

Statistical Methodologies for Group Comparisons of Brain Connectivity using Multimodal Neuroimaging Data

BY

WEIHAN ZHAO

BS, Nanjing University, Nanjing, China, 2002

PhD, University of Illinois at Chicago, Chicago, Illinois, 2009

MS, University of Illinois at Chicago, Chicago, Illinois, 2010

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, 2014

Chicago, Illinois

Defense Committee:

Dulal K. Bhaumik, PhD, Professor of Biostatistics and Psychiatry, Advisor and Chair

Olusola A. Ajilore, MD PhD, Assistant Professor of Psychiatry

Sally A. Freels, PhD, Associate Professor of Biostatistics

Emil F. Coccaro, MD, EC Manning Professor and Chairman of Psychiatry

The University of Chicago

Theresa L. Pape, DrPH, Deputy Assistant Chief of Staff and Clinical Neuroscientist

Edward Hines Jr. Veterans Affairs Hospital

To my wife and my daughters

ACKNOWLEDGMENTS

It is my pleasure to have the opportunity now to express my gratitude to all the people who have been accompanying and supporting me during my thesis work.

First I would like to express my deepest gratitude to my thesis advisor, Dr. Dulal Bhaumik, for providing a stimulating environment that has helped me grow as a biostatistician. I have benefited greatly from Dr. Bhaumik's enthusiasm, inspiration, and patience and I am especially grateful for his willingness to provide help when needed, no matter what the circumstance.

I would like to thank other members of my thesis committee, Dr. Olusola Ajilore, Dr. Emil Coccaro, Dr. Sally Freels, and Dr. Theresa Pape, for their guidance and help during my thesis work. This thesis would not have become what it is today without their valuable insights and suggestions. Especially I want to thank Dr. Ajilore for providing the neuroimaging data and sharing methods and ideas for analyzing this type of data.

I want to acknowledge the faculty and staff members of the Epidemiology and Biostatistics Division at UIC School of Public Health for their great support that has made my graduate studies an enjoyable experience. I also want to thank all the fellow students in the division for their friendship and help.

I am grateful to my family, especially my wife, Yamin, for her unconditional love, patience, and continuous confidence in me. And I want to thank her for the birth of our daughters, who bring additional joys to our life.

WZ

TABLE OF CONTENTS

<u>CHAPTER</u>	<u>PAGE</u>
1. INTRODUCTION	1
2. BRAIN CONNECTIVITY	11
2.1 Three Types of Connectivities	12
2.2 Measures of Structural Connectivity	13
2.2.1 Diffusion Tensor Imaging	14
2.2.2 High Angular Resolution Diffusion Imaging	19
2.3 Measures of Functional Connectivity	21
2.4 Methods for Functional Magnetic Resonance Imaging Study	23
2.4.1 Model-Based Functional Magnetic Resonance Imaging Methods	23
2.4.2 Data-Driven Functional Magnetic Resonance Imaging Methods	25
2.5 Combining Structural and Functional Neuroimaging Data	26
2.5.1 Structural Connectivity Analysis with Functional Data	28
2.5.2 Functional Connectivity Analysis with Structural Data	30
3. NEUROIMAGING DATA	33
3.1 Subject Recruitment	33
3.2 Image Acquisition	34
3.3 Data Processing	35
3.4 Networking Analysis Based on Degrees of Similarity in Connectivity	37
4. BIVARIATE MIXED-EFFECTS MODEL	39
4.1 The Use of Linear Mixed-Effects Model for Neuroimaging Data	39
4.2 Model Definition	42
4.3 Parameter Estimation	43
4.3.1 Empirical Bayes Estimation	45

TABLE OF CONTENTS (continued)

<u>CHAPTER</u>	<u>PAGE</u>
4.3.2 Maximum Marginal Likelihood Estimation	46
4.4 Simulation Study	49
4.5 Hypothesis Testing	51
4.6 Multiple Comparisons	54
4.6.1 False Discovery Rate Assuming Independent Comparisons	54
4.6.2 False Discovery Proportion under Empirical Dependence	55
5. BAYESIAN MIXED-EFFECTS MODEL	57
5.1 Overview of Bayesian Methods	57
5.2 Markov Chain Monte Carlo	59
5.3 Markov Chain Monte Carlo Algorithms	61
5.3.1 The Metropolis-Hastings Algorithm	62
5.3.2 The Gibbs Sampler	64
5.4 Model Definition	65
5.5 Model Update and Convergence Diagnostics	67
6. MACHINE LEARNING AND CLASSIFICATION	69
6.1 Supervised Learning	69
6.2 Support Vector Machine	71
6.3 Application to Analysis of Neuroimaging Data	76
7. DATA ANALYSIS	78
7.1 Mixed-Effects Models	79
7.1.1 Independent Link-Specific Analyses	80
7.1.2 Classical Mixed-Effects Models	81
7.1.3 Bayesian Mixed-Effects Models	99

TABLE OF CONTENTS (continued)

<u>CHAPTER</u>	<u>PAGE</u>
7.1.4 Age Effect on Connectivity	104
7.2 Network Structure Analysis	106
7.3 Machine Learning and Classification	114
7.3.1 Linear Support Vector Machine	114
7.3.2 Support Vector Machine with Other Kernels	118
8. SUMMARY AND CONCLUSION	121
REFERENCES	127
APPENDICES	150
APPENDIX A	151
APPENDIX B	155
APPENDIX C	158
APPENDIX D	161
VITA	163

LIST OF TABLES

<u>TABLE</u>	<u>PAGE</u>
I. COMBINING INFORMATION IN TWO CONNECTIVITIES	27
II. BRAIN REGIONS ANALYZED	37
III. PARAMETER ESTIMATES IN BMEM SIMULATION	51
IV. INDEPENDENT T-TESTS RESULT	81
V. SUMMARY OF LINK-SPECIFIC MEAN DTI AND FMRI	83
VI. SIGNIFICANT CONNECTIVITIES FOR DTI—FDR=0.30	84
VII. SIGNIFICANT CONNECTIVITIES FOR FMRI—FDR=0.30	85
VIII. SIGNIFICANT CONNECTIVITIES FOR BIVARIATE—FDR=0.20	87
IX. SIGNIFICANT CONNECTIVITIES FOR CORRELATION—FDR=0.30	89
X. CMEM: COMPARISON ACROSS ANALYSES	91
XI. FALSE DISCOVERY PROPORTION—EMPIRICAL	97
XII. FALSE DISCOVERY PROPORTION—INDEPENDENT	97
XIII. COMPARISON OF RESULTS ACROSS METHODS	101
XIV. BMEM RESULTS WITH DIFFERENT PRIORS	103
XV. AGE EFFECT ON CONNECTIVITY	104
XVI. SIGNIFICANT CONNECTIVITIES CONTROLLING FOR AGE	105
XVII. CONNECTIVITY SIMILARITY IN LLD AND HC	107
XVIII. COMMON CONNECTIVITIES IN LLD AND HC	109
XIX. COMPARISON ACROSS METHODS	114
XX. CLASSIFICATION USING SVM	116
XXI. SVM CLASSIFICATION USING CMEM RESULT	117
XXII. SVM CLASSIFICATION USING CMEM RESULT—UNBIASED	118
XXIII. SVM WITH THIRD DEGREE POLYNOMIAL KERNEL	119

LIST OF TABLES (continued)

<u>TABLE</u>	<u>PAGE</u>
XXIV. SVM WITH GAUSSIAN KERNEL	120

LIST OF FIGURES

<u>FIGURE</u>		<u>PAGE</u>
1.	Diffusion tensor along a fiber tract.	16
2.	Support vector machine.	73
3.	DTI between left cerebellum cortex and left thalamus proper.	80
4.	Histogram of mean DTI and fMRI across all links.	82
5.	Significant connectivities for DTI	92
6.	Significant connectivities for fMRI	93
7.	Significant connectivities for DTI-fMRI jointly	94
8.	Significant connectivities for DTI-fMRI correlation	95
9.	Convergence diagnostics of Bayesian Models: trace (A), auto correlation (B), density (C), and BGR (D) plots.	100
10.	Gamma and beta fit to the mean DTI and fMRI measures.	106
11.	Strong structural connectivities in both HC and LLD subjects.	108
12.	Strong functional connectivities in both HC and LLD subjects.	108
13.	Network diagram of connectivities in HC but not LLD subjects.	111
14.	Image of structural connectivities in HC but not LLD subjects.	112
15.	Image of functional connectivities in HC but not LLD subjects.	113
16.	Diagram of classification using SVM.	115

LIST OF ABBREVIATIONS

ADC	Apparent Diffusion Coefficient
BGR	Brooks-Gelman-Rubin statistic
BMEM	Bayesian Mixed-Effects Model
BOLD	Blood Oxygenation Level Dependent
BUGS	Bayesian Inference Using Gibbs Sampler
CB	Cingulate Bundle
CCA	Cross-Correlation Analysis
CMEM	Classical Mixed-Effects Model
DTI	Diffusion Tensor Imaging
EB	Empirical Bayes
EEG	Electroencephalography
EM	Expectation-Maximization
EPI	Echo-Planar Imaging
FA	Fractional Anisotropy
FCA	Fuzzy Clustering Analysis
FDP	False Discovery Proportion
FDR	False Discovery Rate
fMRI	Functional Magnetic Resonance Imaging
FOV	Field Of View
HARDI	High Angular Resolution Diffusion Imaging
HC	Healthy Control
LLD	Late-life Depression
MCMC	Markov Chain Monte Carlo

LIST OF ABBREVIATIONS (continued)

MDD	Major Depressive Disorder
MML	Maximum Marginal Likelihood
MRI	Magnetic Resonance Imaging
MTR	Magnetization Transfer Ratio
PFA	Principal Factor Approximation
ROI	Region Of Interest
SPM	Statistical Parametric Mapping
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TBI	Traumatic Brain Injury
UF	Uncinate Fasciculus

SUMMARY

Recent advances in mental health research have led to novel treatments that target specific brain networks involved in the pathophysiology of psychiatric disorders. This has made it increasingly important for researchers to characterize the networks that are associated with these disorders. Although neuroscientists in independent studies have identified abnormalities in structural and functional brain connectivities associated with these diseases, the relationship between the two remains elusive. As complex psychiatric disorders (like depression) are not likely to be simply related to isolated brain regions of interest, we assessed structural (measured using high angular resolution diffusion imaging) and functional (measured using resting-state functional magnetic resonance imaging) connectivity in a model of geriatric depression using whole-brain graph theory-based connectivity analyses. To appropriately account for both the between-modality and within-subject correlations of the connectivity measures, we fitted bivariate linear mixed-effect models with random subject intercepts and heteroscedastic errors to neuroimaging data containing measures for structural and functional connectivities. The models were estimated using an expectation-maximization like approach by iterating between the empirical Bayes estimates of the random effects and maximum marginal likelihood estimates of the fixed and covariance parameters. By comparing study and control subjects using the model estimates, we performed unimodal, bimodal, and correlation analyses, and detected brain region links (connectivities) that demonstrated significant differences in single and both connectivity modalities, as well as in connectivity coherence between the two groups of subjects. We observed consistent results in similar analyses using Bayesian mixed-effect models, as well as in approaches that utilized machine learning techniques to determine the importance of each connectivity in group classification. Our approach represents a first attempt in simultaneous modeling of the two connectivity modalities

SUMMARY (continued)

in a population study while considering both the between-modality and within-subject correlations of the measures.

1. INTRODUCTION

The human brain is a complex system where neural populations function by acting upon each other via various forms of connections to allow cognitive tasks to be achieved. When part of the system fails to function at a certain required level, the related tasks will not be able to be performed well, leading to aging or disease conditions. A detailed understanding of how our brain differs during aging and neurological or psychiatric diseases is fundamental to the development of treatments to these conditions. To this end, numerous efforts have been made to decipher how brain connections change as we age and how they are damaged during illness. For example, the Human Connectome Project (<http://www.humanconnectome.org/consortia/>) represents the most ambitious effort to map the neural pathways that underlie the function of the human brain and will pave the way toward our better understanding of how brain connectivity is involved in the pathophysiology of aging and disease.

Brain connectivity usually refers to a pattern of anatomical links (structural connectivity) and statistical dependencies (functional connectivity). Structural connectivity is the structural link or neural pathways between two areas. This typically corresponds to white matter tracts that run between pairs of brain regions. Functional connectivity is defined as the “temporal correlation between spatially remote neurophysiological events (Friston et al., 1993b, 5).” In other words, functional connectivity indicates the synchronized and correlated patterns of activity that may even occur between pairs of structurally unconnected regions. The third type of connectivity, effective connectivity, is defined as “the influence that one neural system exerts over another either directly or indirectly (Friston et al., 1993a, 69).” The three types of connectivity are closely related to each other.

Modern neuroimaging techniques have been well developed to study the structural and functional connectivities in the human brain. Among all available techniques, diffusion tensor imaging (DTI), which measures the diffusion of water molecules between two brain regions along a tensor in a three-dimensional space, and functional magnetic resonance imaging (fMRI), which measures the correlation between activities of two brain regions through monitoring the blood oxygenation level-dependent signals, are the most widely used methods for structural and functional connectivity, respectively. Using these two techniques, numerous studies have discovered brain regions that are involved in the pathophysiology of neurodegenerative and psychiatric disorders such as Alzheimer’s disease, epilepsy, traumatic brain injury, dementia, and depression (see chapter 2, section 2.2 and 2.3 for reference).

Major depressive disorder (MDD) is a type of mental disorder “characterized by [episodes of] low mood, low self-esteem, and loss of interest in enjoyable activities (<http://en.wikipedia.org/>).” Psychiatric disorders like MDD are “a leading cause of death and disability” and “exert a tremendous economic burden worldwide (<http://www.latestcnnnews.com/>).” The World Health Organization foresees that MDD alone “will be the second leading contributor to the global burden of disease across all ages (<http://www.latestcnnnews.com/>).” It is reported that in the United States, 9.1% of adults meet the criteria for current depression (major depression or “other depression” that meet criteria for a depressive disorder according to the Diagnostic and Statistical Manual of Mental Disorders), including 4.1% who meet the criteria for major depression (CDC, 2010). Moreover, up to 60% of people who commit suicide had depression or other mood disorders (Barlow and Durand, 2005).

Late-life depression (LLD) refers to “major depressive [episodes that occur] for the first time in people older than 50 or 60 years of age (http://en.wikipedia.org).” Although

depression is not a “natural part of aging”, older adults face increased risk for developing the disease. Risk factors for LLD include “a history of depression, chronic medical illness, being single or divorced, brain disease, alcohol abuse, use of certain medications, and stressful life events (<http://www.aafp.org/>).” According to a report by the United States Institute of Medicine, nearly 2 million older adults in 2010 were affected by depressive disorders (Institute of Medicine, 2012). Since the population of Americans over the age of 65 is expected to nearly double and hit 70 million by the year 2030 (Vincent and Velkoff, 2010), it is increasingly important for researchers to develop accurate diagnosis tools and effective treatment regimens for LLD. Neuroscientists believe that early diagnosis and intervention can help people at-risk and thus reduce the damage caused by depression.

Current treatment of LLD generally has poor response and is complicated by concurrent medical problems and low functionality of older patients. The recent finding of white matter hyperintensities in people with LLD has argued that depression may occur as a result of vascular disease in a significant subpopulation of older adults (Kales et al., 2005). The vascular depression hypothesis, defined by Krishnan et al. (1997) and Alexopoulos et al. (1997), states that “cerebrovascular disease may predispose, precipitate, or perpetuate some geriatric depressive syndromes (Alexopoulos et al., 1997, 915).” This hypothesis is supported by indirect evidence including the high rate of depression in patients with vascular disease (Musselman et al., 1998; Broadley et al., 1995; Godfrey et al., 1990), frequent “silent stroke” and white-matter hyperintensities in LLD (Fujikawa et al., 1993), and absence of family histories of psychosocial stressors in such patients (Fujikawa et al., 1994). More recent findings include the association between vascular conditions and risk factors and subsequent risk of depression (Mast et al., 2008; Kivimäki et al., 2012), the association of the vascular risk factors with poorer response to

treatment for depression (Sheline et al., 2010), and the predictability of depression onset using white-matter changes (Firbank et al., 2012).

Despite the indirect evidence, the exact white-matter changes that cause late life onset of depression is largely unknown. However, it is believed that “brain changes can be triggered by experiencing stresses of life events (<http://en.wikipedia.org>),” and that cardiovascular risk in older adults “may increase vulnerability to depression by disrupting mood regulation circuits in the brain, decreasing its ability to respond to stressful events (NARSAD, 2013).” Moreover, recent advances in mental health research have been demanding novel treatments for depression that target specific brain networks involved in the pathophysiology of the disease. This has made it increasingly important for researchers to relate the disruptions in structural brain connectivity networks defined by white-matter alterations, with perturbations in functional connectivity that impact cognitive and affective processing.

Abnormalities in structural and functional brain connectivity associated with LLD have been identified in independent studies (Taylor et al., 2007; Sheline et al., 2008; Dalby et al., 2010; Mettenberg et al., 2012). The uncinate fasciculus (UF) and cingulate bundle (CB) tracts are the two most studied brain regions for disruptions in structural connectivity in LLD since they are considered to be involved in emotion processing, attention, and memory. The UF tract is a “ventral associative bundle that connects the anterior temporal lobe, including the amygdala and hippocampus, with the medial and lateral orbitofrontal cortex (Zhang et al., 2012, 960).” The CB tract is a “medial associative bundle that runs within the cingulate gyrus superior to the corpus callosum (Zhang et al., 2012, 960).” Decreased white-matter integrity has been demonstrated in these tracts in different studies. For example, Taylor et al. (2007) and Sheline et al. (2008) found decreased white-matter integrity in UF with geriatric depression. Dalby et al.

(2010) further showed that the whiter-matter alterations correlated with the severity of depression. Mettenburg et al. (2012) observed greater age effects on CB integrity in LLD subjects compared to controls.

For functional connectivity, previous studies have identified resting-state networks involved in functions such as cognitive control, default mode, and visual and affective processing (Albert et al., 2009). These networks have been shown to be involved in modulating task-related activities and be modified by experience (Albert et al., 2009). Using resting-state fMRI, a number of studies have identified activity changes in some resting-state networks. For example, Greicius et al. (2007) showed that default model network activity, especially in the subgenual cingulate, was significantly increased in people with MDD. For people with LLD, Kenny et al. (2010) demonstrated increased caudate connectivity, particularly on the right; Wu et al. (2011) found decreased subgenual anterior cingulate connectivity and increased dorsomedial and orbitofrontal connectivity; and Alexopoulos et al. (2012) showed increased default mode network activity and decreased cognitive control network activity.

Although numerous studies have successfully linked structural or functional connectivity changes with LLD, the relationship between the two connectivities and how they are involved in the pathophysiology of the disorder remains elusive. Relatively little progress has been made in combining information from structural and functional imaging modalities and previous attempts have been made to determine only the consistency between independent structural and functional analyses. For example, using graphs and cortical connection matrices, Sporns et al. (2000) examined the relationship between structural features of networks and the patterns of functional connectivity and found the linkage between complex functional dynamics and dense structural networks. Koch et al. (2002) showed that there is no simple correlation between structural and functional connectiv-

ity, except that “low values of functional connectivity were not found together with high values of [structural] connectivity”. Honey et al. (2007) demonstrated on different time scales that in the cerebral cortex, functional connectivity is shaped by network structure. And finally, using computational modeling, Honey et al. (2009) showed that systems-level spatial and temporal properties of functional networks can be accounted for by properties of underlying structural network. In clinical studies, population differences are identified with regard to structural and functional connectivities and compared a posteriori. For instance, Ke et al. (2008) assessed structural and functional connectivity separately and found that decreased functional connectivity in early-onset schizophrenia is correlated with the regions of structural changes. Similarly, Zhou et al. (2008) showed concurrence of the decreased functional connectivity and damaged anatomical connectivity between the hippocampus and other regions in schizophrenia.

Research is lacking in joint statistical modeling of structural and functional connectivity for population studies. The only attempt to date is the work by Venkataraman et al. that merged structural and functional information in a probabilistic framework and identified connectivity patterns in the brain (Venkataraman et al., 2012). These authors modeled the interaction between latent structural and functional connectivity from an underlying process in the brain not observed directly from the data, and showed that the model exhibited improved classification and better generalization accuracy than similar models fitting only one of the modalities (Venkataraman et al., 2012). However, there are some drawbacks of this work. First, it models latent connectivity via discrete random variables, which may oversimplify the situation and fail to detect subtle between-group differences. Next, the model assumed that all subject data are drawn from the same distribution, whereas in practice subject variation can be high and both structural and functional connectivity can vary across subjects. And finally, the model does not include

parameters that measure the interactions between structural and functional connectivity and thus cannot provide information about how functional abnormality may be related to structural damage, which is critical for the development of treatments addressing alterations in brain networks associated with LLD and other mental disorders.

Identification of hubs of brain networks is critical in understanding the underlying mechanism of neural activities in the brain and in the development of novel treatments to psychiatric disorders. Modern neuropsychotherapies to psychiatric disorders often require one to provide stimulation to hubs of certain brain networks according to the type of disorders and tasks. In order to obtain the maximum benefit of the stimuli, it is important to target hubs of some specific networks that are primarily involved in the pathophysiology of the disorder and related tasks. It is unfortunate that the statistical methods used in developing such important networks are often inadequate as they fail to incorporate appropriate correlation structures, and thus inferences from this procedure will be based on ill-defined assumptions. The goal of this thesis is to develop appropriate statistical methodologies to identify brain regions, connections, and hubs of networks under certain conditions, and use those results for group comparisons.

In this work, we proposed a bivariate linear mixed-effects model with random subject intercepts and heteroscedastic errors to jointly fit neuroimaging data with both structural and functional connectivity modalities. We used a neuroimaging data set consisting of DTI and fMRI measures for the 87 brain regions of 10 LLD and 13 healthy control (HC) subjects to illustrate our model. The model allowed us to incorporate both the between-modality and within-subject correlations of the connectivity measures. By using a multimodal approach trying to link structural connectivity alterations associated with vascular risk to functional connectivity changes associated with cognitive and affective impairment, we hoped to provide a mechanism for the vascular depression hy-

pothesis. Since complex disorders like LLD are not likely to be related to isolated regions of interest, we assessed structural and functional connectivities of depressed subjects with non-depressed control subjects following a whole-brain analysis (87 brain regions). When comparing connectivities of all links among brain regions, we followed a false discovery rate approach to control for the type I error rate. The proportion of our false discoveries were estimated in scenarios with different correlation assumptions.

In addition, we employed Bayesian mixed-effects models (BMEMs) with the same structure and noninformative priors in analyzing these neuroimaging data. Bayesian inference treats parameters as random variables with prior distributions. These prior distributions are then combined with data likelihood to obtain posterior distributions and we make inferences based on these posterior distributions. The introduction of prior distributions in Bayesian methods allows researchers to easily incorporate prior knowledge (e.g., historical data, expert’s knowledge) into the analysis. Thus the flexibility of this procedure goes beyond the boundary of completely relying on one specific sample and as a result it makes the inference more robust. Bayesian inference is based on posterior distributions and provides the conditional probability of parameters given the data. In contrast, traditional non-Bayesian methods make use of P-values, which are essentially the probability of obtaining a sample that is more extreme than the current data when the null hypotheses about the parameters are true. Compared to classical approaches, Bayesian methods make direct inference on parameters and thus deal with the research problems in a more natural way. By using BMEMs in the analysis of multimodal neuroimaging data, we provided a platform where inferences can be made using experimental data combined with any prior knowledge about the parameters of interest.

We also compared the study and control subjects by evaluating the network structure dissimilarities between the two group of subjects. By classifying each connectivity according to its relative strength and determining the concordant and discordant pairs between the diseased and control groups, we were able to identify connectivities that are strong in both groups and those that are weaker in one group compared to the other. Brain images and network diagrams were generated for these sets of connectivities. The focus was on identifying damaged connectivities that were present in the control but weak in the diseased subjects and determining brain regions that were most affected in the study group.

Although the above-mentioned methods were potentially able to identify abnormalities in structural and functional brain connectivity, they would not be able to provide a diagnostic measure that is both sensitive and specific, and that can accurately diagnose or predict future development of psychiatric disorders in patients at-risk. Machine learning and classification methods are potential tools to predict cognitive states and group membership of subjects using structural and functional connectivity data. The main idea of this technique is to let the machine learn from a training data set so that it gains knowledge about the characteristics or inputs that are specific to those groups. Then the trained machine can be used to classify new subjects into either of the groups based on their inputs. In addition, feature (variable) selection procedures can potentially increase the performance of the machine learning algorithms. Combining popular machine learning algorithms and feature selection procedures in the analysis of multimodal neuroimaging data can potentially provide not only powerful diagnostic tools that help us effectively classify individuals into groups, but also prognostic devices that help us predict future outcomes. Using machine learning techniques, we were able to identify brain connectivities that were important in classifying LLD versus HC subjects. We

also evaluated the performance of several popular machine learning and classification techniques using the neuroimaging data.

The remainder of this thesis is organized as follows: In chapter 2 we describe different types of brain connectivity and popular brain imaging techniques used by neuroscientists to measure them. In chapter 3 we provide details about the collection, processing, and features of the neuroimaging data, as well as the techniques that we used to determine the community structure of the data. In chapter 4 we propose the bivariate linear mixed-effects models with heteroscedastic errors that we used to model the neuroimaging data. In chapter 5 we describe the BMEMs we employed in modeling the same data. In chapter 6 we provide a brief introduction to the machine learning techniques we utilized in our classification studies. In chapter 7 we demonstrate in detail the results across all our analyses. Finally in chapter 8 we summarize the work and discuss about limitations and future works.

2. BRAIN CONNECTIVITY

In the human brain, information is represented by patterns of activity occurring over neural populations. Instead of functioning as separate units, neural populations in the human brain act on each other via various forms of connections to allow cognitive tasks to be performed. Brain mapping and connectivity research provide insights into how different regions of the brain interact with each other and how these interactions are altered during the course of the disease. Work in this field during the past two decades has fostered the development of novel diagnostic and therapeutic interventions to recognize and treat “neurodegenerative and psychiatric disorders such as Alzheimer’s disease, attention-deficit hyperactivity disorder, [post-traumatic] stress disorder, epilepsy, traumatic brain injury, stroke, dementia, and depression (<http://www.findhealthnews.com/>).”

One of the most significant recent trends in brain connectivity research is that the focuses have gradually expanded from specific region-of-interest-based assessments to broader whole-brain analyses (Watanabe and Tanaka, 2007). Instead of identifying the isolated brain regions activated under certain conditions, recent work has paid more attention to how different parts of the brain connect, interact, and coordinate with each other to perform certain kinds of cognitive functions (Fox et al., 2005; Greicius et al., 2003; Munk et al., 2002). It is now believed that “most cognitive functions, such as vision, attention, memory, language, and motor planning, are implemented in the brain via large-scale networks [that comprise] different brain regions (Rykhlevskaia et al., 2008, 173).”

2.1 Three Types of Connectivities

A network is “a mathematical representation of a real-world complex system defined by a collection of nodes and links between pairs of nodes (Rubinov and Sporns, 2010, 1060).” In brain networks, the “nodes” are brain regions, whereas the “links” are the connections between regions. Current conceptualization of the interrelationship between brain network nodes classifies brain connections into three different connectivities: the structural (or anatomical), functional, and effective connectivity.

Structural connectivity is the structural link or neural pathway between two brain regions. This is typically characterized by white-matter tracts between pairs of brain regions. Functional connectivity is defined as the “temporal correlation of a neurophysiological index measured in different brain areas” (Friston et al., 1993b, 5). In other words, functional connectivity measures the synchronized and correlated patterns of activity between two brain regions. This may even occur between brain region pairs that are not structurally connected. Depending on circumstances, functional connectivity may refer to “linear or nonlinear interactions, as well as interactions at different time scales (Zhou et al., 2009, 125).” Functional connectivity does not necessarily have to be direct or causal. On the contrary, effective connectivity is defined as “the influence that one neural system exerts over another either directly or indirectly” (Friston et al., 1993a, 69). This causal relationship may be estimated from observed perturbations (Friston et al., 1993a).

The three types of connectivity are closely related to each other. An example of one-to-one mapping between functional and structural connectivity is observed at synaptic connections, whereas the lack of structural connections between two brain regions makes the existence of a direct functional connection impossible. Effective connectivity can be

viewed as a special functional connectivity. Instead of bidirectional relationships between components in typical functional models, the relationships are unidirectional and causal in effective models.

Neuroimaging techniques have evolved rapidly during the past two decades due to the development and refinement of image acquisition, representation, analysis, visualization, and interpretation techniques. Modern neuroimaging techniques, especially the non-invasive ones (such as fMRI and DTI) have made it possible for us to effectively and safely study the anatomy and function of the brain. These techniques provide measures for two main aspects of brain mapping: the structural and functional neuroimaging.

2.2 Measures of Structural Connectivity

Structural connectivity reflects the presence of fiber (white matter) tracts that run between brain regions. These tracts contain neurons that transmit neural signals to allow communication between brain regions. Before the application of imaging techniques in visualizing the human brain, structural connectivity had been studied using histological methods in animals. Modern in vivo neuroimaging techniques have been developed to allow the visualization of white matter in humans. Standard magnetic resonance imaging (MRI) has been used to measure white-matter integrity. Au Duong et al. (2005) assessed white-matter abnormalities in multiple sclerosis using conventional MRI (T1- and T2-weighted images) and magnetization transfer ratio (MTR) imaging. The size of the structures on MRI can be determined by segmenting the tissue to isolate the structure of interest. For example, Allen et al. (1991) obtained the structure of corpus callosum by manually tracing it from multiple slices of structural MRI image. On the other hand, computerized procedures such as dynamic contour models (McInerney and Terzopoulos,

1996) and model-guided segmentation (Lundervold et al., 1999) can be used to synthesize information along the tract. Determination of white-matter integrity usually involves “segmentation of three-dimensional brain images into three tissue types: white matter, grey matter, and cerebrospinal fluid (Rykhlevskaia et al., 2008, 175).” Smith et al. (2004) reviewed methods to separate these three tissue types in MRI (Smith et al., 2004). These methods provide probabilistic maps of white matter, which can then be used to determine the degree of structural connection in the whole brain or specific regions. For instance, several studies have employed the volumetric white-matter estimates obtained using these methods and showed that the size of the corpus callosum is correlated to age-related performance reduction (Gunning-Dixon and Raz, 2000; Colcombe et al., 2005).

2.2.1 Diffusion Tensor Imaging

During the past decade, a new imaging technique called DTI has been developed and quickly become a widely used method for detecting structural connectivity. In a DTI analysis, the diffusion of water molecules is monitored in living tissues and the directionality of this movement is determined. The diffusion process causes “water molecules to move out from a central point and gradually reach the surface of an ellipsoid if the medium is anisotropic (or sphere if medium is isotropic) (<http://en.wikipedia.org/>).” Because the dielectric (electrically insulating) myelin sheath around the axons is a barrier to the diffusion of water, water diffuses mostly along the direction of the axons. The measured directionality of the diffusion therefore provides indirect information about brain structure. Compared with the traditional T1- or T2-weighted MRI images, DTI images provide more extensive information about white-matter structures (Rykhlevskaia et al., 2008), and thus have been used in many procedures of in vivo fiber tracking and brain

visualization. Mori and van Zijl (2002) provides a good review of fiber tracking methods and strategies.

By monitoring the diffusion gradients, a formula can be generated for the numerical measure of the diffusion—the diffusion coefficient. Since multiple barriers to the diffusion, such as axon membranes and myelin sheaths, are present in the brain, apparent diffusion coefficient (ADC, α) is measured by treating all movements as Brownian motion. The ADC in an anisotropic tissue depends on the direction and is high when diffusion is along the axon and low when diffusion is perpendicular to the direction of the axon. Once diffusion is measured in six or more directions, an ellipsoid can be constructed so that its “diameter in any direction estimates the diffusivity in that direction, and that its major principle axis is oriented in the direction of maximum diffusivity (<http://www.ajnr.org/>) (see Figure 1).” The information of diffusivity lies in the diffusion tensor, which is “a mathematical model of diffusion in a three-dimensional space” and “a 3×3 matrix derived from diffusion coefficients in several different directions (<http://appliedradiology.com>).” It is symmetric with six degrees of freedom (six of the nine elements are independent), which means diffusion measures in a minimum of six directions are required to construct a diffusion tensor. Measures on more directions will improve the accuracy of the tensor (Jellison et al., 2004). Diagonalization of the tensor matrix will give a matrix containing three eigenvectors and a diagonal matrix containing three eigenvalues ($\lambda_1, \lambda_2, \lambda_3$). These three eigenvalues represent the degree of diffusivity along the three axes determined by the eigenvectors.

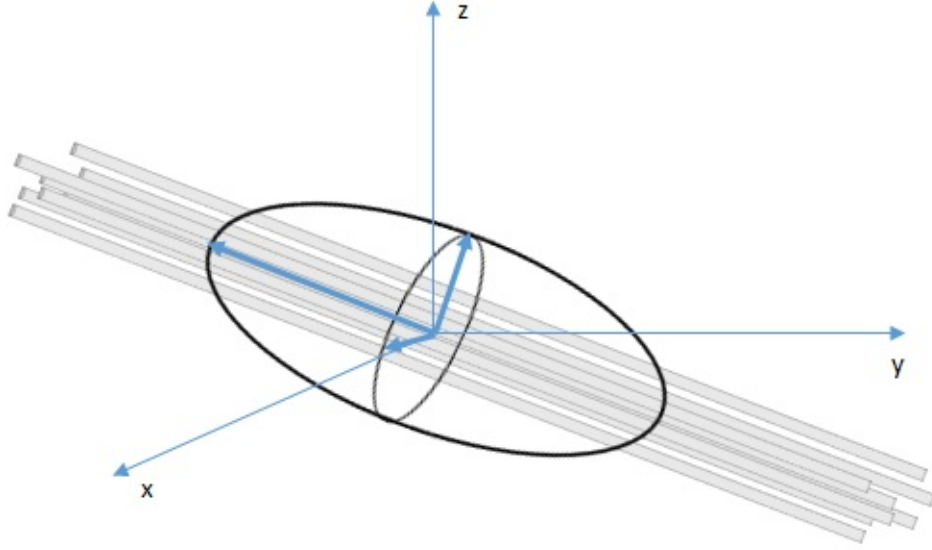


Figure 1. Diffusion tensor along a fiber tract.

From the tensor, one can determine the direction of maximum diffusivity or estimate diffusivity in any given direction. For example, if the ellipsoid is sitting in a three-dimensional Cartesian coordinate system, the projections of the ellipsoid onto the axes will provide the ADC for each of the three axes, α_x , α_y , and α_z . Then the average ADC α_m in the voxel can be written as

$$\alpha_m = \frac{\alpha_x + \alpha_y + \alpha_z}{3}. \quad (2.1)$$

On the other hand, if we consider the axes of the ellipsoid itself, we have a principal axis where diffusion is maximum, and two smaller axes. All three axes are perpendicular to each other and cross at the center point of the ellipsoid. These axes are the eigenvectors and the diffusion measures along these axes are the corresponding eigenvalues. Denote the largest eigenvalue (maximum diffusivity) as λ_1 and the two smaller values λ_2 and λ_3 . The difference of λ from ADC is that ADC is a projection of the diffusivity on the main axes whereas λ is the actual diffusivity measure. The eigenvalue along the principal axis,

λ_1 , is also called axial diffusivity (or parallel diffusivity) $\lambda_{||}$, where radial diffusivity (or perpendicular diffusivity) refers to the average of λ_2 and λ_3 that are along the two minor axes $\lambda_{\perp} = (\lambda_2 + \lambda_3)/2$.

Compared to individual measures of diffusion, it is usually more important to measure the relative degree of anisotropy in a voxel. The simplest measure for this would be to divide the axial diffusivity by the smallest diffusivity λ_1/λ_3 . However, this measure is very sensitive to measurement noise. To overcome this, the square-root of the sum of squares of diffusivity differences $\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2}$ has been used. This value is dominated by the largest component and is not susceptible to measurement noise like the simple diffusivity ratio. The most commonly used measure in structural connectivity analysis is the measure of fractional anisotropy (Conturo et al., 1996; Pierpaoli and Basser, 1996, FA), which can be written as:

$$FA = \frac{\sqrt{3((\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2)}}{\sqrt{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}, \quad (2.2)$$

where $\bar{\lambda} = (\lambda_1 + \lambda_2 + \lambda_3)/3$, and the constant $\sqrt{3/2}$ on the right hand side of the above equation is to make FA to have a range of (0, 1). An FA value of 0 indicating the voxel is spherical and the diffusion is equally restricted in all directions, and 1 referring to the elongated voxel where the diffusion is exclusively restricted in one direction.

