Statistics Methodologies for Neuroconnectivity Analysis Using fMRI Data in Autism

by

Fei Jie
Bachelor of Medicine (Shanghai Medical University, Shanghai, China) 1990
PhD (Iowa State University, Ames, Iowa) 2005
MS (Iowa State University, Ames, Iowa) 2005

Thesis submitted in 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, 2016

Chicago, Illinois

Defense Committee:
Dulal Bhaumik, PhD, Professor of Biostatistics and Psychiatry, UIC, Advisor and Chair
Sanjib Basu, PhD, Professor and Director of the Division of Statistics, NIU
Sally Freels, PhD, Associate Professor of Biostatistics, UIC
Scott Langenecker, PhD, Associate Professor of Psychiatry and Psychology, UIC
Theresa Pape, DrPH, Deputy ACOS & Clinical Neuroscientist, Edward Hines Jr. VAH
To my family
ACKNOWLEDGMENTS

As I am approaching a major milestone in my life, I am very glad to finally have a chance to thank many people who have accompanied me on this exciting journey.

My deepest gratitude goes to my advisor, Dr. Dulal Bhaumik, for his guidance and encouragement throughout my thesis work. I have particularly enjoyed his insightful advice that stimulated my intellectual curiosity. I also wanted to thank my dissertation committee members, Drs. Sanjib Basu, Sally Freels, Scott Langenecker and Theresa Pape, for taking time to critically review my thesis and provide comments. Dr. Bikas Sinha from India Statistical Institute provided some important comments which helped improve the quality of this dissertation. I am grateful to Dr. Runa Bhaumik for her assistance in MATLAB and ABIDE data pre-processing. I wanted to thank my fellow graduate students, Weihan Zhao and Yuanbo Song, for helpful discussions.

I am extremely grateful to my family, especially my wife, Jun, for unconditional love, patience and sacrifice in difficult times. And I want my kids, Christine and Albert, to know 'where there is a will, there is a way'.

To Mom and Dad, thank you for having confidence in me throughout these years.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Autism</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Functional Magnetic Resonance Imaging</td>
<td>3</td>
</tr>
<tr>
<td>1.2.1 Principles of Functional Magnetic Resonance Imaging</td>
<td>3</td>
</tr>
<tr>
<td>1.2.2 Functional Connectivity and Resting State Functional Magnetic Resonance Imaging</td>
<td>4</td>
</tr>
<tr>
<td>1.2.3 Functional Magnetic Resonance Imaging and Neurological Disorders</td>
<td>6</td>
</tr>
<tr>
<td>1.3 Neuroimaging Analysis in Autism</td>
<td>7</td>
</tr>
<tr>
<td>1.3.1 Structural Changes in Autism</td>
<td>7</td>
</tr>
<tr>
<td>1.3.2 Functional Changes in Autism</td>
<td>8</td>
</tr>
<tr>
<td>1.3.2.1 Task-specific Functional Changes in Autism</td>
<td>9</td>
</tr>
<tr>
<td>1.3.2.2 Resting-state Functional Changes in Autism</td>
<td>11</td>
</tr>
<tr>
<td>2 DATASET</td>
<td>15</td>
</tr>
<tr>
<td>2.1 Autism Brain Imaging Data Exchange</td>
<td>15</td>
</tr>
<tr>
<td>3 LINEAR MIXED EFFECT MODEL</td>
<td>22</td>
</tr>
<tr>
<td>3.1 Model Specification</td>
<td>22</td>
</tr>
<tr>
<td>3.2 Parameter Estimation</td>
<td>24</td>
</tr>
<tr>
<td>3.2.1 Empirical Bayes Estimation</td>
<td>24</td>
</tr>
<tr>
<td>3.2.2 Maximum (Marginal) Likelihood Estimation</td>
<td>29</td>
</tr>
<tr>
<td>3.3 Simulation</td>
<td>33</td>
</tr>
<tr>
<td>3.3.1 Bias</td>
<td>34</td>
</tr>
<tr>
<td>3.3.2 Standardized bias</td>
<td>34</td>
</tr>
<tr>
<td>3.3.3 Root mean squared error</td>
<td>34</td>
</tr>
<tr>
<td>3.4 Hypothesis Testing</td>
<td>35</td>
</tr>
<tr>
<td>3.5 Overview of False Discovery Rate</td>
<td>37</td>
</tr>
<tr>
<td>3.6 Data Analysis Using Mixed Effect Model</td>
<td>43</td>
</tr>
<tr>
<td>3.6.1 Transformation of Data</td>
<td>43</td>
</tr>
<tr>
<td>3.6.2 Mixed Effect Analysis on Data from Each Site</td>
<td>47</td>
</tr>
<tr>
<td>3.6.3 Mixed Effect Analysis on Pooled Data</td>
<td>48</td>
</tr>
<tr>
<td>4 PRINCIPAL FACTOR APPROXIMATION</td>
<td>54</td>
</tr>
<tr>
<td>4.1 Overview of Principal Factor Approximation</td>
<td>55</td>
</tr>
<tr>
<td>4.2 Proof of Proposition 2</td>
<td>58</td>
</tr>
<tr>
<td>4.3 Application to Data Analysis</td>
<td>78</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.3.1 Principal Factor Approximation Analysis with Empirical Variance</td>
<td>78</td>
</tr>
<tr>
<td>4.3.2 Principal Factor Approximation Analysis with Random Subject Effect Removed</td>
<td>80</td>
</tr>
<tr>
<td>4.3.3 Principal Factor Approximation Analysis Assuming Independence</td>
<td>81</td>
</tr>
<tr>
<td>5 BAYESIAN ANALYSIS</td>
<td>84</td>
</tr>
<tr>
<td>5.1 Bayesian Hierarchical Model</td>
<td>85</td>
</tr>
<tr>
<td>5.2 Implementation of Hierarchical Bayesian</td>
<td>89</td>
</tr>
<tr>
<td>5.2.1 Markov chain Monte Carlo</td>
<td>89</td>
</tr>
<tr>
<td>5.2.2 Metropolis-Hasting Algorithm</td>
<td>91</td>
</tr>
<tr>
<td>5.2.3 Gibbs Sampler</td>
<td>92</td>
</tr>
<tr>
<td>5.2.4 WinBUGS</td>
<td>93</td>
</tr>
<tr>
<td>5.2.5 R2WinBUGS</td>
<td>94</td>
</tr>
<tr>
<td>5.3 Data Analysis Using Bayesian Hierarchical Model</td>
<td>94</td>
</tr>
<tr>
<td>5.3.1 Model Specification</td>
<td>94</td>
</tr>
<tr>
<td>5.3.2 Model Update and Diagnosis</td>
<td>96</td>
</tr>
<tr>
<td>5.3.3 Analysis Results</td>
<td>98</td>
</tr>
<tr>
<td>5.4 Result Comparison: Hierarchical Bayesian vs Empirical Bayesian</td>
<td>100</td>
</tr>
<tr>
<td>6 META ANALYSIS</td>
<td>104</td>
</tr>
<tr>
<td>6.1 Vote Counting</td>
<td>105</td>
</tr>
<tr>
<td>6.2 Overview of Meta Analysis</td>
<td>106</td>
</tr>
<tr>
<td>6.3 Meta Analysis Applied to Study Data</td>
<td>110</td>
</tr>
<tr>
<td>6.4 Meta Analysis Results</td>
<td>112</td>
</tr>
<tr>
<td>6.5 Meta Analysis Discussion</td>
<td>115</td>
</tr>
<tr>
<td>7 SUMMARY AND CONCLUSION</td>
<td>119</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>127</td>
</tr>
<tr>
<td>Appendix A</td>
<td>128</td>
</tr>
<tr>
<td>Appendix B</td>
<td>132</td>
</tr>
<tr>
<td>Appendix C</td>
<td>136</td>
</tr>
<tr>
<td>Appendix D</td>
<td>140</td>
</tr>
<tr>
<td>CITED LITERATURE</td>
<td>144</td>
</tr>
<tr>
<td>VITA</td>
<td>152</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>16</td>
</tr>
<tr>
<td>II</td>
<td>17</td>
</tr>
<tr>
<td>III</td>
<td>35</td>
</tr>
<tr>
<td>IV</td>
<td>38</td>
</tr>
<tr>
<td>V</td>
<td>47</td>
</tr>
<tr>
<td>VI</td>
<td>49</td>
</tr>
<tr>
<td>VII</td>
<td>52</td>
</tr>
<tr>
<td>VIII</td>
<td>82</td>
</tr>
<tr>
<td>IX</td>
<td>83</td>
</tr>
<tr>
<td>X</td>
<td>83</td>
</tr>
<tr>
<td>XI</td>
<td>99</td>
</tr>
<tr>
<td>XII</td>
<td>102</td>
</tr>
<tr>
<td>XIII</td>
<td>112</td>
</tr>
<tr>
<td>XIV</td>
<td>122</td>
</tr>
<tr>
<td>XV</td>
<td>124</td>
</tr>
<tr>
<td>XVI</td>
<td>125</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I NUMBER OF SUBJECTS AT EACH SITE</td>
<td>16</td>
</tr>
<tr>
<td>II REGIONS ANALYZED</td>
<td>17</td>
</tr>
<tr>
<td>III SIMULATION TO EVALUATE E-M ALGORITHM</td>
<td>35</td>
</tr>
<tr>
<td>IV OUTCOMES AMONG A FAMILY OF TESTS</td>
<td>38</td>
</tr>
<tr>
<td>V NUMBER OF SIGNIFICANT LINKS BY SITE</td>
<td>47</td>
</tr>
<tr>
<td>VI SUMMARY OF LINK SPECIFIC MEAN FMRI</td>
<td>49</td>
</tr>
<tr>
<td>VII SIGNIFICANT LINKS FDR=0.3</td>
<td>52</td>
</tr>
<tr>
<td>VIII FALSE DISCOVERY PROPORTION-EMPIRICAL COVARIANCE</td>
<td>82</td>
</tr>
<tr>
<td>IX FDP WITH RANDOM SUBJECT EFFECT REMOVED</td>
<td>83</td>
</tr>
<tr>
<td>X FDP ASSUMING INDEPENDENCE</td>
<td>83</td>
</tr>
<tr>
<td>XI GEWEKE DIAGNOSTICS TEST</td>
<td>99</td>
</tr>
<tr>
<td>XII SIGNIFICANT LINKS BY HIERARCHICAL BAYESIAN</td>
<td>102</td>
</tr>
<tr>
<td>XIII SIGNIFICANT LINKS BY META ANALYSIS</td>
<td>112</td>
</tr>
<tr>
<td>XIV NUMBER OF SIGNIFICANT LINKS (FDR=0.3, 3 METHODS)</td>
<td>122</td>
</tr>
<tr>
<td>XV NUMBER OF SIGNIFICANT LINKS (FDR=0.3, 4 METHODS)</td>
<td>124</td>
</tr>
<tr>
<td>XVI SIGNIFICANT LINKS IDENTIFIED BY ALL 4 METHODS</td>
<td>125</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Blood-Oxygen-Level Dependent (BOLD) response to neural activity.</td>
</tr>
<tr>
<td>2</td>
<td>Resting-state fMRI and functional connectivity.</td>
</tr>
<tr>
<td>3</td>
<td>fMRI measurement for the first 2 links</td>
</tr>
<tr>
<td>4</td>
<td>fMRI measurement for the middle 2 links</td>
</tr>
<tr>
<td>5</td>
<td>fMRI measurement for the last 2 links</td>
</tr>
<tr>
<td>6</td>
<td>Histogram of Mean fMRI across all Links.</td>
</tr>
<tr>
<td>7</td>
<td>Histogram of p values from mixed analysis.</td>
</tr>
<tr>
<td>8</td>
<td>Plots produced by qvalue package</td>
</tr>
<tr>
<td>9</td>
<td>Brain network mixed analysis.</td>
</tr>
<tr>
<td>10</td>
<td>Empirical distribution of Z statistics</td>
</tr>
<tr>
<td>11</td>
<td>Histogram of p values.</td>
</tr>
<tr>
<td>12</td>
<td>The impact of threshold t</td>
</tr>
<tr>
<td>13</td>
<td>Non-hierarchical model vs hierarchical model</td>
</tr>
<tr>
<td>14</td>
<td>Diagnostic plots</td>
</tr>
<tr>
<td>15</td>
<td>Brain network Bayesian analysis.</td>
</tr>
<tr>
<td>16</td>
<td>Network hub plot.</td>
</tr>
<tr>
<td>17</td>
<td>Brain network from meta analysis.</td>
</tr>
<tr>
<td>18</td>
<td>Forest plot for mean difference at each site.</td>
</tr>
<tr>
<td>19</td>
<td>Venn plot for significant links (3 methods).</td>
</tr>
<tr>
<td>20</td>
<td>Venn plot for significant links (4 methods).</td>
</tr>
<tr>
<td>21</td>
<td>Network hub plot for significant links identified by all 4 methods.</td>
</tr>
</tbody>
</table>
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASD</td>
<td>Autism Spectrum Disorders</td>
</tr>
<tr>
<td>BGR</td>
<td>Brooks-Gelman-Rubin statistic</td>
</tr>
<tr>
<td>DMN</td>
<td>Default Mode Network</td>
</tr>
<tr>
<td>EB</td>
<td>Empirical Bayes</td>
</tr>
<tr>
<td>EM</td>
<td>Expectation-Maximization</td>
</tr>
<tr>
<td>FDP</td>
<td>False Discovery Proportion</td>
</tr>
<tr>
<td>FDR</td>
<td>False Discovery Rate</td>
</tr>
<tr>
<td>fMRI</td>
<td>Functional Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>HB</td>
<td>Hierarchical Bayes</td>
</tr>
<tr>
<td>MCMC</td>
<td>Markov Chain Monte Carlo</td>
</tr>
<tr>
<td>MML</td>
<td>Maximum Marginal Likelihood</td>
</tr>
<tr>
<td>PFA</td>
<td>Principal Factor Approximation</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>ROI</td>
<td>Region Of Interest</td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphism</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
</tbody>
</table>
SUMMARY

The human brain is an amazingly complex network. Aberrant activities in this network can lead to various neurological disorders such as multiple sclerosis, Parkinson’s disease, Alzheimer’s disease and autism.

We are particularly interested in defects in neuroconnectivity in autism. As autism is a behavior disorder, no biomarker or genetic tests have been developed. fMRI has emerged as an important tool to delineate the neural networks affected by autism.

Previous research has identified a series of impairments in neuroconnectivity in autism, leading to the under-connectivity theory. However, several recent studies provided evidence that neural network was actually over-connected, but not under-connected as previously thought, in autism. These apparent discrepancies can be attributed to the drawbacks in study design and statistical methods. First of all, most previous studies suffered from small sample size, a problem that rendered these studies to be underpowered. Secondly, these studies failed to account for two critical sources of variation: difference among subjects and variation among sites. As a result, the estimated difference between patient and control and its standard error were not accurate. This not only decreased the efficiency of analysis but also lead to biased inferences. The reason that previous studies elected to run simple t-test(and thus ignored subject and site effect) was most likely due to computational complexity in dealing with neuroimaging data.

In this work we took advantage of the largest autism database, the Autism Brain Imaging Data Exchange(ABIDE). Our analyses included 361 subjects from 8 medical centers. We
SUMMARY (Continued)

believed that the sample size in this work avoided the 'underpower' issue in most previous studies. More importantly, we were able to properly account for subject and site effect. By incorporating a random subject effect in mixed effect analysis, we explicitly partitioned total variation into three parts: difference between patient and control, different due to subject effect and random error. As we were dealing with huge datasets, no commercially available software could fit a mixed effect model using the ABIDE data. By implementing EM algorithm, which iterated between E step that used empirical Bayesian to estimate random effect and M step which evaluated fixed effect via MLE, we were able to obtain accurate estimate of difference between patient and control and its standard error. After applying false discovery rate control, we identified 12 links with significant difference between patient and control.

The second statistical methodology that we implemented in this work was Bayesian hierarchical modeling. Although empirical Bayesian in mixed effect model helped obtain parameter estimates, it suffered from several pitfalls. For example, it was accused of using data more than once. Moreover it was considered as an approximation to the fully Bayesian approach. In the work, we developed Bayesian hierarchical model, a fully Bayesian model. With the help of modern computational tools such as MCMC, we fitted this model to ABIDE data.

The FDR control approach in the above-mentioned methods failed to account for the correlation among statistical tests. As fMRI data was obtained from different regions of the same brain, it was highly unlikely that the statistical tests were independent. Principal factor approximation was recently proposed to implement FDR control in the presence of dependence structure. As the original abbreviated proof in the pioneering paper contained major mistake,
we developed a full proof. It turned out that the main conclusion in the original paper still held. We applied PFA to the ABIDE dataset and identified links with significant difference in autism.

Lastly, the three above-mentioned methods did not take into account site effect, which could be an important source of variation. We performed meta analysis by conducting separate analysis by site and then combined analysis with weight equal to the inverse of standard error. Meta analysis was able to confirm all links identified by mixed effect analysis, Bayesian hierarchical modeling and Principal factor approximation. It also identified additional significant links.

Taken together, in this work we have not only implemented several novel analyses but also derived detailed theoretical proof. The methods developed in this work can be easily applied to other analyses involving large datasets.
CHAPTER 1

INTRODUCTION

The human brain is a complex network with a large number of brain regions that are constantly communicating with each other. The disruption of brain network has been shown to be associated with many diseases, such as multiple sclerosis, Parkinson’s disease, Alzheimer’s disease and autism (1).

In this work, we will focus on statistical methodologies in the analysis of neuroconnectivity using fMRI data in autism. As autism is a severe mental disease with no biomarker or genetic diagnostic tools, resting-state fMRI is becoming an increasingly important technique in autism research. In this Chapter the basic concept of autism and fMRI will be reviewed.

1.1 Autism

Autism is a neurodevelopmental disorder that is characterized by poor social communication abilities in addition to repetitive behaviors or restricted interests. The first two cases of autism were described in the early 1940s by (2) in the United States and (3) in Austria. It is estimated that autism strikes 1-2 per 1000 people (4).

The autism disease is usually considered to include autistic disorder, Asperger disorder and pervasive development disorder not otherwise specified (PPD-NOS) according to DSM-IV, published by American Psychiatric Association (5). As autism is an etiologically and clinically diverse group of disorders, it is commonly referred to as the autism spectrum disorders (ASD).
The cause of autism has not been fully elucidated. Recent studies indicated that individual genetic makeup might be important in the development of autism. High concordance rate of autism has been observed in twins (6). In fact, the concordance rate for identical twins was 64% vs only 9% for fraternal twins. Secondly, high recurrence risk in siblings was observed in family history studies. For example, it is reported that the latter sibling recurrence risk was 8.6% (7). Moreover, for families with two or more children with autism, the recurrence risk could reach 35%. Additional evidence to support the notion that autism is genetically determined comes from genetic testing. A chromosomal or Mendelian cause or at least predisposition could be identified in 15% to 40% of ASD children (8). Given all these findings, it is generally accepted nowadays that genetic factors play a critical role in autism.

The link between vaccination and autism has been issue of hot debate over last 20 years. The first report indicated that 8 children started to develop symptoms of autism within 1 month after MMR vaccination (9). It was later discovered that all 8 of these children had gastrointestinal symptoms. Additionally, endoscopy of these children revealed lymphoid nodular hyperplasia. It was postulated that MMR vaccine might contribute to the development of autism by causing GI inflammation. As a result, usually nonpermeable peptides entered the bloodstream and, subsequently, to the brain where they induced autism. Later research disapproved the association between MMR vaccination and autism. First of all, study (9) did not have a control group hence it was impossible to determine whether autism after receipt of MMR vaccine was causal or coincidental. Secondly, several recent large scale, well controlled
studies found no evidence to support the association between autism and vaccination (10), (11), (12), (13).

1.2 Functional Magnetic Resonance Imaging

Functional magnetic resonance imaging or functional MRI (fMRI) is "a functional neuroimaging procedure using MRI technology that measures brain activity by detecting associated changes in blood flow" (Wikipedia).

1.2.1 Principles of Functional Magnetic Resonance Imaging

Functional magnetic resonance imaging relies on the measurement of blood oxygen-level dependent (BOLD) response, which is originated from the change in the amount of deoxyhemoglobin in the tissue.

