymbol{Z}_j\tilde{\boldsymbol{\Sigma}}_{\gamma_j|\boldsymbol{y}_j}\boldsymbol{Z}_j^T)\right] \\
\implies \hat{\boldsymbol{\Sigma}}_{Grp_j} &= \frac{1}{N}\sum_{j=1}^{N}\left[\hat{\boldsymbol{\epsilon}}_j\hat{\boldsymbol{\epsilon}}_j^T + \mathrm{tr}\big(\boldsymbol{Z}_j\tilde{\boldsymbol{\Sigma}}_{\gamma_j|\boldsymbol{y}_j}\boldsymbol{Z}_j^T\big)\right], \qquad \text{(A.17)}
\end{aligned}$$

where $\hat{\boldsymbol{\epsilon}}_j = \boldsymbol{y}_j - \boldsymbol{X}_j\hat{\boldsymbol{\beta}} - \boldsymbol{Z}_j\tilde{\gamma}_j$.

Finally, the first derivative of the marginal log-likelihood function in Equation (3.38) with respect to σ_γ^{-2} is,

$$\begin{aligned}
\frac{\partial \log L}{\partial \sigma_\gamma^{-2}} &= \frac{\partial}{\partial \sigma_\gamma^{-2}}\sum_{j=1}^{N}\log\big[f(\boldsymbol{y}_j)\big] \\
&= \sum_{j=1}^{N}\frac{\partial \log f(\boldsymbol{y}_j)}{\partial \sigma_\gamma^{-2}} \\
&= \sum_{j=1}^{N}\frac{1}{f(\boldsymbol{y}_j)}\frac{\partial f(\boldsymbol{y}_j)}{\partial \sigma_\gamma^{-2}} \\
&= \sum_{j=1}^{N}\frac{1}{f(\boldsymbol{y}_j)}\frac{\partial\big[\int f(\boldsymbol{y}_j \mid \gamma_j)f(\gamma_j)d\gamma_j\big]}{\partial \sigma_\gamma^{-2}} \\
&= \sum_{j=1}^{N}\frac{1}{f(\boldsymbol{y}_j)}\int f(\boldsymbol{y}_j \mid \gamma_j)\frac{\partial f(\gamma_j)}{\partial (\sigma_\gamma^2)^{-1}}d\gamma_j \\
&= \sum_{j=1}^{N}\int \frac{f(\boldsymbol{y}_j \mid \gamma_j)f(\gamma_j)}{f(\boldsymbol{y}_j)}\frac{\partial \log f(\gamma_j)}{\partial \sigma_\gamma^{-2}}d\gamma_j \\
&= \sum_{j=1}^{N}\int f(\gamma_j \mid \boldsymbol{y}_j)\left\{\frac{1}{2}\left[\frac{\partial \log \sigma_\gamma^{-2}}{\partial \sigma_\gamma^{-2}} - \frac{\partial \gamma_j^2\sigma_\gamma^{-2}}{\partial \sigma_\gamma^{-2}}\right]\right\}d\gamma_j \\
&= \frac{1}{2}\sum_{j=1}^{N}\int f(\gamma_j \mid \boldsymbol{y}_j)(\sigma_\gamma^2 - \gamma_j^2)d\gamma_j \\
&= \frac{1}{2}\sum_{j=1}^{N}\left[\sigma_\gamma^2 - \int \gamma_j^2 f(\gamma_j \mid \boldsymbol{y}_j)d\gamma_j\right]
\end{aligned}$$

$$= \frac{1}{2}\sum_{j=1}^{N}\left[\sigma_\gamma^2 - \int\left[\tilde{\gamma}_j + (\gamma_j - \tilde{\gamma}_j)\right]^2 f(\gamma_j \mid \boldsymbol{y}_j)d\gamma_j\right]$$

$$= \frac{1}{2}\sum_{j=1}^{N}\big(\sigma_\gamma^2 - \tilde{\gamma}_j^2 - \tilde{\boldsymbol{\Sigma}}_{\gamma_j|\boldsymbol{y}_j}\big). \tag{A.18}$$

Setting Equation (A.18) equal to zero yields

$$\frac{\partial \log L}{\partial \sigma_\gamma^{-2}} = \frac{1}{2}\sum_{j=1}^{N}\big(\sigma_\gamma^2 - \tilde{\gamma}_j^2 - \tilde{\boldsymbol{\Sigma}}_{\gamma_j|\boldsymbol{y}_j}\big) = \mathbf{0}$$

$$\implies \sum_{j=1}^{N}\sigma_\gamma^2 = \sum_{j=1}^{N}\big(\tilde{\gamma}_j^2 - \tilde{\boldsymbol{\Sigma}}_{\gamma_j|\boldsymbol{y}_j}\big)$$

$$\implies \hat{\sigma}_\gamma^2 = \frac{1}{N}\sum_{j=1}^{N}\big(\tilde{\gamma}_j^2 - \tilde{\boldsymbol{\Sigma}}_{\gamma_j|\boldsymbol{y}_j}\big). \tag{A.19}$$

CITED LITERATURE

S. Afyouni, S. M. Smith, and T. E. Nichols. Effective degrees of freedom of the pearson's correlation coefficient under autocorrelation. *NeuroImage*, 199(1):609–625, 2019.

G. S. Alexopoulos. Mechanisms and treatment of late-life depression. *Translational Psychiatry*, 9:188, 2019.

G. S. Alexopoulos, M. J. Hoptman, D. Kanellopoulos, C. F. Murphy, K. O. Lim, and F. M. Gunning. Functional connectivity in the cognitive control network and the default mode network in late-life depression. *Journal of Affective Disorders*, 139(1):56–65, 2012.

Z. Y. Algamal and M. H. Lee. Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification. *Computers in Biology and Medicine*, 67:136–145, 2015.

C. Andreescu, M. Wu, M. A. Butters, J. Figurski, C. F. Reynolds, and H. J. Aizenstein. The default mode network in late-life anxious depression. *American Journal of Geriatric Psychiatry*, 19(11):980–983, 2011.

C. E. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2(6):1152–1174, 1974.

D. Attwell and C. Iadecola. The neural basis of functional brain imaging signals. *Trends in Neurosciences*, 25(12):621–625, 2002.