Depending on the “shape” of the diffusion ellipsoid, FA can also be separated into linear, planar, and spherical measures (Westin et al., 1981). Assume $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$, then for the linear case, we have $\lambda_1 \gg \lambda_2 \simeq \lambda_3$, for the planar case, we have $\lambda_1 \simeq \lambda_2 \gg \lambda_3$, and for the spherical case, we have $\lambda_1 \simeq \lambda_2 \simeq \lambda_3$. Denote linear, planar, and spherical measures by ζ_l , ζ_p , and ζ_s , respectively, these measures can be computed as follows (Bi

et al., 2010):

$$\begin{aligned}\zeta_t &= \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3}, \\ \zeta_p &= \frac{2(\lambda_2 - \lambda_3)}{\lambda_1 + \lambda_2 + \lambda_3}, \\ \zeta_s &= \frac{3\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}.\end{aligned}\tag{2.3}$$

All these measures range between 0 and 1 and they sum to unity. From these measures the deviation from sphericity, which is another measure of anisotropy, can be written as

$$\zeta_m = \zeta_t + \zeta_p = 1 - \zeta_s = \frac{\lambda_1 + \lambda_2 - 2\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}.\tag{2.4}$$

And finally, yet another measure of anisotropy is the so-called relative anisotropy (Conturo et al., 1996; Pierpaoli and Basser, 1996), which is defined as

$$RA = \frac{\sqrt{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2}}{\sqrt{3\bar{\lambda}}}.\tag{2.5}$$

Using the measures mentioned above, degree of anisotropy and local fiber direction can be determined. These can be summarized in a summary DTI map. However, despite its power in in vivo brain visualization, DTI has its limitations, especially in resolution. In DTI, spatial and orientation resolution is lacking and the information about the directionality (left versus right, retrograde versus anterograde, and so on) is not provided. The poor spatial resolution of DTI has long been the major obstacle to the efficient reconstruction of fibers as branching and crossing of fibers are common in a single voxel. In addition, false positive connections can be generated when the probability of the existence of a connection is misestimated due to partial volume effects.

2.2.2 High Angular Resolution Diffusion Imaging

Traditional DTI methods can only deal with cases when there is a single ellipsoid in the voxel, because the tensor model only assumes a single orientational maximum, which is the major eigenvalue of the diffusion tensor. This means that the axons in this voxel would all travel in exactly the same direction, which is often the case. However, at the resolution of DTI, which is in the scale of millimeters, the chance of intravoxel orientation heterogeneity of white-matter tracts is high. In fact, it is estimated that more than 30% of the voxels have more than one axon traveling in different directions and passing through each other (Meyer et al., 1999; Makris et al., 1997, 1999), and these crossing tracts are treated as noise in traditional DTI. To solve this problem, Tuch et al. (2002) developed a new technique called high angular resolution diffusion imaging (HARDI).

These authors used high angular resolution, high b-value (the value that summarizes the influence of the gradients on the diffusion weighted images with higher value indicating stronger diffusion weighting) “diffusion gradient sampling, [and were] able to detect diffusion signals with multiple, discrete maxima/minima as a function of gradient orientation, indicating the presence of multiple underlying fiber populations [within the same voxel] (Tuch et al., 2002, 577).” To resolve the underlying directions of the diffusion along these fibers, they went on to model the diffusion signal as a discrete mixture of Gaussian diffusion processes in slow exchange (a mixture of tensors) (Tuch et al., 2002).

If diffusion is a Gaussian process, then the signal from a single compartment is

$$E(\mathbf{q}_k) = \exp(-\mathbf{q}_k^T \mathbf{D} \mathbf{q}_k \tau), \quad (2.6)$$

where “ $E(\mathbf{q}_k)$ is the normalized diffusion signal magnitude for the diffusion gradient wave-vector $\mathbf{q}_k = \gamma\delta\boldsymbol{\gamma}_k$, γ is the gyromagnetic ratio, δ is the diffusion gradient duration, $\boldsymbol{\gamma}_k$ is the k th diffusion gradient, τ is the effective diffusion time, and $\mathbf{D}$ is the apparent diffusion tensor (Tuch et al., 2002, 578).” The diffusion function for multiple compartments assuming a finite mixture of Gaussians is

$$E(\mathbf{q}_k) = \sum_i f_i \exp(-\mathbf{q}_k^T \mathbf{D}_i \mathbf{q}_k \tau), \quad (2.7)$$

where f_i is the apparent volume fraction of the voxel with diffusion tensor $\mathbf{D}_i$. The assumptions for this Gaussian mixture model “are that (1) the mixture consists of a discrete number of homogeneous regions; (2) the regions are in slow exchange (i.e., separated by a distance much greater than the diffusion mixing length); and (3) the diffusion within each region is Gaussian (Tuch et al., 2002, 578).” This Gaussian mixture problem has to be solved with physiological constraints on the eigenvalues and the authors employed a gradient descent scheme with multiple restarts (Tuch et al., 2002).

The multiple models observed in HARDI diffusion signals allow us to characterize different intravoxel fiber populations corresponding to known fiber anatomy. However, HARDI requires a model of the underlying diffusion process. Recently a model-independent version of HARDI has been developed. The technique uses a spherical tomographic inversion called the “Funk-Radon transform”, also known as the “spherical Radon transform (Tuch, 2004).” The resulting “Q-Ball” method can “resolve [multiple] intravoxel fiber orientations and does not require any assumptions on the diffusion process such as Gaussianity or multi-Gaussianity (Tuch, 2004, 1358).”

2.3 Measures of Functional Connectivity

Functional connectivity “refers to the concurrent activity of different cortical regions of the brain and is typically measured by analyzing the patterns of synchronization of brain activity measured at different regions” (Rykhlevskaia et al., 2008, 75). During the past two decades, many measures of brain activity have been used to study functional connectivity. For example, Friston et al. (1993b) performed principal component analysis using large sets of positron emission tomography data and showed that the variance in neuropsychological measurements was accounted for by these analysis. Astolfi et al. (2007) used electroencephalography (EEG) combined with linear inverse estimation and multivariate methods and studied the functional connectivity patterns of cortical activity. Schnitzler and Gross (2005) reviewed magnetoencephalography and its applications in answering questions in neurosciences. Pfurtscheller and Andrew (1999) used both EEG and MEG to show event-related synchronization and de-synchronization in neurological disorders. Rykhlevskaia et al. (2006) established a statistical framework for studying functional connectivity using the event-related optical signal method and demonstrated that the integrity of anatomical connections between the two hemispheres explains different patterns of cross-hemispheric interactions.

Among all imaging modalities, fMRI remains the most commonly used tool for mapping human brain functions. This technique takes advantage of the linkage between brain function and its blood flow. The first step to measure blood flow to the brain dates back to Linus Pauling and Charles Coryell’s discovery in 1936 (Pauling and Coryell, 1936) that “oxygen-rich blood (oxygenated hemoglobin) was weakly repelled by magnetic fields, while oxygen-depleted blood (deoxygenated hemoglobin) was attracted to magnetic fields.” However, it was not until in the early 90s that this property was found

to be able to improve MRI. Ogawa et al. (1990a) at AT&T Bell Labs found that “different magnetic properties of [oxygenated and deoxygenated hemoglobins] caused by blood flow to activated brain regions would cause measurable changes in MRI signals (Ogawa et al., 1990a, 1990b).” These authors obtained in vivo images of brain microvasculature with image contrast reflecting the blood oxygen level using gradient-echo techniques and showed that blood oxygenation level-dependent (BOLD) techniques can be used in MRI to measure regional neural activity. In a seminal study, Ogawa et al. (1990b) obtained the MRIs of brains of live mice and rats using a strong magnetic field. In this study, the authors manipulated blood oxygen level by changing the amount of oxygen the animals breathed and showed that a map of blood flow in the brain could be seen in MRI as the oxygen level fell. They “changed the level of oxygen in the air breathed by rats and scanned them while monitoring brain activity using EEG”, and showed that these blood flow changes were related to functional brain activity (Ogawa et al., 1990b).

Compared to other imaging techniques, fMRI allows noninvasive yet in vivo (i.e., in living human) representation of brain states with high spatial resolution, and thus has gained huge popularity in neuroscience research since its emergence. Bandettini et al. (1992) used fMRI and observed a local signal increase of $4.3 \pm 0.3\%$ in the human brain during task activation. Kwong et al. (1992) obtained images demonstrating activity-induced changes in brain MRI signals observed at a temporal resolution of seconds. These authors also showed that the stimulation frequency dependence of visual activation agrees with previous positron emission tomography observations and that “similar signal changes occur within the human primary motor cortex during a hand-squeezing task and in animal models of increased blood flow by hypercapnia.” Ogawa et al. (1992) reported that visual stimulation produces a transient increase in the intensity of water proton MRI signals in human primary visual cortex in fMRI images and that the “observed changes

predominantly occurred in areas containing gray matter and could be used to produce high-resolution functional brain maps in humans.” For a review of the early development of fMRI, including image acquisition, noise reduction, relation to neuronal events, merits, and limitations, as well as examples of this technique, we refer to the review by Ogawa et al. (1998).

2.4 Methods for Functional Magnetic Resonance Imaging Study

Current methods for investigating functional connectivity using fMRI can generally be grouped into two types: the model-based methods and the data-driven methods. In this section we briefly describe these methods and their advantages and limitations. We refer to Li et al. (2009) for details about the methods for functional brain connectivity detection using fMRI.

2.4.1 Model-Based Functional Magnetic Resonance Imaging Methods

The majority of currently available functional connectivity methods are model-based. These methods usually rely on prior neuroscience knowledge. In these studies, some brain regions are selected as regions of interest (ROIs), or the so-called “seeds.” Then certain measures are used to determine whether other regions are connected to these seeds. In the end, a functional map of the brain can be generated. According to the measures used, these methods can usually be classified as cross-correlation analysis (CCA), coherence analysis (CA), and statistical parametric mapping (SPM).

The most widely used model-based method is CCA introduced by Cao and Worsley (1999). In CCA, one estimates the correlation between two BOLD time courses that

correspond to the two regions at a certain lag. If the correlation is more than a pre-determined threshold, then the two regions are considered functionally connected. The correlation of two BOLD time courses (T_x and T_y) at lag k can be estimated as

$$Corr_{xy}(k) = \frac{Cov_{xy}(k)}{\sqrt{Var_x Var_y}}, \quad (2.8)$$

where $Corr_{xy}(k)$ is the covariance, Var_x and Var_y are the variances of the two time courses T_x and T_y . Due to the high computational burden to obtain correlations at all lags, correlations for limited time points are generally computed, and many studies use zero lag correlations (Ogawa et al., 1990b; Blamire et al., 1992; Friston et al., 1994).

The use of zero-lag in CCA as the connectivity measurement has been questioned (Cecchi et al., 2007) because of the fact that correlation tends to vary across different subjects and brain regions (Buckner et al., 1998; Lee et al., 2001; Miezin et al., 2000; Saad et al., 2001), and it is very sensitive to noises from cardiac and brain blood vessel activities (Friston et al., 1995). To overcome these problems, Sun et al. (2004) proposed CA that uses the spectral representation of correlation in frequency domain:

$$Coh_{xy}(\mu) = \frac{|F_{xy}(\mu)|^2}{F_{xx}(\mu)F_{yy}(\mu)}, \quad (2.9)$$

where $F_{xy}(k)$, $F_{xx}(k)$, and $F_{yy}(k)$ are spectra defined by the Fourier transforms of the corresponding covariance and variances:

$$\begin{aligned} F_{xy}(\mu) &= \sum_k Cov_{xy}(k) e^{-jk\mu}, \\ F_{xx}(\mu) &= \sum_k Var_x(k) e^{-jk\mu}, \\ F_{yy}(\mu) &= \sum_k Var_y(k) e^{-jk\mu}. \end{aligned} \quad (2.10)$$

The appearance of CA allows scientists to study the real time course relationship of two regions. In addition, this method provides good separation of signal from noises based on their frequency differences.

Yet another model-based method is the SPM. This method usually uses general linear models and Gaussian random fields jointly to build up statistical parametric maps and make spatial inferences. Details about SPM theory can be found in Friston et al. (1995).

2.4.2 Data-Driven Functional Magnetic Resonance Imaging Methods

Although widely used in the detection of functional connectivity, model-based methods do have limitations that may require special attention when used in real studies. First, since these methods are seed-based, the result connectivity may be sensitive to seed selection (Ma et al., 2007). Second, the requirement for prior knowledge constrains studies to focus primarily on previously studied brains regions.

To overcome these limitations, data-driven methods that are independent of assumed models or prior information have been developed. These methods can be generally classified into two types. The first type of data-driven method relies on decomposition techniques to represent the original data space as a linear combination of base vectors or independent components. The most commonly used techniques include principal component analysis, singular value decomposition (SVD), and independent component analysis. The other type of data-driven method is based on clustering analysis to partition the data into different clusters based on the intensity proximity of the time course (Li et al., 2009). Example of these methods include fuzzy clustering analysis (FCA), vector quantization,

self-organizing maps, and neural gas network (Martinetz et al., 1993; Chuang et al., 1999; Ngan and Hu, 1991).

The choice of model-based versus data-driven methods depends on the aims and goals of the study as none of the methods outperforms the other in every situation. Baumgartner et al. (1998) compared CCA and FCA using simulated and in vivo fMRI data. These authors found that, even without prior knowledge, the performance of FCA was comparable to that of CCA where extensive prior knowledge has to be added, and that FCA yields a more particular description of fMRI data (Baumgartner et al., 1998). In contrast, Worsley et al. (2005) compared thresholding correlations and SVD and found that “thresholding correlations are better at detecting focal regions of correlated voxels, whereas SVD is better at detecting extensive regions of correlated voxels.” If prior knowledge is available, it would be more reasonable to use the model-based methods where prior information can be incorporated. On the other hand, if the goal is to detect extensive regions of correlated voxels, a data-driven method would make much more sense.

2.5 Combining Structural and Functional Neuroimaging Data

As structural and functional properties of the brain can be measured simultaneously, how to effectively combine the two measures has long been an important topic of brain connectivity research. Different approaches have been used for combining structural and functional information in brain connectivity research. Rykhlevskaia et al. (2008) provides an excellent review of these approaches (see Table I). Sui et al. (2012) is another good review of multivariate methods for multimodal fusion of brain imaging data.

TABLE I. COMBINING INFORMATION IN TWO CONNECTIVITIES

		Functional Connectivity	
		Separate Analysis of Regions	Combined Analysis of Regions
Structural Connectivity	Separate Analysis of Regions	Univariate functional analysis using anatomy as reference.	Multivariate functional study on the basis of anatomical information.
	Combined Analysis of Regions	Univariate functional study informed by white matter tract analysis.	Multivariate functional Connectivity study informed by structural connectivity analysis.

In traditional functional brain imaging studies, each region is analyzed separately for both structural and functional connectivity (row 3, column 3 in Table I). This type of analysis provides little direct information on combining the two measures aside from showing effects on the two connectivities and suggesting possible networks linked with the condition. The widely used functional connectivity studies have been conducted for more than two decades using the seeded ROIs based on structural information (row 3, column 4 in Table I). A set of structural imaging data is usually obtained first to serve as a basis for localizing seeds for later CCA of functional connectivity. It would be ideal to assess all possible connections among various brain regions and identify those with high functional connectivity. However, this approach is usually computationally demanding. In reality, most of the studies start by extracting ROIs defined based on structural data and then studying the functional relationship among the selected ROIs. For example, Stein et al. (2000) studied seed voxels located in the hippocampus or thalamus using structural information and found highly significant correlations with voxels in the ipsilateral and contralateral hippocampus (Stein et al., 2000). In many studies a large number of ROIs

are examined for strong functional relationships, forming functional networks involved in particular tasks (Rykhlevskaia et al., 2008).

2.5.1 Structural Connectivity Analysis with Functional Data

The remaining two approaches in Table I represent more complex analyses of structural and functional connectivity. “Structural analysis of brain connectivity with univariate analysis of functional imaging data” (Rykhlevskaia et al., 2008, 178) (row 4, column 3 in Table I) provides valuable information on how structural connectivity is involved in the activities of various cortical areas during processing. These studies “do not provide direct information on functional connectivity (Rykhlevskaia et al., 2008, 174).” The DTI and fMRI data are usually collected sequentially from the same subject using the same scanner within the same time period. The first study of this type was conducted by Werring et al. (1999) and demonstrated the feasibility of combining structural and functional data to study the structural properties of activated brain regions. These authors collected DTI and fMRI data on healthy subjects and showed that a BOLD response in areas with low FA corresponded to the relatively isotropic cortical gray matter (Werring et al., 1999). Both this and a later study conducted by Seghier et al. (2004) addressed the question of the mechanisms of structure-function interaction at a very broad level and discussed the potential of the integration of the two connectivity measures. Since then, a number of studies combined DTI and fMRI data in clinical applications (Rykhlevskaia et al., 2008). Most of these studies were focused on either the motor or occipital cortex because of the fact that DTI is only capable of detecting major white-matter structures possessing specific orientations.

Studies of structural connectivity using univariate functional data can be classified into three categories based on their main objectives. The first type of studies has been focused on improving the imaging of structural connectivity. In these studies fMRI results served as auxiliary information aiding the analysis of DTI data. Because fMRI is a well-established imaging method, this approach allows for a substantial reduction in the number of potential structural connections for analysis. For example, Guye et al. (1997) used seed regions adjacent to the location of maximum fMRI activity and constructed three-dimensional connectivity maps of DTI images within the whole brain. The second type of studies has tried to investigate the structure-function relationships. For instance, Johansen-Berg et al. (2004) studied the functional connections in cortex defined by the structural connectivity profile. And later studies by Toosy et al. (2004) and Seghier et al. (2004) addressed the relationships between measures of neural structures in optic radiations and neural activities in the visual cortex. Understanding the structural-functional relationships is important in determining the linkage between structural brain damage and functional impairment. So far relatively little progress has been made in this area of research. The third type of studies has been focused on neurosurgical applications. Combining DTI and fMRI data can provide valuable information for presurgical brain mapping. For example, Berman et al. (2004) showed that tracking fibers using DTI and fMRI has the potential to reduce surgical morbidity by revealing subcortical connections of the functional cortex. Recent studies have also started to investigate the potential applications of multimodal imaging in real-time neuronavigation systems (Krings et al., 2001).

2.5.2 Functional Connectivity Analysis with Structural Data

Structural connections between different brain regions provide a picture of multicomponent structural networks and neural pathways. Traditionally information was summarized from literature about structural anatomical pathways. Recently, as DTI emerged as a powerful method for assessing structural connectivity, more and more studies have been combining measures of functional connectivity, white-matter integrity, and sometimes even behavior within one study. This is particularly meaningful in deepening our understanding on how functional connectivity may mediate the relationship between brain structure and human behavior. A good example of these studies is by Cook et al. (2002). These authors examined structural, functional, and behavioral data and showed that cognitive performance was best explained by coherence, and only secondarily by subclinical structural brain disease, which is a disorder characterized by alterations in brain structure-cortical atrophy, central atrophy, and deep white-matter hyperintensities (Cook et al., 2002). Another work by Rykhlevskaia et al. (2006) applied dynamic factor analysis for testing various structural models on the lagged covariance matrices derived from the event-related optical signal data, and showed that the age-related reduction of corpus callosum size is linked to changes in functional connectivity patterns that further affect preparatory processes (Rykhlevskaia et al., 2006). These results provide insights into how function may generally relate to the underlying structure. However, they are not enough to inform on the one-to-one linkage between structural and functional connections between two brain regions.

Several studies have tried to address the issue of structural-functional relationships. For example, in a brain-mapping meta-analysis of 107 studies, Koski and Paus (2000) found that “activity in different parts of the anterior cingulate cortex was correlated with

different areas of the frontal cortex, consistent with known differences in the anatomical connectivity of these regions.” This result supports the idea of independent functional and structural connectivity. However, the details behind this potential independence between the two modalities are largely not clear. It is speculated that direct structural connectivity is neither a sufficient nor a necessary condition for observing functional activity (Rykhlevskaia et al., 2008). For the first part, it is not necessary because structural connections only provide neural pathways for functional connectivity. The actual occurrence of coordinated activities between two brain regions is much more complex and it is not uncommon that no information exchange happens even though a number of neural pathways exist between two brain regions. A good example of this is shown in the work of Hampson et al. (2004). These authors examined correlations between the middle temporal area and other brain regions in a resting or active state and found that, during active tasks, only a subset of the areas in the visual network was correlated with activity in the middle temporal area (Hampson et al., 2004). Another issue to consider is the subject effect. As there are possibly many neural pathways existing between two brain regions, the actual pathways engaged for a specific task may vary from subject to subject. The work by Schoenfeld et al. (2002) tested this idea in a patient with homonymous hemianopsia and Riddoch syndrome, and found that “visual perception after the V1 lesion in Riddoch syndrome was mediated by subcortical pathways that bypass V1,” feeding directly into higher order visual areas. It is highly likely that these pathways originally exist and are secondary under normal conditions.

Structural connectivity may not be necessary for observing functional activity either. Actually in a lot of cases the number of observed functional connections are more than that of existing neural pathways. Current methods for assessing functional connectivity measure the cross-correlations of the activities between brain regions. These

unconstrained analyses are unable to distinguish true direct functional connections from indirect and spurious relations. The latter are classified as functional connectivity by current methods. However, they cannot be explained by the presence of underlying structural connections and are thus not meaningful in representing information exchange. Eggermont and Smith (1996) analyzed cross-correlogram peaks for spontaneous and stimulated activity in the auditory cortex and showed that spontaneous and secondary effects can obscure the majority of true functional relationships produced by actual structural connections (Eggermont and Smith, 1996). Moreover, Koch et al. (2002) studied the relationships between functional and structural connectivity using fMRI and DTI data and failed to observe a simple linear relationship between the two connectivity measures. These authors state that “the regions which are clearly directly linked by white matter fiber tracts should show high functional connectivity, but that the inverse need not to be true as functional connectivity may also be indirectly mediated via more distant gray matter regions” (Koch et al., 2002, 241). Taken together, studies have shown that structural connection is a good predictor of functional relationship. However, the observed functional relationships can originate from various pathways and direct structural connections may not be necessary. Future directions in this area will be focused on how to quantitatively integrate multimodal neuroimaging data in a probabilistic way so that a better understanding of the structural-functional relationships within the brain can be achieved.

3. NEUROIMAGING DATA

Our neuroimaging data contain two measures, DTI for structural connectivity, and fMRI for functional connectivity. We recruited 23 subjects (13 healthy control, HC, and 10 adults with LLD) and analyzed connectivity among 87 brain regions in each subject.

3.1 Subject Recruitment

Subject recruitment is the same as described in Zhang et al. (2012) and took place via community outreach (e.g., newspaper, radio, and television advertisements) and relevant outpatient clinics. The inclusion criteria for all subjects were “55 years of age and older, medication-naïve or anti-depressant free for at least two weeks (in the case of subjects with LLD) and no history of unstable cardiac or neurological diseases. The exclusion criteria included: schizophrenia, bipolar, or any psychotic disorders; history of anxiety disorder outside of major depressive episodes; history of head trauma; history of substance abuse; contraindications to MRI such as metal implants. This study was approved by the University of Illinois-Chicago Institutional Review Board, and written informed consent was obtained from each participant.”

According to Zhang et al. (2012), “all eligible subjects were assessed by a trained research assistant with the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. The severity of depression was quantified by a board-certified/board-eligible psychiatrist using the 17-item Hamilton Depression Rating Scale (Hamilton, 1960). At the time of enrollment, LLD subjects met the criteria set forth in the Structured Clinical Interview for Diagnostic and Statistical Manual

of Mental Disorders for MDD and required a score of 15 or greater on the Hamilton Depression Rating Scale.”

3.2 Image Acquisition

According to Zhang et al. (2012), “brain MRIs were acquired on a Philips 3.0T Achieva scanner (Philips Medical Systems, Best, The Netherlands) using an 8-channel SENSE (Sensitivity Encoding) head coil. Participants were positioned comfortably on the scanner bed and fitted with soft ear plugs; foam pads were used to minimize head movement. Participants were instructed to remain still throughout the scan. High resolution three-dimensional T1-weighted images were acquired with a Magnetization Prepared Rapid Acquisition Gradient Echo sequence (field of view: [FOV] = 240 mm; 134 contiguous axial slices; TR/TE = 8.4/3.9 ms; flip angle = 8°; voxel size = $1.1 \times 1.1 \times 1.1$ mm). Resting-state data were acquired with the following parameters: Single-shot gradient-echo echo-planar imaging (EPI) sequence, TR/TE = 2000/30 ms, Flip angle = 80°, EPI factor = 47, FOV = $23 \times 23 \times 15$ cm³, in-plane resolution = 3×3 mm², slice thickness/gap = 5/0 mm, slice number = 30, SENSE reduction factor = 1.8, NEX = 200, total scan time = 6 : 52. Subjects were instructed to keep their eyes closed and “not think of anything in particular.” Single-shot spin-echo EPI sequence (FOV = 240 mm; voxel size = $0.83 \times 0.83 \times 2.2$ mm; TR/TE = 6,994/71 ms; Flip angle = 90°) were used when acquiring DTI images. Sixty-seven contiguous axial slices aligned to the anterior commissure-posterior commissure line were collected in 32 gradient directions with $b = 700$ s / mm² and one acquisition without diffusion sensitization (B0 image). Parallel imaging technique was utilized with factor at 2.5 to reduce scanning time to approximately four minutes.”

3.3 Data Processing

Structural brain networks were generated using a pipeline that integrates multiple image analysis techniques and has been reported elsewhere (GadElkarim et al., 2012; Leow et al., 2013). In brief, diffusion-weighted MRI images were eddy-current corrected using the automatic image registration tool embedded in DtiStudio software (<http://www.mristudio.org>) by “registering all diffusion-weighted MRI images to their corresponding b0 images with 12-parameter affine transformations.” This was followed by “the computation of diffusion tensors and deterministic tractography” using the DtiStudio program. The T1-weighted images were used to generate label maps using the Freesurfer software (<http://surfer.nmr.mgh.harvard.edu>). Brain networks formed by the 87 cortical/subcortical grey-matter regions were generated using an in-house program in Matlab by counting the number of fibers connecting each pair of nodes.

Functional brain networks were generated using the resting-state fMRI toolbox, CONN (<http://www.nitrc.org/projects/conn>; Whitfield-Gabrieli and Nieto-Castanon, 2012). In brief, raw EPI images were realigned, co-registered, normalized, and smoothed before analyses. Confound effects from motion artifact, white matter, and cerebrospinal fluid were regressed out of the signal. Using the same 87 labels as the structural brain networks, functional brain networks were derived using pairwise BOLD signal correlations.

For example, each subject would have the structural connectivity data matrix for the 87 brain regions as

$$\begin{pmatrix} - & & & & \\ DTI_{1-2} & - & & & \\ DTI_{1-3} & DTI_{2-3} & - & & \\ .. & ... & ... & - & \\ DTI_{1-87} & DTI_{2-87} & ... & ... & - \end{pmatrix}$$

Therefore, for each subject, a structural/functional connectivity data set for 87 brain regions contains measures for the $87 \times 86/2 = 3471$ connectivities. The 87 brain regions analyzed in this work are listed in Table II.

TABLE II. BRAIN REGIONS ANALYZED

L- & R-Cerebellum Cortex	L- & R-Thalamus Proper
L- & R-Caudate	L- & R-Putamen
L- & R-Pallidum	L- & R-Hippocampus
L- & R-Amygdala	L- & R-Accumbens area
L- & R-VentralDC	L- & R-bankssts
L- & R-caudalanteriorcingulate	L- & R-caudalmiddlefrontal
L- & R-cuneus	L- & R-entorhinal
L- & R-fusiform	L- & R-inferiorparietal
L- & R-inferiortemporal	L- & R-isthmuscingulate
L- & R-lateraloccipital	L- & R-lateralorbitofrontal
L- & R-lingual	L- & R-medialorbitofrontal
L- & R-middletemporal	L- & R-parahippocampal
L- & R-paracentral	L- & R-parsopercularis
L- & R-parsorbitalis	L- & R-parstriangularis
L- & R-pericalcarine	L- & R-postcentral
L- & R-posteriorcingulate	L- & R-precentral
L- & R-precuneus	L- & R-rostralanteriorcingulate
L- & R-rostralmiddlefrontal	L- & R-superiorfrontal
L- & R-superiorparietal	L- & R-superiortemporal
L- & R-supramarginal	L- & R-frontalpole
L- & R-temporalpole	L- & R-transversetemporal
L- & R-insula	Brain Stem

3.4 Networking Analysis Based on Degrees of Similarity in Connectivity

We employed an approach in analyzing differences of network structure between LLD and HC subjects based on degree of similarity in connectivities. In this approach, gamma and beta distributions were fitted to DTI and fMRI measures (absolute values taken for fMRI). Then, for both the LLD and HC groups, the strengths of each connectivity in the LLD and HC groups were assigned a level of either high, medium high, medium low, or low according to the quartiles of the gamma or beta distributions. A 4×4 table was constructed by using the connectivity strength level in the LLD and HC as the two margins. Each link was placed into the cell corresponding with its connectivity

strength level in LLD and HC. We then analyzed the the degrees of similarity between the strength of connections in the two groups. We were specifically interested in those links that exhibit stronger connectivity in the HC group, but lower connectivity in the LLD group (i.e., damaged connectivity in LLD). Network diagrams were created for these links and compared between the two groups.

In addition, we performed a node-level analysis using the same idea. For each node, we assigned its connectivities with all other nodes into the four categories according to their strengths in the LLD and HC groups. Then these connectivities were placed into the cells of the 4×4 table and the degrees of similarity were analyzed between the strength of connections in the two groups. Again, we were focusing on links that exhibit stronger connectivity in the HC group, but lower connectivity in the LLD group.

4. BIVARIATE MIXED-EFFECTS MODEL

4.1 The Use of Linear Mixed-Effects Model for Neuroimaging Data

Data collected in research studies often contain hierarchical structures. For example, in a study of the effect of gender and class size on reading achievement in the third grade, students are nested within schools that are further nested within school districts. Another good example is a longitudinal trial, where multiple measures are taken from each study subject (cluster). As we can imagine, the observations within a cluster are not independent, but rather correlated with each other. Ignoring this correlation causes inefficient use of information contained in the data and often leads to biased inferences. To account for the correlation among observations within a cluster, random effects are introduced into models. These so called mixed-effects regression models can explain the hierarchy of data and thus enable us to make unbiased inferences regarding our hypotheses. Mixed-effects models for longitudinal and clustered outcomes are well established. For details about model specification, estimation, and application, we refer to an excellent book by Hedeker and Gibbons (2006).

In practice, however, it is not uncommon that multiple outcomes are measured simultaneously in a study. When this happens, for simplicity, a common way to analyze the data is to model each outcome and interpret the result separately. However, if outcomes are interrelated, which is the case most of the time, modeling them separately will lose efficiency since the correlation of the outcomes will be ignored. Simultaneous analysis of multiple outcomes involves multivariate models that make more efficient use of information and are yet more complicated to construct and estimate.

Recently a number of studies have made use of bivariate mixed-effects models to model outcomes measured at a single time point in a study. For example, Sitholea and Jones (2007) used bivariate mixed-effects models to detect changes in the prescribing habits in the United Kingdom due to an educational intervention. Inoue et al. (2008) modeled Baltimore Longitudinal Study of Aging data using a joint longitudinal Bayesian natural history model. Hubbard et al. (2009) proposed “a joint model for self-rated health and physical functioning that describes the relationship between perceptions of health and the rate of change of physical functioning or disability.” Ghosh and Tu (2008) proposed a pair of simultaneous models for the zero-inflated event counts to analyze longitudinally measured sexual behavioral data.

Research is lacking in simultaneous statistical modeling of structural and functional connectivity for population studies. The only attempt to date is the work by Venkataraman et al. that merged structural and functional information in a probabilistic framework and identified connectivity patterns in the brain (Venkataraman et al., 2012). These authors modeled the interaction between latent structural and functional connectivity from an underlying process in the brain not observed directly from the data, and employed an expectation-maximization (EM) algorithm to estimate the model parameters maximizing the data likelihood. By using separate latent variables, the model “simultaneously infers the templates of latent connectivity for each group and the differences in connectivity between the groups.” The authors illustrated the method using a schizophrenia study and identified “significant increases in functional connectivity between the parietal/posterior cingulate region and the frontal lobe, and reduced functional connectivity between the parietal/posterior cingulate region and the temporal lobe in schizophrenia.” It was also shown that the simultaneous modeling of the two modalities yielded better results than analyzing each one separately.

Venkataraman et al. (2012) showed that the model exhibited improved classification and better generalization accuracy than similar models fitting only one of the modalities, indicating that simultaneous modeling of the two connectivity modalities does provide better efficiency. However, there are some drawbacks of this work. First, it models latent connectivity via discrete random variables, which simplifies the connectivity into all-or-none categories. This marginalization may oversimplify the situation and fail to detect subtle between-group differences. Next, the model assumed that all subject data are drawn from the same distribution. In practice, especially in the realm of mental health research, subject variation can be high and both structural and functional connectivity can vary across subjects. Failure to account for the between-subject variability can inflate standard error estimates and lead to biased inferences. And finally, the model does not provide parameters that measure the interactions between structural and functional connectivity. A thorough knowledge of how functional abnormality may be related to structural damage will shed light on our understanding of how impaired pathways may hamper normal treatment responses and how intact pathways can be employed in the development of novel treatments. This is crucial for the development of treatments addressing alterations in brain networks associated with LLD and other mental disorders.

In our case, structural and functional connectivity data, as measured by DTI and fMRI, of links among various brain regions are collected from the same subject at the same time. To appropriately account for the correlation between connectivity measures taken from the same subject, we would like to use two-level models with random subject effects. In addition, both structural and functional connectivity measures are taken from the same subject for the same link, which requires simultaneous modeling of the two outcomes. Moreover, since it is our aim to study the relationship between the structural and functional connectivities, our model must allow us to not only correctly incorporate

but also efficiently estimate the correlations between the two connectivity modalities for each link. For these purposes, we have constructed bivariate mixed-effects models with random subject intercepts and heteroscedastic errors (i.e., error distributions differ among links). In this chapter we provide details about the definition, estimation, and inference of the bivariate mixed-effects model specifically constructed for modeling structural and functional neuroimaging data obtained for a large number of brain regions.

4.2 Model Definition

We would like to obtain the mean and covariance structure of fMRI and DTI measures for each of the links. Hence we allow the mean and covariance to differ for each link. Observations are correlated within-subject since multiple measures were taken on the same subject. Two random intercepts, one for fMRI and the other for DTI, were introduced to our model to incorporate such within-subject correlations. The model can be written as:

$$Y_{ijk} = \beta_{0ik} \times (1 - Group_j) + \beta_{1ik} \times Group_j + \gamma_{jk} + \epsilon_{ijk}, \quad (4.11)$$

where $i = 1, 2, \dots, m$ denotes the i th link, $j = 1, 2, \dots, n$ denotes the j th subject, $k = 1$ if the measure is fMRI, $k = 2$ if the measure is DTI, $Group_j = 0$ if the j th subject is from the control group, and $Group_j = 1$ if the j th subject is from the experimental group. We assume $\gamma_j = (\gamma_{j1}, \gamma_{j2}) \sim N(\mathbf{0}, \Sigma_\gamma)$, $\epsilon_{ij} \sim N(\mathbf{0}, \Sigma_{0i})$ if the j th subject is from the control group, and $\epsilon_{ij} \sim N(\mathbf{0}, \Sigma_{1i})$ if the j th subject is from the experimental group.

Based on this model, given the random effects, the bivariate distribution of the fMRI and DTI measures of the i th link can be written as

$$\mathbf{y}_{ij}|\boldsymbol{\gamma}_j \sim N(\boldsymbol{\beta}_{0i}, \Sigma_{0i}), \quad (4.12)$$

if the j th subject is from the control group, and

$$\mathbf{y}_{ij}|\boldsymbol{\gamma}_j \sim N(\boldsymbol{\beta}_{1i}, \Sigma_{1i}), \quad (4.13)$$

if the j th subject is from the experimental group.