Hemoglobin, an oxygen carrying protein, exists in two forms: the oxygen-bound form (oxyhemoglobin) or oxygen-not-bound form (deoxyhemoglobin). Oxygenated blood, with high level of oxyhemoglobin which is a diamagnetic protein, is not influenced by external magnetic field. On the other hand, deoxyhemoglobin has the paramagnetic property which causes a disturbance of the main magnetic field (14), (15). When brain activity increases, oxygen consumption also increases. This leads to an elevated level of deoxyhemoglobin. Due to paramagnetic feature of deoxyhemoglobin, increase in deoxyhemoglobin can reduce rather than increase BOLD signal. The observed increase in BOLD during neural activation is due to an increase in brain blood flow that over-responds to increase in deoxyhemoglobin (or decrease in oxygen). This overcompensation results in an oversupply of oxygenated blood (16), (17). The reason for the over-compensation is still under investigation.
The time course of a typical BOLD response is depicted in Figure 1. The BOLD response to neural activity can be broken down into three phases. The initial response after stimulus is a small negative response that reaches its minimum in about 2-3 seconds. This is called fast response or early dip. The main response is characterized by a big positive increase that peaks at about 5 seconds. This the focus of fMRI data analysis. The last phase of BOLD response is another negative response, the so-called post-stimulus undershot.

1.2.2 Functional Connectivity and Resting State Functional Magnetic Resonance Imaging

Functional connectivity is defined as the temporal dependence of activities of anatomically separated regions of the brain (18), (19). With regard to fMRI, functional connectivity is char-
characterized by similar activation patterns of anatomically separated regions, indicating functional communications between these regions.

The first observation that functional connectivity exists during resting state came from (20). In this study, subjects were instructed to close their eyes and refrain from thinking anything in particular but without falling into sleep. Functional magnetic resonance imaging analysis revealed that regions of the sensorimotor cortex were activated with hand movements. The authors analyzed the low frequency (< 0.1Hz) fluctuations in fMRI signal from these regions, which was considered to reflect the resting state brain function. The analysis of time courses of this fluctuation indicated a high degree of temporal correlation (P < 0.001) within these regions. This pioneering study suggested functional connectivity between left and right primary motor cortices of the brain. Since then several studies have identified additional resting state functional networks, including (21), (22) and (23). The studies to identify functional connectivity in resting state fMRI can be schematically illustrated in Figure 2 (modified from (24)).

Subjects in resting state were placed in fMRI scanner Figure 2a. In panel b, a region of interest, a seed voxel, was selected and its BOLD response was measured. To detect functional connectivity between the seed voxel (referred to as voxel i) and other region of the brain, another voxel, voxel j, was also selected. The resting state time series of seed voxel i was correlated with that of voxel j (panel c). A high level of correlation between these two time series indicated high level of functional connectivity. To explore the functional connectivity between seed voxel and
all other regions of the brain, the time series of seed voxel i was correlated with time series of all other voxels, resulting in a functional connectivity map (panel d).

1.2.3 Functional Magnetic Resonance Imaging and Neurological Disorders

In recent years, there has been a growing number of studies that explored the possible relationship between defects in functional connectivity and neurological disorders using resting state fMRI.

Functional disconnectivity has been strongly implicated in the development of Alzheimer’s disease. In a study of 19 subjects (10 Alzheimer’s and 9 control subjects), a parametric reduction in neuroconnectivity in Alzheimer’s patients was reported (25). A subsequent study
(26) reported a decreased connectivity in the default mode network (DMN) which consists of the posterior cortex, the posterolateral parietal cortices and the medial temporal lobes/hippocampi. These results were confirmed by (27), which indicated that functional connectivity defects in Alzheimer’s were not caused by structural changes.

Numerous previous studies have suggested that schizophrenia is best characterized by defects in neuroconnectivity. Recent resting state fMRI studies provided new insights into the impairment in brain network connectivity in schizophrenia. Schizophrenia patients were found to have reduced connectivity in DMN (28). Another study confirmed this finding (29).

In addition, aberrant activities in resting state brain network have also been implied in several other diseases, such as depression, dementia and ADHD. The impairment in functional connectivity seems to be a rather universal phenomenon in neurological diseases.

1.3 Neuroimaging Analysis in Autism

As autism spectrum disorders (ASD) is defined by abnormal behaviors, there is no known biological markers for its diagnosis. Neuroimaging has proven to be an important technique to study changes in structure and connectivity in ASD.

1.3.1 Structural Changes in Autism

One of the hallmarks of ASD is accelerated brain growth in early childhood. Autism children were found to have brain enlargement (2). Additional studies provided further evidence of increased total brain volume (TBV) in autism. For instance, 22 autistic subjects had increased TBV as compared with 36 normal controls in a study (30). Postmortem studies also confirmed increased TBV in autism (31). An MRI study was conducted to examine the TBV of 60 autism
patients and 52 normal subjects (32). The autism group consisted of boys of 2-4 years old and 5-15 years old. Boys in 2-4 age autism group had significantly larger brain than control, while boys of 5-15 years old had normal brain size. The results demonstrated that over-growth occurs early in brain development. The extent of TBV increase has been consistent in several studies which indicated that autism can lead to 5%-10% increase in total brain volume as compared with normal children (33), (34).

In addition to increase in TBV, both gray and white matter abnormalities have also been reported in autism. The most prominent findings in this regard is the increase in frontal lobe volume. Several independent studies have confirmed this observation (35), (33). However, inconsistencies have been identified in the location and direction (increase or decrease) of structure changes which may be due to the fact that ASD is a heterogeneous disease.

1.3.2 Functional Changes in Autism

Functional Magnetic Resonance Imaging (fMRI) is one of the functional neuro-imaging tools for ASD studies. This method played a critical role in establishing ASD as a neurological disorder. Due to its excellent contrast properties, spatial resolution, and temporal resolution, fMRI is ideally suited for autism research. In addition, fMRI is also convenient to patients since it does not need radio-labeling and is non-invasive.

As ASD represents diverse varieties of disease, there is no unifying account of brain dysfunctions. In addition, there is evidence that suggest that autism can affect different areas in different patients. For example, while most autism patients are considered as intellectually
retarded, some ASD patients have very high IQ's. Therefore brain defects in autism may be domain specific.

1.3.2.1 **Task-specific Functional Changes in Autism**

Early fMRI studies examined task-specific changes, which included social cognition and executive functions. Although autism can cause defects in many domains of the brain, studies have shown that it mainly affects social cognition. Task-specific fMRI studies have helped addressing the defects in so called "social brain".

The most studied social cognition defect in autism is face processing. In a study of face recognition (36), 9 ASD subjects (7 autism and 2 Asperger’s syndrome) and 9 typically-developing control subjects were recruited. Control subjects and ASD subjects were matched in age and IQ. On one testing section, the subjects were asked to judge emotional expression (happy versus angry). On the other section, subjects were prompted to judge gender. This study confirmed the findings in (37) in that subjects with ASD showed significantly lower level of activity in FG and greater level of activity in ITG. In another study (38), seven autism subjects and 8 typically-developing controls were matched in sex, age, and handedness. The analysis of fMRI results revealed that there were significantly less activation in FG in autism group as compared with the control. In summary, the studies in face-processing in autism indicate aberrant activation in autistic subjects with significantly less activation in brain areas typically associated in face-recognition and higher activation in alternate brain regions.

One of the most significant feature of autism is that autistic patients often exhibit repetitive interests and activities. A hypothesis to explain this phenomenon is that these repetitive
behaviors may be a result of defects in executive functions which lead to impaired behavioral control (39).

Several studies have been conducted to test this hypothesis. In one study, 12 high-functioning autistic patients and 12 age- and IQ-matched control were evaluated in a response inhibition task and an memory inhibition task (40). A decrease in ACC activation were identified in autistic patients as compared with the control. As a result, autism subjects may mis-interpret normal signals as something of concern and thus trigger repetitive behaviors. In another study, brain activation of high-functioning autism subjects were compared to control group via fMRI (41). The two groups exhibited similar behavioral results. However, the fMRI analysis indicated that autistic subjects might use visual codes to perform memory tasks as compared with control subjects who might use verbal codes.

Planning which is another aspect of executive function was examined in a study (42). Eighteen high-functioning autism subjects and 18 age and IQ-matched healthy controls were included. The brain activation of autistic subjects were compared with the control via fMRI during the performance of a Tower of London task. Although there was little difference in activated cortical areas in general, underconnectivity in autism group were identified in frontal and parietal regions.

There have also been several reports that suggest over-connectivity in ASD. For example, there was partially increased thalamocortical connectivity in ASD (43). In addition, it was reported that in autism the connection between caudate nuclei and cerebral cortex was enhanced (44).
In summary, the studies of task-specific functional changes in autism have revealed many defects in autism. The impaired social cognition in autism subjects are attributed to several regions of the brain, especially left FG and PT. In addition, autistic subjects exhibit reduced response inhibition which may lead to repetitive behavior in autism.

1.3.2.2 Resting-state Functional Changes in Autism

Task specific fMRI studies have identified many defects in ASD. However, there have been concerns that these findings are at least partially driven by activation effect which may be confounded by task performance (35). To alleviate these concerns, recent years has seen the explosive growth of task-independent fMRI studies. The findings that several regions of the brain exhibit higher activities in resting state than in active task state leads to the hypothesis of default mode network (DMN).

The first study to test DMN hypothesis is (22). Three questions are addressed in this study: "Is there a resting-state network in the human brain? How is it modulated during simple sensory processing? How is it regulated during cognitive processing?" This study revealed that PCC and vACC regions exhibited decreased connection during a cognitive (working memory) task than in rest state. The default network might include PCC coupled with vACC and several other brain regions. In addition, the authors showed functional connectivity of PCC and vACC during a visual processing task and were very similar to resting state. However, during cognitive processing the default mode network activity were decreased.

Another report provided support for default mode hypothesis (45). The authors measured brain oxygen extraction fraction (OEF), which is defined as the ratio of oxygen used by the brain
to oxygen delivered by flowing blood. They found OEF decreases in goal-directed behaviors, suggesting the existence of default brain functions which are reduced during active state.

The changes in resting state fMRI during autism have attracted enormous interest in recent years. Many earlier reports suggested widespread under-connectivity in resting state fMRI in autism.

A resting state fMRI study was conducted in 57 autism subjects and 57 age and IQ-matched controls (46). Network connectivity was found to be reduced in autism. This functional under-connectivity was most prominent in the anterior-posterior connections. This was the first report that indicated cortical underconnectivity in autism in the resting state.

In a resting state fMRI study, default mode network (DMN) was found to be reduced in autism patients as compared with the control (47). This reduction in DMN was thought to play important roles in symptoms of autism.

It has been a consensus among researcher that decreased connectivity in resting state fMRI is a hallmark of autism (48). Surprisingly, two recent studies indicated that brain connectivity was actually increased, not decreased as commonly believed, in autism children.

In the first report, the authors studied the local connectivity in 29 autism and 29 and IQ matched controls using resting state fMRI (49). By using methods in graph theory, the authors computed the whole-brain maps of local connectivity and compared this maps between autism and control subjects. They were able to identify regions with enhanced connectivity in autism. More specifically, the occipitotemporal regions showed over-connectivity, which is most pronounced in several form of autism.
In a separate report, an analysis of whole-brain functional connectivity was conducted using resting state fMRI in 55 autism children and 55 age-, gender-, and IQ-matched control (50). Widespread functional brain hyperconnectivity was identified in children with autism. In addition, this over-connectivity was shown to be associated with severity of symptoms.

In summary, earlier reports suggested under-connectivity in resting state fMRI in autism children whereas some recent studies indicated over-connectivity. This discrepancy may be due to the following reasons: (1) most studies have relatively small sample size, which ranges from 28 subjects in (51), 53 in (47), 58 in (49), 110 in (50), to 114 in (46). (2) The statistical analysis methods employed in these studies are questionable. For example, two-sample t-test was used to compare autism and control subjects in (51), (47) and (46). Simple t test fails to account for the correlation between regions of the brains. It also fails to take into account the correlation among subjects from the same sites (medical centers).

In this work, we took advantage of the largest autism resting state fMRI database, the Autism Brain Imaging Data Exchange (ABIDE). Our analysis included 361 subjects from 8 medical centers. In addition, we employed sophisticated statistical methods, which include mixed effect model, principal factor approximation, Bayesian hierarchical model and meta analysis to analyze ABIDE data. In order to obtain a reliable estimate of the difference between autism and control, we explicitly separated difference between autism and control from the variations due to subject and site effect.

The remainder of this thesis is organized as follows: In chapter 2 we describe ABIDE, the datasets used in this work and data pre-processing methods. In Chapter 3, we introduce mixed
effect model, focusing on empirical Bayesian method for random effect estimation. We also provide an introduction to false discovery rate (FDR) control. By implementing mixed effect analysis and FDR control, we identify links that show significant difference between autism and control in resting state fMRI. In Chapter 4, we introduce principal factor approximation (PFA) which takes into account correlation among different hypotheses when controlling FDR. As the original proof contains major mistakes, we provide detailed theoretical derivation. Additionally, we apply PFA to the analysis of ABIDE dataset. We also compare the performance of PFA with other FDR control techniques that ignore correlation among hypotheses. In Chapter 5, we describe a full Bayesian approach, the hierarchical Bayesian method, and apply it in modeling of ABIDE data. The pros and cons of hierarchical Bayesian method versus empirical Bayesian method is also discussed. Chapter 6 includes a brief introduction to meta analysis which is followed by application of this method in the analysis of our data. Finally in Chapter 7 we summarize various statistical methods employed in this work and discuss advantages and disadvantages of each method.
CHAPTER 2

DATASET

The goal of this dissertation is to identify brain regions affected by autism using resting state functional MRI data. Many previous fMRI studies suffered from small sample size which led to lack of replication of findings (48). To avoid this issue, we need to find a dataset with many autistic and control subjects.

2.1 Autism Brain Imaging Data Exchange

We decided to use the fMRI data from the autism brain imaging data exchange (ABIDE), which is part of the 1000 Functional Connectome Project/International Neuroimaging Data-sharing Initiative (INDI) (http://fcon_1000.projects.nitrc.org). It is the largest repository of fMRI data for autism. All datasets in ABIDE are fully anonymized to be consistent with HIPAA (Health Insurance Portability and Accountability) guidelines and 1000 Functional Connectomes Project / INDI protocols with no protected health information included.

Autism Brain Imaging Data Exchange is a large international collaboration among medical centers around the globe. In total, 16 sites participated in ABIDE initiative, sharing 1112 datasets of fMRI data and phenotypic information. The ABIDE consists of fMRI data from 539 autism patients and 573 typically developing controls. Detailed information regarding sample size, subject characteristics (e.g., age), diagnostic criteria, data acquisition, and site-specific protocols at each medical center can be found at:
TABLE I: NUMBER OF SUBJECTS AT EACH SITE

<table>
<thead>
<tr>
<th>Site</th>
<th>Control</th>
<th>Autism</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caltech</td>
<td>21</td>
<td>16</td>
<td>37</td>
</tr>
<tr>
<td>NYU</td>
<td>42</td>
<td>35</td>
<td>77</td>
</tr>
<tr>
<td>Olin</td>
<td>16</td>
<td>20</td>
<td>36</td>
</tr>
<tr>
<td>Pitt</td>
<td>28</td>
<td>30</td>
<td>58</td>
</tr>
<tr>
<td>Sbl</td>
<td>15</td>
<td>15</td>
<td>30</td>
</tr>
<tr>
<td>Sdsu</td>
<td>22</td>
<td>14</td>
<td>36</td>
</tr>
<tr>
<td>Sjh</td>
<td>25</td>
<td>22</td>
<td>47</td>
</tr>
<tr>
<td>Stanford</td>
<td>20</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>189</td>
<td>172</td>
<td>361</td>
</tr>
</tbody>
</table>

http://fcon_1000.projects.nitrc.org/indi/abide/.

For our work, we used fMRI measurements from 8 medical centers (sites). The number of subjects at each site is summarized in Table I. Most of the sites have fewer than 60 subjects except for NYU. In total, we have 361 subjects which include 189 control and 172 autism subjects.

We chose to work on the first 84 regions of the ABIDE datasets. Table II lists region number and corresponding description for each region.
<table>
<thead>
<tr>
<th>Region number</th>
<th>Region description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BA.1 (L). Primary Somatosensory Cortex</td>
</tr>
<tr>
<td>2</td>
<td>BA.1 (R). Primary Somatosensory Cortex</td>
</tr>
<tr>
<td>3</td>
<td>BA.10 (L). Anterior Prefrontal Cortex</td>
</tr>
<tr>
<td>4</td>
<td>BA.10 (R). Anterior Prefrontal Cortex</td>
</tr>
<tr>
<td>5</td>
<td>BA.11 (L). Orbitofrontal Cortex</td>
</tr>
<tr>
<td>6</td>
<td>BA.11 (R). Orbitofrontal Cortex</td>
</tr>
<tr>
<td>7</td>
<td>BA.13 (L). Insular Cortex</td>
</tr>
<tr>
<td>8</td>
<td>BA.13 (R). Insular Cortex</td>
</tr>
<tr>
<td>9</td>
<td>BA.17 (L). Primary Visual Cortex</td>
</tr>
<tr>
<td>10</td>
<td>BA.17 (R). Primary Visual Cortex</td>
</tr>
<tr>
<td>11</td>
<td>BA.18 (L). Secondary Visual Cortex</td>
</tr>
<tr>
<td>12</td>
<td>BA.18 (R). Secondary Visual Cortex</td>
</tr>
<tr>
<td>13</td>
<td>BA.19 (L). Associative Visual Cortex</td>
</tr>
<tr>
<td>14</td>
<td>BA.19 (R). Associative Visual Cortex</td>
</tr>
<tr>
<td>15</td>
<td>BA.2 (L). Primary Somatosensory Cortex</td>
</tr>
<tr>
<td>16</td>
<td>BA.2 (R). Primary Somatosensory Cortex</td>
</tr>
<tr>
<td>17</td>
<td>BA.20 (L). Inferior Temporal Gyrus</td>
</tr>
<tr>
<td>18</td>
<td>BA.20 (R). Inferior Temporal Gyrus</td>
</tr>
<tr>
<td>Region number</td>
<td>Region description</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>19</td>
<td>BA.21 (L). Middle Temporal Gyrus</td>
</tr>
<tr>
<td>20</td>
<td>BA.21 (R). Middle Temporal Gyrus</td>
</tr>
<tr>
<td>21</td>
<td>BA.22 (L). Superior Temporal Gyrus</td>
</tr>
<tr>
<td>22</td>
<td>BA.22 (R). Superior Temporal Gyrus</td>
</tr>
<tr>
<td>23</td>
<td>BA.23 (L). Ventral Posterior Cingulate Cortex</td>
</tr>
<tr>
<td>24</td>
<td>BA.23 (R). Ventral Posterior Cingulate Cortex</td>
</tr>
<tr>
<td>25</td>
<td>BA.24 (L). Ventral Anterior Cingulate Cortex</td>
</tr>
<tr>
<td>26</td>
<td>BA.24 (R). Ventral Anterior Cingulate Cortex</td>
</tr>
<tr>
<td>27</td>
<td>BA.25 (L). Subgenual cortex</td>
</tr>
<tr>
<td>28</td>
<td>BA.25 (R). Subgenual cortex</td>
</tr>
<tr>
<td>29</td>
<td>BA.27 (L). Piriform Cortex</td>
</tr>
<tr>
<td>30</td>
<td>BA.27 (R). Piriform Cortex</td>
</tr>
<tr>
<td>31</td>
<td>BA.28 (L). Posterior Entorhinal Cortex</td>
</tr>
<tr>
<td>32</td>
<td>BA.28 (R). Posterior Entorhinal Cortex</td>
</tr>
<tr>
<td>33</td>
<td>BA.29 (L). Retrosplenial Cingulate Cortex</td>
</tr>
<tr>
<td>34</td>
<td>BA.29 (R). Retrosplenial Cingulate Cortex</td>
</tr>
<tr>
<td>35</td>
<td>BA.3 (L). Primary Somatosensory Cortex</td>
</tr>
<tr>
<td>36</td>
<td>BA.3 (R). Primary Somatosensory Cortex</td>
</tr>
<tr>
<td>37</td>
<td>BA.30 (L). Cingulate Cortex</td>
</tr>
<tr>
<td>Region number</td>
<td>Region description</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>38</td>
<td>BA.30 (R). Cingulate Cortex</td>
</tr>
<tr>
<td>39</td>
<td>BA.31 (L). Dorsal Posterior Cingulate Cortex</td>
</tr>
<tr>
<td>40</td>
<td>BA.31 (R). Dorsal Posterior Cingulate Cortex</td>
</tr>
<tr>
<td>41</td>
<td>BA.32 (L). Dorsal anterior Cingulate Cortex</td>
</tr>
<tr>
<td>42</td>
<td>BA.32 (R). Dorsal anterior Cingulate Cortex</td>
</tr>
<tr>
<td>43</td>
<td>BA.33 (L). Anterior Cingulate Cortex</td>
</tr>
<tr>
<td>44</td>
<td>BA.33 (R). Anterior Cingulate Cortex</td>
</tr>
<tr>
<td>45</td>
<td>BA.34 (L). Anterior Entorhinal Cortex</td>
</tr>
<tr>
<td>46</td>
<td>BA.34 (R). Anterior Entorhinal Cortex</td>
</tr>
<tr>
<td>47</td>
<td>BA.35 (L). Perirhinal cortex</td>
</tr>
<tr>
<td>48</td>
<td>BA.35 (R). Perirhinal cortex</td>
</tr>
<tr>
<td>49</td>
<td>BA.36 (L). Parahippocampal cortex</td>
</tr>
<tr>
<td>50</td>
<td>BA.36 (R). Parahippocampal cortex</td>
</tr>
<tr>
<td>51</td>
<td>BA.37 (L). Fusiform gyrus</td>
</tr>
<tr>
<td>52</td>
<td>BA.37 (R). Fusiform gyrus</td>
</tr>
<tr>
<td>53</td>
<td>BA.38 (L). Temporopolar Area</td>
</tr>
<tr>
<td>54</td>
<td>BA.38 (R). Temporopolar Area</td>
</tr>
<tr>
<td>55</td>
<td>BA.39 (L). Angular gyrus</td>
</tr>
<tr>
<td>56</td>
<td>BA.39 (R). Angular gyrus</td>
</tr>
<tr>
<td>Region number</td>
<td>Region description</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td>57</td>
<td>BA.4 (L). Primary Motor Cortex</td>
</tr>
<tr>
<td>58</td>
<td>BA.4 (R). Primary Motor Cortex</td>
</tr>
<tr>
<td>59</td>
<td>BA.40 (L). Supramarginal Gyrus</td>
</tr>
<tr>
<td>60</td>
<td>BA.40 (R). Supramarginal Gyrus</td>
</tr>
<tr>
<td>61</td>
<td>BA.41 (L). Primary Auditory Cortex</td>
</tr>
<tr>
<td>62</td>
<td>BA.41 (R). Primary Auditory Cortex</td>
</tr>
<tr>
<td>63</td>
<td>BA.42 (L). Primary Auditory Cortex</td>
</tr>
<tr>
<td>64</td>
<td>BA.42 (R). Primary Auditory Cortex</td>
</tr>
<tr>
<td>65</td>
<td>BA.43 (L). Subcentral Area</td>
</tr>
<tr>
<td>66</td>
<td>BA.43 (R). Subcentral Area</td>
</tr>
<tr>
<td>67</td>
<td>BA.44 (L). IFC pars opercularis</td>
</tr>
<tr>
<td>68</td>
<td>BA.44 (R). IFC pars opercularis</td>
</tr>
<tr>
<td>69</td>
<td>BA.45 (L). IFC pars triangularis</td>
</tr>
<tr>
<td>70</td>
<td>BA.45 (R). IFC pars triangularis</td>
</tr>
<tr>
<td>71</td>
<td>BA.46 (L). Dorsolateral Prefrontal Cortex</td>
</tr>
<tr>
<td>72</td>
<td>BA.46 (R). Dorsolateral Prefrontal Cortex</td>
</tr>
<tr>
<td>73</td>
<td>BA.47 (L). Inferior Prefrontal Gyrus</td>
</tr>
<tr>
<td>74</td>
<td>BA.47 (R). Inferior Prefrontal Gyrus</td>
</tr>
<tr>
<td>75</td>
<td>BA.5 (L). Somatosensory Association Cortex</td>
</tr>
<tr>
<td>Region number</td>
<td>Region description</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>76</td>
<td>BA.5 (R). Somatosensory Association Cortex</td>
</tr>
<tr>
<td>77</td>
<td>BA.6 (L). Premotor Cortex</td>
</tr>
<tr>
<td>78</td>
<td>BA.6 (R). Premotor Cortex</td>
</tr>
<tr>
<td>79</td>
<td>BA.7 (L). Somatosensory Association Cortex</td>
</tr>
<tr>
<td>80</td>
<td>BA.7 (R). Somatosensory Association Cortex</td>
</tr>
<tr>
<td>81</td>
<td>BA.8 (L). Dorsal Frontal Cortex</td>
</tr>
<tr>
<td>82</td>
<td>BA.8 (R). Dorsal Frontal Cortex</td>
</tr>
<tr>
<td>83</td>
<td>BA.9 (L). Dorsolateral Prefrontal Cortex</td>
</tr>
<tr>
<td>84</td>
<td>BA.9 (R). Dorsolateral Prefrontal Cortex</td>
</tr>
</tbody>
</table>
CHAPTER 3