F. A. Azevedo, L. R. Carvalho, L. T. Grinberg, J. M. Farfel, R. E. Ferretti, R. E. Leite, W. Jacob Filho, R. Lent, and S. Herculano-Houzel. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *The Journal of Comparative Neurology*, 513(5):532–541, 2009.

R. Aziz and D. C. Steffens. What are the causes of late-life depression? *The Psychiatric clinics of North America*, 36(4):497–516, 2013.

A. Badhwar, A. Tama, C. Dansereaua, P. Orbana, F. Hoffstaedtere, and P. Bellec. Resting-state network dysfunction in Alzheimer's disease: A systematic review and meta-analysis. *Alzheimer's & Dementia (Amsterdam, Netherlands)*, 8:73–85, 2017.

P. A. Bandettini, A. Jesmanowicz, E. C. Wong, and J. S. Hyde. Processing strategies for time-course data sets in functional MRI of the human brain. *Magnetic Resonance in Medicine*, 30 (2):161–173, 1993.

P. J. Basser, J. Mattiello, and D. LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267, 1994.

S. Basu and S. Chib. Marginal likelihood and bayes factors for Dirichlet process mixture models. *Journal of the American Statistical Association*, 98(461):224–235, 2003.

A. Bechara, H. Damasio, and A. R. Damasio. Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex*, 10(3):295–307, 2000.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 57(1):289–300, 1995.

Y. Benjamini and Y. Hochberg. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, 25(1): 60–83, 2000.

D. K. Bhaumik, F. Jie, R. Nordgren, R. Bhaumik, and B. K. Sinha. A mixed-effects model for detecting disrupted connectivities in heterogeneous data. *IEEE Transactions on Medical Imaging*, 37(11):2381–2389, 2018a.

D. K. Bhaumik, Y. Song, D. Zhao, and O. Ajilore. Controlling false disrupted connectivities in neuroimaging studies. *Journal of Biostatistics and Biometrics*, 2018b. doi: 10.29011/JBSB-108.100008.

M. A. A. Binnewijzend, M. M. Schoonheim, E. Sanz-Arigita, A. M. Wink, W. M. van der Flier, N. Tolboom, S. M. Adriaanse, J. S. Damoiseaux, P. Scheltens, B. N. M. van Berckel, and F. Barkhof. Resting-state fMRI changes in Alzheimer's disease and mild cognitive impairment. *Neurobiology of Aging*, 33(9):2018–2028, 2014.

B. Biswal, F. Yetkin, V. Haughton, and J. S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34 (4):537–541, 1995.

D. Blackwell and J. B. MacQueen. Ferguson distribution via pólya urn scheme. *The Annals of Statistics*, 1(2):353–355, 1973.

D. M. Blei and M. I. Jordan. Variational inference for dirichlet process mixtures. *Bayesian Analysis*, 1(1):121–143, 2006.

D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

D. M. Blumberger, F. Vila-Rodriguez, K. E. Thorpe, K. Feffer, Y. Noda, P. Giacobbe, Y. Knyahnytska, S. H. Kennedy, R. W. Lam, Z. J. Daskalakis, and J. Downar. Effectiveness of theta burst versus high-frequency repetitive transcranial magnetic stimulation in patients with depression (THREE-D): a randomised non-inferiority trial. *Lancet*, 391:1683–1692, 2018.

S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1997.

T. T. Brunyé, J. M. Moran, A. Holmes, C. R. Mahoney, and H. A. Taylor. Non-invasive brain stimulation targeting the right fusiform gyrus selectively increases working memory for faces. *Brain and Cognition*, 113:32–39, 2017.

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

S. Cai, T. Chong, Y. Zhang, J. Li, K. M. von Deneen, J. Ren, M. Dong, L. Huang, and for the Alzheimer's Disease Neuroimaging Initiative. Altered functional connectivity of fusiform gyrus in subjects with amnestic mild cognitive impairment: A resting-state fMRI study. *Frontiers in Human Neuroscience*, 9:471, 2015.

T. T. Cai. Comment: Microarrays, empirical Bayes and the two-group model. *Statistical Science*, 23(1):29–33, 2008.

T. T. Cai and W. Sun. Large-scale global and simultaneous inference: Estimation and testing in very high dimensions. *Annual Review of Economics*, 9(1):411–439, 2017.

J. Cao and K. Worsley. The geometry of correlation fields with an application to functional connectivity of the brain. *The Annals of Applied Probability*, 9(4):1021–1057, 1999.

G. Casella and E. I. George. Explaining the Gibbs sampler. *The American Statistician*, 46(3): 167–174, 1992.

F. X. Castellanos, D. S. Margulies, C. Kelly, L. Q. Uddin, M. Ghaffari, A. Kirsch, D. Shaw, Z. Shehzad, A. Di Martino, B. Biswal, J. Sonuga-Barke, Edmund J.S.and Rotrosen, L. A. Adler, and M. P. Milhama. Cingulate-precuneus interactions: A new locus of dysfunction in adult attention-deficit/hyperactivity disorder. *Biological Psychiatry*, 63(3):332–337, 2008.

O. Y. Chén. The roles of statistics in human neuroscience. *Brain Science*, 9(8):194, 2019.

W. Cheng, E. T. Rolls, J. Qiu, W. Liu, Y. Tang, C.-C. Huang, X. Wang, J. Zhang, W. Lin, L. Zheng, J. Pu, S.-J. Tsai, A. C. Yang, C.-P. Lin, F. Wang, P. Xie, and J. Feng. Medial reward and lateral non-reward orbitofrontal cortex circuits change in opposite directions in depression. *Brain*, 139(12):3296–3309, 2016.

V. L. Cherkassky, R. K. Kana, T. A. Keller, and M. A. Just. Functional connectivity in a baseline resting-state network in autism. *Neuroreport*, 17(16):1687–1690, 2006.

V. K. M. Cheung, L. Meyer, A. D. Friederici, and S. Koelsch. The right inferior frontal gyrus processes nested non-local dependencies in music. *Scientific Reports*, 8:3822, 2018.

S. Chiang. *Hierachical Bayesian Models for Multimodal Neuroimaging*. PhD thesis, Rice University, 2016.

T. E. Conturo, R. C. McKinstry, E. Akbudak, and B. H. Robinson. Encoding of anisotropic diffusion with tetrahedral gradients: a general mathematical diffusion formalism and experimental results. *Magnetic Resonance in Medicine*, 35(3):399–412, 1996.