With this model, the mean outcome vectors for the i th link for the control and experimental groups are $\boldsymbol{\beta}_{0i}$ and $\boldsymbol{\beta}_{1i}$, respectively, and the Pearson correlation coefficients of the two measures for the i th link can be computed from Σ_{0i} and Σ_{1i} for the control and experimental group, respectively.

When fixed covariates are included, the model becomes:

$$Y_{ijk} = \beta_{0ik} \times (1 - Group_j) + \beta_{1ik} \times Group_j + \mathbf{X}_{ijk}\boldsymbol{\beta}_c + \gamma_{jk} + \epsilon_{ijk}, \quad (4.14)$$

where $\boldsymbol{\beta}_c$ is the vector of coefficients for the fixed covariates and $\mathbf{X}_{ijk}$ is the corresponding design matrix.

4.3 Parameter Estimation

For simplicity reasons in this section we only consider the model without covariates. Estimation of model parameters with fixed covariates may be conducted in a similar way

by including the corresponding covariate parameters in the coefficient vector $\boldsymbol{\beta}$. Rewrite the mixed-effects model for the j th subject as:

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \boldsymbol{\gamma}_j + \boldsymbol{\varepsilon}_j, \quad (4.15)$$

where $\mathbf{y}_j$ is a $2m \times 1$ vector of the outcome (m DTI measures and m fMRI measures) for the j th subject, $\boldsymbol{\beta}$ is a $4m \times 1$ (for each of the m links there are 4 fixed effects: mean DTI and fMRI for the two groups) vector for the fixed-parameters, $\boldsymbol{\gamma}_j$ is a 2×1 (two random subject effect, one for DTI, and the other for fMRI) vector of random effects for the j th subject, $\mathbf{X}_j$ and $\mathbf{Z}_j$ are $2m \times 4m$ and $2m \times 2$ design matrices of the fixed and random effects for the j th subject, and $\boldsymbol{\varepsilon}_j$ is a $2m \times 1$ vector of errors.

Then the observation $\mathbf{y}_j$ and random effects $\boldsymbol{\gamma}_j$ have joint normal distribution

$$\begin{bmatrix} \mathbf{y}_j \\ \boldsymbol{\gamma}_j \end{bmatrix} \sim MVN \left(\begin{bmatrix} \mathbf{X}_j \boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Z}_j \Sigma_\gamma \mathbf{Z}_j^T + \Sigma_j & \mathbf{Z}_j \Sigma_\gamma \\ \Sigma_\gamma \mathbf{Z}_j^T & \Sigma_\gamma \end{bmatrix} \right), \quad (4.16)$$

where Σ_γ is a 2×2 covariance matrix of random effects and Σ_j is a $2m \times 2m$ block diagonal matrix of the error covariance matrices of all links.

Estimation of the model is based on an EM algorithm. In the E step, with the current values of the other parameters, we compute the “expected a posteriori” or empirical Bayes (EB) estimates of the random effects as well as the conditional variances of the random effects given data. In the M step, given the current values of the random effects, we obtain the maximum marginal likelihood (MML) estimates of the regression coefficients, error variances, and the variances of the random effects. The algorithm iterates between the EB and MML estimates until convergence.

4.3.1 Empirical Bayes Estimation

Two stochastic processes are taken into account when making EB inferences: one for the data $f(\mathbf{y}_j|\boldsymbol{\gamma}_j)$, and the other one for the random effects $g(\boldsymbol{\gamma}_j)$. The posterior distribution of the random effects given data $g(\boldsymbol{\gamma}_j|\mathbf{y}_j)$ contains all information about $\boldsymbol{\gamma}_j$ available in $\mathbf{y}_j$. The Bayes theorem states

$$g(\boldsymbol{\gamma}_j|\mathbf{y}_j) = \frac{f(\mathbf{y}_j|\boldsymbol{\gamma}_j)g(\boldsymbol{\gamma}_j)}{h(\mathbf{y}_j)}. \quad (4.17)$$

We want to obtain the posterior mean $\tilde{\boldsymbol{\gamma}}_j$ and the covariance matrix $\tilde{\Sigma}_{\boldsymbol{\gamma}|\mathbf{y}_j}$ of the random effects. Since $\boldsymbol{\gamma}_j$ and $\mathbf{y}_j$ are jointly normally distributed (4.16), the conditional distribution can be written as

$$\begin{aligned} \boldsymbol{\gamma}_j|\mathbf{y}_j &\sim N(\Sigma_{\boldsymbol{\gamma}}\mathbf{Z}_{ij}^T(\mathbf{Z}_j\Sigma_{\boldsymbol{\gamma}}\mathbf{Z}_j^T + \Sigma_j)^{-1}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta}), \\ &\quad \Sigma_{\boldsymbol{\gamma}} - \Sigma_{\boldsymbol{\gamma}}\mathbf{Z}_j^T(\mathbf{Z}_j\Sigma_{\boldsymbol{\gamma}}\mathbf{Z}_j^T + \Sigma_j)^{-1}\mathbf{Z}_j\Sigma_{\boldsymbol{\gamma}}). \end{aligned} \quad (4.18)$$

The posterior mean provides the EB estimate of $\boldsymbol{\gamma}_j$ and the posterior covariance matrix provides the EB estimate of the uncertainty about $\boldsymbol{\gamma}_j$. Hence we have

$$\begin{aligned}
\tilde{\gamma}_j &= \Sigma_\gamma \mathbf{Z}_j^T (\mathbf{Z}_j \Sigma_\gamma \mathbf{Z}_j^T + \Sigma_j)^{-1} (\mathbf{y}_j - \mathbf{X}_j \beta) \\
&= \Sigma_\gamma \left[\Sigma_\gamma + (\mathbf{Z}_j^T \Sigma_j^{-1} \mathbf{Z}_j)^{-1} \right]^{-1} (\mathbf{Z}_j^T \Sigma_j^{-1} \mathbf{Z}_j)^{-1} \mathbf{Z}_j^T \Sigma_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \beta) \\
&= \mathbf{R} (\mathbf{Z}_j^T \mathbf{Z}_j)^{-1} \mathbf{Z}_j^T (\mathbf{y}_j - \mathbf{X}_j \beta) \\
\tilde{\Sigma}_\gamma | \mathbf{y}_j &= \Sigma_\gamma - \Sigma_\gamma \mathbf{Z}_j^T (\mathbf{Z}_j \Sigma_\gamma \mathbf{Z}_j^T + \Sigma_j)^{-1} \mathbf{Z}_j \Sigma_\gamma \\
&= (\mathbf{Z}_j^T \Sigma_j^{-1} \mathbf{Z}_j + \Sigma_\gamma^{-1})^{-1} \\
&= (\mathbf{I} - \mathbf{R}) \Sigma_\gamma,
\end{aligned} \tag{4.19}$$

where $\mathbf{R} = \Sigma_\gamma \left[\Sigma_\gamma + (\mathbf{Z}_j^T \Sigma_j^{-1} \mathbf{Z}_j)^{-1} \right]^{-1}$. Note that compared to the expression in (4.18) that involves inversion of $2m \times 2m$ matrices, the formula in (4.19) only requires inverting 2×2 matrices.

4.3.2 Maximum Marginal Likelihood Estimation

To obtain the MML estimates, the joint distribution needs to be integrated over the distribution of the random effects to get the marginal distribution

$$h(\mathbf{y}_{ij}) = \int_{\boldsymbol{\gamma}} f(\mathbf{y}_{ij} | \boldsymbol{\gamma}_j) g(\boldsymbol{\gamma}_j) d\boldsymbol{\gamma}, \tag{4.20}$$

where $f(\mathbf{y}_{ij} | \boldsymbol{\gamma}_j) = N(\mathbf{X}_{ij} \beta_i + \mathbf{Z}_{ij} \boldsymbol{\gamma}_j, \Sigma_{0i})$ if the j th subject is from the control group, $f(\mathbf{y}_{ij} | \boldsymbol{\gamma}_j) = N(\mathbf{X}_{ij} \beta_i + \mathbf{Z}_{ij} \boldsymbol{\gamma}_j, \Sigma_{1i})$ if the j th subject is from the experimental group, and $g(\boldsymbol{\gamma}_j) = N(\mathbf{0}, \Sigma_\gamma)$.

To obtain the MML estimates of β_i , Σ_i , and Σ_γ , the marginal log-likelihood needs to be summed over all regions and subjects. Hence

$$\log L = \sum_{i=1}^m \sum_{j=1}^n \log h(\mathbf{y}_{ij}) = \sum_{i=1}^m \sum_{j=1}^n \log \left[\int_{\boldsymbol{\gamma}} f(\mathbf{y}_{ij} | \boldsymbol{\gamma}_j) g(\boldsymbol{\gamma}_j) d\boldsymbol{\gamma} \right]. \quad (4.21)$$

After differentiation with respect to $\boldsymbol{\beta}_i$ we obtain, for fixed covariates, that

$$\hat{\boldsymbol{\beta}}_i = \left(\sum_{j=1}^n \mathbf{X}_{ij}^T \mathbf{X}_{ij} \right)^{-1} \left[\sum_{j=1}^n \mathbf{X}_{ij}^T (\mathbf{y}_{ij} - \mathbf{Z}_{ij} \tilde{\boldsymbol{\gamma}}_j) \right]. \quad (4.22)$$

And the variance of the fixed effects $\boldsymbol{\beta}_i$ can be estimated from

$$\text{Var}(\hat{\boldsymbol{\beta}}_i) = \left[\mathbf{X}_i^T (\Sigma_i + \mathbf{Z}_i \Sigma_{\boldsymbol{\gamma}} \mathbf{Z}_i^T)^{-1} \mathbf{X}_i \right]^{-1}, \quad (4.23)$$

where Σ_i is the block diagonal matrix of the error covariance matrix of the i th link for all subjects.

For residual variance Σ_{0i} for $\hat{\boldsymbol{\epsilon}}_{0ij} = \mathbf{y}_{ij} - \mathbf{X}_{ij} \hat{\boldsymbol{\beta}}_i - \mathbf{Z}_{ij} \tilde{\boldsymbol{\gamma}}_j$ for the control group, and Σ_{1i} for $\hat{\boldsymbol{\epsilon}}_{1ij} = \mathbf{y}_{ij} - \mathbf{X}_{ij} \hat{\boldsymbol{\beta}}_i - \mathbf{Z}_{ij} \tilde{\boldsymbol{\gamma}}_j$ for the experimental group, we obtain

$$\begin{aligned} \hat{\Sigma}_{0i} &= \frac{1}{n_0} \sum_{j=1}^{n_0} (\hat{\boldsymbol{\epsilon}}_{0ij} \hat{\boldsymbol{\epsilon}}_{0ij}^T + \mathbf{Z}_{0ij} \tilde{\Sigma}_{\boldsymbol{\gamma}} | \mathbf{y}_j \mathbf{Z}_{0ij}^T), \\ \text{and } \hat{\Sigma}_{1i} &= \frac{1}{n_1} \sum_{j=1}^{n_1} (\hat{\boldsymbol{\epsilon}}_{1ij} \hat{\boldsymbol{\epsilon}}_{1ij}^T + \mathbf{Z}_{1ij} \tilde{\Sigma}_{\boldsymbol{\gamma}} | \mathbf{y}_j \mathbf{Z}_{1ij}^T), \end{aligned} \quad (4.24)$$

where n_0, n_1 are the sample sizes and $\mathbf{Z}_{0ij}, \mathbf{Z}_{1ij}$ are the design matrices corresponding to random effects for the control and experimental groups, respectively. Pearson correlation coefficients ρ_{0i} and ρ_{1i} can be calculated from the above estimated error covariance matrices.

For variance of the random effects, we have

$$\hat{\Sigma}_\gamma = \frac{1}{n} \sum_{j=1}^n \left(\tilde{\gamma}_j \tilde{\gamma}_j^T + \tilde{\Sigma} \gamma | \mathbf{y}_j \right). \quad (4.25)$$

We see that the total error (or random effect) variance is the sum of the expected variance computed from the estimated residuals (or random effects) and the EB estimates of the uncertainty about the residuals (or random effects). The details about the above derivations can be found in Appendix A. The estimation algorithm iterates between the EB estimates of the random effects and the MML estimates of the fixed effects and error variances until it converges.

Finally, it is known that the covariance of normal random variables follows a Wishart distribution. As the errors in this model are assumed to follow normal distributions, their covariance matrices should follow Wishart distributions. Then we have the covariance estimates of the variance components as can be found in Muirhead (1982), page 90:

$$\begin{aligned} Cov[vec(\hat{\Sigma}_{0i})] &= n_0(\mathbf{I}_4 + \mathbf{K}) \left(\frac{\hat{\Sigma}_{0i}}{n_0} \otimes \frac{\hat{\Sigma}_{0i}}{n_0} \right), \\ \text{and } Cov[vec(\hat{\Sigma}_{1i})] &= n_1(\mathbf{I}_4 + \mathbf{K}) \left(\frac{\hat{\Sigma}_{1i}}{n_1} \otimes \frac{\hat{\Sigma}_{1i}}{n_1} \right), \end{aligned} \quad (4.26)$$

$$\text{where } \mathbf{K} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Using the delta method, we can obtain the variance estimates for the estimates of Pearson correlation coefficients $\hat{\rho}_{0i}$ and $\hat{\rho}_{1i}$.

4.4 Simulation Study

We ran a simulation study to evaluate the performance of our algorithm for estimating the bivariate mixed-effects model with heteroscedastic errors. In this small simulation, we assumed a sample of 50 subjects (25 experimental, 25 control), with structural and functional connectivity data taken for 40 brain regions of each subject. The proposed values for the model parameters were based on estimates obtained from our neuroimaging data. We assumed, for the i th link, that

$$\begin{aligned}
\beta_i &= \begin{bmatrix} \beta_{0i1} \\ \beta_{0i2} \\ \beta_{1i1} \\ \beta_{1i2} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 2 \\ 0.5 \\ 3 \end{bmatrix}, \begin{bmatrix} 0.05 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.05 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right) \\
\Sigma_{0i} &= \begin{bmatrix} \sigma_{01i}^2 & \rho_{0i}\sigma_{01i}\sigma_{02i} \\ \rho_{0i}\sigma_{01i}\sigma_{02i} & \sigma_{02i}^2 \end{bmatrix} \\
\Sigma_{1i} &= \begin{bmatrix} \sigma_{11i}^2 & \rho_{1i}\sigma_{11i}\sigma_{12i} \\ \rho_{1i}\sigma_{11i}\sigma_{12i} & \sigma_{12i}^2 \end{bmatrix} \\
\sigma_{01i}^2 &\sim \Gamma(0.2, 1); \sigma_{02i}^2 \sim \Gamma(1, 1); \rho_{0i} \sim U(-1, 0.5) \\
\sigma_{11i}^2 &\sim \Gamma(0.3, 1); \sigma_{12i}^2 \sim \Gamma(1.5, 1); \rho_{1i} \sim U(-0.5, 1) \\
\Sigma_\gamma &= \begin{bmatrix} \sigma_{\gamma 1}^2 & \rho_\gamma \sigma_{\gamma 1} \sigma_{\gamma 2} \\ \rho_\gamma \sigma_{\gamma 1} \sigma_{\gamma 2} & \sigma_{\gamma 2}^2 \end{bmatrix} = \begin{bmatrix} 0.05 & 0.05 \\ 0.05 & 0.5 \end{bmatrix}
\end{aligned} \tag{4.27}$$

Neuroimaging data were then simulated using the model

$$\begin{aligned}
Y_{ijk} &= \beta_{0ik} \times (1 - Group_j) + \beta_{1ik} \times Group_j + \gamma_{jk} + \epsilon_{ijk} \\
\boldsymbol{\gamma}_j &= (\gamma_{j1}, \gamma_{j2}) \sim N(\mathbf{0}, \Sigma_\gamma) \\
\boldsymbol{\epsilon}_{ij} &\sim N(\mathbf{0}, \Sigma_{0i}) \text{ (for control)} \\
\boldsymbol{\epsilon}_{ij} &\sim N(\mathbf{0}, \Sigma_{1i}) \text{ (for experimental)}
\end{aligned} \tag{4.28}$$

Parameters were estimated using the above-mentioned EM algorithm that iterates between the EB and MML estimates. The process was repeated 1,000 times and the results were summarized in Table III. We assessed the performance of our estimation algorithm by monitoring the parameter estimates obtained for the fixed effects as well as the variance components. Since we had link-specific fixed effects and error variance parameters, we evaluated the accuracy and precision of our estimates by examining the first (mean, μ) and second (variance, σ^2) moments of them. Table III demonstrates that, for a relatively small sample of 50 subjects, both the average means and variances of the fixed effects as well as the error variance components were close to the true values (Table III, column 3). Also shown in the table are the root mean square error values, which were generally very small (Table III, column 4). Similar results were observed for estimates of the random effect variance components. Taken together, the simulation results demonstrate satisfactory performance of our estimation algorithm for the bivariate mixed-effects model with heteroscedastic errors.

TABLE III. PARAMETER ESTIMATES IN BMEM SIMULATION

Parameter	Variable	True Value	Mean (SD)	RMSE
Fixed	$\mu_{\beta_0 i1}$	0	0.0023 (0.0464)	0.0465
	$\mu_{\beta_0 i2}$	2	1.9988 (0.1487)	0.1487
	$\mu_{\beta_1 i1}$	0.5	0.5011 (0.0447)	0.0447
	$\mu_{\beta_1 i2}$	3	2.9882 (0.1425)	0.1430
	$\sigma_{\beta_0 i1}^2$	0.05	0.0578 (0.0032)	0.0084
	$\sigma_{\beta_0 i2}^2$	0.5	0.5401 (0.028)	0.0489
	$\sigma_{\beta_1 i1}^2$	0.05	0.0622 (0.0034)	0.0126
	$\sigma_{\beta_1 i2}^2$	1	1.0615 (0.0512)	0.0801
Error	$\mu_{\sigma_{01i}^2}$	0.2	0.1911 (0.0156)	0.0179
	$\mu_{\sigma_{02i}^2}$	1	0.9604 (0.0371)	0.0543
	$\mu_{\rho_{0i}}$	-0.25	-0.2063 (0.016)	0.0465
	$\mu_{\sigma_{11i}^2}$	0.3	0.2868 (0.0197)	0.0237
	$\mu_{\sigma_{12i}^2}$	1.5	1.4363 (0.0452)	0.0781
	$\mu_{\rho_{1i}}$	0.25	0.2278 (0.0157)	0.0272
	$\sigma_{\sigma_{01i}^2}^2$	0.2	0.2005 (0.0471)	0.0471
	$\sigma_{\sigma_{02i}^2}^2$	1	1.0802 (0.1317)	0.1542
	$\sigma_{\rho_{0i}}^2$	0.1875	0.1772 (0.0071)	0.0126
	$\sigma_{\sigma_{11i}^2}^2$	0.3	0.3023 (0.0592)	0.0593
	$\sigma_{\sigma_{12i}^2}^2$	1.5	1.6527 (0.1794)	0.0081
	$\sigma_{\rho_{1i}}^2$	0.1875	0.1916 (0.007)	0.2356
Random	$\sigma_{\gamma 1}^2$	0.05	0.0487 (0.01)	0.0101
	$\sigma_{\gamma 2}^2$	0.5	0.4833 (0.0976)	0.1349
	ρ_{γ}	0.3162	0.3104 (0.1347)	0.0990

4.5 Hypothesis Testing

The goal of this research was to perform group comparisons of brain connectivities. This was achieved by conducting hypothesis tests using the estimated model parameters. To compare mean connectivity measures between two groups (either univariate or bivariate comparison), the testing was based on the regression coefficients (β_i). For comparing the coherence of the two connectivity measures between two groups, tests

were constructed using estimates of Pearson correlation coefficients (ρ_{0i} and ρ_{1i}). In the following we briefly describe the testing methods we used in this work.

To test the hypothesis $H_0 : \mathbf{H}\beta_i = h_i$, where $\mathbf{H}$ is a $q \times 4$ contrast matrix, we have, under H_0 , that

$$\mathbf{H}\hat{\beta}_i \sim N\left(h_i, \mathbf{H}[\mathbf{X}_i^T(\Sigma_i + \mathbf{Z}_i\Sigma_\gamma\mathbf{Z}_i^T)^{-1}\mathbf{X}_i]^{-1}\mathbf{H}^T\right) \quad (4.29)$$

And thus

$$C^2 = (\mathbf{H}\hat{\beta}_i - h_i)^T \left\{ \mathbf{H}[\mathbf{X}_i^T(\hat{\Sigma}_i + \mathbf{Z}_i\hat{\Sigma}_\gamma\mathbf{Z}_i^T)^{-1}\mathbf{X}_i]^{-1}\mathbf{H}^T \right\}^{-1} (\mathbf{H}\hat{\beta}_i - h_i) \sim \chi^2(q) \quad (4.30)$$

We reject H_0 when $C^2 > \chi_{1-\alpha}^2(q)$.

To test the hypothesis $H_0 : \rho_{0i} = \rho_{1i}$, where ρ_{0i} and ρ_{1i} are Pearson correlation coefficients between fMRI and DTI measures for the i th link for the control and experimental groups, respectively, we use Fisher's z-transformation. In brief, the estimates of the correlation coefficients, $\hat{\rho}_{0i}$ and $\hat{\rho}_{1i}$, are computed from $\hat{\Sigma}_{0i}$ and $\hat{\Sigma}_{1i}$, and then normalized using Fisher's z-transformation

$$\begin{aligned} z_{0i} &= \frac{1}{2} \log\left(\frac{1 + \hat{\rho}_{0i}}{1 - \hat{\rho}_{0i}}\right) \\ z_{1i} &= \frac{1}{2} \log\left(\frac{1 + \hat{\rho}_{1i}}{1 - \hat{\rho}_{1i}}\right) \end{aligned} \quad (4.31)$$

The normalized z_{0i} and z_{1i} are then compared using z-tests.

In another approach, variances were estimated for the correlation coefficient estimates using the covariance matrices of the error covariance components described in

section 4.3. Denote

$$\begin{aligned} C_{0i} &= Cov(\hat{\sigma}_{0i.11}, \hat{\sigma}_{0i.12}, \hat{\sigma}_{0i.22}), \\ C_{1i} &= Cov(\hat{\sigma}_{1i.11}, \hat{\sigma}_{1i.12}, \hat{\sigma}_{1i.22}). \end{aligned} \quad (4.32)$$

where $j, k = 1, 2, 3, 4$, $\hat{\sigma}_{0i.11}$, $\hat{\sigma}_{0i.12}$, and $\hat{\sigma}_{0i.22}$ are the three independent components of $\hat{\Sigma}_{0i}$, and $\hat{\sigma}_{1i.11}$, $\hat{\sigma}_{1i.12}$, and $\hat{\sigma}_{1i.22}$ are the three independent components of $\hat{\Sigma}_{1i}$. Since $\hat{\rho}_{0i} = \hat{\sigma}_{0i.12}/\sqrt{\hat{\sigma}_{0i.11}\hat{\sigma}_{0i.22}}$ and $\hat{\rho}_{1i} = \hat{\sigma}_{1i.12}/\sqrt{\hat{\sigma}_{1i.11}\hat{\sigma}_{1i.22}}$, then using the delta method, we have

$$\begin{aligned} \hat{Var}(\hat{\rho}_{0i}) &= n_0 \times \frac{\partial \hat{\rho}_{0i}}{\partial(\hat{\sigma}_{0i.11}, \hat{\sigma}_{0i.12}, \hat{\sigma}_{0i.22})} C_{0i} \frac{\partial \hat{\rho}_{0i}}{\partial(\hat{\sigma}_{0i.11}, \hat{\sigma}_{0i.12}, \hat{\sigma}_{0i.22})^T}, \\ &= n_0 \times \begin{pmatrix} -\frac{1}{2}\hat{\sigma}_{0i.12}\hat{\sigma}_{0i.11}^{-3/2}\hat{\sigma}_{0i.11}^{-1/2} \\ \hat{\sigma}_{0i.11}^{-1/2}\hat{\sigma}_{0i.22}^{-1/2} \\ -\frac{1}{2}\hat{\sigma}_{0i.12}\hat{\sigma}_{0i.11}^{-1/2}\hat{\sigma}_{0i.22}^{-3/2} \end{pmatrix}^T C_{0i} \begin{pmatrix} -\frac{1}{2}\hat{\sigma}_{0i.12}\hat{\sigma}_{0i.11}^{-3/2}\hat{\sigma}_{0i.11}^{-1/2} \\ \hat{\sigma}_{0i.11}^{-1/2}\hat{\sigma}_{0i.22}^{-1/2} \\ -\frac{1}{2}\hat{\sigma}_{0i.12}\hat{\sigma}_{0i.11}^{-1/2}\hat{\sigma}_{0i.22}^{-3/2} \end{pmatrix} \\ \hat{Var}(\hat{\rho}_{1i}) &= n_1 \times \frac{\partial \hat{\rho}_{1i}}{\partial(\hat{\sigma}_{1i.11}, \hat{\sigma}_{1i.12}, \hat{\sigma}_{1i.22})} C_{1i} \frac{\partial \hat{\rho}_{1i}}{\partial(\hat{\sigma}_{1i.11}, \hat{\sigma}_{1i.12}, \hat{\sigma}_{1i.22})^T}, \\ &= n_1 \times \begin{pmatrix} -\frac{1}{2}\hat{\sigma}_{1i.12}\hat{\sigma}_{1i.11}^{-3/2}\hat{\sigma}_{1i.11}^{-1/2} \\ \hat{\sigma}_{1i.11}^{-1/2}\hat{\sigma}_{1i.22}^{-1/2} \\ -\frac{1}{2}\hat{\sigma}_{1i.12}\hat{\sigma}_{1i.11}^{-1/2}\hat{\sigma}_{1i.22}^{-3/2} \end{pmatrix}^T C_{1i} \begin{pmatrix} -\frac{1}{2}\hat{\sigma}_{1i.12}\hat{\sigma}_{1i.11}^{-3/2}\hat{\sigma}_{1i.11}^{-1/2} \\ \hat{\sigma}_{1i.11}^{-1/2}\hat{\sigma}_{1i.22}^{-1/2} \\ -\frac{1}{2}\hat{\sigma}_{1i.12}\hat{\sigma}_{1i.11}^{-1/2}\hat{\sigma}_{1i.22}^{-3/2} \end{pmatrix} \end{aligned} \quad (4.33)$$

Hypothesis testings were performed by T-tests using the Pearson correlation coefficient estimates and the calculated variance estimates.

4.6 Multiple Comparisons

4.6.1 False Discovery Rate Assuming Independent Comparisons

Since multiple comparisons are involved in making inferences for all the links, the type I error rate needs to be controlled. The commonly used Bonferroni (and Tukey) correction divides the significance level by the number of comparisons and generally results in conservative conclusions. Especially in our case, where there are thousands of comparisons to be made, the Bonferroni approach is rather impractical. Therefore, to address the multiple comparison issue that is special in our case, we employed the false discovery rate (FDR) approach developed by Benjamini and Hochberg (1995). Benjamini and Hochberg (1995; 2000) show that “the FDR procedures have good power compared to the Bonferroni correction,” and that “the method is adaptive, in the sense that the thresholds that are chosen are automatically adjusted to the strength of the signal in the data.” The parameter q “has a definite and clear meaning that is comparable across studies.” Finally, “since the procedures work with p-values, and not test statistics, FDR methods can be applied with any valid statistical test, with equal ease (Benjamini and Hochberg, 1995, 2000).”

Denote the “number of hypotheses tested” by m , then the steps of FDR procedure are:

1. Select “a desired FDR bound q between 0 and 1 (the maximum proportion of false discoveries that the researcher is willing to tolerate, on average).” A q value of 0.15–0.20 is reasonable in many cases (Benjamini, personal communication).

2. Order the p-values from smallest to largest: $p_1 \leq p_2 \leq \dots \leq p_m$ and $H_i/$ is the hypothesis corresponding with p-value p_i .
3. Let r be the largest i for which $p_i \leq q \times i/m$
4. Reject the hypotheses $H_1, H_2, \dots, H_r$.

In this work, we evaluate multiple comparison results using q values ranging from 0.05–0.40.

4.6.2 False Discovery Proportion under Empirical Dependence

The original FDR method was developed assuming independence among tests (Benjamini and Hochberg, 1995). It has been shown that this approach still controls the false discovery rate at the nominal level when dealing with dependent tests (Benjamini and Yekutieli, 2001), but slightly loses control with increasing dependence (Benjamini and Hochberg, 2000). In our case, since the connectivity measures are nested within the brain of the same subject, the independence assumption is rather implausible. To estimate the proportion of false discoveries among rejected hypotheses, we follow an approach developed by Fan et al. (2012) based on principal factor approximation (PFA) and compute the false discovery proportion (FDP) under arbitrary covariance dependence. The idea is briefly as follows:

Let $V(\phi)$, $S(\phi)$, and $R(\phi)$ be the number of false, true, and total discoveries, respectively, depending on p test statistics $\mathbf{Z} = (Z_1, \dots, Z_p)^T$ and the rejection threshold ϕ . By definition, $FDP(\phi) = V(\phi)/R(\phi)$. Assume $\mathbf{Z} = (Z_1, \dots, Z_p)^T \sim N((\mu_1, \dots, \mu_p)^T, \mathbf{\Sigma})$, with $\lambda_1, \dots, \lambda_p$ and $\Gamma_1, \dots, \Gamma_p$ being the eigenvalues and the corresponding eigenvectors of

Σ . For the smallest integer $k < p$ such that $\frac{\sqrt{\lambda_{k+1}^2 + \dots + \lambda_p^2}}{\lambda_1 + \dots + \lambda_p} < \epsilon$, let $\mathbf{L} = (\sqrt{\lambda_1}\Gamma_1, \dots, \sqrt{\lambda_k}\Gamma_k)$, then Σ can be written as:

$$\Sigma = \mathbf{L}\mathbf{L}^T + \mathbf{A}. \quad (4.34)$$

And PFA decomposes dependent normal vectors into a factor model with weakly dependent normal errors:

$$Z_i = \mu_i + \beta_i^T \mathbf{W} + K_i, \quad i = 1, \dots, p, \quad (4.35)$$

where $\beta_i = (b_{i1}, \dots, b_{ip})^T$, $(b_{1j}, \dots, b_{pj})^T = \sqrt{\lambda_j}\Gamma_j$, $\mathbf{W} \sim N(\mathbf{0}, \mathbf{I}_k)$ are the factors, and $(K_1, \dots, K_p)^T \sim N(\mathbf{0}, \mathbf{A})$ are the random errors.

Fan et al. (2012) showed that the realized FDP can be estimated by

$$FDP_A(\phi) = \min \left(\sum_{i=1}^p [\Phi(a_i(z_{\phi/2} + \hat{\eta}_i)) + \Phi(a_i(z_{\phi/2} - \hat{\eta}_i))], R(\phi) \right) / R(\phi), \quad (4.36)$$

where $a_i = (1 - \sum_{j=1}^k b_{ij}^2)^{-1/2}$, $\eta_i = \beta_i^T \mathbf{W}$, and $\Phi(\cdot)$ and z are the cumulative distribution function and lower quantile of a standard normal distribution. Via simulation studies, these authors demonstrated that this FDP estimate performed better than the Efron approach (Efron, 2007), which is based on a parametric model for $V(\phi)$.

In this work, we estimate FDPs for the unimodal DTI and fMRI comparisons using empirical covariance matrices computed from the sample.

5. BAYESIAN MIXED-EFFECTS MODEL

5.1 Overview of Bayesian Methods

The past two decades have seen the fast development of Bayesian statistics. The major difference between the classical statistical theory (i.e., the frequentist approach) and the Bayesian approach is that the latter treats parameters as random variables with prior distributions. These prior distributions are combined with the data likelihood to obtain the posterior distribution of the parameters of interest. It is based on the posterior distributions of the parameters of interest that we make our inferences.

Bayesian statistics obtained its name from the Bayes' theorem, which states that, for the two outcomes, the parameter θ and the data Y ,

$$p(\theta|Y) = \frac{f(Y|\theta)g(\theta)}{h(Y)} \propto f(Y|\theta)g(\theta) \quad (5.37)$$

where $g(\theta)$ is the prior distribution of θ , $f(Y|\theta)$ is the data likelihood given the parameter. From this theorem, we see that the posterior distribution of the parameter, $p(\theta|Y)$, differs from $f(Y|\theta)g(\theta)$, which combines the likelihood and the prior, by only a constant. Therefore, after observing data Y , we can employ this rationale to combine the data information with prior and obtain the posterior distribution in order to make Bayesian inferences.

The Bayesian approach introduces the subjective viewpoint in the analysis via introducing the prior distribution. However, Bayesian methods failed to become widely used data analysis tools until recently due to the intractabilities involved in the computation of posterior distribution. Although asymptotic results were available for specific prob-

lems (especially those having conjugate prior and posterior distributions), a generalized approach was lacking for all problems until 1990.

The Gelfand group (Gelfand and Smith, 1990; Gelfand et al., 1990) implemented the Markov Chain Monte Carlo (MCMC) methods in the framework of Bayesian computation; as a result, Bayesian methods regained attention of the statistical society. Since then, Bayesian statistics became fashionable and a lot of related work appeared in the literature. During the early 1990s, studies focused on implementing MCMC methods in different models including hierarchical and mixed-effects models (Gelman and Rubin, 1992; Gelfand et al., 1992; Gilks and Wild, 1992; Dellaportas and Smith, 1993). The years since the late 1990s saw the development of algorithms that allowed model averaging, selection, and exploration (Green, 1995; Dellaportas and Forster, 1999; Dellaportas et al., 2002; Sisson, 2005; Hans et al., 2007).

These advances are an example of the rapid development of Bayesian statistics. However, Bayesian analysis did not gain wide popularity until the appearance of standard Bayesian analysis software based on MCMC. The first versions of Bayesian inference Using Gibbs Sampling (BUGS) software appeared in the late 1990s. In BUGS, the user only needs to specify the structure of the statistical model. The software will then generate samples from the posterior distribution of the specified model using one of the MCMC methods, the Gibbs sampler. Since its first appearance, the BUGS software has become a popular tool in Bayesian methods and demonstrated its value in the implementation of Bayesian models in various scientific fields. WinBUGS (Lunn et al., 2000) is a stand-alone software developed by the BUGS project team ([http:// www.mrc-bsu.cam.ac.uk/bugs/ welcome.shtml](http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml)). It features a graphical user interface and on-line monitoring and convergence diagnostics and can be called from other statistical software (such as from R using R2WinBUGS).

5.2 Markov Chain Monte Carlo

In this section we provide a brief introduction to the stochastic simulation methods for posterior sampling that are widely used in Bayesian analysis. The focus is on MCMC and, especially, the Gibbs sampler, which is the algorithm behind the computations in WinBUGS. Detailed descriptions and applications of the methods can be found in Ntzoufras (2009), Gilks et al. (1996), Givens and Hoeting (2005); and Gamerman and Lopes (2006).

Properly obtaining a sample from the posterior distribution has always been the main issue that bothers Bayesian statisticians. Unless in rare situations (usually when priors and posteriors are conjugate), it is impossible to obtain closed-form expressions of the posterior distributions. Numerical methods are often required to obtain posterior samples. The MCMC methods are a class of widely used methods that have been shown to be able to efficiently generate posterior samples.

The method of Monte Carlo is a computational technique that provides numerical solutions to problems using repeated random sampling. By running stochastic simulations many times, Monte Carlo methods compute probabilities heuristically as if they are actually happening. A good example will be to approximate the value of π : Using the idea of the Monte Carlo method, one can randomly draw uniform samples from the unit square and count the number of times samples are in the unit circle. That number divided by the sample size and multiplied by four provides a good approximation to the value of π .

Although traditional Monte Carlo methods are popular in many optimization and numerical integration problems where it is impossible to obtain a closed-form expression

or infeasible to apply a deterministic algorithm, they mostly deal with unidimensional distributions and cannot be used in all cases. Especially when the problem involves sampling from posterior distributions, traditional Monte Carlo methods show their limitations (Givens and Hoeting, 2005). In contrast, simulation methods based on Markov chains, namely MCMC, are much more flexible and thus can deal with many problems, such as posterior sampling, that are beyond the traditional Monte Carlo methods.