LINEAR MIXED EFFECT MODEL

The ABIDE dataset that we analyzed contains fMRI data from 361 autism and healthy subjects. For each subject, 3486 fMRI measurements are included in the dataset. As these 3486 observations are collected from the same subject, they are not independent, but rather are correlated with each other. Ignoring this correlation not only decreases the efficiency of the analysis but also leads to biased inferences.

In this work, we adopted a linear mixed effect model to account for correlation among observations within the same subject by introducing random subject effect. For an excellent review of mixed-effect models, please refer to (52).

3.1 Model Specification

For each link, the total variation among different subjects can be partitioned into two types of difference: (1) difference between two groups: autism patient and control, and (2) difference between subjects. As we are not interested in estimating the subject effect, we treat each subject as a random sample from the study population. We adopt the mixed effect model to analyze fMRI data. Specifically,

\[ Y_{ij} = \beta_0 + (1 - G_j) + \beta_1 G_j + \gamma_j + \epsilon_{ij} \]  (3.1)
where $Y_{ij}$ is the fMRI measurement for $i^{th}$ link from $j^{th}$ subject, $i = 1 \ldots m$ and $j = 1 \ldots n$, $G_j = 0$ if subject is a control while $G_j = 1$ if the subject has autism. $\gamma_j$ is the random effect term for $j^{th}$ subject and $\epsilon_{ij}$ is the error term. We further assume that $\gamma_j \sim N(0, \sigma^2_\gamma)$, $\epsilon_{ij} \sim N(0, \sigma^2_{\epsilon_i})$ if subject is from control group and $\epsilon_{ij} \sim N(0, \sigma^2_{\epsilon_1})$ if subject has autism. $\gamma_j$ is independent of $\epsilon_{ij}$.

Just as in standard mixed effect model, the group effect ($\beta_0$ and $\beta_1$) are treated as fixed effect while the subject effect $\gamma_j$ is considered as random effect. This model allows each link to have its own mean and its own variance. In other words, the fMRI measurement for $i^{th}$ link can be expressed as

$$Y_{ij} | \gamma \sim N(\beta_0, \sigma^2_0)$$

(3.2)

if subject is from control group and

$$Y_{ij} | \gamma \sim N(\beta_1, \sigma^2_1)$$

(3.3)

if the subject is from autism group.

The dataset contains fMRI measurements from 361 subjects in 8 medical centers. There are 3486 measurements for each subject. Taken together, there are $361 \times 3486 = 1258446$ data-points in this dataset. Commercial packages, such as SAS proc mixed or R `lm()`, fail to fit the model specified in Equation 3.1 due to insufficient memory. We have created custom R programs to run the analysis. The following sections describe the algorithms in detail.
3.2 Parameter Estimation

The model specified in Equation 3.1 can be translated to the notations of (53). In that model, the mixed effect model is expressed as

\[ y_j = X_j\beta + Z_j\gamma_j + \varepsilon_j, \]  

where \( y_j \) denotes the fMRI measurement for \( j^{th} \) subject. The \( m \times 1 \) vector of \( y_j \) is modeled with both fixed effect \( \beta \) and random effect \( \gamma_j \). \( \beta \) is a \( 2m \times 1 \) vector and \( \gamma_j \) is a scalar. \( X_j \) is a \( m \times 2m \) binary design matrix for fixed effect and \( Z_j \) is a \( m \times 1 \) design matrix for random effect. \( \varepsilon_j \) is a \( m \times 1 \) vector.

The fixed effect term \( \beta \) and random effect term \( \gamma_j \) can be estimated via E-M algorithm as discussed in (52). Briefly, the random effect \( \gamma \) is estimated in the E step via empirical Bayes (EB) estimation whereas the fixed effect \( \beta \), its variance and variance components are estimated in the M step via maximum likelihood estimator (MLE).

3.2.1 Empirical Bayes Estimation

Empirical Bayes (EB) approach can be considered as a compromise between frequentist and Bayesian approaches. Suppose that the observed data \( y \) has some kind of distribution with \( \theta \) as its parameter, or \( y = (y_1, \ldots, y_n) \sim f(y|\theta) \). For example, \( y \sim \alpha + \beta \times x \). \( \theta \) is a vector of \( (\alpha, \beta) \). In statistical analysis, the goal is to make inference about \( \theta \). A frequentist would consider \( \theta \) as a fixed quantity to be estimated from the observed data \( y \). A Bayesian, in contrast, would consider \( \theta \) as a random variable with its own distribution. As a result, a prior distribution
\( \pi(\theta|\eta) \) is placed on \( \theta \). After observed data is obtained, the distribution of \( \theta \) can be updated by combining prior distribution with observed data. The resulting distribution of \( \theta \) is called posterior distribution \( \pi(\theta|y, \eta) \). Posterior distribution is the basis of all Bayesian inference. For example, a Bayesian point estimate of \( \theta \) is the mean of posterior distribution, that is,

\[
E(\theta|y, \eta) = \int \theta \times \pi(\theta|y, \eta) d\theta
\]  

(3.5)

If \( \eta \) is known, the posterior distribution can be derived via the following Bayes’ rule,

\[
p(\theta|y, \eta) = \frac{f(y|\theta) \times \pi(\theta|\eta)}{\int f(y|\theta) \times \pi(\theta|\eta) d\theta} = \frac{f(y|\theta) \times \pi(\theta|\eta)}{m(y|\eta)}
\]  

(3.6)

where \( m(y|\eta) \) is the marginal distribution of \( y \) given \( \eta \).

In almost all cases, however, \( \eta \) is unknown. Empirical Bayes and fully Bayesian approaches differ in how to proceed. Empirical Bayes would estimate \( \varepsilon \) from marginal distribution \( m(y|\eta) \) to get \( \hat{\eta} \) via techniques such as MLE and then plug in \( \hat{\eta} \) to get \( p(\theta|y, \hat{\eta}) \). Empirical Bayes considers \( p(\theta|y, \hat{\eta}) \) as the posterior distribution of \( \theta \).

A statistician who decides to take the fully Bayesian approach would place a distribution on \( \eta \), or \( \eta \sim h(\eta|\lambda) \) where \( \lambda \) is the hyperparameter. The posterior distribution of \( \theta \) can be derived in the following way,

\[
p(\theta|y, \lambda) = \frac{\int f(y|\theta) \times \pi(\theta|\eta)h(\eta|\lambda) d\eta}{\int \int f(y|\theta) \times \pi(\theta|\eta)h(\eta|\lambda) d\theta d\eta} = \int f(\theta|y, \eta) \times h(\eta|y, \lambda) d\eta
\]  

(3.7)
The major difference between EB and fully Bayesian lies in the fact that EB uses data to help determine the prior through the estimation of the hyperparameter $\eta$. Some statisticians argue that EB goes directly against the central dogma in Bayesian statistics that prior should be chosen before data is observed (54).

Empirical Bayesian methods have been widely applied in many areas, such as longitudinal, survival, logistic, and mixed effect analysis (55), (56), (57). For an excellent overview of EB, please refer to (58). As an example of EB analysis modified from (58), let’s suppose that $p$ random variables $x_1, \ldots, x_p$ each follow a normal distribution with different means but the same variance, that is, $x_i \sim n(\theta_i, \sigma^2), i = 1 \ldots p$. Suppose that $\theta_i$ has the following prior distribution $\theta_i \sim n(\mu, \tau^2)$. The posterior distribution of $\theta_i$, denoted as $\pi(\theta_i|x_i, \mu, \tau^2)$, is $n(\left[ \frac{\sigma^2}{\sigma^2 + \tau^2} \right] \mu + \left[ \frac{\tau^2}{\sigma^2 + \tau^2} \right] x_i, \frac{1}{\sigma^2 + \tau^2})$. The Bayes estimator of $\theta_i$, $\theta_i^B$ is given by

$$\theta_i^B = \left[ \frac{\sigma^2}{\sigma^2 + \tau^2} \right] \mu + \left[ \frac{\tau^2}{\sigma^2 + \tau^2} \right] x_i$$  (3.8)

since $\theta_i^B$ is the mean of posterior distribution. To take the empirical Bayesian approach, one would derive the marginal distribution of $x_i$. Standard derivation reveals that $f(x_i) \sim n(\mu, \sigma^2 + \tau^2)$. EB estimates $\mu$ and $\tau$ from marginal distribution of $x_i$, that is,

$$\mu = E(\bar{x}), E[\frac{(p - 3) \times \sigma^2}{\sum (x_i - \bar{x})^2}] = \frac{\sigma^2}{\sigma^2 + \tau^2}$$  (3.9)
The EB estimator of $\theta_i, \theta_i^{EB}$, is obtained by replacing $\mu$ and $\tau^2$ in Equation 3.8 with their estimators in Equation 3.9, that is,

$$
\theta_i^{EB} = \left( \frac{(p - 3) \times \sigma_i^2}{\sum (x_i - \bar{x})^2} \right) \times \bar{x} + \left[ 1 - \frac{(p - 3) \times \sigma_i^2}{\sum (x_i - \bar{x})^2} \right] \times x_i
$$

(3.10)

As we have completed the brief introduction to EB, we are ready to apply EB to estimate $\gamma_i$ in Equation 3.4. In Equation 3.4, observed fMRI measurement $y_j$ and random effect $\gamma_j$ have the following joint distribution,

$$
\begin{bmatrix}
y_j \\
\gamma_j
\end{bmatrix} \sim \text{MVN} \left( \begin{bmatrix} X_j \beta \\ 0 \end{bmatrix}, \begin{bmatrix} Z_j \sigma_i^2 Z_j^T + \Sigma_j & Z_j \sigma_i^2 \\
\sigma_i^2 Z_j^T & \sigma_i^2
\end{bmatrix} \right),
$$

(3.11)

where $\sigma_i^2$ is the variance of random effects and $\Sigma_j$ is a $m \times m$ block diagonal error covariance matrix. The conditional distribution of $\gamma_j$ given $y_j$ can be expressed as

$$
\gamma_j | y_j \sim N(\sigma_i^2 Z_j^T (Z_j \sigma_i^2 Z_j^T + \Sigma_j)^{-1} (y_j - X_j \beta), \sigma_i^2 - \sigma_i^2 Z_j^T (Z_j \sigma_i^2 Z_j^T + \Sigma_j)^{-1} Z_j \sigma_i^2).
$$

(3.12)

The EB posterior distribution of $\gamma_j$ can be obtained after replacing $\sigma_i^2, \Sigma_j$ and $\beta$ in Equation 3.12 with their corresponding estimates. There are several ways to estimate $\sigma_i^2, \Sigma_j$ and $\beta$. One of them is maximum likelihood estimation which will be discussed in the next section.
The posterior mean of $\gamma_j$ and posterior covariance of $\gamma_j$ provide EB estimates of $\gamma_j$ and its covariance, denoted as $\hat{\gamma}_j$ and $\hat{\Sigma}_{\gamma \mid y_j}$, respectively.

$$
\hat{\gamma}_j = \sigma_\gamma^2 Z_j^T (Z_j \sigma_\gamma^2 Z_j^T + \Sigma_j)^{-1} (y_j - X_j \beta)
$$

$$
= \sigma_\gamma^2 [(Z_j \sigma_\gamma^2 Z_j^T)^{-1} + \Sigma_j (Z_j^T)^{-1}]^{-1} (y_j - X_j \beta)
$$

$$
= \sigma_\gamma^2 [Z_j \sigma_\gamma^2 + \Sigma_j (Z_j^T)^{-1}]^{-1} (y_j - X_j \beta)
$$

$$
= \sigma_\gamma^2 [(Z_j)^{-1} Z_j \sigma_\gamma^2 + (Z_j)^{-1} \Sigma_j (Z_j^T)^{-1}]^{-1} (y_j - X_j \beta)
$$

$$
= \sigma_\gamma^2 [(Z_j^T \Sigma_j^{-1} Z_j)^{-1}]^{-1} (Z_j^T Z_j)^{-1} (y_j - X_j \beta)
$$

$$
= \mathbf{R} (Z_j^T \Sigma_j)^{-1} (y_j - X_j \beta)
$$

$$
\hat{\Sigma}_{\gamma \mid y_j} = \sigma_\gamma^2 - \sigma_\gamma^2 Z_j^T (Z_j \sigma_\gamma^2 Z_j^T + \Sigma_j)^{-1} Z_j \sigma_\gamma^2
$$

$$
= [1 - \sigma_\gamma^2 Z_j^T (Z_j \sigma_\gamma^2 Z_j^T + \Sigma_j)^{-1} Z_j] \sigma_\gamma^2
$$

$$
= [1 - \sigma_\gamma^2 Z_j \sigma_\gamma^2 Z_j^T (Z_j^T)^{-1} + \Sigma_j (Z_j^T)^{-1}]^{-1} Z_j \sigma_\gamma^2
$$

$$
= [1 - \sigma_\gamma^2 Z_j \sigma_\gamma^2 + \Sigma_j (Z_j^T)^{-1}]^{-1} Z_j \sigma_\gamma^2
$$

$$
= [1 - \sigma_\gamma^2 (Z_j^{-1} Z_j \sigma_\gamma^2 + Z_j^{-1} \Sigma_j (Z_j^T)^{-1})]^{-1} \sigma_\gamma^2
$$

$$
= [1 - \sigma_\gamma^2 (Z_j^{-1} \Sigma_j)^{-1}]^{-1} \sigma_\gamma^2
$$

$$
= (\mathbf{I} - \mathbf{R}) \sigma_\gamma^2,
$$

where $\mathbf{R} = \sigma_\gamma^2 [(Z_j^T \Sigma_j^{-1} Z_j)^{-1}]^{-1}$. Note that compared to the expression in Equation 3.12 which involves inversion of $m \times m$ matrices, the formula in Equation 3.13 only requires inverting $2 \times 2$ matrices.
3.2.2 Maximum (Marginal) Likelihood Estimation

As noted in the previous section, in order to obtain EB estimates of random effect $\gamma$ and its covariance, $\Sigma_j$ and $\beta$ in Equation 3.13 need to be estimated from the marginal distribution of $Y$. There are several techniques to obtain estimates of $\Sigma_j$ and $\beta$, including maximum likelihood estimation which will be discussed in this section.

The following derivations are modified from (52).

The marginal distribution of fMRI measurements for $j^{th}$ subject, $y_j$, can be derived by integrating out $\gamma_j$ from the joint distribution of $y_j$ and $\gamma_j$, that is

$$m(y_j) = \int f(y_j|\gamma_j)g(\gamma_j)d\gamma$$  \hspace{1cm} (3.14)

where according to Equation 3.11

$$f(y_j|\gamma_j) = N(X_j\beta + Z_j\gamma_j, \Sigma_j)$$  \hspace{1cm} (3.15)

$$g(\gamma_j) = N(0, \sigma^2_\gamma)$$  \hspace{1cm} (3.16)

The marginal distribution of fMRI measurement from $i^{th}$ link of $j^{th}$ subject, $y_{ij}$, is

$$m(y_{ij}) = \int f(y_{ij}|\gamma_j)g(\gamma_j)d\gamma$$  \hspace{1cm} (3.17)
where

\[
f(y_{ij}|\gamma_j) = (2\pi)^{-\frac{1}{2}}|\Sigma_i|^{-\frac{1}{2}} \\
\exp \left[ -\frac{1}{2} (y_{ij} - x_{ij}\beta_i - Z_{ij}\gamma_j)^\top \Sigma_i^{-1} (y_{ij} - x_{ij}\beta_i - Z_{ij}\gamma_j) \right]
\]

(3.18)

\[
g(\gamma_j) = (2\pi)^{-\frac{1}{2}}|\sigma_\gamma|^{-\frac{1}{2}} \exp \left( -\frac{1}{2\sigma_\gamma^2} \gamma_j^\top \gamma_j \right)
\]

\[
\Sigma_i = \sigma^2_{\beta i} \text{ if the subject is a control, and } \Sigma_i = \sigma^2_{\alpha i} \text{ if the subject has autism.}
\]

In order to find the MLE of \(\beta_i, \Sigma_i\), we need to sum marginal log-likelihoods over all subjects for \(i^{th}\) link:

\[
\log L_i = \sum_{j=1}^{n} \log m(y_{ij})
\]

(3.19)

and find out when \(\beta_i, \Sigma_i\) maximize this function.