S. G. Costafreda. Pooling fMRI data: Meta-analysis, mega-analysis and multi-center studies. *Frontiers in Neuroinformatics*, 3:33, 2009.

R. V. Craiu and L. Sun. Choosing the lesser evil: Trade-off between false discovery rate and non-discovery rate. *Statistica Sinica*, 18(3):861–879, 2008.

D. B. Dahl and M. A. Newton. Multiple hypothesis testing by clustering treatment effects. *Journal of the American Statististical Association*, 102:517–526, 2007.

Z. Dai, C. Yan, K. Li, Z. Wang, J. Wang, M. Cao, Q. Lin, N. Shu, M. Xia, Y. Bi, and Y. He. Identifying and mapping connectivity patterns of brain network hubs in Alzheimer's disease. *Cerebral Cortex*, 25:3723–3742, 2015.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 39(1):1–38, 1977.

A. P. Dempster, D. B. Rubin, and R. K. Tsutakawa. Estimation in covariance components models. *Journal of the American Statistical Association*, 78(374):341–353, 1981.

M. J. Denwood. runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software*, 71(9):1–25, 2016. doi: 10.18637/jss.v071.i09.

S. Dudoit, J. P. Shaffer, and J. C. Boldrick. Multiple hypothesis testing in microarray experiments. *Statistical Science*, 18(1):71–103, 2003.

O. J. Dunn. Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52–64, 1961.

B. Efron. Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistical Association*, 99(465):96–104, 2004.

B. Efron. Size, power and false discovery rates. *The Annals of Statistics*, 35(4):1351–1377, 2007.

B. Efron. Microarrays, empirical bayes and two-groups model. *Statistical Science*, 23(1):1–22, 2008.

B. Efron and R. Tibshirani. Using specially designed exponential families for density estimation. *The Annals of Statistics*, 24(6):2431–2461, 1996.

B. Efron and R. Tibshirani. Empirical bayes methods and false discovery rates for microarrays. *Genetic Epidemiology*, 23:70–86, 2002.

B. Efron, R. Tibshirani, J. D. Storey, and V. Tusher. Empirical bayes analysis of a microarray experiment. *Journal of the American Statistical Association*, 96(456):1151–1160, 2001.

B. Efron, B. B. Turnbull, and B. Narasimhan. *locfdr Vignette: Complete Help Documentation Including Usage Tips and Simulation Example*, 2015. URL `https://cran.r-project.org/web/packages/locfdr/`. R package version 1.1-8.

I. Elbinoune, B. Amine, S. Shyen, S. Gueddari, R. Abouqal, and N. Hajjaj-Hassouni. Chronic neck pain and anxiety-depression: prevalence and associated risk factors. *Pan African Medical Journal*, 24(89):8831, 2016.

J. M. Ellison, H. H. Kyomen, and D. G. Harper. Depression in later life: An overview with treatment recommendations. *Psychiatric Clinics of North America*, 35:203–229, 2012.

S. Elmer. Broca pars triangularis constitutes a "hub" of the language-control network during simultaneous language translation. *Frontier in Human Neuroscience*, 10:491, 2016.

A. D. Elster. Questions and Answers in MRI. `http://mri-q.com/index.html`. Accessed: 2020-01-06.

Encyclopaedia Britannica. Hemoglobin. `https://www.britannica.com/science/hemoglobin`, 2020.

T. S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.

P. Fettes, L. Schulze1, and J. Downar. Cortico-striatal-thalamic loop circuits of the orbitofrontal cortex: Promising therapeutic targets in psychiatric illness. *Frontiers in Systems Neuroscienc*, 11, 2017.

B. Fischl. FreeSurfer. *NeuroImage*, 62(2):774–781, 2012.

P. T. Fox and M. E. Raichle. Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proceedings of the National Academy of Sciences*, 83(4):1140–1144, 1986.

P. T. Fox, M. E. Raichle, M. A. Mintun, and C. Dence. Nonoxidaive glucose consumption during focal physiologic neural activit. *Science*, 241(4864):462–464, 1988.

A. R. Franco, J. Ling, A. Caprihan, V. Calhoun, R. E. Jung, G. Heileman, and A. R. Mayer. Multimodal and multi-tissue measures of connectivity revealed by joint independent component analysis. *IEEE Journal of Selected Topics in Signal Processing*, 2:986–997, 2008.

A. D. Friederici, B. Opitz, and D. Y. von Cramon. Segregating semantic and syntactic aspects of processing in the human brain: an fMRI investigation of different word types. *Cerebral Cortex*, 10:698–705, 2000.

J. H. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.

K. J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2:56–78, 1994.

K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. J. Frakowiak. Functional connectivity: The principal-component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow and Metabolism*, 13(1):5–14, 1993.

K. J. Friston, A. P. Holmes, K. J. Worsley, J.-B. Poline, C. D. Frith, and S. J. Richard Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2:189–210, 1994a.

K. J. Friston, P. Jezzard, and R. Turner. Analysis of functional MRI time-series. *Human Brain Mapping*, 1:153–171, 1994b.

J. J. GadElkarim, D. Schonfeld, O. Ajilore, L. Zhan, A. F. Zhang, J. D. Feusner, P. M. Thompson, T. J. Simon, A. Kumar, and A. D. Leow. A framework for quantifying node-level community structure group differences in brain connectivity networks. *Medical Image Computing and Computer-Assisted Intervention*, pages 196–203, 2012.

A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–533, 2006.

A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.

A. Gelman, A. Jakulin, M. G. Pittau, and Y.-S. Su. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4): 1360–1383, 2008.

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, Third edition, 2013.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6 (6):721–741, 1984.

C. R. Genovese and L. Wasserman. Operating characteristics and extensions of the FDR procedure. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64(3): 499–517, 2002a.

C. R. Genovese and L. Wasserman. Bayesian and frequentist multiple testing. In *Bayesian Statistics 7*, pages 145–161. Oxford University Press, 2002b.

C. R. Genovese and L. Wasserman. A stochastic process approach to false discovery control. *The Annals of Statistics*, 32(3):1035–1061, 2004.

C. R. Genovese, N. A. Lazar, and T. Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, 15(4):870–878, 2002.

M. S. George, S. H. Lisanby, D. Avery, W. M. McDonald, V. Durkalski, M. Pavlicova, B. Anderson, Z. Nahas, P. Bulow, P. Zarkowski, P. E. Holtzheimer, T. Schwartz, and H. A. Sackeim. Daily left prefrontal transcranial magnetic stimulation therapy for major depressive disordera sham-controlled randomized trial. *Archives of General Psychiatry*, 67(5):507–516, 2010.

J. Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. Bernardo, J. Berger, A. Dawid, and A. Smith, editors, *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*. Clarendon Press, First edition, 1992.

Z. Ghahramani. Non-parametric Bayesian methods. Uncertainty in Artifical Intelligence Tutorial, 2005. URL `http://mlg.eng.cam.ac.uk/zoubin/talks/uai05tutorial-b.pdf`.

N. Ghosal. *Clustered-Temporal Bayesian Models for Brain Connectivity in Neuroimaging Data*. PhD thesis, University of Illinois at Chicago, 2019.

J. A. Glover and S. Srinivasan. Assessment and treatment of late-life depression. *Journal of Clinical Outcomes Management*, 24(3):135–144, 2017.

R. Gopalan and D. A. Berry. Bayesian multiple comparisons using Dirichlet process priors. *Journal of the American Statistical Association*, 93(443):1130–1139, 1998.

A. Gossmann, P. Zille, V. Calhoun, and Y.-P. Wang. Fdr-corrected sparse canonical correlation analysis with applications to imaging genomics. *IEEE Transactions on Medical Imaging*, 87 (8):1761–1774, 2018.

C. Goutte, P. Toft, E. Rostrup, F. Nielsen, and L. K. Hansen. Multimodal and multi-tissue measures of connectivity revealed by joint independent component analysis. *NeuroImage*, 9: 298–319, 1999.

M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon. Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, 100(1):253–258, 2003.

M. D. Greicius, B. H. Flores, V. Menon, G. H. Glover, H. B. Solvason, H. M. Kenna, A. L. Reiss, and A. F. Schatzberg. Resting-state functional connectivity in major depression: Abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biological Psychiatry*, 62(5):429–437, 2007.

S. Grimm, J. Beck, D. Schuepbach, D. Hell, P. Boesiger, F. Bermpohl, L. Niehaus, H. Boeker, and G. Northoff. Imbalance between left and right dorsolateral prefrontal cortex in major depression is linked to negative emotional judgment: An fMRI study in severe major depressive disorder. *Biological Psychiatry*, 63(4):369–376, 2008.

J. P. Hamilton, A. Etkin, D. J. Furman, M. G. Lemus, R. F. Johnson, and I. H. Gotlib. Functional neuroimaging of major depressive disorder: A meta-analysis and new integration of baseline activation and neural response data. *American Journal of Psychiatry*, 169(7): 693–703, 2012.

M. Hamilton. A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, 23(1):56–62, 1960.

D. S. Hasin, R. D. Goodwin, F. S. Stinson, and B. F. Grant. Epidemiology of major depressive disorder: Results from the national epidemiologic survey on alcoholism and related conditions. *Archives of General Psychiatry*, 62(10):1097–1106, 2005.

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

D. Hedeker and R. D. Gibbons. *Longitudinal Data Analysis*. Wiley, Hoboken, 2006.

G. Herbet, G. Lafargue, F. Almairac, S. Moritz-Gasser, F. Bonnetblanc, and H. Duffau. Disrupting the right pars opercularis with electrical stimulation frees the song: case report. *Journal of Neurosurgery*, 123(6):1351–1614, 2015.

Y. Hochberg. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75 (4):800–802, 1988.

S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

G. Hommel. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75(2):383–386, 1988.

G. Hommel. A comparison of two modified Bonferroni procedures. *Biometrika*, 76(3):624–625, 1989.

C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J. P. Thiran, R. Meuli, and P. Hagmann. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, 106(6):2035–2040, 2009.

J. S. Hyde and A. Jesmanowicz. Cross-correlation: An fMRI signal-processing strategy. *NeuroImage*, 62(2):848–851, 2012.

N. Ihara, K. Wakaizumi, D. Nishimura, J. Kato, T. Yamada, T. Suzuki, S. Hashiguchi, Y. Terasawa, S. Kosugi, and H. Morisaki. Aberrant resting-state functional connectivity of the dorsolateral prefrontal cortex to the anterior insula and its association with fear avoidance belief in chronic neck pain patients. *PLoS One*, 14(8):e0221023, 2019.

H. Ishwaran and L. F. James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173, 2001.

H. Ishwaran and L. F. James. Approximate Dirichlet process computing in finite normal mixtures: Smoothing and prior information. *Journal of Computational and Graphical Statistics*, 11(3):508–532, 2002.

S. Japee, K. Holiday, M. D. Satyshur, I. Mukai, and L. G. Ungerleider. A role of right middle frontal gyrus in reorienting of attention: a case study. *Frontiers in Systems Neuroscience*, 9: 23, 2015.

H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186:453–461, 1946.

B. J. Jellison, A. S. Field, J. Medow, M. Lazar, M. S. Salamat, and A. L. Alexander. Diffusion tensor imaging of cerebral white matter: A pictorial review of physics, fiber tract anatomy, and tumor imaging patterns. *American Journal of Neuroradiology*, 25(3):356–369, 2004.

H. Jiang, P. C. M. van Zijl, J. Kim, G. D. Pearlson, and S. Mori. DtiStudio: Resource program for diffusion tensor computation and fiber bundle tracking. *Computer Methods and Programs in Biomedicine*, 81(2):106–116, 2006.

F. Jie. *Statistics Methodologies for Neuroconnectivity Analysis Using fMRI Data in Autism*. PhD thesis, University of Illinois at Chicago, 2016.

J. Jin and T. T. Cai. Estimating the null and the proportion of nonnull effects in large-scale multiple comparisons. *Journal of the American Statistical Association*, 102:495–506, 2007.

R. H. Kaiser, J. R. Andrews-Hanna, T. D. Wager, and D. A. Pizzagalli. Large-scale network dysfunction in major depressive disorder a meta-analysis of resting-state functional connectivity. *JAMA Psychiatry*, 72(6):603–611, 2015.

N. Kanwisher and G. Yovel. The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 361(1476):2109–2128, 2006.

N. Kanwisher, J. McDermott, and M. M. Chun. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17(11): 4302–4311, 1997.

H. Keren, G. O'Callaghan, P. Vidal-Ribas, G. A. Buzzell, M. A. Brotman, E. Leibenluft, P. M. Pan, L. Meffert, A. Kaiser, S. Wolke, D. S. Pine, and A. Stringaris. Reward processing in depression: A conceptual and meta-analytic review across fMRI and EEG studies. *The American Journal of Psychiatry*, 175(11):1111–1120, 2018.