A Markov chain requires that the distribution of the random variable θ at time $t + 1$ given θ values at all previous times only depends on the θ value at time t , and this distribution is independent of time t . That is

$$f(\theta^{(t+1)}|\theta^{(1)}, \dots, \theta^{(t)}) = f(\theta^{(t+1)}|\theta^{(t)}) \quad \forall t \quad (5.38)$$

According to Gilks et al. (1996), “when the Markov chain is irreducible, aperiodic, and positive-recurrent, as $t \rightarrow \infty$ the distribution of $\theta^{(t)}$ converges to its equilibrium distribution, and this is independent of the initial values of the chain $\theta^{(1)}$.”

In 1953, Metropolis et al first introduced MCMC into physics (Metropolis et al., 1953). Later advances include Hastings “generalization of the Metropolis algorithm in 1970 (Hastings, 1970),” and the “development of the Gibbs sampler by Geman and Geman in 1984 (Geman and Geman, 1984).” Starting in the late 1980s, MCMC methods were implemented in Bayesian methods (Tanner and Wong, 1987; Gelfand and Smith, 1990; Gelfand et al., 1990) and soon became the major computation tools in Bayesian inference. The main idea of MCMC methods is to construct a Markov chain whose equilibrium distribution is the target distribution (i.e., the chain converges to the target distribution). This separates MCMC from other common simulation methods that

sample directly from the target distribution. Another distinction is that, in contrast to independent samples obtained from common simulations, samples from MCMC are dependent since they are from Markov chains. For the details of MCMC methods we refer to Gilks et al. (1996).

In the Bayesian analysis context, the target is the posterior distribution and we need to construct Markov chains so that they converge to the posterior distribution. In addition, it must be easy to generate samples from the conditional distribution $f(\theta^{(t+1)}|\theta^{(t)})$. With these requirements satisfied, we can follow these steps to generate samples from the posterior distribution and make inferences:

1. Set initial value $\theta^{(0)}$ for the Markov chain(s).
2. Generate N values (iterations) until the chain(s) reaches equilibrium.
3. Perform convergence diagnostics by examining the trace, density, auto-correlation, and other plots. If the diagnostics fail, generate more samples.
4. Discard the first M observations (burn-in) and use the rest $(N - M)$ values as posterior sample.
5. Obtain the summaries of the posterior sample, including mean, median, standard deviation, quantiles, correlations, and so on. Make Bayesian inferences using these posterior summaries.

5.3 Markov Chain Monte Carlo Algorithms

Since the introduction of the initial Metropolis algorithm in 1953 (Metropolis et al., 1953), many MCMC algorithms have been developed. The two most popular methods are

Metropolis-Hastings algorithm, which Hastings (Hastings, 1970) developed based on the original Metropolis algorithm, and the Gibbs sampler (Geman and Geman, 1984). Recent advances include the “slice Gibbs sampler” (Higdon, 1998; Damien et al., 1990; Neal, 2003), the “reversible jump MCMC” (Green, 1995), and “perfect sampling” (Propp and Wilson, 1996; Merller, 1999). These variants are all based on the original algorithms and most of them deal with specific problems. In this section we provide brief introductions to the two major MCMC methods, the Metropolis-Hastings algorithms and the Gibbs sampler.

5.3.1 The Metropolis-Hastings Algorithm

Metropolis first applied methods based on Markov chain simulations in physics (Metropolis et al., 1953). Later work by Hastings generalized the original algorithm and developed the Metropolis-Hastings algorithm (Hastings, 1970). This algorithm has been serving as the basis for all MCMC methods. The idea behind the algorithm is that one generates candidates from a proposal distribution and updates samples with a probability determined by the densities of the target and proposal distributions. The theory states that, regardless of the proposal distribution selected, the Metropolis-Hastings algorithm will converge to its equilibrium distribution. In practice, though, the choice of the proposal distribution is important since poorly chosen proposals will significantly slow down the algorithm from converging.

In the Bayesian framework, the posterior $f(\boldsymbol{\theta}|y)$ is the target distribution. With a proposal distribution $q(\boldsymbol{\theta}'|\boldsymbol{\theta})$, the steps of the Metropolis-Hastings algorithm for Bayesian inference may be written as:

1. Take starting values $\boldsymbol{\theta}^{(0)}$.
2. For t from 1 to T iterate:
 - (a) Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$.
 - (b) Generate new candidate values for $\boldsymbol{\theta}'$ from $q(\boldsymbol{\theta}'|\boldsymbol{\theta})$.
 - (c) Calculate probability $\alpha = \min\left(1, \frac{f(\boldsymbol{\theta}'|y)q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{f(\boldsymbol{\theta}|y)q(\boldsymbol{\theta}'|\boldsymbol{\theta})}\right)$.
 - (d) Update $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}'$ with probability α .

Since the normalizing constant $f(y)$ in $f(\boldsymbol{\theta}|y)$ cancels out, the probability α can be rewritten as

$$\alpha = \min\left(1, \frac{f(y|\boldsymbol{\theta}')f(\boldsymbol{\theta}')q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{f(y|\boldsymbol{\theta})f(\boldsymbol{\theta})q(\boldsymbol{\theta}'|\boldsymbol{\theta})}\right) \quad (5.39)$$

There are some special cases of the Metropolis-Hastings algorithm. Random-walk Metropolis uses a special proposal $q(\boldsymbol{\theta}'|\boldsymbol{\theta}) = q(|\boldsymbol{\theta}' - \boldsymbol{\theta}|)$ instead of the symmetric proposal in the original algorithm, resulting in an acceptance probability $\alpha = \min\left(1, \frac{f(\boldsymbol{\theta}'|y)}{f(\boldsymbol{\theta}|y)}\right)$, which depends only on the posterior distribution. In the independence sampler, the proposal distribution does not depend on the previous state of the chain. This algorithm is efficient when the proposal is a good approximation to the posterior distribution. And “in contrast to the random-walk Metropolis, where the optimal acceptance rate is around 0.25, the acceptance rate for an independent sample must be high enough for the algorithm to be efficient (Ntzoufras, 2009, 45).” Component-wise Metropolis-Hastings, which is also called Metropolis within Gibbs, is an algorithm where the parameter vector is divided into subvectors that are updated sequentially using original Metropolis-Hastings procedures. The advantage of this algorithm is that it involves sampling from distri-

butions of lower dimensions that are usually straightforward and computationally less intensive.

5.3.2 The Gibbs Sampler

In 1984 Geman and Geman developed the Gibbs sampler (Geman and Geman, 1984), which uses the full conditional distribution as the proposal. Because of its popularity, Gibbs sampler is usually cited as a separated MCMC technique, even though it is only a special case of the single-component Metropolis-Hastings algorithm. The full conditional distribution $f(\theta_j | \boldsymbol{\theta}_{\setminus j}, \mathbf{y})$ is the distribution of the j th component of $\boldsymbol{\theta}$ given current values of all other parameters and data. Such a proposal distribution results in acceptance probability of $\alpha = 1$ —i.e., the chain moves accepting all iterations. Since at each step random values are generated from univariate distributions, and frequently these distributions have a known and simple form, the computation is straightforward and one can select methods from a wide variety of tools. The algorithm does become ineffective when the parameters are highly correlated or the parameter space is complicated. The algorithm can be summarized as follows:

1. Set initial values $\boldsymbol{\theta}^{(0)}$.
2. For $t = 1, \dots, T$ repeat the following steps
 - (a) Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$.
 - (b) For $j = 1, \dots, d$ update θ_j from $\theta_j \sim f(\theta_j | \boldsymbol{\theta}_{\setminus j}, \mathbf{y})$.
 - (c) Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}$ where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_j, \dots, \theta_d)$.

5.4 Model Definition

The main structure of our Bayesian mixed-effects model is the same as our classical mixed-effects model (CMEM) described in section 4.2.

$$Y_{ijk} = \beta_{0ik} \times (1 - Group_j) + \beta_{1ik} \times Group_j + \gamma_{jk} + \epsilon_{ijk} \quad (5.40)$$

where $i = 1, 2, \dots, m$ denotes the i th link, $j = 1, 2, \dots, n$ denotes the j th subject, $k = 1$ if the measure is fMRI, $k = 2$ if the measure is DTI, $Group_j = 0$ if the j th subject is from the control group, and $Group_j = 1$ if the j th subject is from the experimental group. We assume $\boldsymbol{\gamma}_j = (\gamma_{j1}, \gamma_{j2})^T \sim N(\mathbf{0}, \Sigma_\gamma)$, $\boldsymbol{\epsilon}_{ij} = (\epsilon_{ij1}, \epsilon_{ij2})^T \sim N(\mathbf{0}, \Sigma_{0i})$ if the j th subject is from the control group, and $\boldsymbol{\epsilon}_{ij} \sim N(\mathbf{0}, \Sigma_{1i})$ if the j th subject is from the experimental group.

For our Bayesian model, since we do not have any prior knowledge about the model parameters, we assign flat, noninformative priors to these parameters. Informative priors can always be specified for certain or all parameters if we have a priori knowledge about them. For the fixed parameters, we assume a multivariate normal distribution with zero mean and huge variance (e.g., 10000):

$$\boldsymbol{\beta}_i = \begin{bmatrix} \beta_{0i1} \\ \beta_{0i2} \\ \beta_{1i1} \\ \beta_{1i2} \end{bmatrix} \sim N \left(\mathbf{0}, \begin{bmatrix} 10000 & 0 & 0 & 0 \\ 0 & 10000 & 0 & 0 \\ 0 & 0 & 10000 & 0 \\ 0 & 0 & 0 & 10000 \end{bmatrix} \right) \quad (5.41)$$

And for variance components, we assume the variance parameters follow gamma distributions with unit mean and high variance (e.g., 100), and the correlation coefficients follow uniform distributions on $(-1, 1)$:

$$\begin{aligned}
\Sigma_{0i} &= \begin{bmatrix} \sigma_{01i}^2 & \rho_{0i}\sigma_{01i}\sigma_{02i} \\ \rho_{0i}\sigma_{01i}\sigma_{02i} & \sigma_{02i}^2 \end{bmatrix} \\
&\sigma_{01i}^2, \sigma_{02i}^2 \sim \Gamma(0.01, 0.01) \\
&\rho_{0i} \sim U(-1, 1) \\
\Sigma_{1i} &= \begin{bmatrix} \sigma_{11i}^2 & \rho_{1i}\sigma_{11i}\sigma_{12i} \\ \rho_{1i}\sigma_{11i}\sigma_{12i} & \sigma_{12i}^2 \end{bmatrix} \\
&\sigma_{11i}^2, \sigma_{12i}^2 \sim \Gamma(0.01, 0.01) \\
&\rho_{1i} \sim U(-1, 1) \\
\Sigma_{\gamma} &= \begin{bmatrix} \sigma_{1\gamma}^2 & \rho_{\gamma}\sigma_{1\gamma}\sigma_{2\gamma} \\ \rho_{\gamma}\sigma_{1\gamma}\sigma_{2\gamma} & \sigma_{2\gamma}^2 \end{bmatrix} \\
&\sigma_{1\gamma}^2, \sigma_{2\gamma}^2 \sim \Gamma(0.01, 0.01) \\
&\rho_{\gamma} \sim U(-1, 1)
\end{aligned} \tag{5.42}$$

Or, alternatively, we assume the covariance matrices of the errors and random effects follow inverse Wishart distributions:

$$\begin{aligned}
\Sigma_{0i} &\sim IW(\mathbf{V}, n_0) \\
\Sigma_{1i} &\sim IW(\mathbf{V}, n_1) \\
\Sigma_{\gamma} &\sim IW(\mathbf{V}, n_0 + n_1),
\end{aligned} \tag{5.43}$$

where $\mathbf{V}$ is a 2×2 scale matrix. Without specific prior knowledge about the covariance matrices, a simple diagonal matrix can be used for the value of $\mathbf{V}$.

5.5 Model Update and Convergence Diagnostics

We update our Bayesian models by running two Markov chains. After the model is compiled with data, initial values are set for all the model parameters in the two chains. The initial values of the fixed-effect parameters (or β coefficients) are set to be either zeros or random numbers generated from a standard normal distribution. The variance parameters are set to 1 or 0.01 at the starting point and the correlations are set to 0. For the covariance matrices, diagonal matrix with 1 or 0.5 on the diagonals are used as their starting values.

The model is then updated for 2,000 iterations for burn-in, and for another 3,000 iterations to obtain a posterior sample of 6,000. The convergence of the Markov chains is diagnosed by checking the trace, density, autocorrelation, and Brooks-Gelman-Rubin (BGR, (Gelman and Rubin, 1992; Brooks and Gelman, 1997)) plots for each of the model parameters. Upon convergence, the Markov chains are expected to reach equilibrium on the trace plots and the density plots should show nice bell-shaped curves. In addition, no autocorrelations should be seen for any lags greater than 0. The BGR plot provides another way to evaluate convergence of chains by monitoring the ratio of within- and among-chain variabilities. The blue and green curve on a BGR plot represent the widths of the central 80% intervals of the within- and among-chain variance, and the red curve is the ratio of the two. Upon convergence, the within- and among-chain variance are expected to be about the same and the ratio should be close to 1.

After convergence diagnostics, summary statistics of all parameters of interest are obtained. The 95% credible intervals are used to make inferences about the parameters of interest. To perform the FDR approach on Bayesian model results, P-values are obtained using normal distributions constructed from the means and standard deviations of parameters of interest.

6. MACHINE LEARNING AND CLASSIFICATION

Machine learning and classification is a data mining method concerned with the construction of study systems that can learn from data about the characteristics of different (usually two) groups. After learning, the machine can be used to classify new subjects into any of these groups according to the features of the new subject. Recent years have seen an explosion of interest in the application of machine learning and classification to the analysis of neuroimaging data. Machine learning analysis of MRI data has demonstrated high diagnostic accuracy in some psychiatric disorders. For example, Klöppel et al. (2008) used linear support vector machines to classify MRI scans and successfully separated patients with Alzheimer’s disease from healthy aging subjects (Klöppel et al., 2008). In another occasion, Davatzikos et al. (2005) applied a high-dimensional nonlinear pattern classification technique on MRI data to quantify the degree of separation and showed high sensitivity and specificity of the classifier for the diagnosis of schizophrenia (Davatzikos et al., 2005). However, until today, very few studies have applied machine learning techniques to neuroimaging data sets for automated diagnostics and prognostics of mood disorders, such as depression or bipolar disease. In this section we briefly introduce several techniques that we use in the classification of LLD subjects using neuroimaging data. For further information about machine learning techniques, we refer to an excellent text by Hastie et al. (2003).

6.1 Supervised Learning

The idea of machine learning is that one can predict the output y based on a vector of input variables $\mathbf{X}$. It is usually assumed that the output and input variables follow some functional relationship $y = f(\mathbf{X})$. In supervised learning, this function is fully

specified with both $\mathbf{X}$ and y known. The output here can be a vector but for simplicity a scalar is considered. Depending on the nature of the output variable, machine learning can solve two types of problems: classification and regression. Classification is conducted when the decision is among a small and discrete set of choices (e.g., determining LLD versus HC subjects). On the other hand, when the output variable is rather continuous, then regression is performed to estimate the output value using the input values (e.g., assessing severity of depression).

A supervised learning task generally consists of the following steps: (1) Determine a training set that contains a set of input variables and the corresponding outputs; (2) Determine features that are characteristic of the subject to be used in the learned function; (3) Determine the structure of the learned function and the corresponding algorithm; (4) Determine parameter values in the learned function by running the algorithm on the training set; (5) Assess the performance of the learned function on a test set.

There are some issues to consider when performing supervised learning. The first issue is the trade-off between bias and variance (Geman et al., 1992). Bias is the systematical error a learning algorithm produces on different training sets, whereas variance measures the variation of the output the algorithm produces for different training sets. A good learning algorithm needs to be flexible to have low bias. But on the other hand, it should not be too flexible and produce high variance. Many algorithms have a bias/-variance parameter so that users can adjust the trade-off between the two entities.

Another thing to consider is the dimension of the input space. On occasions when the input feature vectors have high dimension and the learned function only depends on a small number of the features, the unnecessary features can easily confuse the learning algorithm and cause it to have high variance. Feature (variable) selection or other

dimension reduction mechanisms can potentially remove irrelevant features and reduce dimensionality of the input vector before running the machine learning algorithm.

Finally, the performance of a learning algorithm also depends on the complexity of the learned function and the amount of training data available. If the learned function is simple, then a relatively small amount of training data is needed for the algorithm to learn the problem well. However, if the function is rather complex, especially when a lot of interactions among input features are involved, then a fairly large amount of training data may be needed for the algorithm to learn. In this case, a flexible learning algorithm may be advantageous in reducing bias and variance. On occasions when only a small amount of training data is available, one may have to choose simple functions so that the results are more reliable.

There are a variety of supervised machine learning algorithms currently available. Each algorithm has its strengths and weaknesses and generally there is no single algorithm that works best in all situations. Of all available algorithms, support vector machine (SVM) and random forests are among the most popular methods due to their high accuracy and versatility.

6.2 Support Vector Machine

The problem of pattern classification originated in 1936 when R. A. Fisher suggested the first algorithm for pattern recognition. These traditional methods usually involve model determining criteria like least squares and maximum likelihood, which can be adversely influenced by data points that are distant from the boundary. In the early 1970s, Vladimir Vapnik first introduced statistical learning theory (or Vapnik-Chervonenkis theory) (Vapnik and Chervonenkis, 1974) and introduced the first version of SVM as the

group further developed statistical learning theory. The current soft margin version of SVM was introduced by Cortes and Vapnik (1995). In this section we briefly introduce the theory and application behind SVM algorithm. For further details, we refer readers to the book by Scholkopf and Smola (2001).

In SVM, the discriminant boundary (i.e., hyperplane, see Figure 2) is defined only in terms of data points that lie close to the boundary. These data points are called support vectors and by allowing the decision function fully specified by these support vectors, SVM resolves the issue of adverse influence by data points that are far from the hyperplane. In our case, SVM defines a hyperplane that separates LLD subjects from HC subjects with maximum margin. In the linear SVM case, the margin is defined by the distance of the hyperplane to the nearest LLD and HC subjects and SVM maximizes both distances of the two groups from the hyperplane. The formula of a linear SVM can be written as:

$$f(x) = \sum_{j=1}^M w_j x_j + b, \quad (6.44)$$

where $f(x)$ is the discriminant function, $x = (x_1, \dots, x_M)'$ is the support vector, $w = (w_1, \dots, w_M)'$ is the weight vector, and b is the bias. In our context, for each input vector x , $f(x)$ assigns a score that is used to classify the corresponding subject. We assume that the label will be either 1 (LLD), or -1 (HC). For $M = 2$, $\sum_{j=1}^M w_j x_j = 0$ represents a line that passes through the origin. For a general M , $\sum_{j=1}^M w_j x_j$ represents a hyperplane. Thus $f(x)$ represents a hyperplane with origin shifted to b .

The hyperplane divides the entire space into two subspaces (or regions) depending on the sign of $f(x)$. The boundary of the regions is called the decision boundary. When a hyperplane classifies data points correctly, it is most likely that many such hyperplanes exist in such situations. The question of choosing the most suitable hyperplane arises

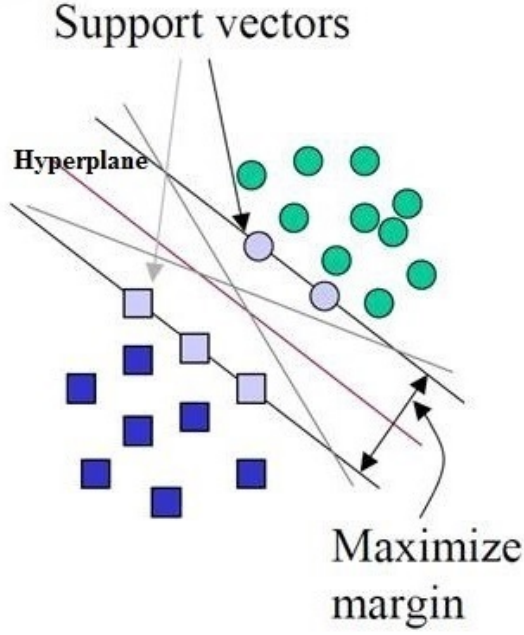


Figure 2. Support vector machine.

in these situations not only for the data in hand but also for correctly classifying future observations. This problem is resolved by the maximum margin. In the linear SVM case, the margin is $m = \frac{1}{\|w\|}$, where $\|w\| = \sqrt{\sum_{j=1}^M w_j^2}$. By a hard margin SVM for linearly separable data, we mean a classifier with maximum margin among all classifiers that correctly classifies all input data. To estimate w and b for the maximum margin hyperplane, we optimize the following:

$$\begin{aligned}
 & \underset{w, b}{\text{minimize}} && \frac{\|w\|^2}{2} \\
 & \text{subject to} && y_i \left(\sum_{j=1}^M w_j x_j + b \right) \geq 1, \text{ for } i=1, \dots, n,
 \end{aligned} \tag{6.45}$$

where y_i is the label that takes a value 1 (LLD), or -1 (HC), and y_i is associated with x_i . In (6.45), the minimization is equivalent to maximum margin, and constraints provide an assurance that all inputs are correctly classified. An optimal solution of (w^*, b^*) from the quadratic of optimization problem in (6.45) would satisfy $y_i(\sum_{j=1}^M w_j^* x_j + b^*) \geq 1$, for each i , and the length of w^* would be as small as possible.

In practice, when all data points cannot be separated linearly, the classifier is allowed with some misclassification, and it is achieved by the idea of the soft margin SVM. In the soft margin SVM, the constraint in (6.45) is changed to $y_i(\sum_{j=1}^M w_j x_j + b) \geq 1 - e_i$, where $e_i \geq 0$ are called slack variables. This procedure makes the margin larger, and the slack variables allow some inputs to fall in the margin or to be misclassified. In order to reduce a large number of inputs to be misclassified, the term $k \sum_{i=1}^n e_i$ is added to the constraint. Hence the new optimization problem is

$$\begin{aligned} & \underset{w, b}{\text{minimize}} \quad \frac{\|w\|^2}{2} + k \sum_{i=1}^n e_i \\ & \text{subject to} \quad y_i \left(\sum_{j=1}^M w_j x_j + b \right) \geq 1 - e_i, \text{ for } i=1, \dots, n, \end{aligned} \quad (6.46)$$

The constant k plays an important role in maximizing the margin and minimizing the amount of slack. Generally, to determine the SVM, the optimization is performed by Lagrangian multipliers using quadratic programmings. The problem can be restated (in terms of β'_i s) as follows:

$$\begin{aligned} & \underset{\beta}{\text{maximize}} \quad \sum_{i=1}^n \beta_i - \frac{\sum_{i=1}^n \sum_{j=1}^n y_i y_j \beta_i \beta_j a_{ij}}{2} \\ & \text{subject to} \quad \left(\sum_{i=1}^n y_i \beta_i \right) = 0, \quad 0 \leq \beta_i \leq k. \end{aligned} \quad (6.47)$$

The weight vector $\mathbf{w}$ in (6.45) can be expressed in terms of inputs (i.e., $\mathbf{x}_i$'s) and the solutions β_i 's. Hence $\mathbf{w} = \sum_{i=1}^n \mathbf{y}_i \beta_i \mathbf{x}_i$. $\mathbf{x}_i$'s with positive weights (i.e., $\beta_i > 0$) are called support vectors, and they lie either on or within the margin.

When a linear classifier fails to provide desired results, nonlinear classifiers are used for better accuracy. However, the mechanism for nonlinear classifiers are often developed by extending the concept of linear classifiers. The idea is to transform the inputs to some other vector space using a function ψ , and then construct the discriminant function $f(x) = \sum_{j=1}^M w_j \psi(x_j) + b$. The discriminant function $f(x)$ is nonlinear to the original space, however it is linear to the transformed space when ψ is nonlinear. The fundamental problem with this approach is with the number of input variables. For high-dimensional data this approach becomes slow and in turn becomes unacceptable. In such cases, SVMs are generalized to nonlinear classifiers using different kernel functions (e.g., Gaussians, polynomials, and so on). The polynomial kernel of degree d is defined as $K_{\kappa,d} = (\sum_{i=1}^M x_i x'_i + \kappa)^d$, where κ is used to denote the deviation from homogeneity. For $d = 1$ and $\kappa = 0$, the kernel becomes a linear discriminant function. The degree of a polynomial kernel provides flexibility to a classifier to classify inputs with a nonlinear relationship. The Gaussian kernel is defined as $K(x, x', \sigma^2) = \exp(-\frac{\|x-x'\|^2}{\sigma^2})$, σ^2 is known as the variance that controls the width of the Gaussian, and in a way it controls the flexibility of the classifier. For large values of σ^2 , almost all support vectors are involved in the discriminant function at a given point x , and as a result the decision boundary becomes smooth and almost linear. As σ^2 decreases, the curvature of the decision boundary becomes prominent. For very small values of σ^2 , the discriminant function becomes non-zero only in a small neighborhood of each support vector. Thus in this scenario, the discriminant function becomes nonsignificant for the bulk of the data points. In the kernel approach also, the Lagrange multipliers β_j 's are estimated by quadratic program.

6.3 Application to Analysis of Neuroimaging Data

Recently SVM has been used in several studies in dealing with classification problems in mental health research using neuroimaging data. Hosseinifard et al. (2011) used SVM on EEG data and showed that, with genetic algorithm for feature selection, SVM can achieve accuracy of 88.6% on classifying LLD patients. Mourao-Miranda et al. (2012) used the Gaussian process classifier on analyzing fMRI data and were able to discriminate healthy low-risk from healthy adolescents genetically at risk for Axis I disorders, including bipolar, depression, and anxiety disorders. Joshi et al. (2013) compared various neural-net classifiers with SVM and evaluated them on real-world clinical data from interactive interviews with LLD and HC subjects. To date, very few studies have investigated the potential use of DTI in conjunction with SVM. O'Dwyer et al. (2012) used DTI measures together with SVM to classify mild cognitive impairment versus control subjects and achieved greater than 90% sensitivity and specificity, "demonstrating the potential of combining DTI and SVM for fast and objective classification of healthy and mild cognitive impairment subjects." This technique "may be useful for large-scale drug trials in Alzheimer's disease where early identification of subjects with mild cognitive impairment is critical (O'Dwyer et al., 2012, 32441)." Adluru et al. (2009) evaluated the utility of shape representations of white-matter tracts extracted from DTI data for classification of autistic and control subjects and showed that SVM "can be built with reasonable prediction accuracies without making heavy use of the spatial normalization machinery needed when using voxel based features."

For analyzing our data set we use SVM with a linear kernel on both DTI and fMRI data. We divide our data set into a training set and a testing data set. We train SVM with the training data set and use the testing set to assess the trained machine. We perform

10-fold cross-validations and calculate sensitivity, specificity, and overall accuracy for each validation.

7. DATA ANALYSIS

In this section we illustrate our statistical methods, including the CMEMs and BMEMs, machine learning and classification methods, and community structure estimates, using a real neuroimaging data set for a population study. The data set contains DTI and fMRI measures of 87 brain regions ($87 \times 86/2 = 3471$ links or connectivities) for 10 LLD and 13 HC subjects. The objective of these analyses was to detect brain regions and/or connectivities that are impaired during depression, and find out whether the identified regions and connectivities can inform us about the development of this disorder. The original research hypothesized that, compared to HC subjects, LLD subjects would demonstrate (1) lower structural connectivity, measured by FA, especially in specific tracts such as cingulate bundle and uncinated fasciculus, (2) increased functional activity in the default mode network and affective network and decreased activity in the cognitive control network, and (3) decoupled structure-function connectivity relationships. However, in this work, the focus was not on proving any of these hypotheses. Rather, statistical methods described in the previous sections were illustrated as we were analyzing this neuroimaging data set, and inferences made using different methods were compared for consistency among results.

7.1 Mixed-Effects Models

We fitted CMEMs and BMEMs to the neuroimaging data. To process the data for this analysis, we applied some data manipulation and transformation to the two connectivity measures in the data set. First, the DTI measures represent fiber counts and there are a number of zero values in the original DTI data, suggesting no structural connectivity between the corresponding brain regions. Since we were estimating the variation within DTI measures as well as the correlation between DTI and fMRI measures in our bivariate mixed-effects models, we have to remove the excess zero values in our data set. To this end, we ignored connectivities with less than two non-zero values of DTI in either the LLD or HC group. After doing this, the total number of connectivities (or links) included in the analysis reduced to 1,433.

Next, since the DTI measures are essentially estimates of fiber counts, we employed the cube-root transformation, which is a well-established transformation method for Poisson count data, to normalize the DTI data. As shown in Figure 3, for the connectivity between left cerebellum cortex and left thalamus proper, although sample sizes are limited in both the LLD and HC groups, we observed that the distribution of the cube-root transformed DTI measures roughly follow normal distributions.

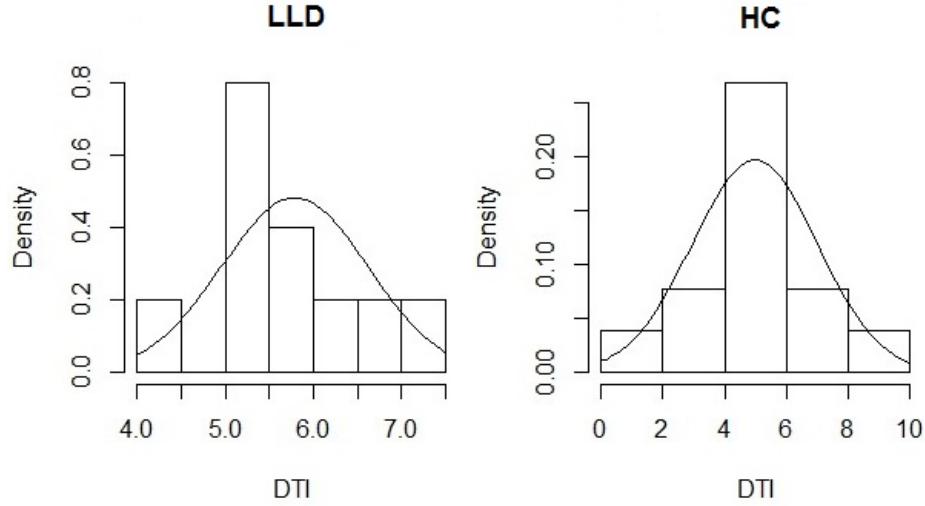


Figure 3. DTI between left cerebellum cortex and left thalamus proper.

Finally, since the fMRI measures are essentially normalized correlation values, we realized that both highly positive and highly negative values indicate strong correlations. We wanted to match both of these strong correlations to high fiber counts. Therefore, we took the absolute values of the fMRI measures when we analyzed the correlations between DTI and fMRI data.

7.1.1 Independent Link-Specific Analyses

A simple and widely used way to analyze these types of data is to perform independent, link-specific analyses to compare the LLD and HC groups for each of the connectivities. While doing this, connectivity measures for different links are assumed to be independent of each other. For each connectivity, we conducted two-sample t-tests for the univariate DTI and fMRI analyses, two-sample Hotelling's T-square tests for the bivariate analysis, and Fisher's z-transformation on Pearson correlation coefficients for

the comparison of DTI-fMRI correlations, between the two groups. The results of these analyses are summarized in Table IV.

TABLE IV. INDEPENDENT T-TESTS RESULT

	Univariate		Bivariate	Correlation
	DTI	fMRI	DTI & fMRI	DTI-fMRI
$P - value < .05$	49	59	55	72
$FDR = 0.30$	0	0	0	0

We observed 49, 59, 55, and 72 links that exhibited significant ($P\text{-value} < .05$) between-group differences in DTI, fMRI, bivariate DTI and fMRI, and DTI-fMRI correlation, respectively. However, after considering the multiplicity problem and controlling the type I error using an FDR rate of 0.30, no significance was detected for each of the four analyses (Table IV, row 4).

7.1.2 Classical Mixed-Effects Models

The classical mixed-effects models (CMEM) that we fitted to our neuroimaging data have been described in chapter 4. Compared to the independent approaches described in the previous section, our mixed-effects models incorporated both the between-modality and within-subject correlations. We obtained estimates of the mean DTI and fMRI measures of each connectivity for each of the two groups using the EM-like parameter estimation approach described in section 4.3. We defined the convergence criteria such that all 14,379 model parameters, including $1,433 \times 4 = 5,732$ fixed effects (regression coefficients), $23 \times 2 = 46$ random subject effects, $1,433 \times 6 = 8,598$ error covariance

components, and three random effect covariance components, are stabilized according to an absolute tolerance level of 10^{-5} . The algorithm converged after five iterations.

Upon convergence of the estimation algorithm, we obtained parameter estimates as well as the corresponding estimates for the variances of the parameters. We summarized the mean DTI and fMRI estimates across all links and compared between the LLD and HC groups. We observed that both DTI and fMRI measures were at the same, if not higher, level in the LLD group compared to the HC group (see Figure 4 and Table V).

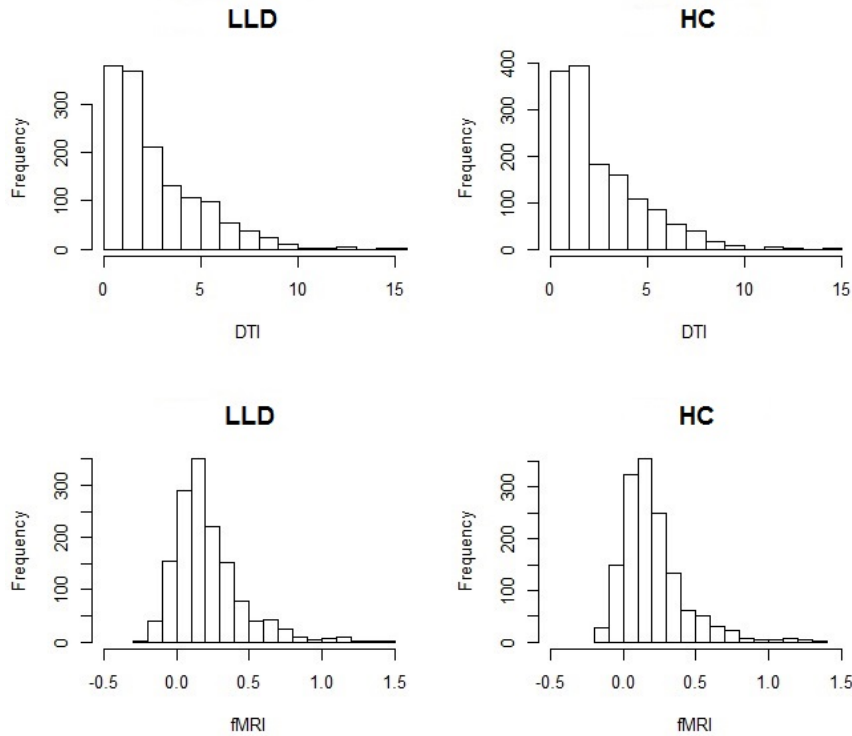


Figure 4. Histogram of mean DTI and fMRI across all links.

TABLE V. SUMMARY OF LINK-SPECIFIC MEAN DTI AND fMRI

	LLD	HC
	Mean (SD)	Mean (SD)
DTI	2.700 (2.268)	2.591 (2.175)
fMRI	0.226 (0.216)	0.217 (0.208)

To compare mean DTI and fMRI measures between LLD and HC groups for each connectivity, we performed t-tests using the link-specific regression coefficients as described in section 4.5. The type I error rate was then controlled using FDR rates between 0.20–0.30. In this univariate analysis, we detected 31 and 28 connectivities that showed significant between-group differences in DTI and fMRI, respectively (see Table VI and VII).