For this, denote density function of \(y_{ij}\) by \(f_{ij}\), the prior density function of \(\gamma_j\) by \(g\), the posterior density of \(\gamma_j\) by \(p_{ij}\), and the marginal likelihood by \(m_{ij}\). We have

\[
p_{ij} = \frac{f_{ij}g_{ij}}{m_{ij}}
\]

\[
\log L_i = \sum_{j=1}^{n} \log m(y_{ij}) = \sum_{j=1}^{n} \log \left[ \int_{\gamma} (f_{ij} \cdot g) d\gamma \right]
\]

(3.20)
To get the MLE of $\beta_i$, we differentiate Equation 3.20 with respect to $\beta_i$

$$\frac{\partial \log L_i}{\partial \beta_i} = \sum_{j=1}^{n} \frac{\partial \log m_{ij}}{\partial \beta_i}$$

$$= \sum_{j=1}^{n} \frac{1}{m_{ij}} \int_{\gamma} \frac{\partial f_{ij}}{\partial \beta_i} \cdot g d\gamma$$

$$= \sum_{j=1}^{n} \int_{\gamma} m_{ij} \frac{\partial \log f_{ij}}{\partial \beta_i} d\gamma$$

$$= \sum_{j=1}^{n} \int_{\gamma} p_{ij} \hat{X}_j^T \Sigma_i^{-1} (y_{ij} - X_{ij} \beta_i - Z_j \gamma_j) d\gamma$$

$$= \sum_{j=1}^{n} \hat{X}_j^T \Sigma_i^{-1} (y_{ij} - X_{ij} \beta_i - Z_j \hat{\gamma}_j)$$

(3.21)

The last step is because $p_{ij}$ is the posterior distribution of $\gamma$, $\int_{\gamma} \gamma_j p_{ij} d\gamma = E(\gamma_j | y) = \hat{\gamma}_j$ and $\int_{\gamma} p_{ij} d\gamma = 1$. Equating the above expression to zero yields

$$\sum_{j=1}^{n} \hat{X}_j^T \Sigma_i^{-1} X_{ij} \beta_i = \sum_{j=1}^{n} \hat{X}_j^T \Sigma_i^{-1} (y_{ij} - Z_j \hat{\gamma}_j)$$

(3.22)

Hence,

$$\hat{\beta}_i = \left( \sum_{j=1}^{n} \hat{X}_j^T \Sigma_i^{-1} X_{ij} \right)^{-1} \left[ \sum_{j=1}^{n} \hat{X}_j^T \Sigma_i^{-1} (y_{ij} - Z_j \hat{\gamma}_j) \right]$$

(3.23)

The covariance of $\beta_i$ is derived in the following steps.

$$y_i = X_i \times \beta_i + Z_i \times \gamma_j + \epsilon_i$$

$$\text{cov}(y_i) = Z_i \sigma_\gamma^2 Z_i^T + \Sigma_i$$

(3.24)
Denote $\Sigma = \text{cov}(y_i)$

Equation 3.24 can be considered as generalized linear model since covariance of $y_i$ is not a identity matrix. Generalized least square theory yields the following results

$$\hat{\beta}_i = (X_i^T \Sigma^{-1} X_i)^{-1} X_i^T \Sigma^{-1} y_i$$

$$\text{Cov}(\hat{\beta}_i) = (X_i^T \Sigma^{-1} X_i)^{-1} = ([X_i^T (Z_i \sigma_y^2 Z_i^T + \Sigma_i)^{-1} X_i]^{-1}$$ (3.25)

For the covariance of $\Sigma_i$, we can refer to (52). In the case of univariate mixed effect model,

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{j=1}^{N} \left[ e_j^T e_j + \text{tr}(Z_j \Sigma \gamma_{\gamma} Z_j^T) \right]$$ (3.26)

Interestingly enough, as $\gamma_j \to 0$, then $\Sigma_{\gamma \gamma} \to 0$, then $\hat{\sigma}^2$ goes to MLE ignoring random subject effect. In other words, $\hat{\sigma}^2$ takes into account both estimated residuals and the EB estimate of the uncertainty about the random effect. The covariance of the random effect can also be found in (52)

$$\hat{\Sigma}_{\gamma} = \frac{1}{n} \sum_{j=1}^{n} \left( \gamma_j \gamma_j^T + \hat{\Sigma}_{\gamma} y_i \right)$$ (3.27)

To obtain estimates of $\beta_i, \text{cov}(\beta_i), \gamma_i, \sigma_{\gamma}^2, \Sigma_i$, our custom R programs iterate the following steps of EM algorithm:

1. Initialize the parameters $\beta_i, \text{cov}(\beta_i), \gamma_i, \sigma_{\gamma}^2, \Sigma_i$ to some random values.

2. Compute MLE for $\beta_i, \text{cov}(\beta_i), \Sigma_i$, given initial values.

3. Obtain EB estimates of $\gamma_i, \sigma_{\gamma}^2$ by plugging in estimated $\beta_i, \text{cov}(\beta_i), \Sigma_i$. 
4. Re-estimate MLE given updated $\gamma_j, \sigma^2_{\gamma}$.

5. Steps 2-4 are repeated until relative change of the estimated values is smaller than some limit of tolerance (i.e., 1e-5).

3.3 Simulation

A simulation study was carried out to evaluate the accuracy and reliability of the parameter estimation algorithm. For this purpose, 1000 data sets, each with 61250 observations (25 control subjects and 25 autism subjects. 50 regions or 1225 links for each subject) were generated according to the following model,

$$Y_{ij} = \beta_0^i \times (1 - \text{Group}_j) + \beta_1^i \times \text{Group}_j + \gamma_j + \epsilon_{ij}$$

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

$$\epsilon_{ij} \sim N(0, \sigma^2_{\epsilon_i}) \text{ (for control)}$$

$$\epsilon_{ij} \sim N(0, \sigma^2_{\epsilon_i}) \text{ (for autism)}$$

We assume the following true values for fixed and random effect parameters:

$$\beta_i = \begin{bmatrix} \beta_0^i \\ \beta_1^i \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0.05 \end{bmatrix}, \begin{bmatrix} 0.1 & 0 \\ 0 & 0.2 \end{bmatrix} \right)$$

$$\sigma^2_{\beta_i} = 0.25$$

$$\sigma^2_{\epsilon_i} = 0.25$$

$$\sigma^2_{\gamma} = 0.01$$
For each simulated dataset, fixed effect and random effect parameters were estimated according to estimation method described in section 3.4. The performance of this method is evaluated using following criteria: biases, standardized biases (SB), and root mean squared errors (RMSE).

The evaluation quantities are defined as follows:

### 3.3.1 Bias

Bias is the expected difference between the estimated value and the true value, i.e., $\text{Bias}(\hat{\theta}, \theta) = E(\hat{\theta} - \theta)$. We took the mean of the estimated parameters as estimators and calculate the difference between the estimators and the true values.

### 3.3.2 Standardized bias

Standardized bias is the ratio between bias and the standard deviation of $\hat{\theta}$, i.e., $\frac{E(\hat{\theta} - \theta)}{SE(\theta)}$, where $SE$ is the standard error associated with the 1000 estimated values from the simulations. According to (59), "if the absolute value of the standardized bias is less than 50 %, then the bias should be considered practically insignificant".

### 3.3.3 Root mean squared error

Mean squared error (MSE), defined as $\text{Var}(\hat{\theta}) + \text{Bias}(\theta, \hat{\theta})$, is the arguably best quantity to evaluate the performance of estimators since it combines accuracy (bias) and precision (variance). Root mean squared error (RMSE), which has the same unit as the quantity being estimated, is obtained by taking the square root of MSE. For an unbiased estimator (bias=0), the RMSE is the standard deviation.

In the following tables, the abbreviations are as follows: EST=estimate, SE=standard error, StdBias=standardized bias, RMSE=root mean squared error.
As shown in Tables Table III, the estimated fixed and random effects $\beta_{0i}$, $\beta_{1i}$, $\text{var}(\beta_{0i})$, $\text{var}(\beta_{1i})$, $\sigma_{0i}^2$, $\sigma_{1i}^2$, and $\sigma_{\gamma}^2$ are very close to their respective true values. The biases in absolute values are small (less than 0.05); standardized biases in for most parameters are also small and RMSEs are less than 0.04. Taken together, this simulation study demonstrates the accuracy and precision of E-M algorithm in estimating parameters.

3.4 Hypothesis Testing

The dataset contains fMRI measurement of association (referred to as link in this dissertation) between 84 regions in the brain. There are $84 \times 83/2 = 3486$ links in total. The goal of this study is to identify those links whose fMRI measurements are significantly changed in autism subjects as compared with control. As $\hat{\beta}_{0i}$ and $\hat{\beta}_{1i}$ are estimated mean fMRI measurements for
$i^{th}$ link in control subjects and autism subjects, repetitively, the comparison will be based on $\hat{\beta}_0i$ and $\hat{\beta}_{1i}$. Specifically, we will test the following hypotheses simultaneously,

$$H_{0i} : \beta_{0i} - \beta_{1i} = 0 \text{ versus } H_{1i} : \beta_{0i} - \beta_{1i} \neq 0 \text{ where } i = 1 \ldots 3486 \hspace{1cm} (3.30)$$

The hypotheses can be expressed in matrix form, that is,

$$H_{0i} : \mathbf{H} \mathbf{\beta}_i = 0 \text{ versus } H_{1i} : \mathbf{H} \mathbf{\beta}_i \neq 0 \text{ where } i = 1 \ldots 3486, \hspace{1cm} (3.31)$$

$$\mathbf{H}^T = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ and } \mathbf{\beta}_i = \begin{bmatrix} \beta_{0i} \\ \beta_{1i} \end{bmatrix} \hspace{1cm} (3.32)$$

As indicated in section 3.2.2, $\hat{\beta}_i$ has the following distribution,

$$\hat{\beta}_i \sim N(\mathbf{\beta}_i, [\mathbf{X}_i^T (\mathbf{Z}_i \sigma_i^2 \mathbf{Z}_i^T + \mathbf{\Sigma}_i)^{-1} \mathbf{X}_i]^{-1}) \hspace{1cm} (3.33)$$

Hence under null hypothesis,

$$\mathbf{H} \hat{\mathbf{\beta}}_i \sim N(0, \mathbf{H}[(\mathbf{X}_i^T (\mathbf{Z}_i \sigma_i^2 \mathbf{Z}_i^T + \mathbf{\Sigma}_i)^{-1} \mathbf{X}_i]^{-1} \mathbf{H}^T) \hspace{1cm} (3.34)$$

The null hypothesis can be tested using Wald’s test, that is,

$$H_{0i} \text{ is rejected when } C_i^2 > \chi^2_{1-a}(q) \hspace{1cm} (3.35)$$
where $C_i^2 = H\hat{\beta}_i \left( H[(X_i^T(Z_i\sigma_i^2Z_i^T + \Sigma_i)^{-1}X_i)^{-1}H^T]^{-1}(H\hat{\beta}_i)^T \sim \chi^2(q) \right)$ and $q$ is the degree of freedom.

### 3.5 Overview of False Discovery Rate

In hypothesis testing, a p-value is computed to determine whether there is enough evidence against null hypothesis. This p-value is defined as the probability of obtaining an observation at least as extreme as the observed results, assuming that the null hypothesis is true. To claim a test to be significant, it is common to choose a cutoff for statistical significance (denoted as $\alpha$). Tests with a p-value smaller than this cutoff are identified as significant. This cutoff value $\alpha$ is also called type I error rate because it is the probability of rejecting a truly null hypothesis. The task of choosing an appropriate cutoff is not a simple one. For a test involving only one null hypothesis, a well-known cutoff value is 0.05. This error rate is called comparisonwise error rate (CWER).

Recent years has seen the rapid emergence of large datasets in many areas of scientific fields, such as genomics, brain imaging, astrophysics, and spatial epidemiology. For example, in a typical microarray experiment, expression level of thousands of genes are quantified. The goal is to identify genes that are differentially expression under different conditions, e.g. patients versus healthy control. As thousands of genes are under study, many hypotheses are tested simultaneously. To control error rate in multiple testing, familywise error rate (FWER), also called the overall type I error rate, is introduced. Familywise error rate is defined as the probability of making at least one type I error among a family of hypothesis tests.
TABLE IV: OUTCOMES AMONG A FAMILY OF TESTS

<table>
<thead>
<tr>
<th></th>
<th>Called significant</th>
<th>Called not significant</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null true</td>
<td>F</td>
<td>(m_0 - F)</td>
<td>(m_0)</td>
</tr>
<tr>
<td>Alternative true</td>
<td>T</td>
<td>(m_1 - T)</td>
<td>(m_1)</td>
</tr>
<tr>
<td>Total</td>
<td>S</td>
<td>(m - S)</td>
<td>(m)</td>
</tr>
</tbody>
</table>

To illustrate the concept of FWER, Table IV lists the outcomes of simultaneously testing several hypotheses (adapted from (60) and (61)).

Let us suppose that \(m\) hypotheses are tested simultaneously. Among those \(m\) hypotheses, \(m_0\) are truly null whereas \(m_1\) are truly alternative. \(F\), \(T\) and \(S\) represent the number of false discoveries, the number of true discoveries and total number of significant results, respectively. Apparently, \(F\) is the number of type I errors. \(m_1 - T\) represents the number of type II errors. In terms of definitions in Table IV, FWER is the probability that \(F\) is larger than 0.

Various methods have been proposed to control FWER. Among them is the famous Bonferroni correction. For a family of \(m\) tests, the Bonferroni correction suggests that the p-value for each significant test should be smaller than \(\frac{\alpha}{m}\). This criterion guarantees that FWER is less than \(\alpha\). However, the Bonferroni approach has proved to be not particularly suitable for analyzing large-dataset experiments, such as microarray. Firstly, in a typical microarray experiment, thousands of genes are included and several are expected to be differentially expressed. When several discoveries are expected, methods controlling FWER tends to be not sensitive enough (61). Secondly, one of the questions that most experimental scientists are interested in is what proportion of the significant results is truly significant. Put in other way, what proportion of
the positive result is repeatable? Familywise error rate criteria fail to address this important question because it controls the error rate of rejecting a truly null hypothesis rather than the error rate of null hypothesis being true given that it is rejected by the test.

In recent years, some novel statistical approaches have been proposed which not only have more power to detect truly significant results but also have a good control over false discovery. Those methods are based on the concept of posterior error rate (PER) which, for a single test, is defined as the probability of null being true given that the test rejected the null.

\[ \text{PER} = P(F = 1 | S = m = 1) \]

Posterior error rate can also be defined in terms of type I error \( \alpha \), type II error \( \beta \) and the probability of discovery \( \pi_1 \), that is,

\[ \text{PER} = \frac{1}{1 + \frac{(1-\beta)\pi_1}{\alpha(1-\pi_1)}} \]

For multiple testing, false discovery rate (FDR) was introduced in a pioneering paper (62). False discovery rate is defined as “the expected proportion of false positive findings among all those rejected hypotheses” (62).

\[ \text{FDR} = E[Q] \quad (3.36) \]

where \( Q = F/S \) if \( S > 0 \) and \( Q = 0 \) otherwise
It is possible that in a given multiple tests, none of the hypotheses is found to be significant. In this case, $S = 0$ and $F/S$ is undefined. To address this issue, a quantity called positive FDR (pFDR) was proposed (60).

\[
pFDR = E[F/S|S > 0]
\]  

(3.37)

Standard statistical methods of controlling type I error are not always associated with a low PER (61). In addition, when the number of test is large and the proportion of alternative hypotheses is high, those method that control FDR rather than FWER have good performance in terms of high power and low PER (61). However, when no type I error can be tolerated an FWER-based should be preferred.

Clearly, the FDR is the useful measure of overall accuracy of a set of significant findings. However, it is more convenient to attach a measure of significance to individual tests. Recall that for the methods that control type I error, each test is given a p-value as a measure of significance. For those methods based on FDR, a similar quantity called q-value was introduced in a groundbreaking work (60). The q-value is defined as ”the minimum pFDR at which the test can be called significant”.

The q-value has a close relationship to the p-value. For a set of hypothesis tests conducted with independent p-values, the q-value of observed p-value $p$ is

\[
q(p) = \inf_{\gamma \leq p} \text{pFDR}(\gamma)
\]  

(3.38)
The following algorithm to calculate the q-values was proposed in (63).

1. For $m$ hypothesis tests, calculate the p-values $p_1, \ldots, p_m$.
2. Let $p(1) \leq \cdots \leq p(m)$ be the ordered p-values.
3. Set $q(p(m)) = \text{pFDR}(p(m))$.
4. Set $q(p(i)) = \min[p\text{FDR}(p(i)), q(p(i + 1))]$ for $i = m - 1, m - 2, \ldots, 1$.

The remaining question is how to estimate the pFDR. It is obvious that because $m$ is always large in large-dataset studies, we have $P(S > 0) \approx 1$ and $\text{pFDR} \approx \text{FDR}$ (64). The following procedure was proposed to estimate FDR (64):

1. A test is called significant if its associated p-value is less than or equal to a given threshold $t$, where $0 < t \leq 1$.
2. Let $F(t) = \#\{\text{null } p_i \leq t; i = 1, \ldots, m\}$ and $S(t) = \#\{p_i \leq t; i = 1, \ldots, m\}$.
3. FDR can be quantified as
   \[
   \text{FDR}(t) = \frac{E[F(t)]}{S(t)} \approx \frac{E[F(t)]}{E[S(t)]} \quad (3.39)
   \]

Of course, a good estimate of $E[S(t)]$ is simply $S(t)$. To estimate $E[F(t)]$, it is obvious that p-values for truly null hypotheses are uniformly distributed. In other words, suppose we want to test $H_{01}, \ldots, H_{0m}$. The corresponding p-values are $p_1, p_2, \ldots, p_m$. If $H_{0i}$ is true, then $p_i \sim \text{Uniform}[0, 1]$. Moreover, if $X \sim \text{Uniform}(0, 1)$, then $F(x) = p(X \leq x) = x$ for $x \in [0, 1]$. Therefore, the probability that a null p-value is smaller than $t$ is simply $t$ and it follows from Table IV that $E[F(t)] = m_0t$, where $m_0$ is the number of truly null hypotheses.
However, $m_0$ is unknown and needs to be estimated. Equivalently, one can also estimate the proportion of hypotheses that are truly null, which is denoted to be $\pi_0 = \frac{m_0}{m}$.

To estimate $\pi_0$, recall that most of the big p-values should correspond to tests with truly null hypotheses. From the histogram of p-values, the turning point beyond which the density of p-values look flat can be identified, which is denoted as $\lambda$. The estimate of $\pi_0$ can be quantified with

$$
\pi_0(\lambda) = \frac{\#\{p_i > \lambda; i = 1, \ldots, m\}}{m(1 - \lambda)}
$$

Another way to estimate $\pi_0$ is as follows:

1. For a range of $\lambda$, say $\lambda = 0, 0.01, 0.02, \ldots, 0.95$, calculate $\pi_0(\lambda) = \frac{\#\{p_i > \lambda; i = 1, \ldots, m\}}{m(1 - \lambda)}$.

2. Plot $\pi_0(\lambda)$ against $\lambda$ and fit a natural cubespline $\hat{f}$.

3. Set the estimate of $\pi_0$ to be $\hat{\pi}_0 = \hat{f}(1)$.

By plugging these quantities into the right side of Equation 3.39, FDR $(t)$ is estimated by

$$
\hat{\text{FDR}}(t) = \frac{\hat{\pi}_0 mt}{S(t)} = \frac{\hat{\pi}_0 mt}{\#\{p_i \leq t\}}.
$$

(3.40)

Just as a p-value is a measure of false positive rate (FPR), the q-value is the measure in terms of FDR. The FPR is the rate that true nulls are called significant whereas FDR is the rate that significant results are actually nulls. The distinction between those two terms may seem subtle but is really important. As mentioned before, an experimental biologist’s main concern about positive results is whether those results can be repeatable. The p-value says little about
this important question. The q-value, on the other hand, directly addresses this question. By definition, q-value is “the minimum FDR that can be attained when calling that test significant”. If we call a test with a q-value of 0.05 significant, then all the tests with smaller q-value are automatically called significant. Within this list of significant tests, we would expect about 5 percent are actually not significant, or put in another way, 5 % significant results are not repeatable. The software for computing q-values based on a list of p-values can be found at http://genomine.org/qvalue/.

After we completed the mixed effect analysis via E-M algorithm and got the p-values for each link, q-value program was used to calculate q-values for each link.