S. Kim, D. B. Dahl, and M. Vannucci. Spiked Dirichlet process prior for Bayesian multiple hypothesis testing in random effects models. *Bayesian Analysis*, 4(4):707–732, 2009.

M. A. Koch, D. G. Norris, and M. Hund-Georgiadis. An investigation of functional and anatomical connectivity using magnetic resonance imaging. *NeuroImage*, 16(1):241–250, 2002.

M. Koenigs and J. Grafman. The functional neuroanatomy of depression: Distinct roles for ventromedial and dorsolateral prefrontal cortex. *Behavioural Brain Research*, 201:239–243, 2009.

T. A. Kolesar, E. Bilevicius, and J. Kornelsen. Salience, central executive, and sensorimotor network functional connectivity alterations in failed back surgery syndrome. *Scandinavian Journal of Pain*, 16:10–14, 2017.

J. Kornelsen, U. Sboto-Frankenstein, T. McIver, P. Gervai, P. Wacnik, N. Berrington, and B. Tomanek. Default mode network functional connectivity altered in failed back surgery syndrome. *The Journal of Pain*, 14(5):483–491, 2013.

S. Kotz, N. Balakrishnan, and N. L. Johnson. Dirichlet and inverted Dirichlet distributions. In *Continuous Multivariate Distributions, Volume 1: Models and Applications*, pages 485–527. John Wiley & Sons, Inc, 2000. ISBN 0-471-18387-3.

A. K. Lebedeva, E. Westman, T. Borza, M. K. Beyer, K. Engedal, D. Aarsland, G. Selbaek, A. K. Haberg, and for the Alzheimer's Disease Neuroimaging Initiative. MRI-based classification models in prediction of mild cognitive impairment and dementia in late-life depression. *Frontier in Aging Neuroscience*, 9:13, 2017.

W. H. Lee and S. Franggou. Linking functional connectivity and dynamic properties of resting-state networks. *Scientific Reports*, 7:16610, 2017.

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

E. L. Lehmann and J. P. Romano. Generalizations of the familywise error rate. *The Annals of Statistics*, 33(3):1138–1154, 2005.

F. C. Leone, L. S. Nelson, and R. B. Nottingham. The folded normal distribution. *Technometrics*, 3(4):543–550, 1961.

B.-J. Li, K. Friston, M. Mody, H.-N. Wang, H.-B. Lu, and D.-W. Hu. A brain network model for depression: From symptom understanding to disease intervention. *CNS Neuroscience & Therapeutics*, 24(11):1004–1019, 2018.

K. Li, L. Guo, J. Nie, G. Li, and T. Liu. Review of methods for functional brain connectivity detection using fMRI. *Computerized Medical Imaging and Graphics*, 33(2):131–139, 2009.

S. Li, N. Hu, W. Zhang, B. Tao, J. Dai, Y. Gong, Y. Tan, D. Cai, and S. Lui. Dysconnectivity of multiple brain networks in schizophrenia: A meta-analysis of resting-state functional connectivity. *Frontiers in Psychiatry*, 10:482, 2019.

M. Liang, Y. Zhou, T. Jiang, Z. Liu, L. Tian, H. Liu, and Y. Hao. Widespread functional disconnectivity in schizophrenia with resting-state functional magnetic resonance imaging. *Neuroreport*, 17(2):209–213, 2006.

J. K. Lindsey. Comparison of probability distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 36(1):38–47, 1974a.

J. K. Lindsey. Construction and comparison of statistical models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 36(3):418–425, 1974b.

J. S. Liu, F. Liang, and W. H. Wong. The multiple-try method and local optimization in metropolis sampling. *Journal of the American Statistical Association*, 95:121–134, 2000.

H. F. Lopes and R. Dias. Bayesian mixture of parametric and nonparametric density estimation: A misspecification problem. *Brazilian Review of Econometrics*, 31(1):19, 2011.

D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, 28:3049–3067, 2009.

Y. Luo, F. Kong, S. Qi, X. You, and X. Huang. Resting-state functional connectivity of the default mode network associated with happiness. *Social Congnitive and Affective Neuroscience*, 11(3):516–524, 2016.

Y. Ma. Neuropsychological mechanism underlying antidepressant effect: a systematic meta-analysis. *Molecular Psychiatry*, 20:311–319, 2015.

G. R. Mangun, M. H. Buonocore, M. Girelli, and A. P. Jha. ERP and fMRI measures of visual spatial selective attention. *Human Brain Mapping*, 6(5-6):383–389, 1998.

M. Martino, P. Magioncalda, Z. Huang, B. Conio, N. Piaggio, N. W. Duncan, G. Rocchi, A. Escelsior, V. Marozzi, A. Wolff, M. Inglese, M. Amore, and G. Northoff. Contrasting variability patterns in the default mode and sensorimotor networks balance in bipolar depression and mania. *Proceedings of the National Academy of Sciences*, 113(17):4824–4829, 2016.

K. Matsuo, D. C. Glahn, M. A. M. Peluso, J. P. Hatch, E. S. Monkul, P. Najt, M. Sanches, F. Zamarripa, J. Li, J. L. Lancaster, P. T. Fox, J.-H. Gao, and J. C. Soares. Prefrontal hyperactivation during working memory task in untreated individuals with major depressive disorder. *Molecular Psychiatry*, 12:158–166, 2007.

V. Menon. Salience network. In A. W. Toga, editor, *Brain Mapping: An Encyclopedic Reference*, volume 2, pages 597–611. Academic Press: Elsevier, First edition, 2015.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6): 1087–1092, 1953.

H. H. Mitchell, T. S. Hamilton, F. Steggerda, and H. W. Bean. The chemical composition of the adult human body and its bearing on the biochemistry of growth. *The Journal of Biological Chemistry*, 158:625–637, 1945.

T. J. Mitchell and J. J. Beauchamp. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032, 1988.

S. Mori, B. J. Crain, V. P. Chacko, and P. C. M. Van Zijl. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology*, 45(2):265–269, 1999.

P. Mukherjee, J. I. Berman, S. W. Chung, C. P. Hess, and R. G. Henry. Diffusion tensor MR imaging and fiber tractography: Theoretic underpinnings. *American Journal of Neuroradiology*, 29(4):632–641, 2008.