TABLE VI. SIGNIFICANT CONNECTIVITIES FOR DTI—FDR=0.30

Num	Region 1	Region 2	P-value
1	L_rostralanteriorcingulate	R_medialorbitofrontal	<0.0001
2	L_Cerebellum_Cortex	L_Pallidum	<0.0001
3	L_Thalamus_Proper	L_superiortemporal	<0.0001
4	L parahippocampal	L_precuneus	<0.0001
5	L_Pallidum	Brain_Stem	<0.0001
6	R_Putamen	R_precuneus	0.0001
7	L_medialorbitofrontal	L_frontalpole	0.0001
8	Brain_Stem	R_temporalpole	0.0001
9	R_paracentral	R_superiorparietal	0.0002
10	R_medialorbitofrontal	R_rostralmiddlefrontal	0.0005
11	L_superiorparietal	R_insula	0.0006
12	L_Thalamus_Proper	L_insula	0.0006
13	R_bankssts	R_insula	0.0011
14	L_inferiorparietal	L_insula	0.0011
15	L_VentralDC	L_inferiorparietal	0.0013
16	L_lateraloccipital	R_insula	0.0013
17	L_caudalanteriorcingulate	R_posteriorcingulate	0.0015
18	R_Cerebellum_Cortex	R_Putamen	0.0017
19	Brain_Stem	R_Putamen	0.0018
20	R_Amygdala	R_VentralDC	0.0021
21	L_insula	R_insula	0.0022
22	L_lateraloccipital	L_superiorparietal	0.0024
23	Brain_Stem	L_insula	0.0025
24	Brain_Stem	L_superiortemporal	0.0026
25	Brain_Stem	R_Amygdala	0.0028
26	L_Caudate	R_Accumbens_area	0.0036
27	L parahippocampal	L_superiorparietal	0.0036
28	L_isthmuscingulate	L parahippocampal	0.0038
29	R_Amygdala	R_fusiform	0.0039
30	R_caudalmiddlefrontal	R_precentral	0.0041
31	L_Pallidum	L_inferiorparietal	0.0061

TABLE VII. SIGNIFICANT CONNECTIVITIES FOR FMRI—FDR=0.30

Num	Region 1	Region 2	P-value
1	L_VentralDC	R_caudalanteriorcingulate	0.0001
2	L_Thalamus_Proper	L_posteriorcingulate	0.0002
3	L_superiorparietal	L_frontalpole	0.0006
4	L_fusiform	L_middletemporal	0.0008
5	R_caudalmiddlefrontal	R_superiorfrontal	0.001
6	R_Thalamus_Proper	R_Caudate	0.0011
7	L_VentralDC	L_caudalanteriorcingulate	0.0014
8	R_isthmuscingulate	R_superiorfrontal	0.0015
9	R_rostralmiddlefrontal	R_insula	0.0016
10	L_Pallidum	L_inferiorparietal	0.0023
11	L_Putamen	R_parstriangularis	0.0026
12	L_Thalamus_Proper	R_Pallidum	0.0029
13	R_caudalmiddlefrontal	R_postcentral	0.0034
14	L_frontalpole	R_precuneus	0.0036
15	L_Thalamus_Proper	R_posteriorcingulate	0.0038
16	L_Thalamus_Proper	R_caudalanteriorcingulate	0.0038
17	R_Hippocampus	R_entorhinal	0.0041
18	L_Putamen	L_parsopercularis	0.0043
19	R_inferiortemporal	R_superiortemporal	0.0044
20	L_Cerebellum_Cortex	L_inferiortemporal	0.0046
21	R_caudalanteriorcingulate	R_rostralmiddlefrontal	0.0047
22	L_Thalamus_Proper	R_Amygdala	0.0047
23	R parahippocampal	R_precuneus	0.005
24	R_posteriorcingulate	R_superiorparietal	0.0056
25	L_fusiform	L_precuneus	0.0056
26	R_precentral	R_rostralmiddlefrontal	0.0056
27	R_Pallidum	R_inferiorparietal	0.0057
28	L_Putamen	L_lateralorbitofrontal	0.0057

Next, to determine links that exhibit significant between-group differences in DTI and fMRI measures jointly, we performed two-sample Hotelling’s T-square tests using the mean DTI and fMRI estimates for the two groups. Using an FDR value of 0.20, we detected 40 connectivities that showed significant between-group differences in both DTI and fMRI (see Table VIII). Ninety significant connectivities were detected with an FDR value of 0.30.

TABLE VIII. SIGNIFICANT CONNECTIVITIES FOR BIVARIATE—FDR=0.20

Num	Region 1	Region 2	P-value
1	L_rostralanteriorcingulate	R_medialorbitofrontal	<0.0001
2	L_Cerebellum_Cortex	L_Pallidum	0.0001
3	L_Cerebellum_Cortex	L_inferiortemporal	0.0001
4	L_parahippocampal	L_precuneus	0.0001
5	L_VentralDC	L_inferiorparietal	0.0001
6	Brain_Stem	R_temporalpole	0.0001
7	L_Pallidum	Brain_Stem	0.0001
8	L_superiorparietal	R_insula	0.0001
9	L_Thalamus_Proper	L_superiortemporal	0.0002
10	L_Thalamus_Proper	L_insula	0.0002
11	L_VentralDC	R_caudalanteriorcingulate	0.0003
12	L_lateraloccipital	R_insula	0.0003
13	R_Putamen	R_precuneus	0.0004
14	L_medialorbitofrontal	L_frontalpole	0.0004
15	R_medialorbitofrontal	R_rostralmiddlefrontal	0.0005
16	L_Pallidum	L_inferiorparietal	0.0007
17	L_Thalamus_Proper	L_posteriorcingulate	0.0009
18	R_paracentral	R_superiorparietal	0.001
19	L_superiorparietal	L_frontalpole	0.0012
20	L_cuneus	L_superiortemporal	0.0013
21	R_caudalmiddlefrontal	R_superiorfrontal	0.0016
22	L_Accumbens_area	R_Accumbens_area	0.0018
23	L_fusiform	L_precuneus	0.0022
24	R_precuneus	R_rostralmiddlefrontal	0.0024
25	L_inferiorparietal	L_insula	0.0026
26	Brain_Stem	R_Putamen	0.0027
27	L_fusiform	L_middletemporal	0.0029
28	R_Caudate	R_precuneus	0.0035
29	R_Thalamus_Proper	R_Caudate	0.0036
30	Brain_Stem	R_Amygdala	0.0037
31	Brain_Stem	R_precuneus	0.0037
32	Brain_Stem	L_superiortemporal	0.0038
33	L_Caudate	L_superiorparietal	0.0039
34	L_VentralDC	L_caudalanteriorcingulate	0.0045
35	L_lateraloccipital	L_superiorparietal	0.0047
36	R_precuneus	R_superiorparietal	0.0049
37	R_bankssts	R_insula	0.0049
38	L_caudalanteriorcingulate	R_posteriorcingulate	0.0052
39	R_Cerebellum_Cortex	R_Putamen	0.0052
40	R_Putamen	R_parstriangularis	0.0053

Finally, to determine links that exhibit altered DTI-fMRI correlations, we first compared the estimated Pearson correlation coefficients for the LLD and HC groups using Fisher’s z-transformation. Using an FDR value of 0.40, we were able to detect only one connectivity (L_paracentral–L_postcentral) that exhibited significant between-group difference in DTI-fMRI correlation. In an alternative approach, we computed the variance (as described in section 4.5) of the estimated correlation coefficients and used them in t-tests comparing the two groups. We were able to detect 33 connectivities that showed different DTI-fMRI relationship between the LLD and HC groups with an FDR value of 0.30 (see Table IX).

TABLE IX. SIGNIFICANT CONNECTIVITIES FOR CORRELATION—FDR=0.30

Num	Region 1	Region 2	P-value
1	L_paracentral	L_postcentral	<0.0001
2	R_parsopercularis	R_parstriangularis	0.0001
3	L_Caudate	R_rostralanteriorcingulate	0.0002
4	R_parstriangularis	R_precuneus	0.0004
5	L_precentral	R_precuneus	0.0004
6	R_Hippocampus	L_superiortemporal	0.0005
7	L_Caudate	L_frontalpole	0.0006
8	L_superiorparietal	L_frontalpole	0.0006
9	R_VentralDC	R_frontalpole	0.0006
10	L_cuneus	L_precuneus	0.0009
11	R_VentralDC	L_paracentral	0.0009
12	L_entorhinal	L_isthmuscingulate	0.0012
13	R_isthmuscingulate	R_pericalcarine	0.0016
14	L_Cerebellum_Cortex	L_Pallidum	0.0018
15	R_Pallidum	L_precentral	0.002
16	L_VentralDC	L_medialorbitofrontal	0.0026
17	R_Thalamus_Proper	R_isthmuscingulate	0.0028
18	R_Putamen	L_rostralanteriorcingulate	0.0031
19	L_caudalmiddlefrontal	L_precentral	0.0033
20	R_Pallidum	R_frontalpole	0.0033
21	L_paracentral	L_superiorparietal	0.0038
22	R_Accumbens_area	R_temporalpole	0.004
23	L_postcentral	R_precuneus	0.0043
24	R_paracentral	R_precuneus	0.0047
25	L_caudalmiddlefrontal	L_parsopercularis	0.0048
26	L_rostralmiddlefrontal	R_insula	0.0054
27	L_caudalanteriorcingulate	L_medialorbitofrontal	0.0054
28	L_superiorfrontal	R_medialorbitofrontal	0.0059
29	L_Putamen	L_pericalcarine	0.0061
30	L_Putamen	L_lingual	0.0062
31	R_parsopercularis	R_superiortemporal	0.0063
32	R_bankssts	R_inferiortemporal	0.0064
33	L_Thalamus_Proper	L_lateraloccipital	0.0064

To find out how results from the univariate, bivariate, and correlation analysis relate to each other, we compared the significant connectivities detected in these analyses and assessed the consistency among them. The result is shown in Table X. We observed almost no overlapping (3.2%) between the univariate DTI and fMRI results, suggesting distinct patterns of disruption of structural and functional connectivity in LLD subjects. In addition, we observed that the bivariate result is influenced to a greater extent by DTI (57.5%) than fMRI (25.0%). This is likely due to higher variability of the DTI data (see Table VI). The DTI-fMRI correlation showed another distinct disruption pattern compared to univariate DTI (0%), fMRI (3.0%), and bivariate (2.5%) analyses.

TABLE X. CMEM: COMPARISON ACROSS ANALYSES

	Univariate DTI	Univariate fMRI	Bivariate	Correlation
Univariate DTI	-			
Univariate fMRI	3.2% (1/31)	-		
Bivariate	57.5% (23/40)	25.0% (10/40)	-	
Correlation	0% (0/33)	3.0% (1/33)	2.5% (1/40)	-

Shown in Figures 5–8 are brain images of the identified connectivities that exhibited significant between-group differences in unimodal DTI and fMRI, bimodal, and correlation analyses. Nodes that are on the left and right sides of the brain are shown in red and green, respectively. The brain stem is shown in blue. The size of the node is proportional to the number of identified significant connections linked to it, and the width of the edge is proportional to the significance level (P-value) of the corresponding comparison between the connectivity in the LLD and HC groups. We can see from these figures that, although the number of common connectivities identified by the unimodal, bimodal, and correlation analyses is low, there exist quite a few common nodes that were involved in the identified connectivities.

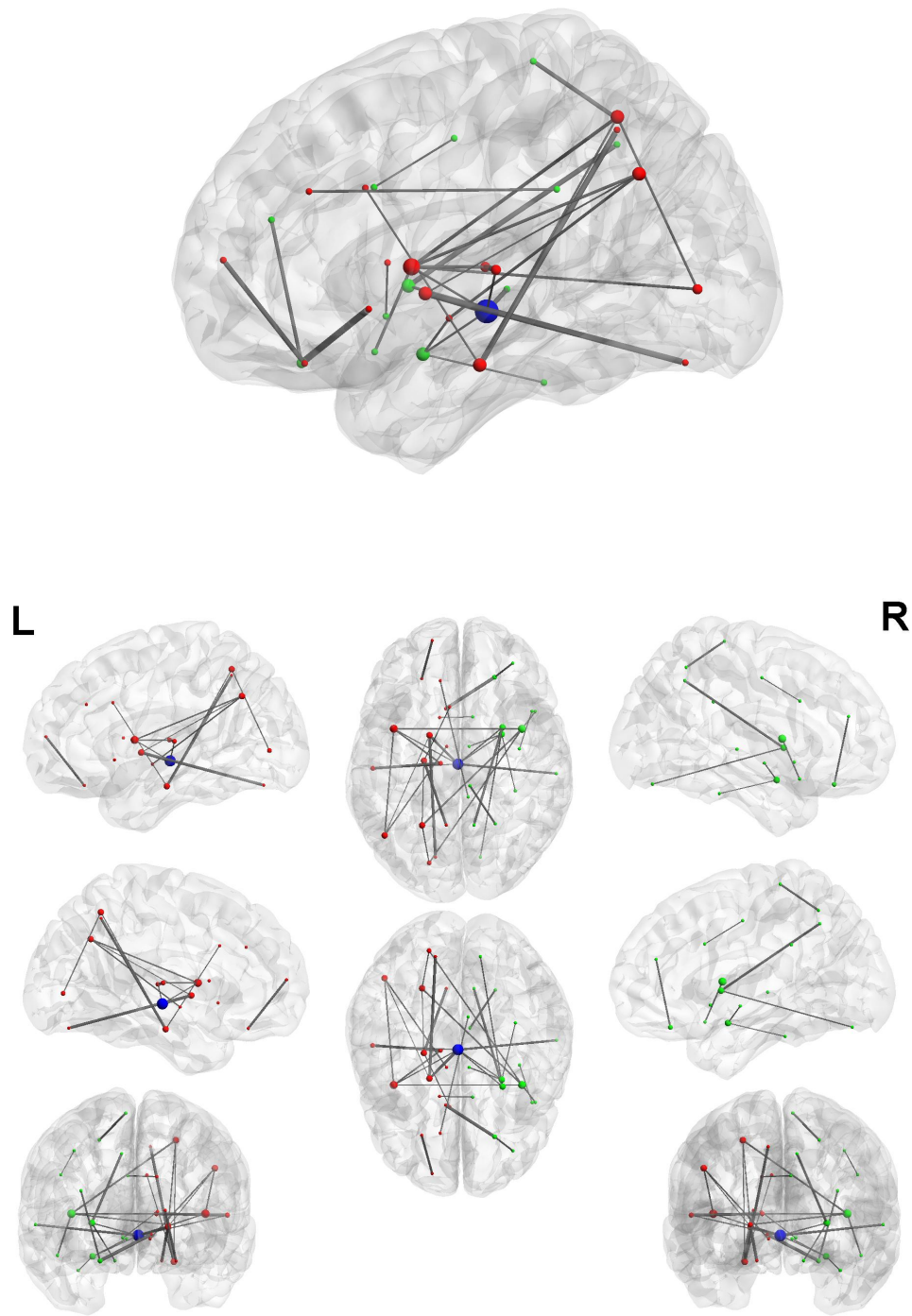


Figure 5. Connectivities with significant between-group difference in DTI.

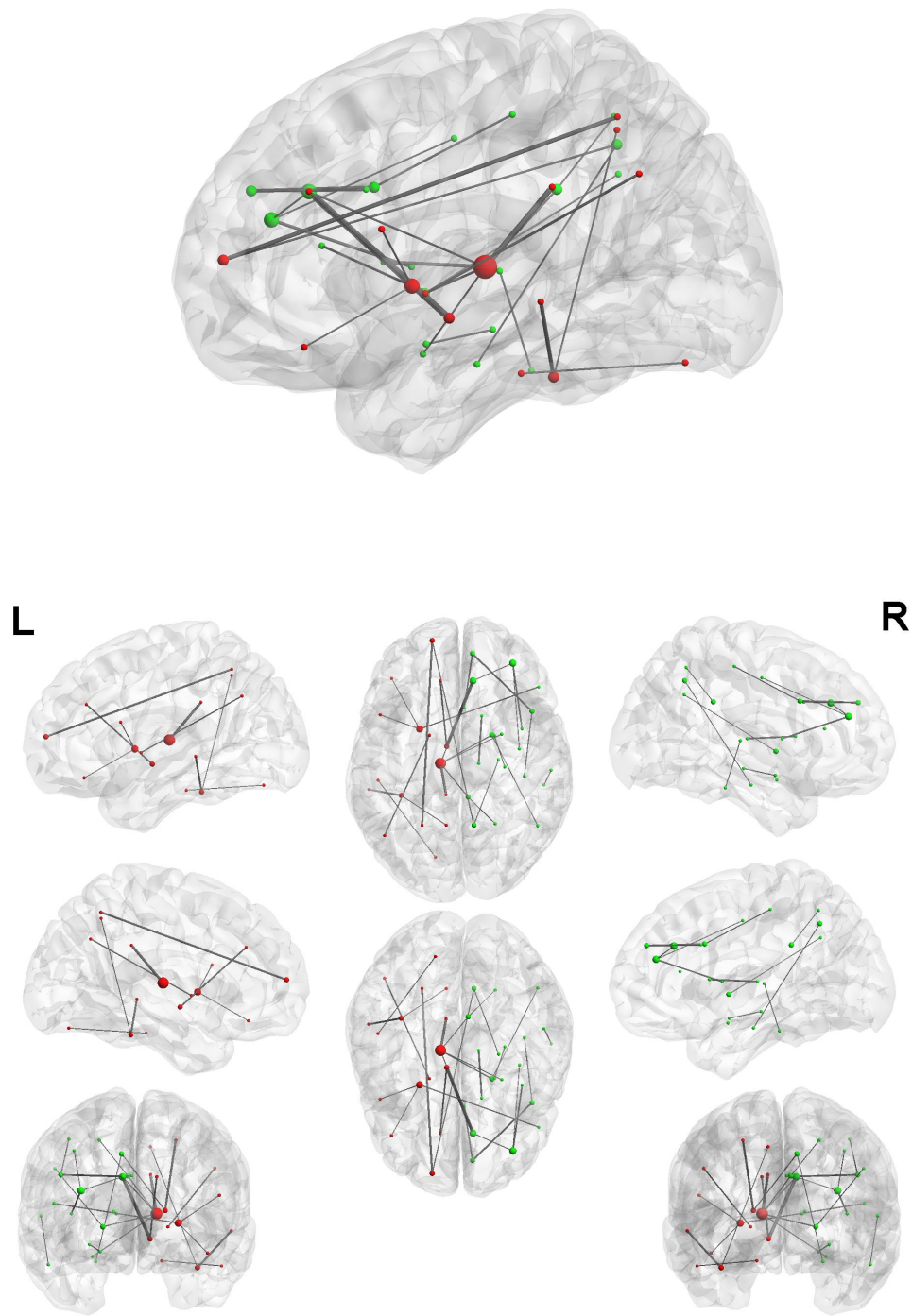


Figure 6. Connectivities with significant between-group difference in fMRI.

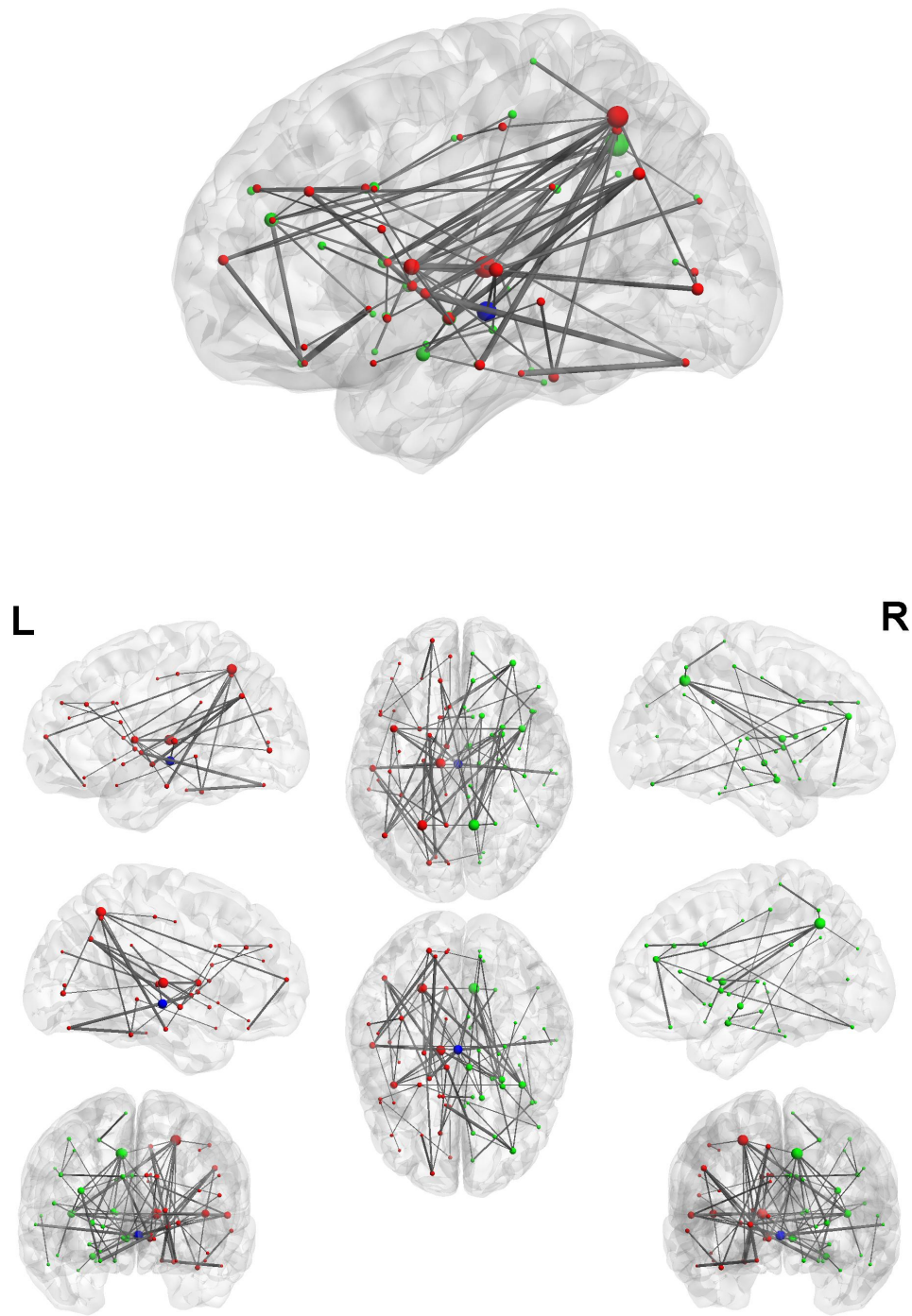


Figure 7. Connectivities with significant between-group difference in DTI and fMRI jointly.

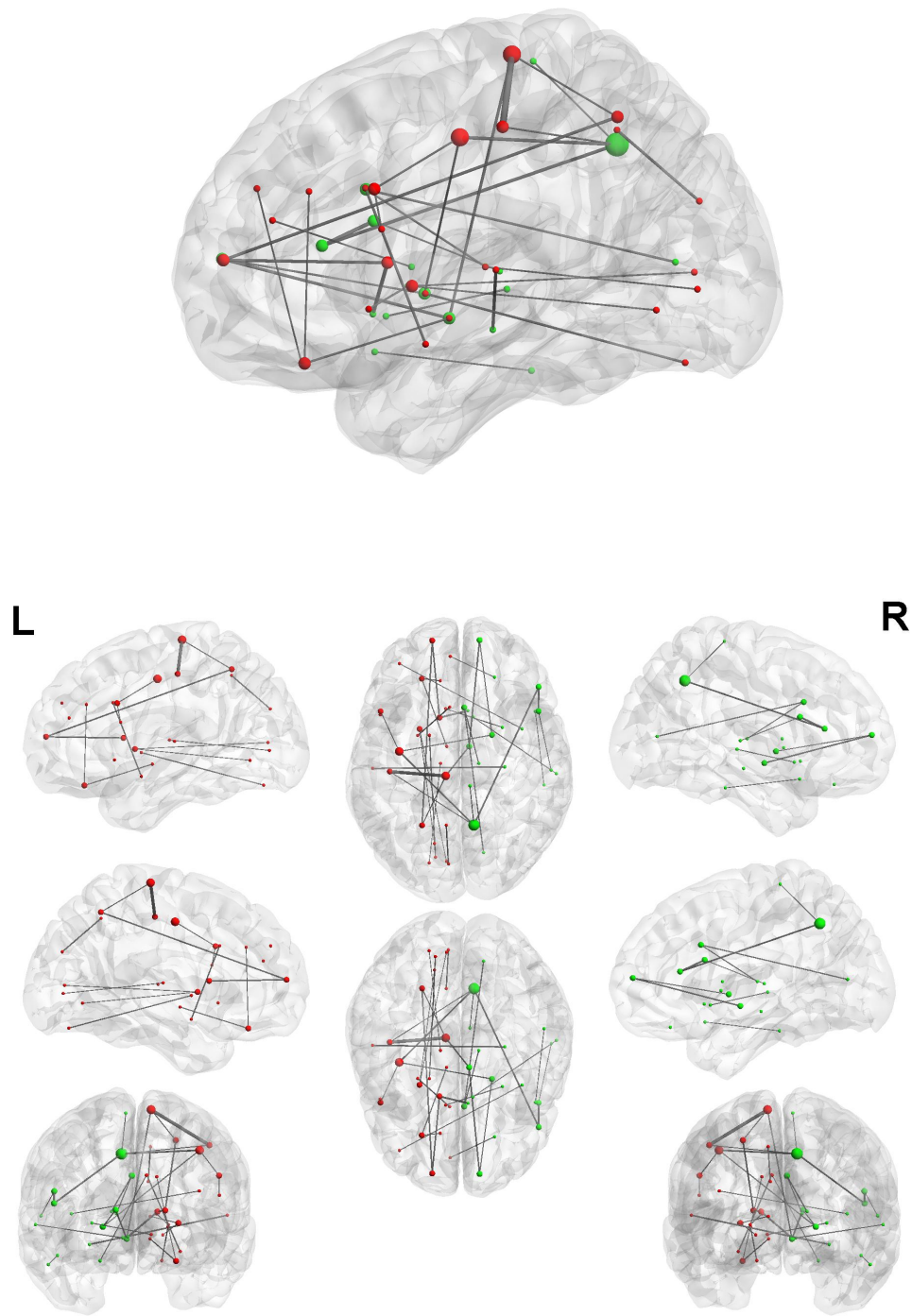


Figure 8. Connectivities with significant between-group difference in DTI-fMRI correlation.

So far we have identified the connectivities exhibiting significant between-group differences using the FDR approach. As connectivities are nested within the brain of the same subject, the comparisons are thus correlated. To address this issue and estimate the proportion of false discoveries we make, we employed an approach developed by Fan et al. (2012). Compared to the FDR approach, in which an FDR bound is prespecified and then a P-value threshold is determined, the rate of false discovery is computed in the FDP approach and the acceptable level of false discovery is determined. We estimated FDPs for the unimodal DTI and fMRI analyses using the empirical covariance matrices computed from the sample, and compared to the FDPs calculated when assuming independence among comparisons. The FDP computation was performed in R using the “pfa” package. The result is summarized in Table XI. Shown in Table XII is the FDP result when assuming independent covariance matrices.

TABLE XI. FALSE DISCOVERY PROPORTION—EMPIRICAL

Rejection	DTI			fMRI		
	Threshold	False Rejection	FDP	Threshold	False Rejection	FDP
1	2e-05	0.001	0.001	7e-06	3e-04	3e-04
...		...			...	
28	0.0052	0.9948	0.0355	0.0089	1.8409	0.0657
29	0.0059	1.1582	0.0399	0.0095	2.0204	0.0697
30	0.0065	1.3389	0.0446	0.0099	2.1369	0.0712
31	0.0066	1.3626	0.0440	0.0101	2.191	0.0707
...		...			...	
104	0.0484	19.5448	0.1879	0.0439	16.2911	0.1566
105	0.0512	21.1510	0.2014	0.0448	16.7238	0.1593
...		...			...	
113	0.0565	24.2362	0.2145	0.0496	19.2727	0.1706
114	0.0586	25.4803	0.2235	0.0500	19.5008	0.1711
...		...			...	

TABLE XII. FALSE DISCOVERY PROPORTION—INDEPENDENT

Rejection	DTI			fMRI		
	Threshold	False Rejection	FDP	Threshold	False Rejection	FDP
1	3e-05	0.0421	0.0421	1e-05	0.0168	0.0168
...		...			...	
28	0.0063	9.0148	0.322	0.0105	15.0417	0.5372
29	0.0070	10.0832	0.3477	0.0112	16.0812	0.5545
30	0.0078	11.2170	0.3739	0.0117	16.7410	0.5580
31	0.0079	11.3626	0.3665	0.0119	17.0433	0.5498
...		...			...	
94	0.0498	71.3175	0.7587	0.0444	63.6029	0.6766
95	0.0500	71.6210	0.7539	0.0454	64.9673	0.6839
...		...			...	
105	0.0565	80.9629	0.7711	0.0497	71.1485	0.6776
106	0.0581	83.2564	0.7854	0.0506	72.4638	0.6836
...		...			...	

From Table XI we see that if we had not addressed the multiple comparison issue and used $P\text{-value} < .05$ as the rejection criteria, we would have been rejecting 104 and 114 comparisons for the DTI and fMRI analyses. According to the table, these results are associated with FDPs of 0.1879 and 0.1711, which are very high false discovery rates. We see from the table that the reduction of the number of false rejections is not proportional to the decrease of the number of total rejections. The number of false discoveries decreases

at a faster rate when the number of rejections is high. As we reject fewer hypotheses, those hypotheses that favor the null are more likely to be retained. Table XI shows that if we reject 31 comparisons for DTI and 28 for fMRI, just like what we reported above when we used an FDR bound of 0.30, the FDPs are 0.0440 and 0.0657, which are reasonable and close to the traditional type I error rate limit of 0.05. Shown in Table XII are FDP results when independent covariance matrices were assumed. We see that FDP increases dramatically when comparisons are assumed to be not correlated whereas in reality they are.

One possible explanation to this is as follows. When tests are correlated and we assume the appropriate structure of dependency in the computation of the FDP, then all highly correlated hypotheses that favor the alternative are likely to be rejected together. Similarly, the hypotheses that favor the null are likely to be retained together because of their dependency. Under this scenario, if the value of the FDP is small, it indicates that the number of false discoveries is not large, and hence any hypothesis that is not rejected is likely not to be significant. However, under the assumption of an uncorrelated structure, if the number of rejections remains the same, the rejected hypotheses are more likely to be randomly selected from all hypotheses, and hence many rejected hypotheses will be falsely detected. This will lead to an inflated FDP. Taken together, this result suggests that when dealing with multiple testing problems, the underlying correlation structure plays an important role to control the FDP, and ignoring such correlations will increase the likelihood of the false discoveries.

7.1.3 Bayesian Mixed-Effects Models

The BMEMs that we employed in analyzing our neuroimaging data share the same model structure as the CMEMs that we used. The prior distributions for model parameters were specified in section 5.4. We used two chains with different starting values. After updating the models for 2,000 iterations of burn-in and further 3,000 iterations, a total of 6,000 samples were collected from the posterior distributions of model parameters. Model convergence diagnostics were performed using trace, density, autocorrelation, and BGR plots. We observed stabilization of both chains on trace plots (Figure 9 panel A), little to no autocorrelation for all lags greater than 0, and nice bell-shaped density plots (Figure 9 panel C). On BGR plots there are three colored lines for the width of 80% intervals of pooled chains (between-chain, green), the average width of 80% intervals for chains (within-chain, blue), and the ratio between the between- and within-chain variations (red). Upon convergence, the green and blue lines are expected to stabilize and the red line should be near one. All of these were observed on the BGR plots (Figure 9 panel D).

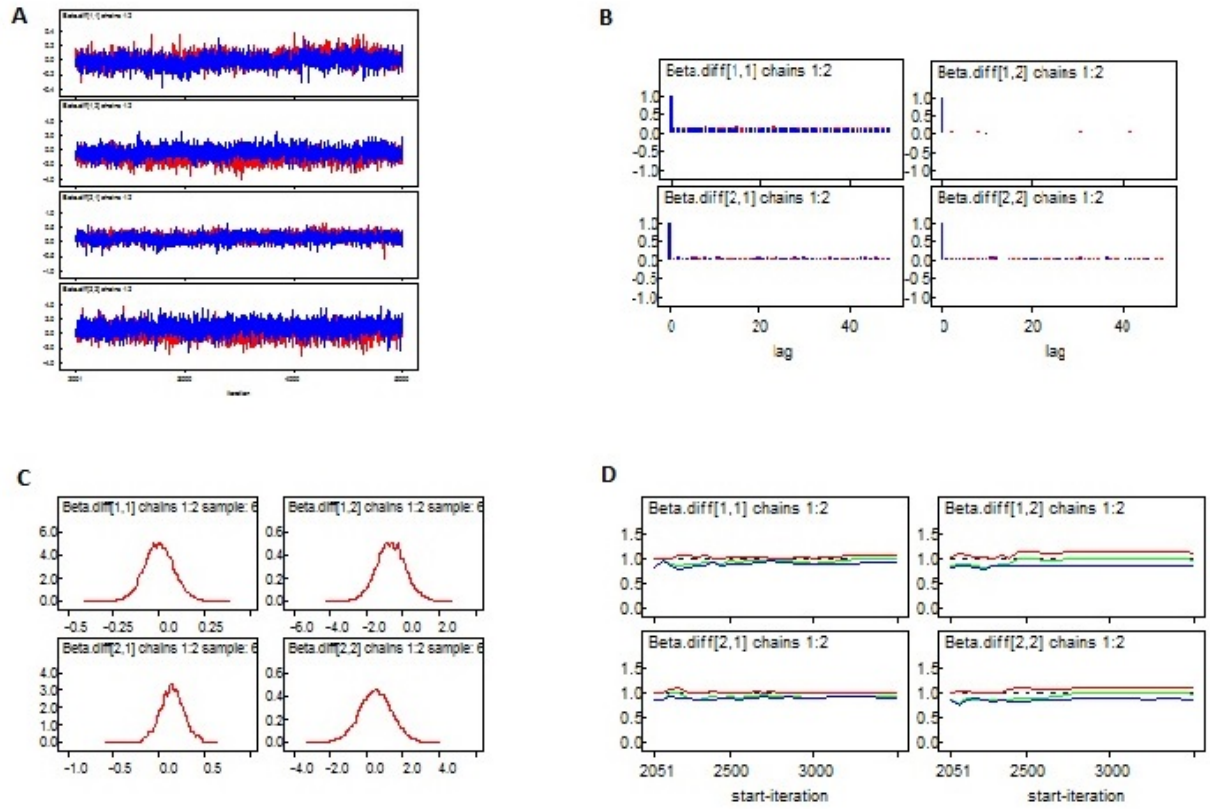


Figure 9. Convergence diagnostics of Bayesian Models: trace (A), auto correlation (B), density (C), and BGR (D) plots.

Shown in Table XIII are results from independent analysis without modeling, from CMEM analysis, and from BMEM analysis using Gamma and uniform priors for covariance parameters. We see that compared to the approach without modeling, where no between-group differences were detected with an FDR rate of 0.30 (Table XIII row 4), we were able to detect more than 30 connectivities that exhibited between-group differences using CMEM and an FDR value of 0.30 (Table XIII row 6). This indicates that by including both the between-modality and within-subject correlations in CMEM, we were able to partition the total variability into error and random effect variations and thus reduce the variation due to error, yielding more significant results. In contrast, we observed no significant between-group differences using BMEM and an FDR rate of 0.30 (Table XIII row 8).

TABLE XIII. COMPARISON OF RESULTS ACROSS METHODS

		Univariate		Bivariate	Correlation
		DTI	fMRI	DTI & fMRI	DTI-fMRI
No Model	P-value<.05	49	59	55	72
	FDR=.30	0	0	0	0
Classical Model	P-value<.05	106	114	160	133
	FDR=.30	31	28	90	33
Bayesian Model	95% CI	28	33	58	65
	FDR=.30	0	0	0	0

The significance in the BMEM result shown above was to a lesser extent compared to that in the CMEM result. This is likely due to different parameterizations we used in the two models. Although the structures of the main models are the same, the error and random effect covariances were estimated as a whole in CMEM, whereas components of the covariance were used as individual parameters in the BMEM analysis shown above.

To address this issue, we assumed inverse Wishart priors for the error and random effect covariances and updated the entire covariance matrices as a whole to obtain posterior samples. Testings were performed in the same way for comparing connectivities between the two groups and the results are summarized in Table XIV. In the table, “Variance Components” refers to the parameterization using individual variance components (variances and correlations) as parameters, whereas the others are cases when Wishart priors were used for the whole covariance matrices.