3.6 Data Analysis Using Mixed Effect Model

In this section, we will apply the methods discussed in this Chapter to analyze fMRI data from Autism Brain Imaging Data Exchange(ABIDE).

3.6.1 Transformation of Data

One of the key assumptions of linear mixed effect model is that $Y_i'$s follow normal distribution. To check this assumption, the empirical distributions of the fMRI measurements for the first 2, middle 2 and last 2 links are presented in Figures 3-5 (left column), together with the fitted normal density curves. It seems that $Y_i'$s roughly follow normal distribution. The right column of Figures 3-5 are normal probability plots, which confirm that normal distribution seems to be an appropriate assumption. Therefore we decided that no data transformation was necessary. The analyses described in the section were performed on the original scale.
Figure 3: fMRI measurement for the first 2 links in control subjects (top two rows) and autism subjects (bottom two rows).
Figure 4: fMRI measurement for the middle 2 links in control subjects (top two rows) and autism subjects (bottom two rows).
Figure 5: fMRI measurement for the last 2 links in control subjects (top two rows) and autism subjects (bottom two rows).
3.6.2 Mixed Effect Analysis on Data from Each Site

Our dataset contains fMRI measurements from 8 medical centers (sites). As can be seen in Table I, the number of subjects in each site is relatively small. Except for NYU, all sites have fewer than 60 subjects.

The first analysis was to fit mixed effect model for each site. As a comparison, we also carried out two-sample t-test, which ignores random subject effect. Table V summarizes the number of significant links, detected by t test or mixed analysis, with or without FDR control.

<table>
<thead>
<tr>
<th>Site</th>
<th>P-value &lt; 0.05</th>
<th>FDR=0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t Test</td>
<td>Mixed Analysis</td>
</tr>
<tr>
<td>Caltech</td>
<td>199</td>
<td>247</td>
</tr>
<tr>
<td>NYU</td>
<td>175</td>
<td>201</td>
</tr>
<tr>
<td>Olin</td>
<td>134</td>
<td>196</td>
</tr>
<tr>
<td>Pitt</td>
<td>242</td>
<td>279</td>
</tr>
<tr>
<td>Sbl</td>
<td>200</td>
<td>272</td>
</tr>
<tr>
<td>Sdsu</td>
<td>166</td>
<td>214</td>
</tr>
<tr>
<td>Sjh</td>
<td>213</td>
<td>265</td>
</tr>
<tr>
<td>Stanford</td>
<td>119</td>
<td>157</td>
</tr>
</tbody>
</table>

The numbers of significant links with FDR controlled at 0.3 are much smaller than the numbers of links without FDR control. This finding highlights the importance of FDR control.
By taking into account random subject effect, mixed effect analysis is able to partition total variation into two type of variations: (1) variation due to difference between autism subject and control subjects, and (2) variation due to random subject effect. Due to this reason, mixed effect analysis is more powerful than t-test. This is confirmed by smaller p values for all links (data not shown) and more significant links (with or without FDR control) detected by mixed effect analysis when compared to t test. Moreover, all links declared as significant by t-test are also detected by mixed model analysis. Therefore, mixed effect analysis is chosen as the analysis in the remaining part of this dissertation.

Due to small sample size at each site, the number of significant links detected by mixed effect model is different from site to site. Moreover, the links deemed significant are not all the same among all sites. An analysis that combines data from all sites is better than by-site analysis since it can produce more reliable estimate of difference between two groups due to increased sample size.

### 3.6.3 Mixed Effect Analysis on Pooled Data

To get a rough estimation of the proportion of links with significant difference between control and autism subjects, we produced histograms for mean fMRI of each link for two groups. As shown in Figure 6, these two histograms are very similar, suggesting that the number of significant links may not be very big. Table VI presents the mean and SD of fMRI in control and autism subjects. Compared with control, the mean in autism subjects is smaller, suggesting decreased association between regions in autism subjects.
TABLE VI: SUMMARY OF LINK SPECIFIC MEAN fMRI

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Autism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>0.0293 (0.119)</td>
<td>0.0272 (0.120)</td>
</tr>
</tbody>
</table>

We fitted the mixed effect model described in section 3.1 to pooled data. The parameters in the model were estimated using E-M algorithm that iterated though the E step (EB) and M step (MLE). There were 14306 parameters to be estimated. The algorithm was considered to converge when difference between two adjacent estimated value was smaller than tolerance limit of $10^{-5}$. It took six iterations for E-M to converge. After estimates of parameters were obtained, hypotheses testing were carried out using Wald’s test according to section 3.4. The resulting p values were subject to FDR adjustment as described in section 3.5.

Figure 7 is the histogram of p values from mixed effect analysis. The majority of p values are large, indicating that most of the hypotheses are probably truly null hypotheses.
Figure 8 are plots produced by qvalue package. Figure 8(a) is a plot of $\pi_0$ versus $\lambda$ where $\pi_0$ is the estimated proportion of truly null hypotheses and $\lambda$ is the cutoff value for p values. A hypothesis with a p value larger than $\lambda$ is considered to be truly null hypothesis. The estimated proportion of truly null hypothesis $\pi_0$ decreases with p value cutoff $\lambda$. $\pi_0$ stabilizes at around 0.924 which is the final estimate of proportion of truly null hypothesis. Figure 8(b) is a plot of q values versus p values. It can be inferred from this plot the expected proportion of false discoveries for different p-value cutoffs. Figure 8(c) shows the number of significant links for each q value. The number of significant links increases with q value cutoff. A sharp increase in number of significant links occur when q value becomes slightly greater than 0.16. Notice that there is a similar jump in number of significant links at q value cutoff of 0.3. This plot allows one to evaluate the sensitivity of number of significant findings to the q value cutoff. Finally, Figure 8(d) shows the expected number of false discoveries versus the number of links called
Figure 8: Plots produced by qvalue package
(a) $\pi_0$ versus the tuning parameter $\lambda$
(b) The q values versus their respective p values
(c) The number of significant links versus the respective q value
(d) The expected number of false positive links versus total number of significant links.
TABLE VII: SIGNIFICANT LINKS FDR=0.3

<table>
<thead>
<tr>
<th>Num</th>
<th>Link</th>
<th>Control Mean</th>
<th>Autism Mean</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63-66</td>
<td>0.12445</td>
<td>-0.00022</td>
<td>3.58E-05</td>
</tr>
<tr>
<td>2</td>
<td>63-64</td>
<td>0.07344</td>
<td>-0.04197</td>
<td>4.02E-05</td>
</tr>
<tr>
<td>3</td>
<td>08-35</td>
<td>0.02158</td>
<td>-0.07116</td>
<td>4.76E-05</td>
</tr>
<tr>
<td>4</td>
<td>14-54</td>
<td>0.05311</td>
<td>-0.02674</td>
<td>0.000117776</td>
</tr>
<tr>
<td>5</td>
<td>63-68</td>
<td>0.0516</td>
<td>-0.04816</td>
<td>0.000296115</td>
</tr>
<tr>
<td>6</td>
<td>07-66</td>
<td>0.19518</td>
<td>0.09752</td>
<td>0.000330611</td>
</tr>
<tr>
<td>7</td>
<td>14-27</td>
<td>0.01138</td>
<td>-0.05342</td>
<td>0.000360923</td>
</tr>
<tr>
<td>8</td>
<td>57-82</td>
<td>-0.06261</td>
<td>0.021</td>
<td>0.000410612</td>
</tr>
<tr>
<td>9</td>
<td>61-66</td>
<td>0.20255</td>
<td>0.10401</td>
<td>0.000494136</td>
</tr>
<tr>
<td>10</td>
<td>20-38</td>
<td>0.0441</td>
<td>-0.02286</td>
<td>0.000697267</td>
</tr>
<tr>
<td>11</td>
<td>08-74</td>
<td>0.19128</td>
<td>0.27571</td>
<td>0.000845696</td>
</tr>
<tr>
<td>12</td>
<td>07-08</td>
<td>0.52881</td>
<td>0.40353</td>
<td>0.000961975</td>
</tr>
</tbody>
</table>

significant. In summary, these plots allow a researcher to have a comprehensive view of how FDR control can affect study results.

Table VII presents the links with significant difference between control and autism subjects with FDR controlled at 0.3. Mixed effect model was able to detect 12 links with significant difference in fMRI between autism and control subjects. Interesting enough, for most of the links except links 68 and 11, the association between regions are decreased in autism subjects when compared to control subjects. The significance of this finding need further confirmation by medical experts.

Figure 9 is a brain network plot that illustrates the links identified as significantly different between autism and control at FDR of 0.3 by mixed effect analysis.
Figure 9: Brain network mixed analysis.
CHAPTER 4

PRINCIPAL FACTOR APPROXIMATION

False discovery rate is defined as the expected proportion of falsely rejected null hypotheses among all rejected null hypotheses (63). False discovery rate control has been a hot topic as it has wide applications in many areas of research, such as genomics, neuroscience, and economics.

Previously several methods have been proposed to control FDR. These methods generally fall into two categories: (1) One is to determine acceptable FDR and then find the corresponding cutoff value so that the expected FDR is less than or equal to pre-chosen FDR (65; 62). (2) The other approach is to fix the p-value threshold \( t \) and estimate FDR whose expectation is no larger than the true FDR at that particular threshold (63). It has been shown that these two approaches are equivalent (66; 67).

The major drawback of all these approaches is that they fail to account for dependence among the hypotheses being tested. For example, originally B-H method assumes independent test statistics (62). Later research has demonstrated that (62) and (63) are able to control FDR under some special dependence structure such as positive regression dependence (68) and weak dependence structure (66). Even if this is the case, these procedures suffer from loss of efficiency since the actual dependence information is not taken into account in the FDR control. Hence, multiple testing under general dependence structure remains a very important and challenging question.
4.1 Overview of Principal Factor Approximation

An elegant solution to multiple testing under arbitrary covariance structure was proposed in a seminal paper (69). The unique feature of their method is that instead of controlling FDR, which is the average of FDP if an experiment is repeated many times, it controls the realized FDP in a given experiment. One can argue that controlling FDP is more relevant since it is directly related to the current study.

Fan’s method is called Principal Factor Approximation(PFA). The basic idea is as follows: (1) apply spectral decomposition to covariance (2) take out largest common factors so that the remaining ones have weak dependence (3) derive the expression of FDP that accounts for strong dependence (4) estimate principal factors (5) plug in the consistent estimate of principal factors to estimate the realized FDP.

The motivation of the PFA method is a genomic study of association between phenotypes, such as weight and blood type, and genotype of SNP. The independent variable $X$ is a $n \times p$ matrix where each row corresponds to a subject and each column is a SNP. Let $X_i$, $X_j$ and $Y$ denote $i$th subject, $j$th SNP and the outcome, respectively. The first step of PFA is to run a marginal linear regression between $Y$ and $X_j$ to get the coefficient $\hat{\beta}_j$. The multiple testing problem can be expressed as follows:

$$H_{0j} : \beta_j = 0 \text{ versus } H_{1j} : \beta_j \neq 0, j = 1 \ldots p$$ (4.1)
Due to the sample correlation among \( \{X_i^j\}_{j=1}^{n;j=1:p} \) and \( \{\hat{\beta}_i^j\}_{j=1}^{p} \) are also correlated. The joint distribution of \( \{\hat{\beta}_i^j\}_{j=1}^{p} \) is specified in Proposition 1.

**Proposition 1.** Assume that the distribution of \( Y_i \) given \( \{X_i^1, \ldots, X_i^p\} \) is \( N(\mu, \sigma^2) \). Then, conditioning on \( \{X_i^j\}_{j=1}^{i=n;j=1:p} \), the joint distribution of \( \{\hat{\beta}_i^j\}_{j=1}^{p} \) is \( \{\hat{\beta}_i^j\}_{j=1}^{p} \sim N(\beta, \Sigma^*) \) where the \( (k,l) \)th element in \( \Sigma^* \) is \( \Sigma_{kl}^* = \frac{\sigma^2 r_{kl}}{n s_{kk}s_{ll}} \). \( r_{kl}, s_{kk}, s_{ll} \) are sample correlation between \( X_k \) and \( X_l \), sample standard deviation of \( X_k \), sample standard deviation of \( X_l \), respectively.

Let \( Z_i \) denote the standardized version of \( \hat{\beta}_i^j \), that is

\[
Z_i = \frac{\hat{\beta}_i^j}{\text{SD}(\hat{\beta}_i^j)}. \tag{4.2}
\]

The distribution of \( Z \) is \( (Z_1 \ldots Z_p)^t \sim N((\mu_1 \ldots \mu_p)^t, \Sigma) \) where \( \Sigma \) has the \( (k,l) \)th element of \( r_{kl} \).

This implies that the diagonals of \( \Sigma \) is 1. As \( Z \) is the standardized version of \( \hat{\beta} \), the simultaneous testings in Equation 4.1 is equivalent to the multiple testings based on \( Z \).

\[
H_{0j} : \mu_j = 0 \quad H_{1j} : \mu_j \neq 0, j = 1, \ldots, p \tag{4.3}
\]

The first step of PFA is to apply spectral decomposition to \( \Sigma \), that is \( \Sigma = \sum_{i=1}^{p} \lambda_i \gamma_i \gamma_i^t \) where \( \lambda_1, \ldots, \lambda_p \) are eigenvalues of \( \Sigma \) in decreasing order and \( \gamma_1, \ldots, \gamma_p \) are corresponding orthonormal eigenvectors. Let \( A = \sum_{i=k+1}^{p} \lambda_i \gamma_i \gamma_i^t \) and \( L = (\sqrt{\lambda_1} \gamma_1, \ldots, \sqrt{\lambda_k} \gamma_k) \), then \( \Sigma \) can be written as

\[
\Sigma = LL^t + A. \tag{4.4}
\]
and $Z_1, \ldots, Z_p$ can be expressed as $Z_i = \mu_i + b_i^T W + K_i$ where $b_i = (b_{i1}, \ldots, b_{ik})^T$, $(b_{i1}, \ldots, b_{ip})^T = \sqrt{n}Y_j$, $W \sim N_k(0, I_k)$ and $(K_1, \ldots, K_p)^T \sim N(0, A)$. In addition, $W_1, \ldots, W_k$ are independent of each other and independent of $K_1, \ldots, K_p$.

The key step of PFA is to choose smallest $k$ such that $(K_1, \ldots, K_p)^T$ is weakly dependent. Theorem 1 in (69) demonstrates how to estimate realized FDP via PFA.

**Theorem 1.** Suppose $(Z_1, \ldots, Z_p)^T \sim N((\mu_1 \ldots \mu_p)^T, \Sigma)$. Choose an appropriate $k$ such that

$$
\frac{\sqrt{\lambda_{k+1}^2 + \cdots + \lambda_p^2}}{\lambda_1 + \cdots + \lambda_p} = O(p^{-\delta}) \text{ for } \delta > 0. \tag{4.5}
$$

let $\sqrt{\lambda_j}Y_j = (b_{1j} \ldots b_{pj})^T$ for $j = 1 \ldots k$, then,

$$
\lim_{p \to \infty} \left\{ \text{FDP}(t) - \frac{1}{p} \sum_{i \in \text{[truenull]}} \left[ \Phi(a_i(z_{1i} + \eta_i)) + \Phi(a_i(z_{1i} - \eta_i)) \right] \right\} = 0 \text{ a.s.} \tag{4.6}
$$

where $a_i = (1 - \sum_{h=1}^{k} b_{ih}^2)^{-\frac{1}{2}}$, $\eta_i = b_i^T W$, and $\Phi(\cdot)$ and $z_{1i}$ are the cumulative distribution function and lower $\frac{1}{2}$ quantile of a standard normal distribution.

The proof of Theorem 1 is based on the following result

**Proposition 2.**

$$
\lim_{p \to \infty} \left( \frac{1}{p} R(t) - \frac{1}{p} \sum_{i=1}^{p} (\Phi(a_i(z_{1i} + \eta_i + \mu_i)) + \Phi(a_i(z_{1i} - \eta_i - \mu_i))) \right) = 0 \text{ a.s.} \tag{4.7}
$$

$$
\lim_{p \to \infty} \left( p_0^{-1} V(t) - p_0^{-1} \sum_{i \in \text{[truenull]}} (\Phi(a_i(z_{1i} + \eta_i)) + \Phi(a_i(z_{1i} - \eta_i))) \right) = 0 \text{ a.s.} \tag{4.8}
$$
where \( V(t) = \#\{\text{true null } P_i : P_i \leq t\} \), \( S(t) = \#\{\text{false null } P_i : P_i \leq t\} \), and \( R(t) = \#\{P_i : P_i \leq t\} \). It is obvious that \( V(t), S(t), R(t) \) are the number of false discoveries, true discoveries, total discoveries, respectively. Among all \( p \) null hypotheses, \( p_0 \) of them are true and \( p_1 \) hypotheses are false \((p_1 = p - p_0)\), and \( p_1 \) is supposed to be very small compared to \( p \).

The performance of PFA is compared with other methods that control FDP, such as B-H procedure in (62) and Storey’s procedure in (70) under six different dependence structures. It is demonstrated that PFA performed better in term of both mean and standard deviation of the FDP, especially under strong dependent structure. In all the cases examined, B-H procedure and Storey’s procedure tend to underestimate the FDP. In addition, these authors also compared the estimated FDP between PFA and Efron’s method (71). Via simulation, it is shown that PFA has smaller relative error. In fact the relative error of PFA is close to 0, indicating the accuracy of PFA in estimating FDP.

### 4.2 Proof of Proposition 2

Theorem 1 is the key result of (69). The proof of Theorem 1 is based on Proposition 2. The paper includes an very brief proof of proposition 2. In this section, we will identify major errors in the original proof of proposition 2 and provide a full proof.

**Theorem 4.2.1.** Let \( a_l = (1 - \sum_{h=1}^{k} b_{lh}^2)^{-\frac{1}{2}} \) as defined in Equation 4.6 and \( K_i \) as defined in Equation 4.4, \( a_l K_i \sim N(0, 1) \)

**Proof.** The following steps are to derive the distribution of \( K_i \).

From Equation 4.2, \( Z_i \) is the standardized form of \( \hat{\beta}_i, Z \sim N(\mu, \Sigma) \) where \( \Sigma \) has the \((k, l)\) element
as $\gamma_{kl}$ and $\gamma_{kl}$ is the correlation between $x_k$ and $x_l$. Hence $\text{Var}(Z_i) = 1$. From Equation 4.4 and Equation 4.6,

\[ Z_i = \mu_i + \eta_i + K_i = \mu_i + \sum_{h=1}^{k} b_{ih} \times w_h + K_i \]

where $\eta_i = \sum_{h=1}^{k} b_{ih} \times w_h$

\[ \text{Var}(Z_i) = \text{Var}(\sum_{h=1}^{k} b_{ih} \times w_h) + \text{Var}(K_i) = \sum_{h=1}^{k} b_{ih}^2 + \text{Var}(K_i) = 1 \]  \hfill (4.9)

since $K_i$ and $W_i$ are independent.