J. W. Murrough, K. A. Collins, J. Fields, K. E. DeWilde, M. L. Phillips, S. J. Mathew, E. Wong, C. Y. Tang, D. S. Charney, and D. V. Iosifescu. Regulation of neural responses to emotion perception by ketamine in individuals with treatment-resistant major depressive disorder. *Translational Psychiatry*, 5:e509, 2015.

S. Nakagawa. A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral Ecology*, 15(6):1044–1045, 2004.

National Institute of Mental Health. Major Depression among Adults. `https://www.nimh.nih.gov/health/statistics/major-depression.shtml`, 2017. Accessed: 2020-01-06.

R. M. Neal. Bayesian mixture modeling. In C. Smith, G. J. Erickson, and P. O. Neudorfer, editors, *Maximum Entropy and Bayesian Methods. Fundamental Theories of Physics*, pages 197–211. Springer, Dordrecht, 1992.

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.

R. Nuzzo. Scientific method: statistical errors. *Nature*, 506:150–152, 2014.

S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990.

T. L. B. Pape, D. K. Bhaumik, and R. Bhaumik. A pilot trial examining the merits of combining amantadine and repetitive transcranial magnetic stimulation as an intervention for persons with disordered consciousness after TBI. Under review, 2020.

H.-J. Park and K. Friston. Structural and functional brain networks: From connections to cognition. *Science*, 342:1238411, 2013.

R. S. Patel, F. D. Bowman, and J. K. Rilling. A bayesian approach to determining connectivity of the human brain. *Human Brain Mapping*, 27(3):267–276, 2006.

L. Pauling and C. D. Coryell. The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. *Proceedings of the National Academy of Sciences*, 22(4):210–216, 1936.

Y. Pawitan, S. Michiels, S. Koscielny, A. Gusnanto, and A. Ploner. Multiple hypothesis testing in microarray experiments. *Bioinformatics*, 21(13):3017–3024, 2005.

J. S. Perrin, S. Merzb, D. M. Bennetta, J. Curriea, D. J. Steelec, I. C. Reida, and C. Schwarzbauerb. Electroconvulsive therapy reduces frontal cortical connectivity in severe depressive disorder. *Proceedings of the National Academy of Sciences*, 109(14):5464–5468, 2012.

S. K. Peters, K. Dunlop1, and J. Downa. Cortico-striatal-thalamic loop circuits of the salience network: A central pathway in psychiatric disease and treatment. *Frontiers in Systems Neuroscienc*, 10:104, 2016.

C. Pierpaoli and P. J. Basser. Toward a quantitative assessment of diffusion anisotropy. *Magnetic Resonance in Medicine*, 36(6):893–906, 1996.

Radiopaedia. BOLD imaging. `http://radiopaedia.org`. Accessed: 2020-01-06.

M. E. Raichle. The brain's default mode network. *Annual Review of Neuroscience*, 38:433–447, 2015.

M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman. A default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(2): 676–682, 2001.

C. E. Rasmussen. The infinite gaussian mixture model. In *In Advance in Neural Information Prcessing Systems 12*, pages 554–560. MIT Press, 2000.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, Third edition, 1999.

V. Ročková and E. I. George. The spike-and-slab LASSO. *Journal of the American Statistical Association*, 113(521):431–444, 2018.

H. W. Roh, C. H. Hong, H. K. Lim, K. J. Chang, H. Kim, N.-R. Kim, J. W. Choi, K. S. Lee, S.-M. Cho, B. Park, and S. J. Son. A 12-week multidomain intervention for late-life depression: a communitybased randomized controlled trial. *Journal of Affective Disorders*, 263:437–444, 2020.

E. T. Rolls. A non-reward attractor theory of depression. *Neuroscience and Biobehavioral Reviews*, 68:47–58, 2016.

E. T. Rolls. The orbitofrontal cortex and emotion in health and disease, including depression. *Neuropsychologia*, 128:14–43, 2019.

J. B. Rowe. Connectivity analysis is essential to understand neurological disorders. *Frontiers in Systems Neuroscience*, 4:144, 2010.

E. Rykhlevskaia, G. Gratton, and M. Fabiani. Combining structural and functional neuroimaging data for studying brain connectivity: A review. *Psychophysiology*, 45:173–187, 2008.

A. Salles, J. G. Bjaalie, K. Evers, M. Farisco, B. Tyr Fothergill, M. Guerrero, H. Maslen, J. Muller, T. Prescott, B. C. Stahl, H. Walter, K. Zilles, and K. Amunts. The human brain project: Responsible brain research for the benefit of society. *Neuron*, 101(3):380–384, 2019.

W. W. Seeley, V. Menon, A. F. Schatzberg, J. Keller, G. H. Glover, H. Kenna, A. L. Reiss, and M. D. Greicius. Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience*, 27(9):2349–2356, 2007.

J. Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

Y. I. Sheline, J. L. Price, Z. Yan, and M. A. Mintun. Resting-state functional MRI in depression unmasks increased connectivity between networks via the dorsal nexus. *Proceedings of the National Academy of Sciences*, 107(24):11020–11025, 2009.

S. M. Sherman and R. W. Guillery. The role of the thalamus in the flow of information to the cortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 357(1428):1695–1708, 2002.

C. E. Snow and R. C. Abram. The indirect costs of late-life depression in the united states: A literature review and perspective. *Geriatrics*, 1(4), 2016.

Y. Song. *Sample Size Determination for High Dimensional Neuroimaging Studies Controlling False Discovery Rate*. PhD thesis, University of Illinois at Chicago, 2016.

R. L. Spitzer, J. B. W. Williams, M. Gibbon, and M. B. First. The structured clinical interview for DSM-III-R (SCID). i: History, rationale, and description. *Archives of General Psychiatry*, 49(8):624–629, 1992.

O. Sporns, G. Tononi, and R. Kötter. The human connectome: A structural description of the human brain. *PLoS Computational Biology*, 1(4):e42, 2005.

D. Sridharan, D. J. Levitin, and V. Menon. A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences*, 105(34):12569–12574, 2008.

D. C. Steffens, I. Skoog, M. C. Norton, A. D. Hart, J. T. Tschanz, B. L. Plassman, B. W. Wyse, K. A. Welsh-Bohmer, and J. C. Breitner. Prevalence of depression and its treatment in an elderly population: the Cache county study. *Archives of General Psychiatry*, 57(6):601–607, 2000.