TABLE XIV. BMEM RESULTS WITH DIFFERENT PRIORS

		Univariate		Bivariate	Correlation
		DTI	fMRI	DTI & fMRI	DTI-fMRI
CMEM	P-value<.05	106	114	160	133
	FDR=.30	31	28	90	33
Variance Components	95% CI	28	33	58	65
	FDR=.30	0	0	0	0
$\mathbf{V} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	95% CI	136	29	186	8
	FDR=.30	76	0	74	0
$\mathbf{V} = \begin{pmatrix} 0.1 & 0 \\ 0 & 1 \end{pmatrix}$	95% CI	136	150	199	164
	FDR=.30	74	35	111	99
$\mathbf{V} = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}$	95% CI	146	108	226	302
	FDR=.30	78	15	206	386

We found that when we assumed Wishart priors and sampled the whole covariance matrices from the posterior distributions for our BMEM analysis, we observed higher extent of significance in comparing the two groups (Table XIV). For example, when we used BMEM with an FDR bound of 0.30 and $\mathbf{V} = \begin{pmatrix} 0.1 & 0 \\ 0 & 1 \end{pmatrix}$ as the scale matrix for the Wishart priors, we identified 74, 35, 111, and 99 connectivities that exhibited significant between-group differences in univariate, bivariate, and correlation analyses (Table XIV row 8). The numbers of significant connectivities are comparable to those found in analyses using CMEM shown previously. We also found that the number of significant connectivities changed as we used Wishart priors with different scale matrices. In other words, BMEM result is sensitive to the priors chosen.

7.1.4 Age Effect on Connectivity

We have identified connectivities that show significant between-group differences. The next thing we wanted to do was to investigate possible covariate effects on the between-group differences in connectivities. Since multiple studies have shown that brain connectivity is related to age (Gong et al., 2011; Mettenburg et al., 2012), it would be interesting to see whether age has any effects on brain connectivities in LLD and HC subjects. To address this, we included four parameters in CMEM for possible age effects: β_{a01} for the age effect on fMRI in the HC group, β_{a11} for the age effect on fMRI in the LLD group, β_{a02} for the age effect on DTI in the HC group, and β_{a12} for the age effect on DTI in the LLD group. We estimated the model using the same algorithm and obtained estimates of the four age effects, as well as the between-group differences in age effects. These estimates are shown in Table XV.

TABLE XV. AGE EFFECT ON CONNECTIVITY

		Estimate	SE	t	P-value
DTI	LLD	-0.1754	0.0741	-2.367	0.0455
	HC	-0.0206	0.0222	-0.924	0.3751
	LLD vs. HC	0.1548	0.0255	-6.066	0.0001
fMRI	LLD	-0.00283	0.0088	-0.321	0.7565
	HC	-0.00142	0.0027	-0.536	0.6027
	LLD vs. HC	0.00141	0.0030	0.4641	0.6535

Significant age effect was observed only for DTI in LLD subjects (P-value=.0455). The estimate for this effect is -0.1754, suggesting that when age increases by one year, the cube-root structural connectivity measured by DTI fiber count decreases by -0.1754

unit. Because of this significant age effect, the difference in age effects between the LLD and HC groups was also significant (P-value=.0001).

To investigate how age affects the number of connectivities exhibiting significant between-group differences, we compared the connectivities between LLD and HC subjects using the mean connectivity estimates obtained from the CMEM model controlling for age. Assuming an FDR of 0.30, we identified 32, 18, 122, and four connectivities that exhibited significant between-group differences in unimodal DTI, unimodal fMRI, bimodal, and correlation analyses (Table XVI row 3). These numbers are similar (except for the correlation analysis) with those determined by the previous CMEM model without controlling for age. Moreover, the identified connectivities by CMEM controlling for age match with those identified by the model not controlling for age (Table XVI row 5). Taken together, these results suggest that age does not have a great influence on functional and structural connectivities of the subjects studied.

TABLE XVI. SIGNIFICANT CONNECTIVITIES CONTROLLING FOR AGE

	Univariate		Bivariate	Correlation
	DTI	fMRI	DTI & fMRI	DTI-fMRI
CMEM with Age	32	18	122	4
CMEM without Age	31	28	90	33
Percent Match	77.4% (24/31)	100% (18/18)	85.6% (77/90)	100% (4/4)

7.2 Network Structure Analysis

We assessed network structure abnormalities in LLD subjects based on degrees of connectivity similarity between these and HC subjects. We obtained mean connectivity estimates from our CMEM model results (regression coefficient estimates), and fitted gamma distributions to the mean DTI measures and beta distributions to the absolute mean fMRI measures of the 1,433 links in the LLD and HC groups. Figure 10 shows the overall good fit of these distributions to the mean connectivity estimates.

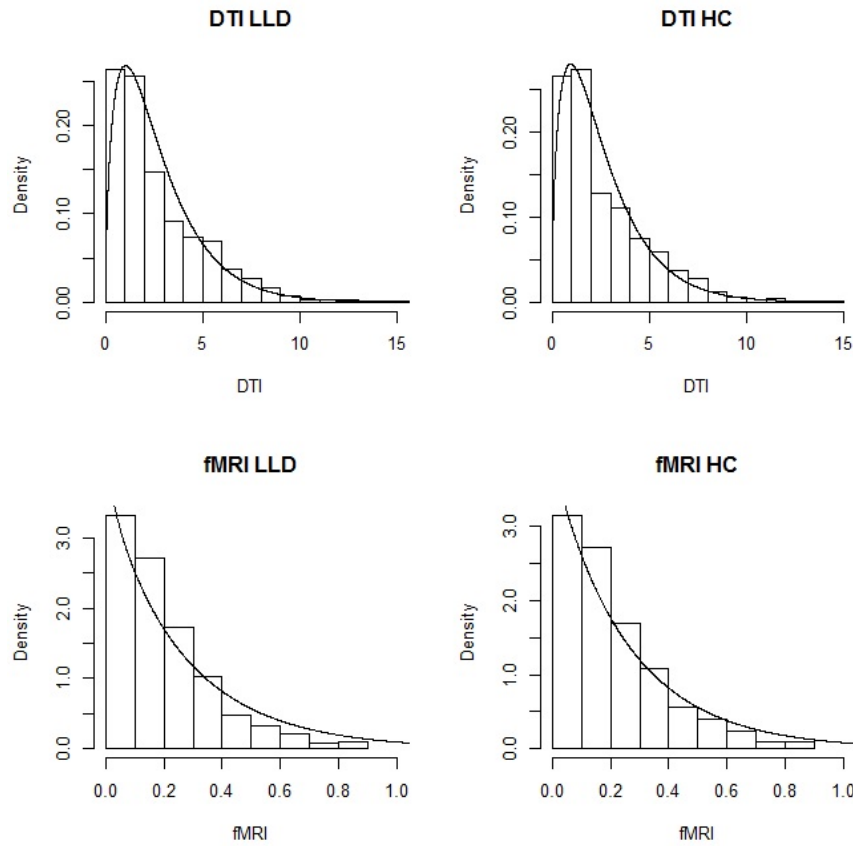


Figure 10. Gamma and beta fit to the mean DTI and fMRI measures.

Next, we placed the 1,433 links into 4×4 tables according to their connectivity strength in the LLD and HC groups and the quartiles determined by the fitted gamma and beta distributions. The result for DTI and fMRI are shown in Table XVII.

TABLE XVII. CONNECTIVITY SIMILARITY IN LLD AND HC

		LLD						LLD					
		1st	2nd	3rd	4th			1st	2nd	3rd	4th		
[DTI]	HC	1st	332	94	4	0	[fMRI]	HC	1st	188	121	20	0
		2nd	110	174	68	4			2nd	138	173	106	1
		3rd	12	54	159	38			3rd	19	74	211	65
		4th	0	3	37	344			4th	0	4	54	259

We observed that, for both DTI and fMRI, most of the links are either along the diagonal, or in a cell that is next to the diagonal of the 4×4 matrix, indicating consistent connectivity strengths of these links in the LLD and HC groups. When we computed the proportion of the cells of the matrix, the cells that are on the diagonal showed much higher observed than expected proportions whereas those that are on the off-diagonal showed much lower observed than expected proportions. This is not surprising because in a whole-brain analysis the majority of connectivities are not likely to be damaged in LLD, and hence connectivity in HD and LLD should show high level of concordance. Shown in Figure 11 and 12 are the links in the $[4, 4]$ cells of the DTI and fMRI matrices, respectively. These are the links exhibiting strong structural and functional connectivities in both LLD and HC subjects—i.e., the connectivities that are not damaged in LLD subjects.

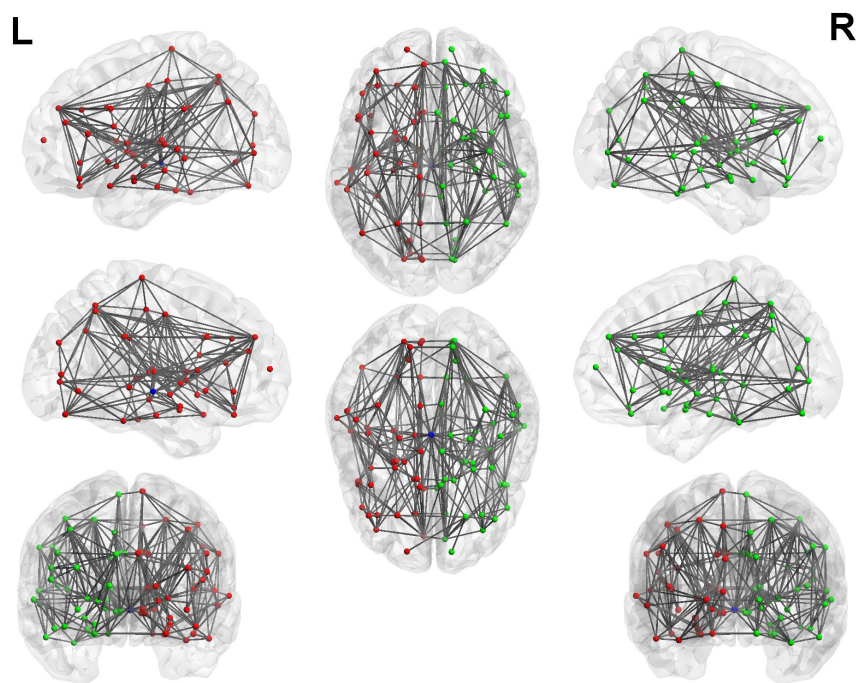


Figure 11. Strong structural connectivities in both HC and LLD subjects.

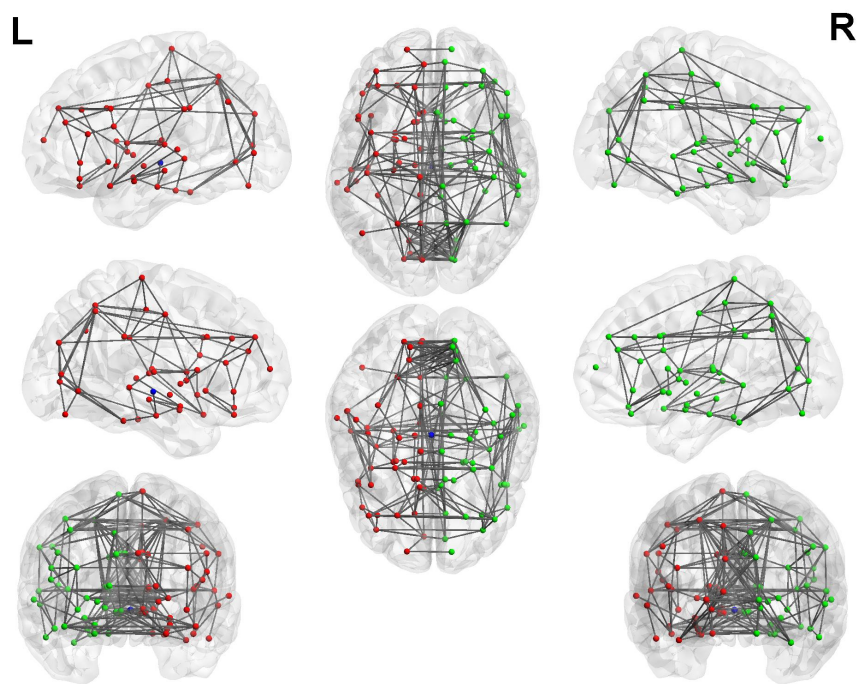


Figure 12. Strong functional connectivities in both HC and LLD subjects.

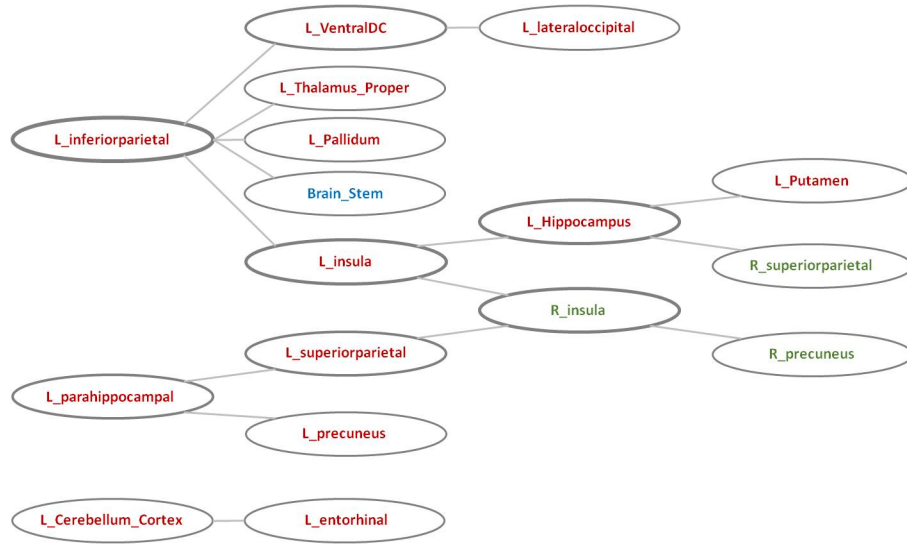
To see whether these strong connectivities are consistent for DTI and fMRI, we counted the number of common links for each cell in the 4×4 matrix. The result is shown in Table XVIII. We observed that, for each cell, only a portion of the links are common between the structural and functional results. This observation is consistent with our previous result that showed nonoverlapping sets of connectivities exhibiting between-group differences in DTI and fMRI (see result in section 7.1).

TABLE XVIII. COMMON CONNECTIVITIES IN LLD AND HC

		LLD			
		1st	2nd	3rd	4th
HC	1st	64	12	0	0
	2nd	15	33	3	0
	3rd	0	5	37	2
	4th	0	0	4	149

We then examined the links that belonged to the off-diagonal entries in the matrices. We were especially interested in those links that exhibited disrupted connectivity in LLD compared to HC. We found that 15 links exhibited lower (more than one quartile difference) DTI in LLD than in HC (12 in 3rd HC quartile and 1st LLD quartile; 3 in 4th HC quartile and 2nd LLD quartile) and 23 links exhibited lower fMRI in LLD than in HC (19 in 3rd HC quartile and 1st LLD quartile; 4 in 4th HC quartile and 2nd LLD quartile), and there were no common links in these two sets (see Table XVII and XVIII). Shown in Figure 13 are the network diagrams and in Figure 14 and 15 are the brain images for these structural and functional connectivities that were potentially damaged in LLD subjects. From these figures, we see that both the damaged structural and functional connectivities were primarily on the left ventral part of the brain. The damaged structural

connectivities were mostly located toward the posterior of the brain whereas the disrupted functional connectivities were more toward the anterior. In addition, we observed left inferiorparietal and left fusiform as network hubs for damaged structural and functional connectivity, respectively.



(a) Structural Connectivity (DTI)



(b) Functional Connectivity (fMRI)

Figure 13. Network diagram of connectivities in HC but not LLD subjects.

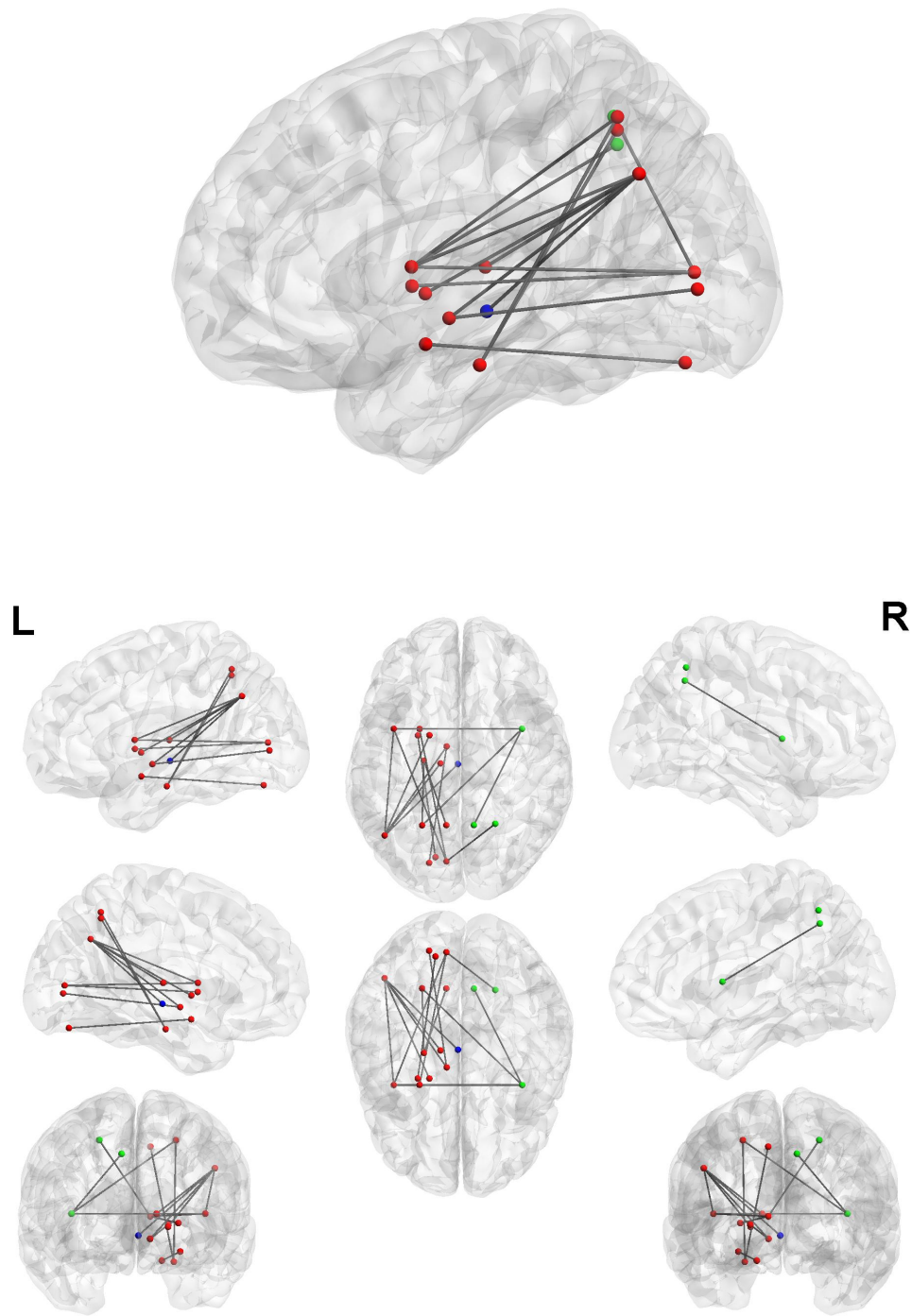


Figure 14. Image of structural connectivities in HC but not LLD subjects.

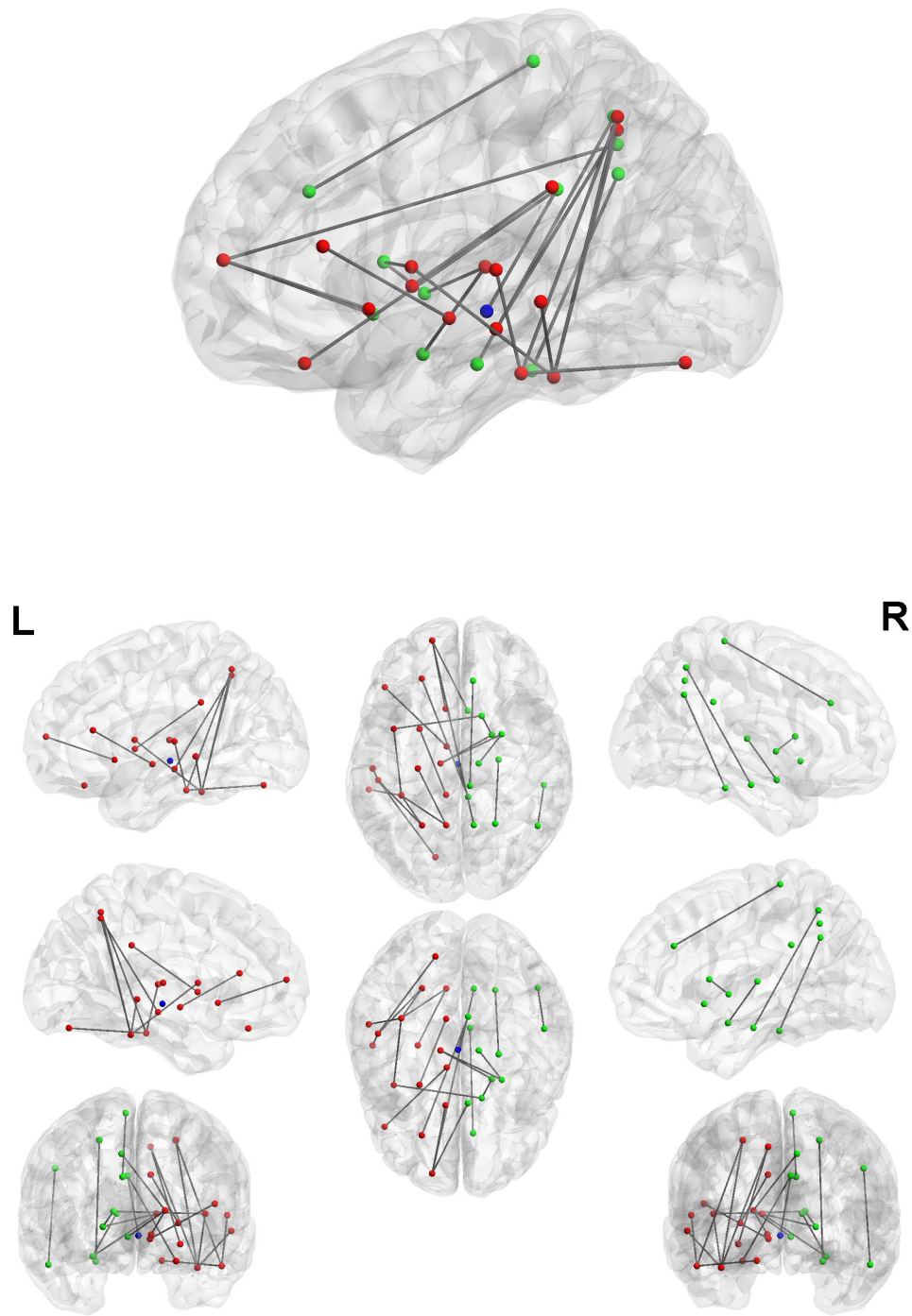


Figure 15. Image of functional connectivities in HC but not LLD subjects.

7.3 Machine Learning and Classification

7.3.1 Linear Support Vector Machine

Using machine learning techniques, we first detected features (connectivities) that are important in discriminating between LLD and HC subjects. We trained SVM using the whole data set (23 subjects). Then all the connectivities were ranked according to their relative importance in maximizing the margin to the decision boundary. We compared the most important features with the significant connectivities that we detected using the mixed-effects models (in section 7.1). The result is shown in Table XIX.

TABLE XIX. COMPARISON ACROSS METHODS

		CMEM	BMEM	SVM
DTI	CMEM	-		
	BMEM	71.0% (22/31)	-	
	SVM	77.4% (24/31)	61.3% (19/31)	-
fMRI	CMEM	-		
	BMEM	64.3% (18/28)	-	
	SVM	64.3% (18/28)	64.3% (18/28)	-

We observed fairly consistent result in terms of the discriminating power among CMEM, BMEM, and SVM. The percentages of overlapping connectivities were more than 60% for all cases for both DTI and fMRI (see Table XIX).

We then employed linear SVMs to classify subjects into LLD versus HC groups using two different approaches (see Figure 16). In the first “Leave One” approach, we chose one subject to be used as the test set and used the remaining 22 subjects as the training

set. In the other “Leave Pair” approach, we chose two subjects, one from the LLD group and the other from the HC group, to be used as the test set and treated the remaining 21 subjects as the training set. For cross-validation of the model, we repeated the process for all combinations of subjects in the training and test sets. Sensitivity (true positive, which is the proportion of correctly identified LLD subjects), specificity (true negative, which is the proportion of correctly identified HC subjects), and overall accuracy were recorded to evaluate the performance of the classification.

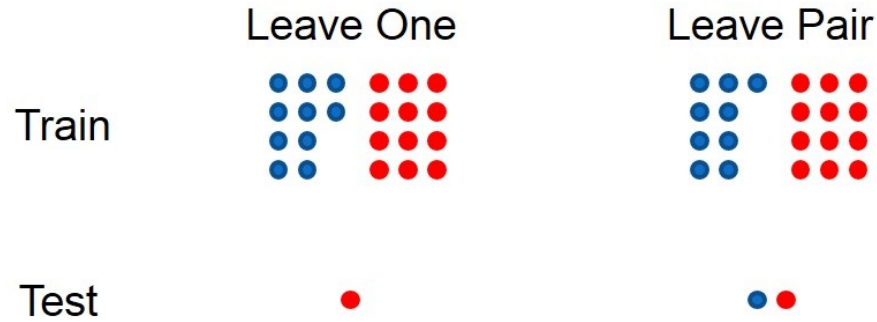


Figure 16. Diagram of classification using SVM.

We observed unsatisfactory accuracy rates for DTI and fMRI measures using both approaches mentioned above (Table XX column 3), mainly due to the low sensitivities we had in these analyses (Table XX column 4).

TABLE XX. CLASSIFICATION USING SVM

		Accuracy	Sensitivity	Specificity
DTI	Leave One	0.3913	0.1000	0.6154
	Leave Pair	0.4500	0.1500	0.7500
fMRI	Leave One	0.6522	0.3000	0.9231
	Leave Pair	0.4750	0.1500	0.8000

Next we wanted to investigate whether the performance of SVM can be improved using feature selection procedures. Specifically, we wanted to find out whether using the important connectivities determined by CMEM could increase the sensitivity and specificity of the machine. We selected connectivities according to the P-values for the between-group comparisons. Cut-offs were defined and only those connectivities whose P-values are within the cut-off were selected as input features for SVM. We applied the “Leave One” and “Leave Pair” approaches but only using the selected connectivities as inputs. The result is shown in Table XXI.

TABLE XXI. SVM CLASSIFICATION USING CMEM RESULT

	P-value Cut-off	No. of Features	Accuracy	Sensitivity	Specificity
DTI	-	1433	0.3913	0.1000	0.6154
	0.9	1313	0.5652	0.4000	0.6923
	0.8	1172	0.6957	0.6000	0.7692
	0.7	1046	0.8261	0.8000	0.8462
	0.6	905	0.8696	0.9000	0.8462
	0.5	771	0.9130	0.9000	0.9230
	0.4	646	0.9565	0.9000	1.0000
	0.3	509	0.9565	0.9000	1.0000
	0.2	355	0.9565	0.9000	1.0000
	0.1	175	1.0000	1.0000	1.0000
fMRI	-	1433	0.6522	0.3000	0.9231
	0.9	1301	0.6957	0.4000	0.9231
	0.8	1152	0.7826	0.5000	1.0000
	0.7	1038	0.9130	0.8000	1.0000
	0.6	896	0.9565	0.9000	1.0000
	0.5	763	0.9565	0.9000	1.0000
	0.4	643	1.0000	1.0000	1.0000
	0.3	488	1.0000	1.0000	1.0000
	0.2	345	1.0000	1.0000	1.0000
	0.1	197	1.0000	1.0000	1.0000

We see in Table XXI that using CMEM result dramatically improved the performance of classification by SVM. For DTI, with a P-value cut-off of 0.5, we obtained more than 90% sensitivity, specificity, and accuracy. And for fMRI, with a P-value cut-off of 0.6, we obtained more than 90% sensitivity, specificity, and accuracy. However, we used the whole data set in the CMEM modeling step in the above-mentioned analysis. This can potentially cause over-fitting and produce biased results. A more appropriate way would be to use only the training set in estimating the CMEM and performing variable selection. To address this, we separated the training and test sets from the beginning, fitted CMEM to only data in the training set, and used the connectivities selected from corresponding models as inputs in SVM. The result is shown in Table XXII.

TABLE XXII. SVM CLASSIFICATION USING CMEM RESULT—UNBIASED

	P-value Cut-off	Accuracy	Sensitivity	Specificity
DTI	0.9	0.5217	0.3000	0.6923
	0.8	0.5217	0.3000	0.6923
	0.7	0.5652	0.3000	0.7692
	0.6	0.5652	0.3000	0.7692
	0.5	0.5652	0.4000	0.6923
	0.4	0.5217	0.3000	0.6923
	0.3	0.5652	0.4000	0.6923
	0.2	0.4348	0.3000	0.6153
	0.1	0.5652	0.4000	0.6923
fMRI	0.9	0.6522	0.4000	0.8462
	0.8	0.6522	0.4000	0.8462
	0.7	0.6522	0.4000	0.8462
	0.6	0.6522	0.4000	0.8462
	0.5	0.6522	0.4000	0.8462
	0.4	0.6957	0.4000	0.9231
	0.3	0.6957	0.4000	0.9231
	0.2	0.6522	0.4000	0.8461
	0.1	0.6522	0.3000	0.7692

We see in Table XXII that feature selection based on CMEM model result did not improve the classification accuracy by SVM. The improved SVM accuracy shown in Table XXII is likely due to over-fitting.

7.3.2 Support Vector Machine with Other Kernels

We have shown that the CMEM model result does not improve the classification power using SVM with linear kernel. To investigate whether the classification power can be improved using other kernels, we ran SVMs with third-degree polynomial and Gaussian kernels for our classification problem and the results are shown in Tables XXIII and XXIV.

TABLE XXIII. SVM WITH THIRD DEGREE POLYNOMIAL KERNEL

	P-value Cut-off	Accuracy	Sensitivity	Specificity
DTI	0.9	0.3913	0.1000	0.6154
	0.8	0.4348	0.1000	0.6923
	0.7	0.5217	0.2000	0.7692
	0.6	0.5217	0.3000	0.6923
	0.5	0.5217	0.3000	0.6923
	0.4	0.5217	0.3000	0.6923
	0.3	0.5652	0.4000	0.6923
	0.2	0.4348	0.2000	0.6154
	0.1	0.5217	0.3000	0.6923
fMRI	0.9	0.6087	0.3000	0.8462
	0.8	0.6087	0.3000	0.8462
	0.7	0.6522	0.4000	0.8462
	0.6	0.6522	0.4000	0.8462
	0.5	0.6522	0.4000	0.8462
	0.4	0.6957	0.4000	0.9231
	0.3	0.6957	0.4000	0.9231
	0.2	0.6522	0.4000	0.8461
	0.1	0.5652	0.3000	0.7692

TABLE XXIV. SVM WITH GAUSSIAN KERNEL

	P-value Cut-off	Accuracy	Sensitivity	Specificity
DTI	0.9	0.5652	0.0000	1.0000
	0.8	0.5652	0.0000	1.0000
	0.7	0.5652	0.0000	1.0000
	0.6	0.5652	0.0000	1.0000
	0.5	0.5652	0.0000	1.0000
	0.4	0.4783	0.0000	0.8462
	0.3	0.4783	0.0000	0.8462
	0.2	0.3913	0.1000	0.6154
	0.1	0.4348	0.1000	0.6923
fMRI	0.9	0.5652	0.0000	1.0000
	0.8	0.5217	0.0000	0.9231
	0.7	0.5652	0.0000	1.0000
	0.6	0.5217	0.0000	0.9231
	0.5	0.5217	0.0000	0.9231
	0.4	0.5217	0.0000	0.9231
	0.3	0.5217	0.0000	0.9231
	0.2	0.4783	0.0000	0.8461
	0.1	0.5217	0.1000	0.8461

We found that, similarly to what we observed when we used linear SVMs, feature selection based on the CMEM model result did not improve the classification accuracy by SVM with either third degree polynomial or Gaussian kernels.

8. SUMMARY AND CONCLUSION

We fitted CMEMs to bimodal neuroimaging data containing both structural (DTI) and functional (fMRI) connectivity measures. The bivariate model incorporated between-modality correlations of the DTI and fMRI measures, as well as the within-subject correlations by including random subject intercepts. We showed that, compared to approaches that analyze each link separately, the CMEM were more powerful in detecting links with significant between-group differences in univariate, bivariate, and correlation analyses. Using CMEMs and correcting for multiple comparisons with an FDR rate of 0.30, we were able to detect 31, 28, 90, and 33 links that exhibited significant differences between the LLD and HC groups in univariate DTI, univariate fMRI, bivariate, and DTI-fMRI correlation analysis, respectively.

Interestingly, we found distinct patterns of disruption in LLD subjects for structural and functional connectivity. Work by Koski and Paus (2000) supports the idea of independent structural and functional connectivity. In fact, it has long been speculated that direct structural connectivity is neither a sufficient nor a necessary condition for observing functional activity (Rykhlevskaia et al., 2008). Structural connections only provide neural pathways between two brain regions for functional connectivity. The actual occurrence of interactive activities between the two regions is more complicated. It is not uncommon that no information exchange would happen even though a number of neural pathways exist between two brain regions (Hampson et al., 2004; Schoenfeld et al., 2002). On the other hand, functional connection can also be present even though there is no direct structural linkage between two brain regions. Although structural connection is a good predictor of functional relationship, studies have shown that functional relationships can originate from various pathways and direct structural connections may

not be necessary (Eggermont and Smith, 1996; Koch et al., 2002). In fact, our result did show that although different structural and functional connectivities might be damaged in LLD subjects, multiple common nodes were associated with the structural and functional disruptions, suggesting the involvement of indirect functional pathways.

Our application of the FDR method provided good controls over the false discovery rates. We showed that, whereas the unadjusted result (using $P\text{-value} < .05$ only) corresponded with FDPs of more than 0.15, the overall FDPs for the univariate analyses were estimated to be around 0.05 when empirical covariance dependence estimated from the sample and an FDR of 0.30 were used. The estimation of FDPs for the bivariate and correlation analyses requires the development of FDP methods for test statistics that follow non-normal distributions and is under development by our group.

No significant links were identified using our BMEMs with the variance component parameters after correcting for multiple comparisons with an FDR. This is likely due to the different parameterizations in this model. In fact, we showed that when we treated all the covariance matrices as a whole and used Wishart priors for them, we observed similar numbers of significant connectivities for BMEM as compared to CMEM. We also saw that the number of significance depended on the Wishart prior chosen.

We identified connectivities that were present in the HC subjects but damaged in the LLD subjects using an approach based on degrees of quartile similarities. We observed that the damaged connectivities were mostly on the left side of the brain, and the damaged structural connectivities were mostly in the posterior part whereas the disrupted functional connectivities were more toward the anterior part of the brain.

We found that, using the connectivities identified by the CMEMs based on the whole data set as inputs, the accuracy of the classification of the linear SVM was greatly

improved. However, this was not the case when we ran separate CMEMs for different cross-validation iterations. Moreover, the same feature selection procedure based on the CMEM model result did not improve the classification accuracy by SVM with either third-degree polynomial or Gaussian kernels. Although other algorithms are available, it is unlikely that feature selection methods will improve the classification performance, mainly due to the small sample size of the data.

We included random subject intercepts in our models to account for the within-subject correlations of the connectivity measures because multiple measures were taken for the same subject. Since in fact each brain region is associated with many connectivities, including random region intercepts would be meaningful as well. However, since two regions are involved for each connectivity, things would be very complicated if these random region effects were considered. Without knowledge about specific region effects on connectivities, it would be difficult to determine the appropriate ways to incorporate the two random effects (through addition or multiplication, for example) for the two regions into the model. For this reason and also model parsimony, we ignored them in our models. If such information is available, random region effects can be easily included and estimated in our models.

We observed age effect on only the structural connectivity in LLD subjects. And we identified similar numbers of significant connectivities for all but the correlation analysis after adjusting for age, suggesting age is not playing a major role in affecting connectivity differences between the LLD and HC groups. We did not study the impact of other covariate effects on the connectivity difference between the two groups. It has been shown by multiple studies that brain connectivity is related to gender (Gong et al., 2011; Lopez-Larson et al., 2011). For depression, “duration of illness (years since first mania and years since first major depression) and the number of mood episodes (manic

and depressive)” are also highly associated with brain connectivities (GadElkarim et al., 2013). Further analyses using CMEM and BMEM including these covariates will not only provide covariate-adjusted inferences on the between-group connectivity differences, but also shed light on the covariate effects on multimodal brain connectivities.