Hence $\text{Var}(K_i) = 1 - \sum_{h=1}^{k} b_{ih}^2 = a_i^{-2}$ where $a_i = (1 - \sum_{i=1}^{k} b_{ih}^2)^{-1/2}$

Therefore $K_i \sim N(0, a_i^{-2})$ and $a_i K_i \sim N(0, 1)$ \hfill \(\square\)

**Theorem 4.2.2.** $P(p_i \leq t | w_1 \ldots w_k) = \Phi[a_i(z_i^2 + \eta_i + \mu_i)] + \Phi[a_i(z_i^2 - \eta_i - \mu_i)]$
Proof.

\[
P(p_i \leq t | w_1 \ldots w_k) = 1 - P(|z_i| \leq -\Phi^{-1}(\frac{t}{2}) | w_1 \ldots w_k)
\]

\[
= 1 - P(z_i < \mu_i + \eta_i + k_i < -z_i)
\]

\[
= 1 - P(z_i - \eta_i - \mu_i < k_i < -z_i - \eta_i - \mu_i)
\]

\[
= 1 - P(a_i(z_i - \eta_i - \mu_i) < a_i k_i < a_i(-z_i - \eta_i - \mu_i))
\]

By Theorem 4.2.1  \(a_i k_i \sim N(0,1)\)

\[
= 1 - \{\Phi[a_i(-z_i - \eta_i - \mu_i)] - \Phi[a_i(z_i - \eta_i - \mu_i)]\}
\]

\[
= 1 - \{1 - \Phi[a_i(z_i + \eta_i + \mu_i)] - \Phi[a_i(z_i - \eta_i - \mu_i)]\}
\]

\[
= \Phi[a_i(z_i + \eta_i + \mu_i)] + \Phi[a_i(z_i - \eta_i - \mu_i)]
\]

\[\square\]

**Theorem 4.2.3.** Define

\[
c_{1i} = a_i(-z_i - \eta_i - \mu_i) \\
c_{2i} = a_i(z_i - \eta_i - \mu_i) \\
c_{1j} = a_i(-z_j - \eta_j - \mu_j) \\
c_{2j} = a_i(z_j - \eta_j - \mu_j) \\
\rho_{ij} = \text{Corr}(K_i, K_j)
\]
\[ P[|z_i| \leq -\Phi^{-1}(\frac{t}{2})|w_i \ldots w_k|, |z_j| \leq -\Phi^{-1}(\frac{t}{2})|w_1 \ldots w_k] = P(\frac{c_{1i}}{a_i} < K_i < \frac{c_{2i}}{a_i}, \frac{c_{1j}}{a_j} < K_j < \frac{c_{2j}}{a_j}) \]  

\begin{equation}
= \int_{-\infty}^{\infty} [\Phi(\frac{\sqrt{\rho_{ij}z + a_1}}{\sqrt{1 - \rho_{ij}}}) - \Phi(\frac{\sqrt{\rho_{ij}z + a_2}}{\sqrt{1 - \rho_{ij}}})] \times [\Phi(\frac{\sqrt{\rho_{ij}z + a_3}}{\sqrt{1 - \rho_{ij}}}) - \Phi(\frac{\sqrt{\rho_{ij}z + a_4}}{\sqrt{1 - \rho_{ij}}})] \phi(z) \, dz
\end{equation}

where

\[
\begin{align*}
a_1 &= \frac{c_{1i}}{a_i^{-1}} = c_{1i} \\
a_2 &= \frac{c_{2i}}{a_i^{-1}} = c_{2i} \\
a_3 &= \frac{c_{1j}}{a_j^{-1}} = c_{1j} \\
a_4 &= \frac{c_{2j}}{a_j^{-1}} = c_{2j}
\end{align*}
\]

Proof. We have

\[
\rho_{ij} = \rho(K_i, K_j), \text{cov}(K_i, K_j) = \rho_{ij} \times \alpha_i^{-1} \times \alpha_j^{-1}
\]

\[
\begin{pmatrix} K_i \\ K_j \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} a_i^{-2} & \rho_{ij} a_i^{-1} a_j^{-1} \\ \rho_{ij} a_i^{-1} a_j^{-1} & a_j^{-2} \end{pmatrix} \right)
\]

The following steps are based on Fact 2.1.1 on page 13 in Chapter 2.1 of (72)

Fact 2.1.1: Let \(Z_0, Z_1, Z_2\) be independent \(N(0,1)\) variables. For arbitrary but fixed \(\mu\) and \(\Sigma\) such that \(|\rho| < 1\), let

\[ X_1 = \sigma_1(\sqrt{1-|\rho|}Z_1 + \sqrt{|\rho|}Z_0) + \mu_1 \] and \(X_2 = \sigma_2(\sqrt{1-|\rho|}Z_2 + \sqrt{|\rho|}Z_0) + \mu_2 \)  

\begin{equation}
(4.12)
\end{equation}
Conditioned on $Z_0$, $X_1$ and $X_2$ are independent, hence by conditioning on $Z_0 = z$ then unconditional and by $\phi(z) = \phi(-z)$, we have

$$F(x_1, x_2) = \int_{-\infty}^{\infty} \Phi\left(\frac{\sqrt{|\rho|}z + a_1}{\sqrt{1 - |\rho|}}\right)\Phi\left(\frac{\sqrt{|\rho|}z + a_2}{\sqrt{1 - |\rho|}}\right)\phi(z)dz \quad (4.13)$$

where $z$ is standard normal distribution and $a_i = \frac{x_i - m_i}{\sigma_i}$

To prove Equation 4.10, we can write $K_i$ and $K_j$ in similar way as in Equation 4.12 and treat them as independent variables conditioning on $z$ and then integrate out $z$. \qed

**Theorem 4.2.4.** $p^{-2} \sum_{i,j=1}^{p} \sqrt{|\text{cov}_{ij}|}$ tends to 0, where $\text{cov}_{ij}$ is the covariance between $K_i$ and $K_j$.

**Proof.** By Holder inequality

$$\frac{1}{c} + \frac{1}{d} = 1, \quad c, d > 1, \quad \sum_{k=1}^{n} |a_kb_k| \leq \left(\sum_{k=1}^{n} |a_k|^c\right)^{\frac{1}{c}} \times \left(\sum_{k=1}^{n} |b_k|^d\right)^{\frac{1}{d}} \quad (4.14)$$

Let $c = 4, d = \frac{4}{3}, b_k = p^{-\frac{3}{2}}$ and $a_k = (p^{-1} |\text{cov}_{ij}|)^{\frac{1}{2}}$
\[ p^{-2} \sum_{i,j=1}^{p} \sqrt{\text{cov}_{ij}} = \sum_{k=1}^{n} |a_k b_k| \]

\[ = \sum_{i,j=1}^{p} \left( (p^{-1} |\text{cov}_{ij}|)^{\frac{1}{2}} p^{-\frac{3}{2}} \right) \]

\[ \leq \left( \sum_{i,j=1}^{p} \left( p^{-1} |\text{cov}_{ij}| \right)^{\frac{1}{2}} \times \left( \sum_{i,j=1}^{p} p^{-\frac{3}{2}} \right)^{\frac{1}{2}} \right) \]

\[ = p^{-\frac{1}{2}} \left[ \sum_{i,j=1}^{p} |\text{cov}_{ij}|^{\frac{1}{2}} \right] \]

\[ \sum_{i,j=1}^{p} |\text{cov}_{ij}|^2 = \text{trace}(A \times A^T) \]

\[ = \text{trace}(UDU^T \times UDU^T) \]

\[ = \text{trace}(UDDU^T) \]

\[ = \sum_{i=k+1}^{p} \lambda_i^2 \]

Where \( U=\text{eigenvectors}, D=\text{diag}(\lambda_i) \) of \( A, i = k + 1, \ldots, p \)

According to Equation 4.5

\[ \sqrt{\frac{\lambda_{k+1}^2 + \lambda_{k+2}^2 + \cdots + \lambda_p^2}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}} = o(p^{-\frac{1}{2}}) \text{ for } \delta > 0 \] (4.16)

\[ \lambda_1 + \lambda_2 + \cdots + \lambda_p = \text{trace}(\Sigma) = p \]

\[ \sqrt{\frac{\lambda_{k+1}^2 + \lambda_{k+2}^2 + \cdots + \lambda_p^2}{p}} \to 0 \]

Hence \( p^{-2} \sum_{i,j=1}^{p} |\text{cov}_{ij}| \to 0 \)
Next, we will use Taylor expansion to analyze the $\Phi$ function in Equation 4.11 with respect to $\text{cov}_{ij}^{1/2}$ where $\text{cov}_{ij}$ is the covariance of $K_i, K_j$.

**Theorem 4.2.5.**

\[
\int_{-\infty}^{\infty} \left[ \Phi \left( \frac{\sqrt{\rho_{ij}} z + a_1}{\sqrt{1 - \rho_{ij}}} \right) - \Phi \left( \frac{\sqrt{\rho_{ij}} z + a_2}{\sqrt{1 - \rho_{ij}}} \right) \right] \times \left[ \Phi \left( \frac{\sqrt{\rho_{ij}} z + a_3}{\sqrt{1 - \rho_{ij}}} \right) - \Phi \left( \frac{\sqrt{\rho_{ij}} z + a_4}{\sqrt{1 - \rho_{ij}}} \right) \right] \phi(z) dz
\]

\[= \left[ \Phi(c_{1i}) - \Phi(c_{2i}) \right] \times \left[ \Phi(c_{1j}) - \Phi(c_{2j}) \right] + b^{-1} cv [\Phi(c_{1i}) - \Phi(c_{2i})][\Phi(c_{1j}) - \Phi(c_{2j})] + R(cv^2). \tag{4.18} \]

where

\[
a_1 = \frac{c_{1i}}{a_{1i}} = c_{1i} \\
a_2 = \frac{c_{2i}}{a_{1i}} = c_{2i} \\
a_3 = \frac{c_{1j}}{a_{1j}} = c_{1j} \\
a_4 = \frac{c_{2j}}{a_{1j}} = c_{2j}
\]

c_{1i}, c_{2i}, c_{1j} and c_{2j} as defined in Theorem 4..3.
Proof. In the following steps, we denote cov\(_{ij}\) by \(cv\), \(b_{ij}\) by \(b\) and \(c_{1i}\) by \(c\). Since \(k_i = a_i^{-1}\), \(k_j = a_j^{-1}\), \((\sigma_{k_i} \times \sigma_{k_i})^{-\frac{1}{2}} = b_{ij}^{-\frac{1}{2}}\), we have

\[
\Phi \left( \frac{\sqrt{pz} + c_{1i}}{\sqrt{1 - \rho}} \right) = \Phi \left( \frac{\sqrt{\frac{cv}{\sigma_{k_i} \times \sigma_{k_i}}} z + c}{\sqrt{1 - \frac{cv}{\sigma_{k_i} \times \sigma_{k_i}}}} \right) = \Phi \left( \frac{cv_{1i} z + cb_{1i}^{-\frac{1}{2}}}{\sqrt{b - cv}} \right)
\]

Taylor expansion of \(\Phi \left( \frac{cv_{1i} z + cb_{1i}^{-\frac{1}{2}}}{\sqrt{b - cv}} \right)\) with respect to \(cv_{1i}^{\frac{1}{2}}\) around 0. Let \(g = \frac{cv_{1i} z + cb_{1i}^{-\frac{1}{2}}}{\sqrt{b - cv}}\) and \(f = \Phi(g)\).

The first term of Taylor expansion \(f(0) = \Phi(c)\) To derive the second term:

\[
f' = \frac{df}{d cv_{1i}^{\frac{1}{2}}} = \Phi(g) \times \frac{dg}{d cv_{1i}^{\frac{1}{2}}}
\]

\[
= \Phi(g) \times \frac{z \sqrt{b - cv} - (\sqrt{cv}z + \sqrt{b}c) \times \frac{1}{2} \times (b - cv)^{-\frac{1}{2}} \times (-2cv_{1i}^{\frac{1}{2}})}{b - cv}
\]

\[
= \Phi(g) \times \frac{z \sqrt{b - cv} + (\sqrt{cv}z + \sqrt{b}c) \times cv_{1i}^{\frac{1}{2}} \times (b - cv)^{-\frac{1}{2}}}{b - cv}
\]

\[
= \Phi(g) \times \frac{z \sqrt{b - cv} + (cv \times z + b_{1i}^{\frac{1}{2}} \times cv_{1i}^{\frac{1}{2}} \times c) \times (b - cv)^{-\frac{1}{2}}}{b - cv}
\]

(4.19)

It is obvious that \(f'(0) = \Phi(c) \times \frac{z \sqrt{b}}{b} = \Phi(c) \times b^{-\frac{1}{2}} \times z\). Hence the second term of Taylor expansion is: \(f'(0) \times (cv_{1i}^{\frac{1}{2}} - 0) = \Phi(c) \times b^{-\frac{1}{2}} \times cv_{1i}^{\frac{1}{2}} \times z\). To derive the third term:

\[
f'' = \left[ \Phi(g) \right]' \times \frac{d g}{d cv_{1i}^{\frac{1}{2}}} + \Phi(g) \times \frac{d^2 g}{d cv_{1i}^{\frac{1}{2}}}
\]

(4.20)
Next we evaluate \([\phi(g)]'\)

\[
[\phi(g)]' = \frac{d\phi(g)}{dcv^2} = \frac{d\phi(g)}{dg} \times \frac{dg}{dcv^2}
= \phi(g) \times (-g) \times \frac{dg}{dcv^2}
= \phi(g) \times (-g) \times \frac{z \times \sqrt{b - cv} + (cv \times z + \sqrt{c} b \sqrt{cv}) \times \sqrt{b - cv}}{b - cv}
\]

\[
\frac{d - \frac{dg}{dcv^2}}{dcv^2} = \frac{[z \times (b - cv)^{-\frac{1}{2}} \times (\frac{1}{2} \times 2\sqrt{cv} + (2\sqrt{cv}z + \sqrt{bc})(b - cv)^{-\frac{1}{2}})(b - cv)]}{(b - cv)^2}
+ \frac{(z \times cv + \sqrt{b} \sqrt{cvc})(\frac{1}{2})(b - cv)^{-\frac{1}{2}} \times (\frac{1}{2})(b - cv)}{(b - cv)^2}
- \frac{[z \times \sqrt{b - cv} + (z \times cv + \sqrt{b} \sqrt{vc}) \times (b - cv)^{-\frac{1}{2}}] \times (-2\sqrt{cv})}{(b - cv)^2}
\]

(4.21)

Setting \(\sqrt{cv} \) to 0,

\[
[\phi(g)]' = \phi(c) \times (-c) \times z \times b^{-\frac{1}{2}} = -\phi(c)b^{-\frac{1}{2}}cz
\]

\[
\frac{d - \frac{dg}{dcv^2}}{dcv^2} = \frac{(b^{-\frac{1}{2}} \times c \times b^{-\frac{1}{2}}) \times b}{b^2} = b^{-1}c
\]

\[
\frac{dg}{dcv} = b^{-\frac{1}{2}} \times z
\]

\[
\phi(g)' \times \frac{dg}{dcv^2} = -\phi(c)b^{-\frac{1}{2}}cz \times b^{-\frac{1}{2}} \times z = -\phi(c)b^{-1}c z^2
\]

\[
f'' = -\phi(c)b^{-1}c z^2 + \phi(c)b^{-1}c = \phi(c)b^{-1}c(1 - z^2)
\]

(4.22)
Hence the third term is $\frac{1}{2} \times \phi(c) \times b^{-1}c(1 - z^2) \times (\sqrt{cv} - 0)^2 = \frac{1}{2} \times \phi(c) \times b^{-1}c(1 - z^2) \times cv$.

This completes the derivation of Taylor expansion.

Next we will apply Taylor expansion to Equation 4.17.

Equation 4.17 = \[
\int_{-\infty}^{\infty} \left\{ \Phi(c_{1i}) - \Phi(c_{2i}) + b^{-\frac{1}{2}}cv^\frac{1}{2}[\phi(c_{1i}) - \phi(c_{2i})]z \\
+ \frac{1}{2}b^{-1}cv[\phi(c_{1i})c_{1i} - \phi(c_{2i})c_{2i}](1 - z^2)] \times (\Phi(c_{1i}) - \Phi(c_{2i})) \\
+ b^{-\frac{1}{2}}cv^\frac{1}{2}[\phi(c_{1i}) - \phi(c_{2i})]z + \frac{1}{2}b^{-1}cv[\phi(c_{1i})c_{1i} - \phi(c_{2i})c_{2i}](1 - z^2)]\phi(z)dz \right\}
\] (4.23)

Letting

$\Phi(c_{1i}) - \Phi(c_{2i}) = t1$

$\frac{1}{2}b^{-\frac{1}{2}}cv^\frac{1}{2}[\phi(c_{1i}) - \phi(c_{2i})] = t2$

$\frac{1}{2}b^{-1}cv[\phi(c_{1i})c_{1i} - \phi(c_{2i})c_{2i}] = t3$

$\Phi(c_{1j}) - \Phi(c_{2j}) = t4$

$\frac{1}{2}b^{-\frac{1}{2}}cv^\frac{1}{2}[\phi(c_{1j}) - \phi(c_{2j})] = t5$

$\frac{1}{2}b^{-1}cv[\phi(c_{1j})c_{1j} - \phi(c_{2j})c_{2j}] = t6$
Hence

Equation 4.17 = \( \int_{-\infty}^{\infty} t1 \times t4 \times \phi(z)dz \)

\[ + \int_{-\infty}^{\infty} t1 \times t5 \times z\phi(z)dz \]

\[ + \int_{-\infty}^{\infty} t1 \times t6 \times (1 - z^2)\phi(z)dz \]

\[ + \int_{-\infty}^{\infty} t2 \times t4 \times z \times \phi(z)dz \]

\[ + \int_{-\infty}^{\infty} t2 \times t5 \times z^2\phi(z)dz \]

\[ + \int_{-\infty}^{\infty} t2 \times t6 \times z \times (1 - z^2)\phi(z)dz \]

\[ + \int_{-\infty}^{\infty} t3 \times t4 \times (1 - z^2)\phi(z)dz \]

\[ + \int_{-\infty}^{\infty} t3 \times t5 \times z \times (1 - z^2)\phi(z)dz \]

\[ + \int_{-\infty}^{\infty} t3 \times t6 \times (1 - z^2)^2\phi(z)dz \]

Since \( z \) is standard normal, \( \int_{-\infty}^{\infty} z \times \phi(z)dz = Ez = 0 \)

Similarly

\[ \int_{-\infty}^{\infty} z^2 \times \phi(z)dz = Ez^2 = 1 \text{ Hence } \int_{-\infty}^{\infty} (1 - z^2) \times \phi(z)dz = 0 \]

And \( \int_{-\infty}^{\infty} z^3 \times \phi(z)dz = Ez^3 = 0 \)

Hence

\[ \text{Equation 4.17 = t1 \times t4 + t2 \times t5 + 2t3 \times t6} \]

\[ = [\Phi(c_{1i}) - \Phi(c_{2i})] \times [\Phi(c_{1j}) - c_{2j}] \]

\[ + b^{-1}\phi(c_{1i}) - \phi(c_{2i})][\phi(c_{1j}) - \phi(c_{2j})] + R(cv^2) \quad (4.24) \]
The last term is dropped in later derivation.

Next we will derive Cov[I(\(i \leq t|w_l \ldots w_k\)), I(p_j \leq t|w_l \ldots w_k)]

**Theorem 4.2.6.**

\[
\text{Cov}[I(\{i \leq t|w_l \ldots w_k\}), I(p_j \leq t|w_l \ldots w_k)] = a_1 a_j \text{cv}[\phi(c_{1i}) - \phi(c_{2i})][\phi(c_{1j}) - \phi(c_{2j})]
\] (4.25)
Proof.

\[
\text{Cov}[I(p_i \leq t|w_1 \ldots w_k), I(p_j \leq t|w_1 \ldots w_k)]
\]

\[= E[I(p_i \leq t|w_1 \ldots w_k), I(p_j \leq t|w_1 \ldots w_k)]
\]

\[-E[I(p_i \leq t|w_1 \ldots w_k)] \times E[I(p_j \leq t|w_1 \ldots w_k)]
\]

\[= P[I(p_i \leq t|w_1 \ldots w_k), I(p_j \leq t|w_1 \ldots w_k)]
\]

\[-P[I(p_i \leq t|w_1 \ldots w_k)] \times P[I(p_j \leq t|w_1 \ldots w_k)]
\]

\[= P[|z_i| \geq -\Phi^{-1}(\frac{t}{2})|w_1 \ldots w_k, |z_j| \geq -\Phi^{-1}(\frac{t}{2})|w_1 \ldots w_k]
\]

\[-P[|z_i| \geq -\Phi^{-1}(\frac{t}{2})|w_1 \ldots w_k] \times P[|z_j| \geq -\Phi^{-1}(\frac{t}{2})|w_1 \ldots w_k]
\]

\[= \int_{-\infty}^{\infty} \{1 - [\Phi(\frac{\sqrt{\rho}z + a_1}{\sqrt{1-\rho}}) - \Phi(\frac{\sqrt{\rho}z + a_2}{\sqrt{1-\rho}})]
\]

\[\times \{1 - [\Phi(\frac{\sqrt{\rho}z + a_3}{\sqrt{1-\rho}}) - \Phi(\frac{\sqrt{\rho}z + a_4}{\sqrt{1-\rho}})]\} \phi(z) dz
\]

\[-P[|z_i| \geq -\Phi^{-1}(\frac{t}{2})|w_1 \ldots w_k] \times P[|z_j| \geq -\Phi^{-1}(\frac{t}{2})|w_1 \ldots w_k]
\]

(4.26)

where

\[a_1 = \frac{c_{ij}}{a_i} = c_{1i}
\]

(4.27)

\[a_2 = \frac{c_{ij}}{a_i} = c_{2i}
\]

(4.28)

\[a_3 = \frac{c_{ij}}{a_j} = c_{1j}
\]

(4.29)

\[a_1 = \frac{c_{ij}}{a_j} = c_{2j}
\]

(4.30)
We will analyze the integral part in the next step.

\[
\int_{-\infty}^{\infty} \left[ 1 - \left\{ \Phi\left( \frac{\sqrt{T\rho} + a_1}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_2}{\sqrt{T - \rho}} \right) \right\} \right] \cdot \phi(z) \, dz
\]