J. D. Storey. A direct approach to false discovery rates. *Journal of the Royal Statistics Society. Series B (Statistical Methodology)*, 64(3):479–498, 2002.

J. D. Storey. The positive false discovery rate: A BAYESIAN interpretation and the q-value. *The Annals of Statistics*, 31(6):2013–2035, 2003.

J. D. Storey and R. Tibshirani. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445, 2003.

J. D. Storey, A. J. Bass, A. Dabney, and D. Robinson. *qvalue: Q-value estimation for false discovery rate control*, 2019. URL `http://github.com/jdstorey/qvalue`. R package version 2.14.1.

A. Stuhrmann, T. Suslow, and U. Dannlowski. Facial emotion processing in major depression: a systematic review of neuroimaging findings. *Biology of Mood and Anxiety Disorders*, 1: 1–17, 2011.

A. Stuhrmann, T. Suslow, and U. Dannlowski. Redefining the role of Broca's area in speech. *Proceedings of the National Academy of Sciences*, 112(9):2871–2875, 2015.

H. Su, C. Zuo, H. Zhang, F. Jiao, B. Zhang, W. Tang, D. Geng, Y. Guan, and S. Shi. Regional cerebral metabolism alterations affect resting-state functional connectivity in major depressive disorder. *Quantitative Imaging in Medicine and Surgery*, 8(9):910–924, 2018.

F. T. Sun, L. M. Miller, and M. D'Esposito. Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. *NeuroImage*, 21:647–658, 2004.

W. Sun and T. T. Cai. Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association*, 102(479):901–912, 2007.

W. Sun and T. T. Cai. Large-scale multiple testing under dependence. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 71(2):393–424, 2009.

D. Szucs and J. P. Ioannidis. Sample size evolution in neuroimaging research: an evaluation of highly-cited studies (1990-2012) and of latest practices (2017-2018) in high-impact journals. *bioRxiv*, 2019.

S. Tang, L. Lu, L. Zhang, X. Hu, X. Bu, H. Li, X. Hu, Y. Gao, Z. Zeng, Q. Gong, and X. Huang. Abnormal amygdala resting-state functional connectivity in adults and adolescents with major depressive disorder: A comparative metaanalysis. *EBioMedicine*, 36:436–445, 2018.

Y. W. Teh. Dirichlet process. In C. Sammut and G. I. Webb, editors, *Encyclopedia of Machine Learning*, pages 280–287. Springer US, Boston, MA, 2010. ISBN 978-0-387-30164-8. doi: 10.1007/978-0-387-30164-8_219. URL `https://doi.org/10.1007/978-0-387-30164-8_219`.

Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101:1566–1581, 2006.

S. J. Teipel, M. J. Grothe, C. D. Metzger, T. Grimmer, C. Sorg, M. Ewers, N. Franzmeier, E. Meisenzahl, S. Klöppel, V. Borchardt, M. Walter, and M. Dyrba. Robust detection of impaired resting state functional connectivity networks in Alzheimer's disease using elastic net regularized regression. *Frontiers in Aging Neuroscience*, 8:318, 2017.

J. B. Thomas, M. R. Brier, R. J. Bateman, A. Z. Snyder, T. L. Benzinger, C. Xiong, M. Raichle, D. M. Holtzman, R. A. Sperling, R. Mayeux, B. Ghetti, J. M. Ringman, S. Salloway, E. McDade, M. N. Rossor, S. Ourselin, P. R. Schofield, C. L. Masters, R. N. Martins, M. W. Weiner, P. M. Thompson, N. C. Fox, R. A. Koeppe, J. Jack, Clifford R., C. A. Mathis, A. Oliver, T. M. Blazey, K. Moulder, V. Buckles, R. Hornbeck, J. Chhatwal, A. P. Schultz, A. M. Goate, A. M. Fagan, N. J. Cairns, D. S. Marcus, J. C. Morris, and B. M. Ances. Functional connectivity in autosomal dominant and late-onset Alzheimer disease. *JAMA Neurology*, 71 (9):1111–1122, 2014.

R. Tibshirani. Regularization shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58(1):267–288, 1996.

G. Tononi, O. Sporns, and G. M. Edelman. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences*, 91(11):5033–5037, 1994.

A. Torkamani, A. A. Scott-Van Zeeland, E. J. Topol, and N. J. Schork. Annotating individual human genomes. *Genomics*, 98(4):233–241, 2011.

M. Trimble and D. Hesdorffer. Music and the brain: the neuroscience of music and musical appreciation. *BJPsych International*, 14(2):28–31, 2017.

L. Q. Uddin. Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscienc*, 16:55–61, 2015.

L. Q. Uddin and V. Menon. The anterior insula in autism: Under-connected and under-examined. *Neuroscience and Biobehavioral Reviews*, 33(8):1198–1203, 2009.

L. Q. Uddin and V. Menon1. Introduction to special topic – resting-state brain activity: Implications for systems neuroscience. *Frontiers in Systems Neuroscience*, 4:37, 2010.

L. Q. Uddin, A. Kelly, B. B. Biswal, F. X. Castellanos, and M. P. Milham. Functional connectivity of default mode network components: Correlation, anticorrelation, and causality. *Human Brain Mapping*, 30:625–637, 2009.

L. Q. Uddin, K. Supekar, C. J. Lynch, A. Khouzam, J. Phillips, C. Feinstein, S. Ryali, and V. Menon. Salience network–based classification and prediction of symptom severity in children with autism. *JAMA Psychiatry*, 70(8):869–879, 2013.

M. P. van den Heuvel and H. E. Hulshoff Pol. Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, 20(8):519–534, 2010.

D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, K. Ugurbil, and WU-Minn HCP Consortium. The WU-Minn human connectome project: An overview. *NeuroImage*, 80(4):62–79, 2013.

Q. Wang, J. S. Poh, D. J. Wen, B. F. P. Broekman, Y.-S. Chong, F. Yap, L. P. Shek, P. D. Gluckman, M. V. Fortier, and A. Qi. Functional and structural networks of lateral and medial orbitofrontal cortex as potential neural pathways for depression in childhood. *Depression & Anxiety*, 36(4):365–374, 2018.