One major limitation of this work is the limited sample size in the data set we used to illustrate our models. The data set has a sample size of only 23, with 10 subjects for the LLD and 13 subjects for the HC group. Therefore, the inferences we made in our analyses may not have high reliability due to this limited sample size.

Recent connectivity research for depression has been focused on the UF and CB tracts for damages in structural connectivity (Taylor et al., 2007; Sheline et al., 2008; Dalby et al., 2010; Mettenburg et al., 2012), and default mode, cognitive control, and affective networks for disruptions of functional connectivity (Albert et al., 2009). Using a neuroimaging data set containing both the structural and functional connectivity measures, we identified brain region links that are disrupted in LLD patients and found that the nodes and edges involved were primarily on the left ventral side of the brain. The ventral localization of the damaged connectivities is consistent with what has been found in literature, as UF and CB are ventral and medial associative bundles. The fact that the damaged connectivities were mostly found on the left side of the brain is also consistent with literature, as it has been shown that left and right sides of the brain have differential activation patterns by tasks (Demonet et al., 1994; O’Doherty et al., 2001; Swick and Ashley, 2008), and that depression decreases functional coupling in the left brain hemisphere (Tao et al., 2013). Compared to other studies, our work not only determined the parts of the brain that were affected in depression, but also identified specific brain region nodes and links, and especially, the brain network hubs that are potentially involved in the pathophysiology of the disease.

Our work represents a first attempt in simultaneous statistical modeling of two connectivity modalities in a population study while considering both the between-modality and within-subject correlations of the measures. To illustrate our methods, we used a data set from a depression research that contained bimodal neuroimaging data with DTI and fMRI measures. Other neuroimaging measures, such as EEG, PEG, MTR, and so on, can also be modeled using our methods with appropriate normalizing transformations. Moreover, the bivariate model can be easily generalized to multivariate systems where multiple neuroimaging measures can be simultaneously examined.

We used a depression study as an example to illustrate the methods proposed in this work. However, the outlined methods are ready to be applied to brain connectivity analysis using neuroimaging data for population studies of other diseases. For example, we are in collaboration with a research group that has been developing treatments for traumatic brain injury (TBI, (Pape et al., 2006; Patil et al., 2011; Pape et al., 2013)), to use the methods we propose in this work to investigate damage of brain connectivities after TBI, and how neuro-therapeutics may be developed to help TBI patients regain their mental and physical functions. Another interesting structure-functional relationship to investigate is in personality disorders (McCloskey et al., 2005). Abnormalities have been identified for white-matter in UF (Phan et al., 2009) and for amygdala and orbitofrontal reactivity (Coccaro et al., 2007) for social anxiety disorders. Using methods outlined in this work, one can investigate the relationship between structural and functional connectivity disruption in these regions during this disease and gain a better understanding of the underlying mechanism of pathophysiological changes in the brain due to this disorder. Other possible areas where our methods can be applied include population studies of dementia, schizophrenia, Alzheimer’s disease, and bipolar disease. It is our hope that this work will provide a general and yet powerful platform for statistical modeling of

multimodal neuroimaging data to help delineate the relationship between structural and functional connectivity in various neurological disorders.

References

- Adluru, N., C. Hinrichs, J. E. Chung, M. K. amd Lee, V. Singh, E. D. Bigler, N. Lange, J. E. Lainhart, and A. L. Alexander (2009). “Classification in DTI using Shapes of White Matter Tracts”. *Conference Proceedings of the IEEE Engineering in Medicine and Biology Society*: 2719–22. doi:10.1109/IEMBS.2009.5333386.
- Albert, N. B., E. M. Robertson, and R. C. Miall. “The Resting Human Brain and Motor Learning”. *Current Biology* 19, no. 12 (2009): 1023–27. doi:10.1016/j.cub.2009.04.028.
- Alexopoulos, G. S., 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, no. 1 (2012): 56–65. doi:10.1016/j.jad.2011.12.002.
- Alexopoulos, G. S., B. S. Meyers, R. C. Young, S. Campbell, D. Silbersweig, and M. Charlson. “‘Vascular Depression’ Hypothesis”. *Archives of General Psychiatry* 54, no. 10 (1997): 915–22.
- Allen, L. S., M. F. Richey, Y. M. Chain, and R. A. Gorski. “Sex Differences in the Corpus Callosum of the Living Human Being”. *Journal of Neuroscience* 11, no. 4 (1991): 933–42.
- Astolfi, L., F. Cincotti, D. Mattia, M. G. Marciani, L. A. Baccala, F. de Vico Fallani, S. Salinari, M. Ursino, M. Zavaglia, L. Ding, J. C. Edgar, M. G. A., B. He, and F. Babiloni. “Comparison of Different Cortical Connectivity Estimators for High-Resolution EEG Recordings”. *Human Brain Mapping* 28, no. 2 (2007): 143–57.
- Au Duong, M. V., B. Audoin, K. Boulanouar, D. Ibarrola, I. Malikova, S. Confort-Gouny, P. Celsis, J. Pelletier, P. J. Cozzzone, and J. P. Ranjeva. “Altered Functional

- Connectivity Related to White Matter Changes inside the Working Memory Network at the Very Early Stage of MS". *Journal of Cerebral Blood Flow and Metabolism* 25, no. 10 (2005): 1245–53.
- Bandettini, P. A., E. C. Wong, R. S. Hinks, R. S. Tikofsky, and J. S. Hyde. "Time Course EPI of Human Brain Function during Task Activation". *Magnetic Resonance in Medicine* 25, no. 2 (1992): 390–97.
- Barlow, D. H. and V. M. Durand. "*Abnormal Psychology: An Integrative Approach (5th ed.)*". Belmont, CA: Thomson Wadsworth, 2005.
- Baumgartner, R., C. Windischberger, and E. Moser. "Quantification in Functional Magnetic Resonance Imaging: Fuzzy Clustering vs. Correlation Analysis". *Magnetic Resonance Imaging* 16, no. 2 (1998): 115–25.
- Benjamini, Y. and Y. Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing". *Journal of the Royal Statistical Society. Series B (Methodological)* 57, no. 1 (1995): 289–300.
- Benjamini, Y. 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, no. 1 (2000): 60–83.
- Benjamini, Y. and D. Yekutieli. "The Control of the False Discovery Rate in Multiple Testing under Dependency". *Annals of Statistics* 29, no. 4 (2001): 1165–88.
- Berman, J. I., M. S. Berger, P. Mukherjee, and R. G. Henry. "Diffusion Tensor Imaging-Guided Tracking of Fibers of the Pyramidal Tract Combined with Intraoperative Cortical Stimulation Mapping in Patients with Gliomas". *Journal of Neurosurgery* 101, no. 1 (2004): 66–72.

- Bi, C., S. Takahashi, and I. Fujishiro. “Interpolating 3D Diffusion Tensors in 2D Planar Domain by Locating Degenerate Lines”. *Advances in Visual Computing* 6543 (2004): 328–37.
- Blamire, A. M., S. Ogawa, K. Ugurbil, D. Rothman, G. McCarthy, J. M. Ellermann, F. Hyder, Z. Rattner, and R. G. Shulman. “Dynamic Mapping of the Human Visual Cortex by High-Speed Magnetic Resonance Imaging”. *Proceedings of the National Academy of Sciences* 89, no. 22 (1992): 11069–73.
- Broadley, A. J. M., A. Korszun, C. J. H. Jones, and M. P. Frenneaux. “Arterial Endothelial Function Is Impaired in Treated Depression”. *Heart* 88, no. 5 (1995): 521–23.
- Brooks, S. P. and A. Gelman. “General Methods for Monitoring Convergence of Iterative Simulations”. *Journal of Computational and Graphical Statistics* 7, no. 4 (1997): 434–55.
- Buckner, R. L., W. Koutstaal, D. L. Schacter, A. M. Dale, M. Rotte, and B. R. Rosen. “Functional-Anatomic Study of Episodic Retrieval. II. Selective Averaging of Event-Related fMRI Trials to Test the Retrieval Success Hypothesis”. *Neuroimage* 7, no. 3 (1998): 163–75.
- Cao, J. and K. Worsley. “The Geometry of Correlation Fields with an Application to Functional Connectivity of the Brain”. *Annals of Applied Probability* 9 (1999): 1021–57.
- CDC. “*MMWR: Current Depression Among Adults—United States, 2006 and 2008*”. Atlanta, GA: Centers for Disease control and Prevention, 2010.
- Cecchi, G. A., A. R. Rao, M. V. Centeno, M. Baliki, A. V. Apkarian, and D. R. Chialvo. “Identifying Directed Links in Large Scale Functional Networks: Application to Brain fMRI”. *BMC Cell Biology* 8, Suppl. 1 (2007): S5. doi:10.1186/1471-2121-8-S1-S5.

- Chuang, K., M. Chiu, C. Lin, and J. Chen. “Model-Free Functional MRI Analysis using Kohonen Clustering Neural Network and Fuzzy C-Means”. *IEEE Trans Med Imaging* 18, no. 12 (1999): 1117–28.
- Coccaro, E. F., M. S. McCloskey, D. A. Fitzgerald, and K. L. Phan. “Amygdala and Orbitofrontal Reactivity to Social Threat in Individuals with Impulsive Aggression”. *Biological Psychiatry* 62, no. 2 (2007): 168–78.
- Colcombe, S. J., A. F. Kramer, K. I. Erickson, and P. Scalf. “The Implications of Cortical Recruitment and Brain Morphology for Individual Differences in Inhibitory Function in Aging Humans”. *Psychology of Aging* 20, no. 3 (2005): 363–75.
- Conturo, T. E., 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, no. 3 (1996): 399–12.
- Cook, I. A., A. F. Leuchter, M. L. Morgan, E. W. Conlee, S. David, R. Lufkin, A. Babaie, J. J. Dunkin, R. O’Hara, S. Simon, A. Lightner, S. Thomas, D. Broumandi, N. Badjatia, L. Mickes, R. K. Mody, S. Arora, Z. Zheng, M. Abrams, and S. Rosenberg-Thompson. “Cognitive and Physiologic Correlates of Subclinical Structural Brain Disease in Elderly Healthy Control Subjects”. *Archives of Neurology* 59, no. 10 (2002): 1612–20.
- Cortes, C. and V. Vapnik. “Support Vector Networks”. *Machine Learning* 20, no. 3 (1995): 273–97.
- Dalby, R. B., J. Frandsen, M. M. Chakravarty, J. Ahdidan, L. Sorensen, R. Rosenberg, P. Videbech, and L. Ostergaard. “Depression Severity Is Correlated to the Integrity of

- White Matter Fiber Tracts in Late-Onset Major Depression”. *Psychiatry Research* 184, no. 1 (2010): 38–48. doi:10.1016/j.psychresns.2010.06.008.
- Damien, P., J. Wakefield, and S. Walker. “Gibbs Sampling for Bayesian Non-Conjugate and Hierarchical Models by using Auxiliary Variable”. *Journal of the Royal Statistical Society B* 61, no. 2 (1999), 331–44. doi:10.1111/1467-9868.00179.
- Davatzikos, C., D. Shen, R. C. Gur, X. Wu, D. Liu, Y. Fan, P. Hughett, B. I. Turetsky, and R. E. Gur. “Whole-brain Morphometric Study of Schizophrenia Revealing a Spatially Complex Set of Focal Abnormalities”. *Archives of General Psychiatry* 62, no. 11 (2005): 1218–27.
- Dellaportas, P. and J. Forster. “Markov Chain Monte Carlo Model Determination for Hierarchical and Graphical Log-Linear Models”. *Biometrika* 86, no. 3 (1999): 615–33.
- Dellaportas, P., J. Forster, and I. Ntzoufras. “On Bayesian Model and Variable Selection using MCMC”. *Statistics and Computing* 12, no. 1 (2002): 27–36.
- Dellaportas, P. and A. Smith. “Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling”. *Journal of the Royal Statistical Society* 42, no. 3 (1993), 443–60.
- Demonet, J. F., C. Price, R. Wise, and R. Frackowiak. “Differential Activation of Right and Left Posterior Sylvian Regions by Semantic and Phonological Tasks: A Positron-Emission Tomography Study in Normal Human Subjects”. *Neuroscience Letter* 182, no. 1 (1994): 25–28.
- Efron, B.. “Correlation and Large-Scale Simultaneous Significance Testing”. *Journal of the American Statistical Association* 102, no. 477 (2007): 93–103. doi:10.1198/016214506000001211.

- Eggermont, J. J. and G. M. Smith. “Neural Connectivity Only Accounts for a Small Part of Neural Correlation in Auditory Cortex”. *Experimental Brain Research* 110, no. 3 (1996): 379–91.
- Fan, J., X. Han, and W. Gu. “Estimating False Discovery Proportion under Arbitrary Covariance Dependence”. *Journal of the American Statistical Association* 107, no. 499 (2012): 1019–35. doi:10.1080/01621459.2012.720478.
- Firbank, M. J., A. Teodorczuk, W. M. van der Flier, A. A. Gouw, A. Wallin, T. Erkinjuntti, D. Inzitari, L. O. Wahlund, L. Pantoni, A. Poggesi, G. Pracucci, P. Langhorne, and J. T. O’Brien. “Relationship between Progression of Brain White Matter Changes and Late-Life Depression: 3-year Results from the LADIS Study”. *British Journal of Psychiatry* 201, no. 1 (2012): 40–45. doi:10.1192/bjp.bp.111.098897.
- Fox, M. D., A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen, and M. E. Raichle. “The Human Brain Is Intrinsically Organized into Dynamic, Anticorrelated Functional Networks”. *Proceedings of the National Academy of Sciences* 102, no. 27 (2005): 9673–78. doi:10.1073/pnas.0504136102.
- Friston, K. J., C. D. Frith, and R. S. J. Frackowiak. “Time-Dependent Changes in Effective Connectivity Measured with PET”. *Human Brain Mapping* 1, no. 1 (1993): 69–80. doi:10.1002/hbm.460010108.
- Friston, K. J., C. D. Frith, P. F. Liddle, and R. S. J. Frackowiak. “Functional Connectivity: The Principal-Component Analysis of Large (PET) Data Sets”. *Journal of Cerebral Blood Flow and Metabolism* 13, no. 1 (1993): 5–14.
- Friston, K. J., A. P. Holmes, K. J. Worsley, J. P. Poline, C. D. Frith, and R. S. J. Frackowiak. “Statistical Parametric Maps in Functional Imaging: A General Linear Approach”. *Human Brain Mapping* 2, no. 4 (1995): 189–210. doi:10.1002/hbm.460020402.

- Friston, K. J., P. Jezzard, and R. Turner. “Analysis of Functional MRI Time-Series”. *Human Brain Mapping* 1, no. 2 (1994): 153–71. doi:10.1002/hbm.460010207.
- Fujikawa, T., S. Yamawaki, and Y. Touhouda. “Incidence of Silent Cerebral Infarction in Patients with Major Depression”. *Stroke* 24, no. 11 (1993): 1631–34.
- Fujikawa, T., S. Yamawaki, and Y. Touhouda. “Background Factors and Clinical Symptoms of Major Depression with Silent Cerebral Infarction”. *Stroke* 25, no. 4 (1994): 798–801.
- GadElkarim, J., O. Ajilore, D. Schonfeld, L. Zhan, P. Thompson, J. Feusner, L. L. Kumar, A. and Altshuler, and A. Leow. “Investigating Brain Community Structure Abnormalities in Bipolar Disorder using Path Length Associated Community Estimation”. *Human Brain Mapping* Epub (2013): doi:10.1002/hbm.22324.
- GadElkarim, J., D. Schonfeld, O. Ajilore, L. Zhan, A. Zhang, J. Feusner, P. Thompson, T. Simon, A. Kumar, and A. Leow. “A Framework for Quantifying Node-Level Community Structure Group Differences in Brain Connectivity Networks”. *Medical Image Computing and Computer-Assisted Intervention* 15, no. 2 (2012): 196–203.
- Gamerman, D. and H. Lopes. “*Markov Chain Monte Carlo*”. New York: Chapman and Hall, 2006.
- Gelfand, A., S. Hills, A. Racine-Poon, and A. Smith. “Illustration of Bayesian Inference in Normal Data Models using Gibbs Sampling”. *Journal of the American Statistical Association* 85, no. 412 (1990): 972–85.
- Gelfand, A. and A. Smith. “Sampling-Based Approaches to Calculation Marginal Densities”. *Journal of the American Statistical Association* 85, no. 410 (1990): 398–409.

- Gelfand, A. E., A. Smith, and T. M. Lee. “Bayesian Analysis of Constrained Parameter and Truncated Data Problems using Gibbs Sampling”. *Journal of the American Statistical Association* 87, no. 418 (1992): 523–32.
- Gelman, A. and D. B. Rubin. “Inference from Iterative Simulation using Multiple Sequences”. *Statistical Science* 7, no. 4 (1992): 457–72.
- Geman, S., E. Bienenstock, and R. Doursat. “Neural Networks and the Bias/Variance Dilemma”. *Neural Computation* 4, no. 1 (1992): 1–58. doi:10.1162/neco.1992.4.1.1.
- Geman, S. and D. Geman. “Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, no. 6 (1984): 721–41.
- Ghosh, P. and W. Tu. “Assessing Sexual Attitudes and Behaviors of Young Women: A Joint Model with Nonlinear Time Effects, Time Varying Covariates, and Dropouts”. *Journal of the American Statistical Association* 103, no. 484 (2008): 1496–1507. doi:10.1198/016214508000000850.
- Gilks, W., S. Richardson, and D. Spiegelhalter. “*Markov Chain Monte Carlo in Practice*”. Suffolk, UK: Chapman and Hall, 1996.
- Gilks, W. and P. Wild. “Adaptive Rejection Sampling for Gibbs Sampling”. *Journal of the Royal Statistical Society* 41, no. 2 (1992), 337–48.
- Givens, G. H. and J. A. Hoeting. “*Computational Statistics*”. New Jersey: Wiley, 2005.
- Godfrey, P. S. A., B. K. Toone, M. W. Carney, T. G. Flynn, T. Bottiglieri, M. Laundry, I. Chanarin, and E. H. Reynolds. “Enhancement of Recovery from Psychiatric Illness by Methylfolate”. *Lancet* 336, no. 8712 (1990): 392–95.

- Gong, G., Y. He, and A. C. Evans. “Brain Connectivity: Gender Makes a Difference”. *The Neuroscientist* 17, no. 5 (2011): 575–91. doi:10.1177/1073858410386492.
- Green, P.. “Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination”. *Biometrika* 82, no. 4 (1995): 711–32. doi:10.1093/biomet/82.4.711.
- Greicius, M. D., B. H. Flores, V. Menon, G. H. Glover, H. B. Solvason, H. 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, no. 5 (2007): 429–37.
- Greicius, M. D., 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, no. 1 (2003): 253–58. doi:10.1073/pnas.0135058100.
- Gunning-Dixon, F. M. and N. Raz. “The Cognitive Correlates of White Matter Abnormalities in Normal Aging: A Quantitative Review”. *Neuropsychology* 14, no. 2 (2000): 224–32.
- Guye, M., G. J. M. Parker, M. Symms, P. Boulby, C. A. M. Wheeler-Kingshott, A. Salek-Haddadi, G. J. Barker, and J. S. Duncan. “Combined Functional MRI and Tractography to Demonstrate the Connectivity of the Human Primary Motor Cortex in vivo”. *NeuroImage* 19, no. 4 (1997): 1349–60.
- Hamilton, M.. “A Rating Scale for Depression”. *Journal of Neurology and Neurosurgical Psychiatry* 23, no. 1 (1960): 56–62.

- Hampson, M. C. A., I. R. Olson, H. C. Leung, P. Skudlarski, and J. C. Gore. “Changes in Functional Connectivity of Human MT/V5 with Visual Motion Input”. *NeuroReport* 15, no. 8 (2004): 1315–19.
- Hans, C., A. Dobra, and M. West. “Shotgun Stochastic Search for Large p Regression”. *Journal of the American Statistical Association* 102, no. 478 (2007): 507–16. doi:10.1198/016214507000000121.
- Hastie, T., R. Tibshirani, and J. H. Friedman. “*The Elements of Statistical Learning: Data Mining, Inference, and Prediction: With 200 Full-Color Illustrations*”. New York: Springer, 2003.
- Hastings, W.. “Monte Carlo Sampling Methods using Markov Chains and Their Applications”. *Biometrika* 57, no. 1 (1970): 97–109.
- Hedeker, D. and R. D. Gibbons. “*Longitudinal Data Analysis*”. Hoboken, NJ: Wiley, 2006.
- Higdon, D.. “Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications”. *Journal of the American Statistical Association* 93, no. 442 (1998): 585–95.
- Honey, C. J., R. Kotter, M. Breakspear, and O. Sporns. “Network Structure of Cerebral Cortex Shapes Functional Connectivity on Multiple Time Scales”. *Proceedings of the National Academy of Sciences* 104, no. 24 (2007): 10240–45. doi:10.1073/pnas.0701519104.
- Honey, C. J., 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, no. 6 (2009): 2035–40. doi:10.1073/pnas.0811168106.

- Hosseinifard, B., M. H. Moradi, and R. Rostami. “Classifying Depression Patients and Normal Subjects using Machine Learning Techniques”. *Conference: 19th Iranian Conference on Electrical Engineering* 109, no. 3 (2011): 1–4. doi:10.1016/j.cmpb.2012.10.008.
- Hubbard, R. A., L. Y. T. Inoue, and P. Diehr. “Joint Modeling of Self-Rated Health and Changes in Physical Functioning”. *Journal of the American Statistical Association* 104, no. 487 (2009): 912–28. doi:10.1198/jasa.2009.ap08423.
- Inoue, L. Y. T., R. Etzioni, C. Morrell, and P. Muller. “Modeling Disease Progression With Longitudinal Markers”. *Journal of the American Statistical Association* 103, no. 481 (2008): 259–70. doi:10.1198/016214507000000356.
- Institute of Medicine. “*The Mental Health and Substance Use Workforce for Older Adults: In Whose Hands?*”. Washington, DC: The National Academies Press, 2012.
- Jellison, B. J., 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, no. 3 (2004): 356–69.
- Johansen-Berg, H., T. E. J. Behrens, M. D. Robson, I. Drobnjak, M. F. S. Rushworth, J. M. Brady, S. M. Smith, D. J. Higham, and P. M. Matthews. “Changes in Connectivity Profiles Define Functionally Distinct Regions in Human Medial Frontal Cortex”. *Proceedings of the National Academy of Sciences* 101, no. 36 (2004): 13335–40. doi:10.1073/pnas.0403743101.
- Joshi, J., A. Dhall, R. Goecke, M. Breakspear, and G. Parker. “Neural-Net Classification For Spatio-Temporal Descriptor Based Depression Analysis”. *Proceedings of the 3rd ACM conference on multimedia retrieval* (2013): 352–28.

- Kales, H. C., D. F. Maixner, and A. M. Mellow. “Cerebrovascular Disease and Late-Life Depression”. *American Journal of Geriatric Psychiatry* 13, no. 2 (2005), 88–98.
- Ke, M., X. Huang, H. Shen, Z. Zhou, X. Chen, and D. Hu. “Combined Analysis for Resting State FMRI and DTI Data Reveals Abnormal Development of Function-Structure in Early-Onset Schizophrenia”. *Lecture Notes in Artificial Intelligence* 5009 (2008): 628–35.
- Kenny, E. R., J. T. O’Brien, D. A. Cousins, J. Richardson, A. J. Thomas, M. J. Firbank, and A. M. Blamire. “Functional Connectivity in Late-Life Depression using Resting-State Functional Magnetic Resonance Imaging”. *American Journal of Geriatric Psychiatry* 18, no. 7 (2010), 643–51. doi:10.1097/JGP.0b013e3181cabd0e.
- Kivimaki, M., M. J. Shipley, C. L. Allan, C. E. Sexton, M. Jokela, M. Virtanen, H. Tiemeier, K. P. Ebmeier, and A. Singh-Manoux. “Vascular Risk Status as a Predictor of Later-Life Depressive Symptoms: A Cohort Study”. *Biological Psychiatry* 72, no. 4 (2012): 324–30. doi:10.1016/j.biopsych.2012.02.005.
- Klöppel, S., C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack, J. Ashburner, and R. S. Frackowiak. “Automatic Classification of MR Scans in Alzheimer’s Disease”. *Brain* 131, no. 3 (2008): 681–89. doi:10.1093/brain/awm319.
- Koch, M. A., D. G. Norris, and M. Hund-Georgiadis. “An Investigation of Functional and Anatomical Connectivity using Magnetic Resonance Imaging”. *NeuroImage* 16, no. 1 (2002): 241–50.
- Koski, L. and T. Paus. “Functional Connectivity of the Anterior Cingulate Cortex within the Human Frontal Lobe: A brain-Mapping Meta-Analysis”. *Experimental Brain Research* 133, no. 1 (2000): 55–65.

- Krings, T., M. H. Reinges, R. Thiex, J. M. Gilsbach, and A. Thron. “Functional and Diffusion-Weighted Magnetic Resonance Images of Space-Occupying Lesions Affecting the Motor System: Imaging the Motor Cortex and Pyramidal Tracts”. *Journal of Neurosurgery* 95, no. 5 (2001): 816–24.
- Krishnan, K. R., J. C. Hays, and D. G. Blazer. “MRI-Defined Vascular Depression”. *American Journal of Psychiatry* 154, no. 4 (1997): 497–501.
- Kwong, K. K., J. W. Belliveau, D. A. Chesler, I. E. Goldberg, R. M. Weisskoff, and B. P. Poncelet. “Dynamic Magnetic Resonance Imaging of Human Brain Activity during Primary Sensory Stimulation”. *Proceedings of the National Academy of Sciences* 89, no. 12 (1992): 5675–79. doi:10.1073/pnas.89.12.5675.
- Lee, S. P., T. Q. Duong, G. Yang, C. Iadecola, and S. G. Kim. “Relative Changes of Cerebral Arterial and Venous Blood Volumes during Increased Cerebral Blood Flow: Implications for BOLD fMRI”. *Magnetic Resonance in Medicine* 45, no. 5 (2001): 791–800.
- Leow, A., O. Ajilore, L. Zhan, D. Arienzo, J. GadElkarim, A. Zhang, T. Moody, J. Van Horn, J. Feusner, and A. Kumar. “Impaired Inter-Hemispheric Integration in Bipolar Disorder Revealed with Brain Network Analyses”. *Biological Psychiatry* 73, no. 2 (2013): 183–93. doi:10.1016/j.biopsych.2012.09.014.
- Li, K., 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, no. 2 (2009), 131–39. doi:10.1016/j.compmedimag.2008.10.011.
- Lopez-Larson, M. P., J. S. Anderson, M. A. Ferguson, and D. Yurgelun-Todd. “Local Brain Connectivity and Associations with Gender and Age”. *Developmental Cognitive Neuroscience* 1, no. 2 (2011): 187–97. doi:10.1016/j.dcn.2010.10.001.

- Lundervold, A., N. Duta, T. Taxt, and A. K. Jain. “Model-Guided Segmentation of Corpus Callosum in MR Images”. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1 (1999): 1231–36.
- Lunn, D., A. Thomas, N. Best, and D. Spiegelhalter. “WinBUGS—A Bayesian Modelling Framework: Concepts, Structure, and Extensibility”. *Statistics and Computing* 10, no. 4 (2000): 325–37.
- Ma, L., B. Wang, X. Chen, and J. Xiong. “Detecting Functional Connectivity in the Resting Brain: A Comparison between ICA and CCA”. *Magnetic Resonance Imaging* 25, no. 1 (2007): 47–56.
- Makris, N., J. W. Meyer, J. F. Bates, E. H. Yeterian, D. N. Kennedy, and V. S. Caviness. “MRI-Based Topographic Parcellation of Human Cerebral White Matter and Nuclei. II. Rationale and Applications with Systematics of Cerebral Connectivity”. *Neuroimage* 9, no. 1 (1999): 18–45.
- Makris, N., A. J. Worth, A. G. Sorensen, G. M. Papadimitriou, O. Wu, T. G. Reese, V. J. Wedeen, T. L. Davis, J. W. Stakes, V. S. Caviness, E. Kaplan, B. R. Rosen, D. N. Pandya, and D. N. Kennedy. “Morphometry of in vivo Human White Matter Association Pathways with Diffusion-Weighted Magnetic Resonance Imaging”. *Annals of Neurology* 42, no. 6 (1997): 951–62.
- Martinetz, T., S. Berkovich, and K. Schulten. “Neural Gas Network for Vector Quantization and Its Application to Time-Series Prediction”. *IEEE Transactions on Neural Networks* 4 (1993): 558–69. doi:10.1109/72.238311.
- Mast, B. T., T. Miles, B. W. Penninx, K. Yaffe, C. Rosano, S. Satterfield, H. N. Ayonayon, T. Harris, and E. M. Simonsick. “Vascular Disease and Future Risk of Depressive Symptomatology in Older Adults: Findings from the Health, Aging,

- and Body Composition study”. *Biological Psychiatry* 64, no. 4 (2008): 320–26. doi:10.1016/j.biopsych.2008.01.025.
- McCloskey, M. S., K. L. Phan, and E. F. Coccaro. “Neuroimaging and Personality Disorders”. *Current Psychiatry Reports* 7, no. 1 (2005): 65–72.
- McInerney, T. and D. Terzopoulos. “Deformable Models in Medical Images Analysis: A Survey”. *Medical Image Analysis* 1, no. 2 (1996): 91–108.
- Merller, J.. “Perfect Simulation of Conditionally Specified Models”. *Journal of the Royal Statistical Society B* 61, no. 1 (1999), 251–64. doi:10.1111/1467-9868.00175.
- Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. “Equations of State Calculations by Fast Computing Machine”. *Journal of Chemical Physics* 21, no. 6 (1953): 1087–92.
- Mettenburg, J. M., T. L. Benzinger, J. S. Shimony, A. Z. Snyder, and Y. I. Sheline. “Diminished Performance on Neuropsychological Testing in Late Life Depression Is Correlated with Microstructural White Matter Abnormalities”. *Neuroimage* 60, no. 4 (2012): 2182–90. doi:10.1016/j.neuroimage.2012.02.044.
- Meyer, J. W., N. Makris, J. F. Bates, V. S. Caviness, and D. N. Kennedy. “MRI-Based Topographic Parcellation of Human Cerebral White Matter”. *Neuroimage* 9, no. 1 (1999): 1–17.
- Miezin, F. M., L. Maccotta, J. M. Ollinger, S. E. Petersen, and R. L. Buckner. “Characterizing the Hemodynamic Response: Effects of Presentation Rate, Sampling Procedure, and the Possibility of Ordering Brain Activity Based on Relative Timing”. *Neuroimaging* 11, no. 6 (2000): 735–59.

- Mori, S. and P. C. M. van Zijl. “Fiber Tracking: Principles and Strategies—A Technical Review”. *NMR in Biomedicine* 15, no. 7–8 (2002): 468–80.
- Mourao-Miranda, J., L. Oliveira, C. D. Ladouceur, A. Marquand, M. Brammer, B. Birmaher, D. Axelson, and M. L. Phillips. “Pattern Recognition and Functional Neuroimaging Help to Discriminate Healthy Adolescents at Risk for Mood Disorders from Low Risk Adolescents”. *PLoS One* 7, no. 2 (2012): e29482. doi:10.1371/journal.pone.0029482.
- Muirhead, R. J.. “*Aspects of Multivariate Statistical Theory*”. Hoboken, NJ: Wiley, 1982.
- Munk, M. H., D. E. L. Linden, L. Muckli, H. Lanfermann, F. E. Zanella, and W. Singer. “Distributed Cortical Systems in Visual Short-Term Memory Revealed by Event-Related Functional Magnetic Resonance Imaging”. *Cerebral Cortex* 12, no. 8 (2002): 866–76.
- Musselman, D. L., D. L. Evans, and C. B. Nemeroff. “The Relationship of Depression to Cardiovascular Disease”. *Archives of General Psychiatry* 55, no. 7 (1998): 580–92.
- NARSAD. “Late Life Depression” (2013). <http://bbrfoundation.org/userFiles/facts.latelifedep.pdf>.
- Neal, R.. “Slice Sampling”. *The Annals of Statistics* 31, no. 3 (2003): 705–1031.
- Ngan, S. and X. Hu. “Analysis of FMRI Imaging Data using Self-Organizing Mapping with Spatial Connectivity”. *Magnetic Resonance in Medicine* 41, no. 5 (1991): 939–46.
- Ntzoufras. “*Bayesian Modeling using WinBUGS*”. Hoboken, NJ: Wiley, 2009.
- O’Doherty, J., M. L. Kringelbach, E. T. Rolls, J. Hornak, and C. Andrews. “Abstract Reward and Punishment Representations in the Human Orbitofrontal Cortex”. *Nature Neuroscience* 4 (2001): 95–102. doi:10.1038/82959.

- O'Dwyer, L., F. Lamberton, A. L. W. Bokde, M. Ewers, and Y. O. Faluyi. "Using Support Vector Machines with Multiple Indices of Diffusion for Automated Classification of Mild Cognitive Impairment". *PLoS ONE* 7, no. 2 (2012): e32441. doi:10.1371/journal.pone.0032441.
- Ogawa, S., 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, no. 24 (1990): 9968–72.
- Ogawa, S., T. M. Lee, A. S. Nayak, and P. Glynn. "Oxygenation-Sensitive Contrast in Magnetic Resonance Image of Rodent Brain at High Magnetic Fields". *Magnetic Resonance in Medicine* 14, no. 1 (1990): 68–78.
- Ogawa, S., R. S. Menon, S. G. Kim, and K. Ugurbil. "On the Characteristics of Functional Magnetic Resonance Imaging of the Brain". *Annual Review of Biophysics and Biomolecular Structure* 27 (1998): 447–74. doi:10.1146/annurev.biophys.27.1.447.
- Ogawa, S., D. W. Tank, R. Menon, J. M. Ellermann, S. G. Kim, H. Merkle, and K. Ugurbil. "Intrinsic Signal Changes Accompanying Sensory Stimulation: Functional Brain Mapping with Magnetic Resonance Imaging". *Proceedings of the National Academy of Sciences* 89, no. 13 (1992): 5951–55.
- Pape, T. L., A. Guernon, S. Lundgren, V. Patil, A. A. Herrold, B. Smith, M. Blahnik, L. M. Picon, B. Harton, M. Peterson, T. Mallinson, and M. Hoffmann. "Predicting Levels of Independence with Expressing Needs and Ideas 1 Year after Severe Brain Injury". *Rehabilitation Psychology* 58, no. 3 (2013): 253–62. doi:10.1037/a0032610.
- Pape, T. L., J. Rosenow, and G. Lewis. "Transcranial Magnetic Stimulation: A Possible Treatment for TBI". *Journal of Head Trauma Rehabilitation* 21, no. 5 (2006), 437–51.

- Patil, V. K., J. R. St Andre, E. Crisan, B. M. Smith, C. T. Evans, M. L. Steiner, and P. T. L.. “Prevalence and Treatment of Headaches in Veterans with Mild Traumatic Brain Injury”. *Headache* 51, no. 7 (2011): 1112–21. doi:10.1111/j.1526-4610.2011.01946.x.
- Pauling, L. and C. D. Coryell. “The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin”. *Proceedings of the National Academy of Sciences* 22, no. 4 (1936): 210–16.
- Pfurtscheller, G. and C. Andrew. “Event-Related EEG/MEG Synchronization and Desynchronization: Basic Principles”. *Clinical Neurophysiology* 110, no. 11 (1999): 1842–57.
- Phan, K. L., A. Orlichenko, E. Boyd, M. Angstadt, E. F. Coccaro, I. Liberzon, and K. Arfanakis. “Preliminary Evidence of White Matter Abnormality in the Uncinate Fasciculus in Generalized Social Anxiety Disorder”. *Biological Psychiatry* 66, no. 7 (2009): 691–94. doi:10.1016/j.biopsych.2009.02.028.
- Pierpaoli, C. and P. J. Basser. “Toward a Quantitative Assessment of Diffusion Anisotropy [published erratum]”. *Magnetic Resonance in Medicine* 36, no. 6 (1996): 893–906. doi:10.1002/mrm.1910360612.
- Propp, J. and D. Wilson. “Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics”. *Random Structures and Algorithms* 9, no. 1–2 (1996): 223–52. doi:10.1002/(SICI)1098-2418.
- Robert, C. and G. Casella. “*Monte Carlo Statistical Methods*”. New York: Springer-verlag, 2004.
- Rubinov, M., O. Sporns. “Complex Network Measures of Brain Connectivity: Uses and Interpretations”. *Neuroimage* 50, no. 3 (2010): 1059–69.