(4.32)

\[
\times \left[ 1 - \left\{ \Phi\left( \frac{\sqrt{T\rho} + a_3}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_4}{\sqrt{T - \rho}} \right) \right\} \right] \phi(z) \, dz
\]

(4.33)

\[
= \int_{-\infty}^{\infty} \left[ 1 - \left\{ \Phi\left( \frac{\sqrt{T\rho} + a_1}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_2}{\sqrt{T - \rho}} \right) \right\} \right] \cdot \phi(z) \, dz
\]

(4.34)

\[
- \left\{ \Phi\left( \frac{\sqrt{T\rho} + a_3}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_4}{\sqrt{T - \rho}} \right) \right\} \]

(4.35)

\[
+ \left\{ \Phi\left( \frac{\sqrt{T\rho} + a_1}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_2}{\sqrt{T - \rho}} \right) \right\}
\]

(4.36)

\[
\times \left[ \Phi\left( \frac{\sqrt{T\rho} + a_3}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_4}{\sqrt{T - \rho}} \right) \right] \phi(z) \, dz
\]

(4.37)

\[
= 1 - \int_{-\infty}^{\infty} \left\{ \Phi\left( \frac{\sqrt{T\rho} + a_1}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_2}{\sqrt{T - \rho}} \right) \right\} \phi(z) \, dz
\]

(4.38)

\[
+ \left\{ \Phi\left( \frac{\sqrt{T\rho} + a_3}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_4}{\sqrt{T - \rho}} \right) \right\} \phi(z) \, dz
\]

(4.39)

\[
\times \left[ \Phi\left( \frac{\sqrt{T\rho} + a_3}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_4}{\sqrt{T - \rho}} \right) \right] \phi(z) \, dz
\]

(4.40)

\[
= 1 - \int_{-\infty}^{\infty} \left\{ \Phi(c_{11}) - \Phi(c_{21}) + b^{-\frac{1}{2}} cv^{\frac{1}{2}} \phi(c_{11}) - \phi(c_{21}) \right\} \phi(z) \, dz
\]

+ \frac{1}{2} b^{-\frac{1}{2}} cv^{\frac{1}{2}} \phi(c_{11}) - \phi(c_{21}) \right\} \phi(z) \, dz
\]

(4.32)

\[
+ \int_{-\infty}^{\infty} \left\{ \Phi\left( \frac{\sqrt{T\rho} + a_1}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_2}{\sqrt{T - \rho}} \right) \right\}
\]

(4.33)

\[
\times \left[ \Phi\left( \frac{\sqrt{T\rho} + a_3}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_4}{\sqrt{T - \rho}} \right) \right] \phi(z) \, dz
\]

(4.34)

\[
= 1 - \left\{ \Phi(c_{11}) - \Phi(c_{21}) \right\} - \left\{ \Phi(c_{11}) - \Phi(c_{22}) \right\}
\]

(4.35)

\[
+ \int_{-\infty}^{\infty} \left\{ \Phi\left( \frac{\sqrt{T\rho} + a_1}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_2}{\sqrt{T - \rho}} \right) \right\}
\]

(4.36)

\[
\times \left[ \Phi\left( \frac{\sqrt{T\rho} + a_3}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_4}{\sqrt{T - \rho}} \right) \right] \phi(z) \, dz
\]

(4.37)

\[
= 1 - \left\{ \Phi(c_{11}) - \Phi(c_{21}) \right\} - \left\{ \Phi(c_{11}) - \Phi(c_{22}) \right\}
\]

(4.38)

\[
+ \int_{-\infty}^{\infty} \left\{ \Phi\left( \frac{\sqrt{T\rho} + a_1}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_2}{\sqrt{T - \rho}} \right) \right\}
\]

(4.39)

\[
\times \left[ \Phi\left( \frac{\sqrt{T\rho} + a_3}{\sqrt{T - \rho}} \right) - \Phi\left( \frac{\sqrt{T\rho} + a_4}{\sqrt{T - \rho}} \right) \right] \phi(z) \, dz
\]

(4.40)
The second term in Equation 4.26:

\[
= -P[|z_i| \geq -\Phi^{-1}(\frac{t}{2})|w_1 \ldots w_k] \times P[|z_i| \geq -\Phi^{-1}(\frac{t}{2})|w_1 \ldots w_k]
\]

\[
= -(1 - P[|z_i| \leq -\Phi^{-1}(\frac{t}{2})|w_1 \ldots w_k]) \times (1 - P[|z_i| \leq -\Phi^{-1}(\frac{t}{2})|w_1 \ldots w_k])
\]

\[
= -\{1 - [\Phi(c_{1i}) - \Phi(c_{2i})] \} \times \{1 - [\Phi(c_{1j}) - \Phi(c_{2j})] \}
\]

\[
= -1 + [\Phi(c_{1i}) - \Phi(c_{2i})] + [\Phi(c_{1j}) - \Phi(c_{2j})] - [\Phi(c_{1i}) - \Phi(c_{2i})] \times [\Phi(c_{1j}) - \Phi(c_{2j})]
\]

\[\text{(4.42)}\]

Hence

\[\text{Equation 4.26} = \int_{-\infty}^{\infty} \left[ \Phi\left(\frac{\sqrt{\rho z} + a_1}{\sqrt{1 - \rho}}\right) - \Phi\left(\frac{\sqrt{\rho z} + a_2}{\sqrt{1 - \rho}}\right) \right] \times \left[ \Phi\left(\frac{\sqrt{\rho z} + a_3}{\sqrt{1 - \rho}}\right) - \Phi\left(\frac{\sqrt{\rho z} + a_4}{\sqrt{1 - \rho}}\right) \right] \phi(z) \, dz
\]

\[
- [\Phi(c_{1i}) - \Phi(c_{2i})] \times [\Phi(c_{1j}) - \Phi(c_{2j})]
\]

\[\text{(4.43)}\]
By Equation 4.24

\[
\int_{-\infty}^{\infty} \left[ \Phi\left( \frac{\sqrt{\beta}z + a_1}{\sqrt{1-\rho}} \right) - \Phi\left( \frac{\sqrt{\beta}z + a_2}{\sqrt{1-\rho}} \right) \right] \\
\times \left[ \Phi\left( \frac{\sqrt{\beta}z + a_3}{\sqrt{1-\rho}} \right) - \Phi\left( \frac{\sqrt{\beta}z + a_4}{\sqrt{1-\rho}} \right) \right] \phi(z) \, dz
\]

\[= [\Phi(c_{11}) - \Phi(c_{21})] \times [\Phi(c_{1j}) - \Phi(c_{2j})] + b^{-1} cv[\phi(c_{11}) - \phi(c_{21})][\phi(c_{1j}) - \phi(c_{2j})] + R(cv^2) \]

(4.44)

Hence

\begin{align*}
\text{Equation 4.26} & = b^{-1} cv[\phi(c_{11}) - \phi(c_{21})][\phi(c_{1j}) - \phi(c_{2j})] + R(cv^2) \\
& = a_i a_j cv[\phi(c_{11}) - \phi(c_{21})][\phi(c_{1j}) - \phi(c_{2j})] \\
& \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \Quad
Proof. Since \( z_2^{1/2} - \mu_1 - \eta_1 \neq 0 \), \( c_{11} = a_i(z_2^{1/2} - \mu_i - \eta_i) \xrightarrow{\text{a.s.}} \infty \). Hence

\[
\begin{align*}
a_i \phi(c_{11}) &= a_i \times \frac{1}{\sqrt{2\pi}} e^{-\frac{c_{11}^2}{2}} \\
&= \frac{a_i \times \frac{1}{\sqrt{2\pi}} a_i \rightarrow \infty}{e^{\frac{c_{11}^2}{2}}} 0
\end{align*}
\] (4.46)

Similarly \( a_i \phi(c_{21}) \xrightarrow{\text{a.s.}} 0 \). Hence \( [\phi(c_{11}) - \phi(c_{21})]a_i \) is bounded and similarly \( [\phi(c_{11}) - \phi(c_{21})]a_i \) is bounded. Therefore \( a_i a_j [\phi(c_{11}) - \phi(c_{21})][\phi(c_{11}) - \phi(c_{21})] \) is bounded when \( z_2^{1/2} - \mu_1 - \eta_1 \neq 0 \) and \(-z_2^{1/2} - \mu_1 - \eta_1 \neq 0\).

**Theorem 4.2.8.**

\[
\text{Var}(\frac{1}{p} \sum_{i=1}^{p} I(p_i \leq t|w_1 \ldots w_k)) = O_p(p^{-\delta}) \text{ for some } \delta > 0
\] (4.47)

**Proof.**

\[
\text{Var}(\frac{1}{p} \sum_{i=1}^{p} I(p_i \leq t|w_1 \ldots w_k)) = p^{-2} \sum_{i=1}^{p} \text{Var}[I(p_i \leq t|w_1 \ldots w_k)] + 2p^{-2} \sum_{1 \leq i \leq j \leq p} \text{Cov}[I(p_i \leq t|w_1 \ldots w_k), I(p_j \leq t|w_1 \ldots w_k)]
\] (4.48)

Since \( I(p_i \leq t|w_1 \ldots w_k) \) is binary

\[
\text{Var}[I(p_i \leq t|w_1 \ldots w_k)] = p_i(1 - p_i) \leq \frac{1}{4}
\]
Hence the first term of RHS of Equation 4.48 \( \leq \frac{1}{4}pp^{-2} = \frac{1}{4}p^{-1} = O_p(p^{-1}) \)

In Equation 4.48 the second term of RHS is

\[
2p^{-2} \sum_{1 \leq i \leq j \leq p} \text{COV}(I[p_i \leq t|w_1 \ldots w_k], I[p_j \leq t|w_1 \ldots w_k])
= 2p^{-2} \sum_{1 \leq i \leq j \leq p} a_i a_j \times [\phi(c_{i1}) - \phi(c_{21})][\phi(c_{1j}) - \phi(c_{2j})] * cv
\]

(4.49)

Next we prove that \( p^{-2} \sum_{1 \leq i \leq j \leq p} cv \) is bounded by Cauchy-Schwarz inequality.

\[
(\sum_{k=1} a_k b_k)^2 \leq (\sum_{k=1} a_k^2)(\sum_{k=1} b_k^2)
\]

(4.50)

Letting

\[ a_k = |a_{ij}|, \quad \text{where } a_{ij} \text{ is the}(i, j)^{th} \text{ element of } A. \]

\[ b_k = p \]

\[
(\sum_{i,j} |a_{ij}|p)^2 \leq (\sum_{i,j} a_{ij}^2) \sum_{i,t} p^2
\]

\[
p^2(\sum_{i,j} a_{ij})^2 \leq (\sum_{i,j} a_{ij}^2)p^2 \times p^2
\]

\[
(\sum_{i,j} a_{ij})^2 \leq (\sum_{i,j} a_{ij}^2)p^2
\]

(4.51)
Since

\[
\sum_{i,j} a_{ij}^2 = \sum \text{diag}(A \times A^t)
\]

\[
= \sum \text{diag}(Q\Delta Q^t Q \Delta Q^t)
\]

\[
= \sum \text{diag}(Q\Delta Q^t)
\]

\[
= \sum \text{diag}(Q \text{diag}(\lambda_i^2) Q^t)
\]

\[
= \sum_{i=k+1}^p \lambda_i^2
\]

where \(A = Q\Delta Q^t\) and \(Q = (U_{k+1} \ldots U_p), \Delta = \text{diag}(\lambda_{k+1}, \ldots, \lambda_p)\)

Hence

\[
p^{-2} (\sum_{i,j} |a_{ij}|) \leq (\sum_{i,j} a_{ij}^2) = \sum_{i=k+1}^p \lambda_i^2
\]

\[
p^{-1} \sum_{ij} |a_{ij}| \leq \sqrt{\sum_{i=k+1}^p \lambda_i^2}
\]

\[
p^{-2} \sum_{ij} |a_{ij}| \leq p^{-1} \sqrt{\sum_{i=k+1}^p \lambda_i^2}
\]

\[
p^{-2} \sum_{ij} |a_{ij}| \leq \sqrt{\sum_{i=k+1}^p \lambda_i^2} = O(p^{-\delta}) \text{(by Equation 4.5 of Theorem 1)} \quad (4.52)
\]

Therefore, the second term of RHS in Equation 4.48 is bounded. We already proved that the first term is bounded. Hence \(\text{Var}(p^{-1} \sum_{i=1}^p I(p_i \leq t|\omega_1 \ldots \omega_k)) = O_p(p^{-\delta})\) for some \(\delta > 0\) \qed
Theorem 4.2.9.

\[
\lim_{p \to \infty} \left( \frac{1}{p} R(t) - \frac{1}{p} \sum_{i=1}^{p} (\Phi(a_i(z_i + \eta_i + \mu_i)) + \Phi(a_i(z_i - \eta_i - \mu_i))) \right) = 0 \text{ a.s.} \quad (4.53)
\]

Proof. It is obvious that

\[
R(t) = \sum_{i=1}^{p} I[p_i \leq t|w_i \ldots w_k] \quad (4.54)
\]

It is proved in Theorem 4.2.2 that

\[
\Phi(a_i(z_i + \eta_i + \mu_i)) + \Phi(a_i(z_i - \eta_i - \mu_i)) = P(p_i \leq t) \quad (4.55)
\]

Therefore Equation 4.53 can be written as

\[
\lim_{p \to \infty} \left( \frac{1}{p} \sum_{i=1}^{p} [I[p_i \leq t|w_i \ldots w_k] - P(p_i \leq t|w_i \ldots w_k)] \right) = 0 \text{ a.s.} \quad (4.56)
\]

Letting \( x_i = I(p_i \leq t|w_i \ldots w_k) - P(p_i \leq t|w_i \ldots w_k) \) By Lemma 1, Equation 4.56 is correct if we can show

\[
\text{Var} \left( \frac{1}{p} \sum_{i=1}^{p} I[p_i \leq t|w_i \ldots w_k] \right) = O_p(p^{-\delta}) \text{ for some } \delta > 0 \quad (4.57)
\]

By Theorem 4.2.8, Equation 4.57 is true. \( \square \)

We have completed the proof of proposition 2. It is worth noting that the authors made a major mistake in their proof by equating \( P(p_i \leq t|w_i \ldots w_k) \) with \( P(|z_i| < -\Phi(\frac{1}{2})) \). This mistake occurs very early in their proof, rendering their proof irrelevant to the conclusion. In
this section, we have fixed this error and come up with a full proof. It turns out, miraculously, that the conclusion of this paper still holds. The main reason is that the difference of two terms, \( P(p_i \leq t|w_1 \ldots w_k, p_j \leq t|w_1 \ldots w_k) \) and \( P(p_i \leq t|w_1 \ldots w_k) \times P(p_j \leq t|w_1 \ldots w_k) \), is important in the derivation of FDP. By making the same mistake in the derivation of both terms, the mistakes cancels out.

4.3 Application to Data Analysis

4.3.1 Principal Factor Approximation Analysis with Empirical Variance

Since links are nested in the brain of each subject, it is reasonable to assume that they are correlated. As a result of the correlation, the testing statistics (z values and p values) are also correlated. To account for this dependence, we have employed the PFA methodologies discussed in this section by implementing the 'pfa' package in R with some modifications.

The empirical distribution and fitted normal curve of Z statistics are presented in Figure 10. Z statistics roughly follows normal distribution. However, due to dependency among z statistics, the distribution is not \( N(0,1) \) any more, as explained in (71) and (73). The histogram of p values in Figure 11 indicates that the proportion of significant links is probably small based on the intuition that false nulls are associated with small p values. As shown in Figure 12, the number of total discoveries (\( \widehat{R}(t) \)), the estimated number of false discoveries (\( \widehat{V}(t) \)) and the estimated false discovery proportion (\( \widehat{FDP}(t) \)) all decrease when \( (-\log(t)) \) increases. In other words, all three estimates decrease when the threshold t increase.

Table VIII depicts the cutoff Pvalue, the estimated number of false rejection, estimated FDP and significant links when FDP threshold is chosen to be 0.3. At the cutoff=0.3 for FDP,
the number of significant links is 11. In contrast, when cutoff of FDR is set at 0.3, the number of significant links is 12 (see Table VII). These results confirms the results in (69) that B-H tends to under-estimate FDP. Interestingly enough, although the number of significant links is different, the significant links identified by PFA is a subset of links identified by B-H. In other words, PFA results confirms most of the significant linked picked up by B-H method.
4.3.2 Principal Factor Approximation Analysis with Random Subject Effect Removed

Fan’s original method did not take into account random subject effect. We thought that estimation of group effect would be more accurate by partitioning total variation into two parts: (1) difference between autism and control, and (2) random subject effect. To this end, before applying PFA, we removed the random subject effect from fMRI measurement. Shown in Table IX are links deemed significant at FDP cutoff of 0.3. Compared with Table VIII, one more links are determined to be significant. Moreover, most links are associated with smaller p values in Table IX than in Table VIII, confirming that removing random subject effect leads to a smaller standard error of group effect.
4.3.3 Principal Factor Approximation Analysis Assuming Independence

The PFA analysis takes into account the information in covariance structure when estimating FPD. To examine the impact of covariance on the estimation of FPD, we assume independence among the multiple testings. Links considered to be significant by PFA analysis at FDP of 0.3 are shown in Table X.

As seen in Table X, the total number of links determined to be significant is increased to 13, which is similar to the number of links identified by B-H method. Moreover, the name of the links are also the same, indicating that B-H method behaves as PFA when assuming
TABLE VIII: FALSE DISCOVERY PROPORTION-EMPIRICAL COVARIANCE

<table>
<thead>
<tr>
<th>Rejection</th>
<th>Threshold</th>
<th>False Rejection</th>
<th>FDP</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.443971e-05</td>
<td>0.1389437</td>
<td>0.1389437</td>
<td>63-64</td>
</tr>
<tr>
<td>2</td>
<td>3.722867e-05</td>
<td>0.1505131</td>
<td>0.07525655</td>
<td>63-66</td>
</tr>
<tr>
<td>3</td>
<td>4.727702e-05</td>
<td>0.1923040</td>
<td>0.06410134</td>
<td>08-35</td>
</tr>
<tr>
<td>4</td>
<td>9.940049e-05</td>
<td>0.4104297</td>
<td>0.10260743</td>
<td>14-54</td>
</tr>
<tr>
<td>5</td>
<td>2.662898e-04</td>
<td>1.1111470</td>
<td>0.22222941</td>
<td>14-27</td>
</tr>
<tr>
<td>6</td>
<td>2.728632e-04</td>
<td>1.1387159</td>
<td>0.18978598</td>
<td>07-66</td>
</tr>
<tr>
<td>7</td>
<td>2.787491e-04</td>
<td>1.1633980</td>
<td>0.16619972</td>
<td>63-68</td>
</tr>
<tr>
<td>8</td>
<td>3.963025e-04</td>
<td>1.6556069</td>
<td>0.20695086</td>
<td>57-82</td>
</tr>
<tr>
<td>9</td>
<td>4.286914e-04</td>
<td>1.7909627</td>
<td>0.19899586</td>
<td>61-66</td>
</tr>
<tr>
<td>10</td>
<td>6.825214e-04</td>
<td>2.8478712</td>
<td>0.28478712</td>
<td>20-38</td>
</tr>
<tr>
<td>11</td>
<td>7.744417e-04</td>
<td>3.2290131</td>
<td>0.29354664</td>
<td>08-74</td>
</tr>
</tbody>
</table>

independence. These results are not surprising as B-H does not take into account covariance structure.
TABLE IX: FDP WITH RANDOM SUBJECT EFFECT REMOVED

<table>
<thead>
<tr>
<th>Rejection</th>
<th>Threshold</th>
<th>False Rejection</th>
<th>FDP</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.922999e-05</td>
<td>0.1091756</td>
<td>0.10917556</td>
<td>63-66</td>
</tr>
<tr>
<td>2</td>
<td>3.109798e-05</td>
<td>0.1163881</td>
<td>0.05819406</td>
<td>63-64</td>
</tr>
<tr>
<td>3</td>
<td>3.971777e-05</td>
<td>0.1497908</td>
<td>0.04993028</td>
<td>08-35</td>
</tr>
<tr>
<td>4</td>
<td>8.962981e-05</td>
<td>0.3453440</td>
<td>0.08633600</td>
<td>14-54</td>
</tr>
<tr>
<td>5</td>
<td>2.274320e-04</td>
<td>0.8910296</td>
<td>0.17820593</td>
<td>63-68</td>
</tr>
<tr>
<td>6</td>
<td>2.373740e-04</td>
<td>0.9305038</td>
<td>0.15508397</td>
<td>07-66</td>
</tr>
<tr>
<td>7</td>
<td>2.626953e-04</td>
<td>1.0310629</td>
<td>0.14729470</td>
<td>14-27</td>
</tr>
<tr>
<td>8</td>
<td>3.638841e-04</td>
<td>1.4330493</td>
<td>0.17913116</td>
<td>61-66</td>
</tr>
<tr>
<td>9</td>
<td>3.850973e-04</td>
<td>1.5173203</td>
<td>0.16859115</td>
<td>57-82</td>
</tr>
<tr>
<td>10</td>
<td>7.045972e-04</td>
<td>2.7850411</td>
<td>0.27850411</td>
<td>20-38</td>
</tr>
<tr>
<td>11</td>
<td>7.879465e-04</td>
<td>3.1151293</td>
<td>0.28319357</td>
<td>08-74</td>
</tr>
<tr>
<td>12</td>
<td>8.794719e-04</td>
<td>3.4772681</td>
<td>0.28977234</td>
<td>07-08</td>
</tr>
</tbody>
</table>