R. W. Y. Wang, W.-L. Chang, S.-W. Chuang, and I.-N. Liu. Posterior cingulate cortex can be a regulatory modulator of the default mode network in task-negative state. *Scientific Reports*, 9:7565, 2019.

L. Wasserstein, Ronald and N. A. Lazar. The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2):129–133, 2016.

M. M. Weissman, M. L. Bruce, P. J. Leaf, L. P. Florio, and C. Holzer. Affective disorders. In L. N. Robins and D. N. Regier, editors, *Psychiatric disorders in America: The Epidemiologic Catchment Study*, pages 53–80. New York: The Free Press, First edition, 1991.

S. Whitfield-Gabrieli and A. Nieto-Castanon. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connectivity*, 2(3):125–141, 2012.

N. D. Woodward, M. Giraldo-Chica, B. Rogers, and C. J. Cascio. Thalamocortical dysconnectivity in autism spectrum disorder: An analysis of the autism brain imaging data exchange. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 2:76–84, 2017.

World Health Organization. Ageing and Health. `https://www.who.int/news-room/fact-sheets/detail/ageing-and-health`, 2018.

World Health Organization. Depression Fact Sheet. `https://www.who.int/news-room/fact-sheets/detail/depression`, 2019.

W. Xue, F. D. Bowman, A. V. Pileggi, and A. R. Mayer. A multimodal approach for determining brain networks by jointly modeling functional and structural connectivity. *Frontiers in Computational Neuroscience*, 9:22, 2015.

F. Z. Yetkin, R. N. Rosenberg, M. F. Weiner, P. D. Purdy, and C. M. Cullum. FMRI of working memory in patients with mild cognitive impairment and probable Alzheimer's disease. *European Radiology*, 16:193–206, 2006.

G. S. Yuen, F. M. Gunning-Dixon, M. J. Hoptman, B. AbdelMalak, A. R. McGovern, J. K. Seirup, and G. S. Alexopoulos. The salience network in the apathy of late-life depression. *International Journal of Geriatric Psychiatry*, 29(11):1116–1124, 2014.

R. Zablocki, A. Schork, R. Levine, O. Andreassen, A. Dale, C. Fan, and W. Thompson. Covariate-modulated local false discovery rate for genome-wide association studies. *Bioinformatics*, 30(15):2098–2104, 2014.

R. Zablocki, R. Levine, A. Schork, S. Xu, Y. Wang, C. Fan, and W. Thompson. Semi-parametric covariate-modulated local false discovery rate for genome-wide association studies. *bioRxiv*, 2017. doi: 10.1101/183384.

R. W. Zablocki. *Large-Scale Inference Incorporating Covariates and Network Dependence, with Application to Genome-Wide Association Studies*. PhD thesis, San Diego State Universityo, 2016.

A. Zhang, A. Leow, O. Ajilore, M. Lamar, S. Yang, J. Joseph, J. Medina, L. Zhan, and A. Kumar. Quantitative tract-specific measures of uncinate and cingulum in major depression using diffusion tensor imaging. *Neuropsychopharmacology*, 37:959–967, 2012.

M. J. Zhang, F. Xia, and J. Zou. Fast and covariate-adaptive method amplifies detection power in large-scale multiple hypothesis testing. *Nature Communications*, 10:3433, 2019.

Q. Zhao, Z. N. Swati, H. Metmer, X. Sang, and J. Lu. Investigating executive control network and default mode network dysfunction in major depressive disorder. *Neuroscience Letters*, 701:154–161, 2019.

W. Zhao. *Statistical Methodologies for Group Comparisons of Brain Connectivity using Multimodal Neuroimaging Data*. PhD thesis, University of Illinois at Chicago, 2014.

Y. Zhou, C. Yu, H. Zheng, Y. Liu, M. Song, W. Qin, K. Li, and T. Jiang. Increased neural resources recruitment in the intrinsic organization in major depression. *Journal of Affective Disorders*, 121:220–230, 2010.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(2):301–320, 2005.

VITA

NAME Yue Wang

EDUCATION

B.S., Ecology, College of Marine Life Sciences, Ocean University of China, Qingdao, China, 1998

Ph.D., Major in Genetics and Molecular Biology, Minor in Statistics, School of Biological Sciences, University of Nebraska at Lincoln, 2005

M.S., Statistics, Department of Statistics, University of Nebraska at Lincoln, 2007

Ph.D., Biostatistics, School of Public Health, University of Illinois at Chicago, 2020

TEACHING

Teaching Assistant, General Biology, School of Biological Sciences, University of Nebraska at Lincoln, Spring 2003, Spring 2002, Fall 2001

Teaching Assistant, General Genetics, School of Biological Sciences, University of Nebraska at Lincoln, Spring 2005, Fall 2004, Fall 2002

RESEARCH

Research Assistant, School of Biological Sciences, College of Medicine, University of Illinois at Chicago, 2000 - 2001, Summer 2002, 2004 and 2005

Research Assistant, Department of Statistics, University of Nebraska at Lincoln, August 2005 - July 2007

FELLOWSHIP

Bioinformatics Interdisciplinary Research Scholars Fellowship, University of Nebraska at Lincoln, June 2003 – June 2004

Milton E. Mohr Fellowship, University of Nebraska at Lincoln, July 2006

PUBLICATIONS

Y. Wang, A.B. Salmon, and L.G. Harshman. A cost of reproduction: oxidative stress susceptibility is associated with increased egg production in *Drosophila melanogaster*. *Experimental Gerontology*, 36:1349-1359, 2001.

D. S. Richard, R. Rybczynski, T.G. Wilson, Y. Wang, M.L. Wayne, Y. Zhou, L. Partridge, and L.G. Harshman. Insulin signaling is necessary for vitellogenesis in *Drosophila melanogaster* independent of the roles of juvenile hormone and ecdysteroids: female sterility of the chico1 insulin signaling mutation is autonomous to the ovary. *Journal of Insect Physiology*, 51(4): 455-464, 2005.

L.G. Harshman, H.G. Müller, X. Liu, Y. Wang and J.R. Carey. The symmetry of longevity. *The journals of gerontology. Series A, Biological sciences and medical sciences*, 60:1233-1237, 2005.

Y. Wang, D. Pot, S.D. Kachman, S.V. Nuzhdin, and L.G. Harshman. A quantitative trait locus analysis of natural genetic variation for oxidative stress survival of *Drosophila melanogaster*. *Journal of Heredity*, 97(4): 355-366, 2006.