- Rykhlevskaia, E., M. Fabiani, and G. Gratton. “Lagged Covariance Structure Models for Studying Functional Connectivity in the Brain”. *Neuroimage* 30, no. 4 (2006): 1203–18.
- Rykhlevskaia, E., G. Gratton, and M. Fabiani. “Combining Structural and Functional Neuroimaging Data for Studying Brain Connectivity: A Review”. *Psychophysiology* 45, no. 2 (2008): 173–87.
- Saad, Z. S., K. M. Ropella, R. W. Cox, and E. A. DeYoe. “Analysis and Use of fMRI Response Delays”. *Human Brain Mapping* 13, no. 2 (2001): 74–93.
- Schnitzler, A. and J. Gross. “Functional Connectivity Analysis in Magnetoencephalography”. *International Review of Neurobiology* 68 (2005): 173–95.
- Schoenfeld, M. A., T. Noesselt, D. Poppel, C. Tempelmann, J. M. Hopf, M. G. Woldorff, H. J. Heinze, and S. A. Hillyard. “Analysis of Pathways Mediating Preserved Vision after Striate Cortex Lesions”. *Annals of Neurology* 52, no. 6 (2002): 814–24.
- Scholkopf, B. and A. J. Smola. *“Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond”*. Cambridge, MA: MIT Press, 2001.
- Seghier, M. L., F. Lazeyras, S. Zimine, S. E. Maier, S. Hanquinet, J. Delavelle, J. J. Volpe, and P. S. Huppi. “Combination of Event-Related fMRI and Diffusion Tensor Imaging in an Infant with Perinatal Stroke”. *NeuroImage* 21, no. 1 (2004): 463–72.
- Sheline, Y. I., C. F. Pieper, D. M. Barch, K. Welsh-Bohmer, R. C. McKinstry, J. R. Macfall, G. D’Angelo, K. S. Garcia, K. Gersing, C. Wilkins, W. Taylor, D. Steffens, R. R. Krishnan, and P. M. Doraiswamy. “Support for the Vascular Depression Hypothesis in Late-Life Depression: Results of a 2-Site, Prospective, Antidepressant Treatment Trial”. *Archives of General Psychiatry* 67, no. 3 (2010): 277–85. doi:10.1001/archgenpsychiatry.2009.204.

- Sheline, Y. I., J. L. Price, S. N. Vaishnavi, M. A. Mintun, D. M. Barch, A. A. Epstein, C. H. Wilkins, A. Z. Snyder, L. Couture, K. Schechtman, and R. C. McKinstry. “Regional White Matter Hyperintensity Burden in Automated Segmentation Distinguishes Late-Life Depressed Subjects from Comparison Subjects Matched for Vascular Risk Factors”. *American Journal of Psychiatry* 165, no. 4 (2008): 524–32. doi:10.1176/appi.ajp.2007.07010175.
- Sisson, S.. “Trans-Dimensional Markov Chains: A Decade of Progress and Future Perspectives”. *Journal of the American Statistical Association* 100, no. 471 (2005): 1077–89. doi:10.1198/016214505000000664.
- Sitholea, J. S. and P. W. Jones. “Bivariate Longitudinal Model for Detecting Prescribing Change in Two Drugs Simultaneously with Correlated Errors”. *Bernoulli* 2, no. 3 (2007): 145–65. doi:10.1080/02664760601005020.
- Smith, S. M., M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. M. Brady, and P. M. Matthews. “Advances in Functional and Structural MR Image Analysis and Implementation as FSL”. *NeuroImage* 23, no. S1 (2004): S208–S219.
- Sporns, O., G. Tononi, and G. M. Edelman. “Theoretical Neuroanatomy: Relating Anatomical and Functional Connectivity in Graphs and Cortical Connection Matrices”. *Cerebral Cortex* 10, no. 2 (2000): 127–41.
- Stein, T., C. Moritz, M. Quigley, D. Cordes, V. Haughton, and E. Meyerand. “Functional Connectivity in the Thalamus and Hippocampus Studied with Functional MR Imaging”. *American Journal of Neuroradiology* 21, no. 8 (2000), 1397–1401.

- Sui, J., T. Adali, Q. Yu, and V. D. Calhoun. “A Review of Multivariate Methods for Multimodal Fusion of Brain Imaging Data”. *Journal of Neuroscience Methods* 204, no. 1 (2012): 68–81. doi:10.1016/j.jneumeth.2011.10.031.
- Sun, F. T., L. M. Miller, and M. D’Esposito. “Measuring Interregional Functional Connectivity using Coherence and Partial Coherence Analyses of fMRI Data”. *Neuroimage* 21, no. 2 (2004): 647–58.
- Swick, D. and V. Ashley. “Left Inferior Frontal Gyrus Is Critical for Response Inhibition”. *BMC Neuroscience* 9 (2008): 102. doi:10.1186/1471-2202-9-102.
- Tanner, M. and W. Wong. “The Calculation of the Posterior Distributions by Data Augmentation”. *Journal of the American Statistical Association* 82, no. 398 (1987): 528–49.
- Tao, H., S. Guo, T. Ge, K. M. Kendrick, Z. Xue, Z. Liu, and J. Feng. “Depression Uncouples Brain Hate Circuit”. *Molecular Psychiatry* 18, no. 1 (2013): 101–11. doi:10.1038/mp.2011.127.
- Taylor, W. D., J. R. Macfall, G. Gerig, and R. R. Krishnan. “Structural Integrity of the Uncinate Fasciculus in Geriatric Depression: Relationship with Age of Onset”. *Journal of Neuropsychiatric Disease and Treatment* 3, no. 5 (2007): 669–74.
- Toosy, A. T., O. Ciccarelli, G. J. M. Parker, C. A. M. Wheeler-Kingshott, D. H. Miller, and A. J. Thompson. “Characterizing Function-Structure Relationships in the Human Visual System with Functional MRI and Diffusion Tensor Imaging”. *NeuroImage* 21, no. 4 (2004): 1452–63.
- Tuch, D. S.. “Q-Ball Imaging”. *Magnetic Resonance in Medicine* 52, no. 6 (2004): 1358–72.

- Tuch, D. S., T. G. Reese, M. R. Wiegell, N. Markris, J. W. Belliveau, and V. J. Wedeen. "High Angular Resolution Diffusion Imaging Reveals Intravoxel White Matter Fiber Heterogeneity". *Magnetic Resonance in Medicine* 48, no. 4 (2002): 577–82.
- Vapnik, V. and A. Chervonenkis. "*Theory of Pattern Recognition [in Russian]*". Moscow, Russia: Nauka, 1974.
- Venkataraman, A., Y. Rathi, M. Kubicki, C. Westin, and P. Golland. "Joint Modeling of Anatomical and Functional Connectivity for Population Studies". *IEEE Transactions of Medical Imaging* 31, no. 2 (2012), 164–82. doi:10.1109/TMI.2011.2166083.
- Vincent, G. K. and V. A. Velkoff. "*The Next Four Decades, The Older Population in the United States: 2010 to 2050, Current Population Reports*". Washington, DC: U.S. Census Bureau, 2010.
- Watanabe, T. and K. Tanaka. "Cognitive Neuroscience: Editorial Overview". *Current Opinion in Neurobiology* 17 (2007): 1–3.
- Werring, D. J., C. A. Clark, G. J. M. Parker, D. H. Miller, A. J. Thompson, and G. J. Barker. "A Direct Demonstration of Both Structure and Function in the Visual System: Combining Diffusion Tensor Imaging with Functional Magnetic Resonance Imaging". *NeuroImage* 9, no. 3 (1999): 352–61.
- Westin, C. F., S. E. Maier, H. Mamata, A. Nabavi, F. A. Jolesz, and R. Kikinis. "Processing and Visualization of Diffusion Tensor MRI". *Medical Image Analysis* 6, no. 2 (1981): 93–108. doi:10.1016/S1361–8415(02)00053–1.
- Whitfield-Gabrieli, S. and A. Nieto-Castanon. "CONN: A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks". *Brain Connectivity* 2, no. 3 (2012): 125–41. doi:10.1089/brain.2012.0073.

- Worsley, K. J., J. I. Chen, J. Lerch, and A. C. Evans. “Comparing Functional Connectivity via Thresholding Correlations and Singular Value Decomposition”. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, no. 1457 (2005): 913–20.
- Wu, M., C. Andreescu, M. A. Butters, R. Tamburo, C. F. I. Reynolds, and H. Aizenstein. “Default-Mode Network Connectivity and White Matter Burden in Late-Life Depression”. *Psychiatry Research* 194, no. 1 (2011): 39–46. doi:10.1016/j.psychresns.2011.04.003.
- Zhang, A., 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, no. 4 (2012): 959–67. doi:10.1038/npp.2011.279.
- Zhou, D., W. K. Thompson, and G. Siegle. “MATLAB Toolbox for Functional Connectivity”. *Neuroimage* 47, no. 4 (2009): 1590–1607. doi:10.1016/j.neuroimage.2009.05.089.
- Zhou, Y., N. Shu, Y. Liu, M. Song, Y. Hao, H. Liu, C. Yu, Z. Liu, and T. Jiang. “Altered Resting-State Functional Connectivity and Anatomical Connectivity of Hippocampus in Schizophrenia”. *Schizophrenia Research* 100, no. 1–3 (2008): 120–32. doi:10.1016/j.schres.2007.11.039.

APPENDICES

APPENDIX A

Maximum Marginal Likelihood Estimation

The maximum likelihood estimation in the linear mixed-effects model is based on the marginal distribution obtained by integrating over the distribution of the random effects. Shown in the following are the detailed steps adapted from Hedeker and Gibbons (2006) for estimating fixed and variance parameters.

The marginal distribution for the i th link and j th subject can be written as

$$h(\mathbf{y}_{ij}) = \int_{\boldsymbol{\gamma}} f(\mathbf{y}_{ij}|\boldsymbol{\gamma}_j; \boldsymbol{\beta}_i, \Sigma_i) g(\boldsymbol{\gamma}_j|\Sigma_{\boldsymbol{\gamma}}) d\boldsymbol{\gamma} \quad (48)$$

where

$$\begin{aligned} f(\mathbf{y}_{ij}|\boldsymbol{\gamma}_j; \boldsymbol{\beta}_i, \Sigma_i) &= (2\pi)^{-1} |\Sigma_i|^{-\frac{1}{2}} \\ &\quad \exp\left[-\frac{1}{2}(\mathbf{y}_{ij} - \mathbf{X}_{ij}\boldsymbol{\beta}_i - \mathbf{Z}_{ij}\boldsymbol{\gamma}_j)^T \Sigma_i^{-1} (\mathbf{y}_{ij} - \mathbf{X}_{ij}\boldsymbol{\beta}_i - \mathbf{Z}_{ij}\boldsymbol{\gamma}_j)\right] \\ g(\boldsymbol{\gamma}_j|\Sigma_{\boldsymbol{\gamma}}) &= (2\pi)^{-1} |\Sigma_{\boldsymbol{\gamma}}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\gamma}_j^T \Sigma_{\boldsymbol{\gamma}}^{-1} \boldsymbol{\gamma}_j\right) \end{aligned} \quad (49)$$

The marginal log-likelihoods need to be summed over all subjects for the i th link

$$\log L_i = \sum_{j=1}^n \log h(\mathbf{y}_{ij}) \quad (50)$$

We need to solve for $\boldsymbol{\beta}_i$ and Σ_i that maximize the above function, as well as $\Sigma_{\boldsymbol{\gamma}}$ that maximizes the sum of the above functions for all links.

APPENDIX A (continued)

Denote $\log L_i = \sum_{j=1}^n \log h_{ij} = \sum_{j=1}^n \log[\int_{\gamma} (f_{ij} \cdot g) d\gamma]$. Let $p_{ij} = f_{ij} \cdot g / h_{ij}$, we have, for the fixed parameter β_i

$$\begin{aligned}
 \frac{\partial \log L_i}{\partial \beta_i} &= \sum_{j=1}^n \frac{\partial \log h_{ij}}{\partial \beta_i} \\
 &= \sum_{j=1}^n \frac{1}{h_{ij}} \frac{\partial [\int_{\gamma} (f_{ij} \cdot g) d\gamma]}{\partial \beta_i} \\
 &= \sum_{j=1}^n \frac{1}{h_{ij}} \int_{\gamma} \frac{\partial f_{ij}}{\partial \beta_i} \cdot g d\gamma \\
 &= \sum_{j=1}^n \int_{\gamma} \frac{f_{ij} \cdot g}{h_{ij}} \frac{\partial \log f_{ij}}{\partial \beta_i} d\gamma \\
 &= \sum_{j=1}^n \int_{\gamma} p_{ij} \mathbf{X}_{ij}^T \Sigma_i^{-1} (\mathbf{y}_{ij} - \mathbf{X}_{ij} \beta_i - \mathbf{Z}_{ij} \gamma_j) d\gamma \\
 &= \sum_{j=1}^n \mathbf{X}_{ij}^T \Sigma_i^{-1} (\mathbf{y}_{ij} - \mathbf{X}_{ij} \beta_i - \mathbf{Z}_{ij} \tilde{\gamma}_j)
 \end{aligned} \tag{51}$$

since $\int_{\gamma} \gamma_j p_{ij} d\gamma = \tilde{\gamma}_j$ and $\int_{\gamma} p_{ij} d\gamma = 1$. Equating the above expression to zero yields

$$\sum_{j=1}^n \mathbf{X}_{ij}^T \Sigma_i^{-1} \mathbf{X}_{ij} \beta_i = \sum_{j=1}^n \mathbf{X}_{ij}^T \Sigma_i^{-1} (\mathbf{y}_{ij} - \mathbf{Z}_{ij} \tilde{\gamma}_j) \tag{52}$$

Hence,

$$\hat{\beta}_i = \left(\sum_{j=1}^n \mathbf{X}_{ij}^T \Sigma_i^{-1} \mathbf{X}_{ij} \right)^{-1} \left[\sum_{j=1}^n \mathbf{X}_{ij}^T \Sigma_i^{-1} (\mathbf{y}_{ij} - \mathbf{Z}_{ij} \tilde{\gamma}_j) \right] \tag{53}$$

For simplicity reasons, we derive the estimates of the residual and random effect variances based on a univariate mixed-effects model. This derivation is adapted from similar expressions in Hedeker and Gibbons (2006). For the residual variance, with $\varepsilon_j = \mathbf{y}_j - \mathbf{X}_j \beta - \mathbf{Z}_j \tilde{\gamma}_j$, we have

APPENDIX A (continued)

$$\begin{aligned}
\frac{\partial \log L}{\partial \sigma^2} &= \sum_{j=1}^N \frac{\partial \log h_j}{\partial \sigma^2} \\
&= \sum_{j=1}^N \frac{1}{h_j} \frac{\partial [\int_{\gamma} (f_j \cdot g) d\gamma]}{\partial \sigma^2} \\
&= \sum_{j=1}^N \frac{1}{h_j} \int_{\gamma} \frac{\partial f_j}{\partial \sigma^2} \cdot g d\gamma \\
&= \sum_{j=1}^N \int_{\gamma} \frac{f_j \cdot g}{h_j} \frac{\partial \log f_j}{\partial \sigma^2} d\gamma \\
&= \sum_{j=1}^N \int_{\gamma} p_j \left[-\frac{n_j}{2} \sigma^{-2} + \frac{1}{2} \sigma^{-4} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta} - \mathbf{Z}_j \boldsymbol{\gamma})^T (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta} - \mathbf{Z}_j \boldsymbol{\gamma}) \right] d\gamma \\
&= \frac{1}{2} \sigma^{-4} \sum_{j=1}^N \left[-n_j \sigma^2 + \boldsymbol{\epsilon}_j^T \boldsymbol{\epsilon}_j + \text{tr}(\mathbf{Z}_j \boldsymbol{\Sigma}_{\gamma|y_j} \mathbf{Z}_j^T) \right]
\end{aligned} \tag{54}$$

since

$$\begin{aligned}
&\int_{\gamma} p_j (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta} - \mathbf{Z}_j \boldsymbol{\gamma})^T (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta} - \mathbf{Z}_j \boldsymbol{\gamma}) d\gamma \\
&= \int_{\gamma} p_j [(\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta} - \mathbf{Z}_j \tilde{\boldsymbol{\gamma}}_j) - \mathbf{Z}_j (\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}_j)]^T [(\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta} - \mathbf{Z}_j \tilde{\boldsymbol{\gamma}}_j) - \mathbf{Z}_j (\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}_j)] d\gamma \\
&= (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta} - \mathbf{Z}_j \tilde{\boldsymbol{\gamma}}_j)^T (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta} - \mathbf{Z}_j \tilde{\boldsymbol{\gamma}}_j) + \int_{\gamma} p_j [\mathbf{Z}_j (\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}_j)]^T [\mathbf{Z}_j (\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}_j)] d\gamma \\
&= \boldsymbol{\epsilon}_j^T \boldsymbol{\epsilon}_j + \text{tr}(\mathbf{Z}_j \boldsymbol{\Sigma}_{\gamma|y_j} \mathbf{Z}_j^T)
\end{aligned} \tag{55}$$

Setting the above expression to zero yields

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{j=1}^N \left[e_j^T e_j + \text{tr}(\mathbf{Z}_j \boldsymbol{\Sigma}_{\gamma|y_j} \mathbf{Z}_j^T) \right] \tag{56}$$

Or, for a random intercept only model

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{j=1}^N \left[e_j^T e_j + n_i \sigma_{\gamma|y_i}^2 \right] \tag{57}$$

APPENDIX A (continued)

The total error variance is essentially the expected variance computed from the estimated residuals, plus the EB estimates of the uncertainty about the residuals. The error variance estimate in the case of a bivariate model can be easily adapted from the above estimate. Hence, for our bivariate model, we have

$$\begin{aligned}\hat{\Sigma}_{0i} &= \frac{1}{n_0} \sum_{j=1}^{n_0} (\hat{\epsilon}_{0ij} \hat{\epsilon}_{0ij}^T + \mathbf{Z}_{0ij} \tilde{\Sigma} \boldsymbol{\gamma}_{|\mathbf{y}_j} \mathbf{Z}_{0ij}^T) \\ \hat{\Sigma}_{1i} &= \frac{1}{n_1} \sum_{j=1}^{n_1} (\hat{\epsilon}_{1ij} \hat{\epsilon}_{1ij}^T + \mathbf{Z}_{1ij} \tilde{\Sigma} \boldsymbol{\gamma}_{|\mathbf{y}_j} \mathbf{Z}_{1ij}^T)\end{aligned}\tag{58}$$

where n_0, n_1 are the sample sizes and $\mathbf{Z}_{0ij}, \mathbf{Z}_{1ij}$ are random effect design matrices for the control and experimental groups, respectively.

We apply the same idea in deriving the variance estimates for the random effects, that the total random effect variance is the expected variance computed from the estimated random effects, plus the EB estimates of the uncertainty about the random effect. Also see Hedeker and Gibbons (2006), 78, for the expression of the variance estimate. Hence, for our bivariate model, we have

$$\hat{\Sigma}_{\boldsymbol{\gamma}} = \frac{1}{n} \sum_{j=1}^n \left(\tilde{\boldsymbol{\gamma}}_j \tilde{\boldsymbol{\gamma}}_j^T + \tilde{\Sigma} \boldsymbol{\gamma}_{|\mathbf{y}_j} \right)\tag{59}$$

APPENDIX B

R Code for Estimating the Bivariate Linear Mixed-Effects Model

```
# Initializing parameters
Beta <- matrix(0, numRegComb, 4)
Beta.var <- matrix(0, numRegComb, 16)
Beta.se <- matrix(0, numRegComb, 4)
Random <- matrix(0, (numSubj0 + numSubj1), 2)
Sigma <- matrix(0, numRegComb, 6)
Sigma.ran <- c(0.005, 0, 0.02)
sigma.ran.y <- array(0, c(2, 2, (numSubj0 + numSubj1)))
for (iSubj in 1:(numSubj0 + numSubj1)) sigma.ran.y[, , iSubj] <- diag(c(0.005, 0.02))

tol <- 1e-5 # Tolerance limit for convergence checking
diff <- rep(FALSE, numRegComb*4+(numSubj0 + numSubj1)*2+numRegComb*6+4) # Convergence criteria

iter <- 1
while (iter < 100 && sum(diff) < length(diff)) {

  print(paste("Iteration", iter))

  # Update ML Estimates of Beta, Sigma, and Sigma.ran
  sigma.ran.mat <- t(Random)%*%Random / (numSubj0 + numSubj1)
    + apply(sigma.ran.y, c(1, 2), mean)
  Sigma.ran.new <- mattovec(sigma.ran.mat)
  Beta.new <- matrix(0, numRegComb, 4)
  Sigma.new <- matrix(0, numRegComb, 6)
  sigma0.inv <- matrix(0, 2*numRegComb, 2*numRegComb)
  sigma1.inv <- matrix(0, 2*numRegComb, 2*numRegComb)
  for (iRegComb in 1:numRegComb) {
    temp <- ImagingData[ImagingData$R1_R2==R1_R2[iRegComb], ]
    y <- Y[ImagingData$R1_R2==R1_R2[iRegComb]]
    x <- X[ImagingData$R1_R2==R1_R2[iRegComb], ]
    z <- Z[ImagingData$R1_R2==R1_R2[iRegComb], ]
    error0 <- matrix((y-x%*%Beta[iRegComb, ]-as.numeric(Random))[temp$Group==0],
      ncol = 2)
    error1 <- matrix((y-x%*%Beta[iRegComb, ]-as.numeric(Random))[temp$Group==1],
      ncol = 2)
```

APPENDIX B (continued)

```

sigma.mat0 <- (t(error0)%*%error0 + apply(sigma.ran.y[, , GroupSubj==0], c(1, 2),
      sum)) / numSubj0
sigma.mat1 <- (t(error1)%*%error1 + apply(sigma.ran.y[, , GroupSubj==1], c(1, 2),
      sum)) / numSubj1
Sigma.new[iRegComb, ] <- c(mattovec(sigma.mat0), mattovec(sigma.mat1))
sigma0 <- c(Sigma.new[iRegComb, 3], -Sigma.new[iRegComb, 2], Sigma.new[iRegComb, 1])
      /(Sigma.new[iRegComb, 1]*Sigma.new[iRegComb, 3]
      - Sigma.new[iRegComb, 2]^2)
sigma1 <- c(Sigma.new[iRegComb, 6], -Sigma.new[iRegComb, 5], Sigma.new[iRegComb, 4])
      /(Sigma.new[iRegComb, 4]*Sigma.new[iRegComb, 6]
      - Sigma.new[iRegComb, 5]^2)

# Define Error Covariance Matrix
sigma0.inv[iRegComb, iRegComb] <- sigma0[1]
sigma0.inv[iRegComb+numRegComb, iRegComb+numRegComb] <- sigma0[3]
sigma0.inv[iRegComb, iRegComb+numRegComb] <- sigma0[2]
      sigma0.inv[iRegComb+numRegComb, iRegComb] <- sigma0[2]
sigma1.inv[iRegComb, iRegComb] <- sigma1[1]
sigma1.inv[iRegComb+numRegComb, iRegComb+numRegComb] <- sigma1[3]
sigma1.inv[iRegComb, iRegComb+numRegComb] <- sigma1[2]
      sigma1.inv[iRegComb+numRegComb, iRegComb] <- sigma1[2]
sigma.e <- matrix(0, (numSubj0+numSubj1)*2, (numSubj0+numSubj1)*2)
for (iSubj in 1:(numSubj0+numSubj1)) {
  sigma.e[iSubj, iSubj] <- Sigma.new[iRegComb, 1]*(1-GroupSubj[iSubj])
      + Sigma.new[iRegComb, 4]*GroupSubj[iSubj]
  sigma.e[((numSubj0 + numSubj1)+iSubj), ((numSubj0 + numSubj1)+iSubj)] <-
      Sigma.new[iRegComb, 3]*(1-GroupSubj[iSubj])
      + Sigma.new[iRegComb, 6]*GroupSubj[iSubj]
  sigma.e[iSubj, ((numSubj0 + numSubj1)+iSubj)] <-
      Sigma.new[iRegComb, 2]*(1-GroupSubj[iSubj])
      + Sigma.new[iRegComb, 5]*GroupSubj[iSubj]
  sigma.e[((numSubj0 + numSubj1)+iSubj), iSubj] <-
      sigma.e[iSubj, ((numSubj0 + numSubj1)+iSubj)]
}
Beta.new[iRegComb, ] <- solve(t(x)%*%sigma.e%*%x)
      %*%t(x)%*%sigma.e%*%(y-as.numeric(Random))
Beta.var[iRegComb, ] <- as.numeric(solve(t(x)%*%solve(sigma.e
      + z%*%sigma.ran.mat%*%t(z))%*%x))

```

APPENDIX B (continued)

```

Beta.se[iRegComb, ] <- sqrt(diag(solve(t(x)%%solve(sigma.e
+ z%%sigma.ran.mat%%t(z))%%x)))
}

# Update EB Estimates and Variance of Random Effects
Random.new <- matrix(0, (numSubj0 + numSubj1), 2)
for (iSubj in 1:(numSubj0 + numSubj1)) {
  temp <- ImagingData[ImagingData$ID==iSubj, ]
  y <- Y[ImagingData$ID==iSubj]
  x <- X[ImagingData$ID==iSubj, ]
  z <- Z[ImagingData$ID==iSubj, ]
  if (unique(temp$Group) == 0) r <- vectomat(Sigma.ran.new)%%
    solve(vectomat(Sigma.ran.new) + solve(t(z)%%sigma0.inv%%z))
  else r <- vectomat(Sigma.ran.new)%%
    solve(vectomat(Sigma.ran.new)+solve(t(z)%%sigma1.inv%%z))
  Random.new[iSubj, ] <- r%%solve(t(z)%%z)%%t(z)%%
    (y-diag(x%%t(Beta.new[rep(1:numRegComb, 2), ])))
  sigma.ran.y[, , iSubj] <- (diag(2)-r)%%vectomat(Sigma.ran.new)
}

diff <- abs(c(as.numeric(Beta.new - Beta), as.numeric(Random.new - Random),
  as.numeric(Sigma.new - Sigma), as.numeric(Sigma.ran.new - Sigma.ran))) < tol
print(paste(sum(diff), "out_of", length(diff), "parameters_converged"))
Beta <- Beta.new
Random <- Random.new
Sigma <- Sigma.new
Sigma.ran <- Sigma.ran.new
iter <- iter + 1
}

```

APPENDIX C

R Code for Simulating Bivariate Neuroimaging Data

```
library(MASS)
library(MCMCpack)

vectomat <- function(vec) matrix(vec[c(1,2,2,3)], nrow = 2)
mattovec <- function(mat) as.vector(mat)[c(1, 2, 4)]

numSubj0 <- 10
numSubj1 <- 10
numRegion <- 40
numSubj <- numSubj0 + numSubj1
GroupSubj <- c(rep(0, numSubj0), rep(1, numSubj1))
numRegComb <- numRegion*(numRegion-1)/2
numSim <- 1000

R1 <- numeric(numRegComb)
R2 <- numeric(numRegComb)
iRow <- 1
for (iR1 in 1:(numRegion-1)) {
  for (iR2 in (iR1+1):numRegion) {
    R1[iRow] <- iR1; R2[iRow] <- iR2
    iRow <- iRow + 1
  }
}

r1 <- as.character(R1)
r1[R1<10] <- paste("0", r1[R1<10], sep = "")
r2 <- as.character(R2)
r2[R2<10] <- paste("0", r2[R2<10], sep = "")
R1.R2 <- paste(r1, r2, sep = "_")

DataStruc <- data.frame(cbind(rep(rep(1:numSubj, each = numRegComb), 2),
  rep(rep(0:1, numRegComb*c(numSubj0, numSubj1)), 2),
  rep(R1, numSubj*2),
  rep(R2, numSubj*2),
  rep(0, numRegComb*numSubj*2),
  rep(c("fMRI", "DTI"), each = numRegComb*numSubj),
```

APPENDIX C (continued)

```
rep(R1.R2, numSubj*2)), stringsAsFactors = FALSE)
names(DataStruc) <- c("ID", "Group", "R1", "R2", "Value", "Measure", "R1.R2")

mean.beta <- c(0, 2, 0.5, 3)
var.beta <- c(0.05, 0.5, 0.05, 1)
sig.gamma <- matrix(c(0.05, 0.05, 0.05, 0.5), nrow = 2)

SimResult <- vector("list", numSim)

set.seed(728)
for (iSim in 1:numSim) {

  print(paste("Simulation", iSim, "out_of", numSim))
  ImagingData <- DataStruc
  beta <- mvrnorm(numRegComb, mean.beta, diag(var.beta))
  rho0 <- runif(numRegComb, -1, 0.5)
  rho1 <- runif(numRegComb, -0.5, 1)
  sig01 <- rgamma(numRegComb, 0.2, 1)
  sig02 <- rgamma(numRegComb, 1, 1)
  sig11 <- rgamma(numRegComb, 0.3, 1)
  sig12 <- rgamma(numRegComb, 1.5, 1)
  gamma <- mvrnorm((numSubj), c(0, 0), sig.gamma)

  for (iRegComb in 1:numRegComb) {
    emat0 <- matrix(c(sig01[iRegComb], (rho0*sqrt(sig01*sig02))[iRegComb],
      (rho0*sqrt(sig01*sig02))[iRegComb], sig02[iRegComb]), nrow = 2)
    emat1 <- matrix(c(sig11[iRegComb], (rho1*sqrt(sig11*sig12))[iRegComb],
      (rho1*sqrt(sig11*sig12))[iRegComb], sig12[iRegComb]), nrow = 2)
    ImagingData[ImagingData$R1.R2==R1.R2[iRegComb] & ImagingData$Group==0, 5]
      <- as.numeric(mvrnorm(numSubj0, beta[iRegComb, 1:2], emat0)
        + gamma[1:numSubj0, ])
    ImagingData[ImagingData$R1.R2==R1.R2[iRegComb] & ImagingData$Group==1, 5]
      <- as.numeric(mvrnorm(numSubj0, beta[iRegComb, 3:4], emat1)
        + gamma[1:numSubj1+numSubj0, ])
  }

  Y <- as.numeric(ImagingData$Value)
```

APPENDIX C (continued)

```
X <- cbind(as.numeric(ImagingData$Measure=="fMRI" & ImagingData$Group==0),
           as.numeric(ImagingData$Measure=="DTI" & ImagingData$Group==0),
           as.numeric(ImagingData$Measure=="fMRI" & ImagingData$Group==1),
           as.numeric(ImagingData$Measure=="DTI" & ImagingData$Group==1))
Z <- cbind(as.numeric(factor(ImagingData$Measure)) - 1,
           2 - as.numeric(factor(ImagingData$Measure)))

source("Model_Estimate.R")
Sigma[, 2] <- Sigma[, 2]/sqrt(Sigma[, 1]*Sigma[, 3])
Sigma[, 5] <- Sigma[, 5]/sqrt(Sigma[, 4]*Sigma[, 6])
SimResult[[iSim]] <- list(Beta = Beta, Sigma = Sigma, Random = Random, Sigma.ran = Sigma.ran)
}
```

APPENDIX D

WinBUGS Code for the Bayesian Linear Mixed-Effects Model

```
model {

  for (i in 1:m) { # RegComb

    for (j in 1:n) { # Subj

      Y[i, j, 1:2] ~ dmnorm(mu[i, j, ], tau[i, j, , ])
      mu[i, j, 1] <- Beta[i, 1]*(1-Group[j]) + Beta[i, 3]*Group[j] + gamma[j, 1]
      mu[i, j, 2] <- Beta[i, 2]*(1-Group[j]) + Beta[i, 4]*Group[j] + gamma[j, 2]

      tau[i, j, 1, 1] <- tau0[i, 1, 1]*(1-Group[j]) + tau1[i, 1, 1]*Group[j]
      tau[i, j, 1, 2] <- tau0[i, 1, 2]*(1-Group[j]) + tau1[i, 1, 2]*Group[j]
      tau[i, j, 2, 1] <- tau0[i, 2, 1]*(1-Group[j]) + tau1[i, 2, 1]*Group[j]
      tau[i, j, 2, 2] <- tau0[i, 2, 2]*(1-Group[j]) + tau1[i, 2, 2]*Group[j]

    }

    Beta[i, 1:4] ~ dmnorm(mu.beta[ ], tau.beta[ , ])
    Beta.diff[i, 1] <- Beta[i, 1] - Beta[i, 3]
    Beta.diff[i, 2] <- Beta[i, 2] - Beta[i, 4]
    tau0[i, 1:2, 1:2] ~ dwish(I[ , ], 13)
    sigma0[i, 1:2, 1:2] <- inverse(tau0[i, , ] )
    tau1[i, 1:2, 1:2] ~ dwish(I[ , ], 10)
    sigma1[i, 1:2, 1:2] <- inverse(tau1[i, , ] )
  }

  for (j in 1:n) {
    gamma[j, 1:2] ~ dmnorm(mu.gamma[ ], tau.gamma[ , ])
  }

  tau.gamma[1:2, 1:2] ~ dwish(I[ , ], 23)
  sigma.gamma[1:2, 1:2] <- inverse(tau.gamma[ , ] )
}
```

APPENDIX D (continued)

```
I[1, 1] <- 1
I[1, 2] <- 0
I[2, 1] <- 0
I[2, 2] <- 1

}
```

VITA

NAME: Weihan Zhao

EDUCATION: BS, Biochemistry, Nanjing University, Nanjing, China, 2002

PhD, Biochemistry and Molecular Genetics, College of Medicine,
University of Illinois at Chicago, 2009

MS, Biostatistics, School of Public Health, University of Illinois
at Chicago, 2010

PhD, Biostatistics, School of Public Health, University of Illinois
at Chicago, 2014

TEACHING: Teaching Assistant, BCHE 307 Introduction to Biochemistry, Col-
lege of Medicine, University of Illinois at Chicago, 2004

Teaching Assistant, BSTT 400 Biostatistics, School of Public
Health, University of Illinois at Chicago, 2008–2009

Weighted random-effects regression models, BSTT 562 Linear
Models, School of Public Health, University of Illinois at Chicago,
2012

RESEARCH: Research Assistant, Institute of Health Research and Policy, Uni-
versity of Illinois at Chicago, 2009–2011

Research Assistant, Department of Psychiatry, College of
Medicine, University of Illinois at Chicago, 2012–present

- PUBLICATIONS: Zhao, W., T. L. Chen, B. M. Vertel, and K. J. Colley. "The CMP-Sialic Acid Transporter is Localized in the Medial-Trans Golgi and Possesses Two Specific Endoplasmic Reticulum Export Motifs in its Carboxyl-Terminal Cytoplasmic Tail." *The Journal of Biological Chemistry* 281, no. 41 (2006): 31106–18.
- Zhao, W., and K. J. Colley. "Nucleotide Sugar Transporters." In *The Golgi Apparatus: State of the Art 110 Years after Camillo Golgi's Discovery*, edited by Alexander A. Mironov and Margit Pavelka, 190-206. New York: SpringerWien NewYork , 2008.
- Sheean, P. M., S. J. Peterson, W. Zhao, D. P. Gurka, and C. A. Braunschweig. "Intensive Medical Nutrition Therapy: Methods to Improve Nutrition Provision in the Critical Care Setting." *Journal of the Academy of Nutrition and Dietetics* 112, no. 7 (2012): 1073–79.
- Touchette, D., S. Rao, P. Dhru, W. Zhao, Y. K. Choi, I. Bhandari, and G. Stettin. "Identification of and Intervention to Address Therapeutic Gaps in Care." *The American Journal of Managed Care* 18, no. 10 (2012): e364–71.
- Guidotti, A., E. Dong, D. P. Gavin, M. Veldic, W. Zhao, D. K. Bhaumik, S. C. Pandey, and D. R. Grayson. "DNA Methylation/Demethylation Network Expression in Psychotic Patients with a History of Alcohol Abuse." *Alcoholism: Clinical and Experimental Research* 37, no. 3 (2013): 417–24.