TABLE X: FDP ASSUMING INDEPENDENCE

<table>
<thead>
<tr>
<th>Rejection</th>
<th>Threshold</th>
<th>False Rejection</th>
<th>FDP</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.998355e-05</td>
<td>0.1044927</td>
<td>0.10449267</td>
<td>63-66</td>
</tr>
<tr>
<td>2</td>
<td>3.189447e-05</td>
<td>0.1111522</td>
<td>0.05557612</td>
<td>63-64</td>
</tr>
<tr>
<td>3</td>
<td>4.070871e-05</td>
<td>0.1418699</td>
<td>0.04728995</td>
<td>08-35</td>
</tr>
<tr>
<td>4</td>
<td>9.166918e-05</td>
<td>0.3194671</td>
<td>0.07986678</td>
<td>14-54</td>
</tr>
<tr>
<td>5</td>
<td>2.320392e-04</td>
<td>0.8086567</td>
<td>0.16173133</td>
<td>63-68</td>
</tr>
<tr>
<td>6</td>
<td>2.421555e-04</td>
<td>0.8439120</td>
<td>0.14065201</td>
<td>07-66</td>
</tr>
<tr>
<td>7</td>
<td>2.679159e-04</td>
<td>0.9336868</td>
<td>0.13338383</td>
<td>14-27</td>
</tr>
<tr>
<td>8</td>
<td>3.708000e-04</td>
<td>1.2922378</td>
<td>0.16152973</td>
<td>61-66</td>
</tr>
<tr>
<td>9</td>
<td>3.923583e-04</td>
<td>1.3673688</td>
<td>0.15192987</td>
<td>57-82</td>
</tr>
<tr>
<td>10</td>
<td>7.167552e-04</td>
<td>2.4978918</td>
<td>0.24978918</td>
<td>20-38</td>
</tr>
<tr>
<td>11</td>
<td>8.013102e-04</td>
<td>2.7925661</td>
<td>0.25386964</td>
<td>08-74</td>
</tr>
<tr>
<td>12</td>
<td>8.941333e-04</td>
<td>3.1160544</td>
<td>0.25967120</td>
<td>07-08</td>
</tr>
<tr>
<td>13</td>
<td>1.034759e-03</td>
<td>3.6061366</td>
<td>0.27739512</td>
<td>65-66</td>
</tr>
</tbody>
</table>
In section 3.2.1, we implemented empirical Bayesian (EB) method to estimate random effect $\gamma_j$. The EB method has been widely applied in many areas of research. It was very popular in the 1960’s to 1990’s (54). For instance, a search of Current index of Statistics on ‘empirical Bayes” yields a median of 2.5 hits per year, 11 hits per year, 32 hits per year, 46 yields per year during 1964-1969, 1970-1979, 1980 to 1989, and 1990-1996, respectively. Despite its apparent success, EB has some inherent pitfalls. First of all, EB has been criticized of ”using data twice”: first in the estimation of hyperparameters and then in the derivation of posterior distribution. Thus it is possible EB over-fits the data. Secondly, EB is considered as an pseudo-Bayesian analysis. Instead of placing a prior on hyperparameters, EB estimates hyperparameters using observed data, and thus violates the central belief in Bayesian statistics that prior should be independent of observed data. Thirdly, as an approximation to so-called fully Bayesian analysis, EB is less powerful. By using a fixed value estimated from data for prior, EB fails to account for the uncertainty involved in the estimation of unknown priors, and thus tends to underestimate the standard errors.

The major advantage of EB is that it is easy to implement. To illustrate this, let’s revisit Equation 3.6. The EB procedure estimates $\eta$ by maximizing the marginal distribution of $m(y|\eta)$ with respect to $\eta$, and uses $p(\theta|y, \eta)$ as the posterior distribution of $\theta$. On the other hand, a fully Bayesian procedure requires integration with respect to the hyperdistribution...
in Equation 3.7. Essentially, EB replaces an integration with maximization. Hence EB is computational straightforward, which at least partially explains its popularity.

In recent years with the advance of novel computational techniques such as Gibbs sampler, fully Bayesian approach becomes possible. In the next section, I will provide a brief summary of fully Bayesian methods such as Bayesian hierarchical modeling.

5.1 Bayesian Hierarchical Model

Bayesian statistics is becoming extremely useful in recent years, although as late as late 1980s, it is still considered as an interesting alternative to the classical frequentist theory.

The major difference between these two school thoughts stems from the fact that Bayesian considers parameters in a model as random variables with their own distributions. Suppose that we are interested in modeling the outcome of flipping coins. Let \( Y \) be the number of heads out of \( N \) coins flips. \( Y \sim \text{Binomial}(p, N) \), where \( p \), the probability of observing heads, is treated as a free parameter. The posterior distribution of \( p \) can be derived via Baye’s theorem

\[
\pi(p|Y) = \frac{P(Y|p)\pi(p)}{P(Y)}
\]  

(5.1)

where \( \pi(p) \), the prior distribution represents the prior belief before data is observed. The term \( P(Y|p) \) is the likelihood function. The term \( P(Y) \) does not contains \( p \). It is treated as a normalizing constant on \( \pi(p|Y) \) since \( \pi(p|Y) \) should be proper, or in other words, \( \int_0^1 \pi(p|Y)dp = 1 \).
1. Due to this reason, the term $P(Y)$ is usually dropped from Equation 5.1. Hence Equation 5.1 can be simplified to

$$\pi(p|Y) \propto P(Y|p)\pi(p)$$

Baye’s theorem in Equation 5.2 can be succinctly expressed as “posterior distribution is proportional to the product of likelihood and prior distribution”.

The term hierarchical Bayes was introduced by (74). Over the years, Bayesian hierarchical modeling has been very useful in modeling complicated problems with multiple parameters. For instance, in a multi-center clinical trial of AIDS treatment, subjects in $i^{th}$ hospital have log hazard of $\theta_i$. It is expected that estimates of $\theta_i$’s should be related to each other. The relationship can be modeled by a prior distribution which treats $\theta_i$’s as a random sample from a common population distribution. The observed data, $Y_{ij}$, which is the survival time of $j^{th}$ subject from $i^{th}$ hospital, depends on $\theta_i$. Therefore, it is common to model this type of problem hierarchically, with observed outcome modeled conditionally on unknown parameters such as $\theta_i$ which can be further modeled conditionally on additional parameters.

Figure 13 (modified from (75)) illustrates the difference between hierarchical model and non-hierarchical model. Figure 13a shows the usual non-hierarchical model where outcome variable $y$’s are treated as random sample from a distribution parameterized by $\theta$. On the other hand, the plot on the right hand side depicts hierarchical model. The observations $y_{ij}$
are categorized into several clusters: $y_{11}, \ldots, y_{1n_1}; y_{21}, \ldots, y_{2n_2}; \ldots$; and $y_{m1}, \ldots, y_{mn_m}$. Hierarchical model specifies the distribution of $y_{ij}$ in the following two steps:

1. The distribution of $y_{ij}$ is jointly determined by (a) parameter $\theta$ which is shared across all clusters (b) parameters $b_i$’s which are shared within a cluster but different between clusters,

2. The cluster specific parameters $b_i$ is parameterized by $\Sigma_b$.

In many statistical applications, data are correlated or connected in certain way. For example, in a study of tumor in rats in several experiment, each experiment should be treated as a cluster. Hierarchical model allows the assessment of within and between cluster effect. On the other hand, non-hierarchical models are usually not suitable for hierarchical data due to the difficulty in identifying correct number of parameters.
Although hierarchical Bayesian analysis has advantages over non-hierarchical method, its implementation used to be a daunting task. To see this, suppose the study data has the following hierarchical structure:

\[ Y_i \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad (5.3) \]
\[ \mu, \sigma^2 \sim \pi(\mu, \sigma^2) \quad (5.4) \]

where \( \pi(\mu, \sigma^2) \) is the joint prior distribution of \( \mu \) and \( \sigma^2 \).

By Baye’s theorem, the joint posterior distribution of \( \mu \) and \( \sigma^2 \) is given by:

\[ \pi(\mu, \sigma^2|Y) \propto L(\mu, \sigma^2)\pi(\mu, \sigma^2) \quad (5.5) \]

where \( \pi(\mu, \sigma^2) \) is the joint posterior distribution. To make statistical inference, marginal posterior is needed. The marginal distribution of one parameter, say \( \mu \), is obtained by integrating out all other parameters. The two marginal posterior distributions can be expressed as:

\[ f(\mu|Y) = \int_{\sigma^2} \pi(\mu, \sigma^2|Y) d\sigma^2 \quad (5.6) \]
\[ f(\sigma^2|Y) = \int_{\mu} \pi(\mu, \sigma^2|Y) d\mu \quad (5.7) \]

This example contains only two parameters: \( \mu \) and \( \sigma^2 \). In practice, a model can contain hundreds of parameters. To derive marginal posterior distribution of a particular parameter, one needs to integrate out all other parameters from the joint posterior distribution. This
problem of high dimensional integration was generally a formidable analytic problem which had hindered the application of hierarchical Bayesian analysis until late 1980s. See (76) for an excellent review.

5.2 Implementation of Hierarchical Bayesian

The implementation of hierarchical Bayesian model used to be a challenging task due to high-dimensional integration. In recent years, the development of computation tools makes it possible to sample from posterior distribution in the presence of high dimension issue. Most of these techniques are based on Markov chain Monte Carlo which will be introduced in this section.

5.2.1 Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) is a computational tool to sample from posterior distribution. It was first introduced into physics in (77). Later developments include generalization of Metropolis algorithm by (78) and invention of Gibbs sampler by (79). Markov chain Monte Carlo was re-discovered by Bayesian scientists in the late 1980s. Nowadays not only has MCMC become the standard computation tools in Bayesian statistics but it also has made significant contributions to the propagation of Bayesian theory.

The main idea of MCMC is to construct a Markov chain that eventually converge to the target distribution (i.e., stationary or equilibrium distribution). Markov chain Monte Carlo produces dependent sample as it is a iterative procedure.
Markov chain is a stochastic process $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(t)}$ such that

$$f(\theta^{(t+1)}|\theta^{(1)}, \ldots, \theta^{(t)}) = f(\theta^{(t+1)}|\theta^{(t)}) \quad \forall t$$  \hspace{1cm} (5.8)

Equation 5.8 suggests that the distribution of $\theta$ at $t + 1$ given all previous time points depends on only the value at the immediate previous time point (i.e., $\theta^{(t)}$). In addition, $f(\theta^{(t+1)}|\theta^{(t)})$ is independent of time $t$. Lastly, under certain conditions, the distribution of $\theta^{(t)}$ converges to its equilibrium distribution. This convergence is independent of the choice of initial values $\theta^{(0)}$.

In order to generate sample from posterior distribution, the Markov chain should converge to target posterior distribution. In addition, it should be easy to sample from the conditional distribution $f(\theta^{(t+1)}|\theta^{(t)})$. Assuming that the Markov chain satisfies the above-mentioned conditions, MCMC algorithm can be summarized as follows (modified from (79)):

1. Choose an initial value $\theta^{(0)}$ for the Markov chain(s).
2. Generate $T$ values (iterations) until the chain(s) reaches equilibrium.
3. Evaluate convergence of the algorithm by examining convergence diagnostics. If the diagnostics fail, generate more observations.
4. Remove the first $B$ observations (i.e., burn-in process).
5. Use the remaining $T - B$ values $\theta^{(B+1)}, \ldots, \theta^{(T)}$ as posterior sample.
6. Obtain the summary statistics of the posterior sample (e.g., mean, median, standard deviation, quantiles and correlations). Perform Bayesian inferences using posterior sample.
After posterior sample of $\theta, \theta^{(1)}, ..., \theta^{(T)}$ is obtained, statistical inference can be made on any function of $\theta$, say $G(\theta)$. The algorithm is as follows (see (80)):

1. Obtain a sample of $G(\theta)$ by plugging in $\theta^{(1)}, ..., \theta^{(T)}$.

2. Obtain summary statistics of $G(\theta)$ using traditional sample estimates. For instance, posterior mean of $G(\theta)$ is $\frac{1}{T} \sum_{t=1}^{T} G(\theta^{(t)})$. Similarly, one can derive other quantities such as the posterior standard deviation, median or quantiles of $G(\theta)$.

## 5.2.2 Metropolis-Hasting Algorithm

Since Metropolis first introduced MCMC method, there have been several developments in expanding original method. These include Metropolis-Hasting algorithm, Gibbs sampler, slice sampler, reversible jump MCMC and perfect sampling. Most of the later developments are more complicated than the original Metropolis algorithm and focus on specific problems.

In this section, we will briefly summarize the two most popular MCMC methods: Metropolis-Hasting algorithm and Gibbs sampler.

Metropolis proposed the original MCMC method in 1953 by using Markov-chain-based simulation to solve problems in physics. The main idea behind Metropolis algorithm is random walk that uses acceptance/rejection rule to converge to target distribution. The main steps to sample from posterior distribution $p(\theta|y)$ can be summarized as follows:

1. Identify a jumping density $J_t(\theta^*|\theta^{t-1})$. Jumping density has to be symmetric, i.e., $J_t(\theta_a|\theta_b) = J_t(\theta_b|\theta_a)$.

2. Draw $\theta^*$ from $J_t(\theta^*|\theta^{t-1})$. 
3. Compute the ratio \( r = \frac{p(\theta^*|y)}{p(\theta^{(t-1)}|y)} \).

4. If \( r \geq 1 \), set \( \theta^t \) to \( \theta^* \). Otherwise set \( \theta^t \) to \( \theta^* \) with probability \( r \) and to \( \theta^{(t-1)} \) with probability \( 1 - r \).

Metropolis’s method was generalized in what is known as the Metropolis-Hasting algorithm (78). The main improvements include (1) the jumping density does not need to be symmetric. (2) the ratio \( r \) is replaced by

\[
    r = \frac{p(\theta^*|y)}{f_t(\theta^*|\theta^{(t-1)})} \frac{f_t(\theta^{(t-1)}|\theta^*)}{p(\theta^{(t-1)}|y)}
\]  

(5.9)

5.2.3 Gibbs Sampler

Gibbs sampler was developed in (79) as a special case of M-H. Suppose that there are \( d \) components of \( \theta \), \( \theta = (\theta_1, \theta_2, \ldots, \theta_d) \). Each iteration of Gibbs sampler cycles through every component(or subvector) of \( \theta \) conditioned on the values of all the others. In other words, the jumping density in density in M-H is replaced by the full conditional posterior distribution

\[
p(\theta_j|\theta_{-j}^{t-1}, y)
\]  

(5.10)

where \( \theta_{-j}^{t-1} \) represents all components of \( \theta \) except for \( \theta_j \):

\[
    \theta_{-j}^{t-1} = (\theta_1^t, \theta_2^t, \ldots, \theta_{j-1}^t, \theta_{j+1}^t, \ldots, \theta_d^t)
\]  

(5.11)

The algorithm of Gibbs sampler can be summarized as follows:
1. Set initial values $\theta^{(0)}$.

2. For $t = 1, \ldots, T$, repeat the following steps

   (a) Set $\theta^{(t)} = \theta^{(t-1)}$.

   (b) For $j = 1, \ldots, d$, update $\theta_j$ from $\theta_j \sim f(\theta_j | \theta_{\setminus j}, y)$.

   (c) Set $\theta^{(t)} = \theta$ and save it as the generated sample at $t + 1$ iteration.

Gibbs sampler has been very popular since at each step it only requires sampling from uni-dimensional distribution. Many statistical packages such as R and SAS provide standard functions that can easily generate random numbers from univariate distributions.

5.2.4 **WinBUGS**

WinBUGS is a free software that generates random numbers from posterior distribution of parameters in Bayesian models. It was developed by statisticians in Medical Research Council Biostatistics Unit in University of Cambridge. The original version, development on UNIX platform, is called BUGS which stands for Bayesian inference Using Gibbs Sampler. The first Windows version, WinBUGS, was released in 1997. WinBUGS can be downloaded from the following website:

http://www.mrc-bsu.cam.ac.uk/software/bugs/

As indicated in section 5.1, sampling from posterior distribution of high-dimensional model is a daunting task that hinders Bayesian analysis for a long time. As WinBUGS implements Gibbs sampler, it allows for sampling sequentially from each parameter’s full conditional distribution. Users of WinBUGS only need to specify Bayesian model and supply data to WinBUGS, which
runs Gibbs sampler automatically in the background. WinBUGS can handle a wide range of posterior distributions. Given its ease of use and broad applications, WinBUGs is becoming the standard software for Bayesian data analysis.

5.2.5 R2WinBUGS

WinBUGS generates random sample from posterior distribution based on Bayesian model, observed data and initial values. It is inconvenient to supply data to WinBUGS using the default "copy and paste" method when working with large datasets. In addition, oftentimes we need to perform further analysis using posterior distribution outputted by WinBUGS. The R package R2WinBUGS is developed to allow for direct calling of WinBUGS from R and importing WinBUGS outputs to R. With R2WinBUGS, users can manipulate study data in R before sending it to WinBUGS. In addition, users can perform further analyses of WinBUGS outputs, such as graphical display and posterior simulation in R. The R package R2WinBUGs is available from CRAN(Comprehensive R Archive Network) at http://cran.r-project.org/.

5.3 Data Analysis Using Bayesian Hierarchical Model

5.3.1 Model Specification

As explained in chapter 5.1, with hierarchical modeling, the joint distribution of data and parameter can be expressed as

\[ f(y|\theta_1)\pi_1(\theta_1|\theta_2)\pi_2(\theta_2|\theta_3)\ldots\pi_k(\theta_k|\lambda) \]  \( (5.12) \)
where \( f(y|\theta_1), \pi_1(\theta_1|\theta_2), \ldots, \pi_k(\theta_k|\lambda) \) specify first, second, \ldots, \( k^{th} \) level of the hierarchical model, respectively.

The main interest is usually the marginal posterior distribution of first level parameters, \( p(\theta_1|y) \). This can be obtained by Gibbs sampling via WinBUGS after the hierarchical model is specified.

To fit Bayesian hierarchical model to our fMRI data, the first level of the model is the same as mixed-effects model described in Chapter 3.1.

\[
Y_{ij} = \beta_{0i} \times (1 - G_j) + \beta_{1i} \times G_j + \gamma_j + \epsilon_{ij} \quad (5.13)
\]

where \( Y_{ij} \) is the fMRI measurement for \( i^{th} \) link from \( j^{th} \) subject, \( i = 1 \ldots m \) and \( j = 1 \ldots n \), \( G_j = 0 \) if subject is a control while \( G_j = 1 \) if the subject has autism. \( \gamma_j \) is the random effect term for \( j^{th} \) subject and \( \epsilon_{ij} \) is the error term. We further assume that \( \gamma_j \sim N(0, \sigma_{\gamma_j}^2) \), \( \epsilon_{ij} \sim N(0, \sigma_{\epsilon_{ij}}^2) \) if subject is from control group and \( \epsilon_{ij} \sim N(0, \sigma_{\epsilon_{ij}}^2) \) if subject has autism. \( \gamma_j \) is independent of \( \epsilon_{ij} \).

The fMRI measurement for \( i^{th} \) link can be expressed as

\[
Y_{ij}|\gamma \sim N(\beta_{0i}, \sigma_{\gamma_{0i}}^2) \quad (5.14)
\]

if subject is from control group and

\[
Y_{ij}|\gamma \sim N(\beta_{1i}, \sigma_{\gamma_{1i}}^2) \quad (5.15)
\]

if the subject is from autism group.
At the second stage of our model, we suppose

\[ \beta_i = \begin{bmatrix} \beta_{0i} \\ \beta_{1i} \end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\beta_{0i}}^2 & 0 \\ 0 & \sigma_{\beta_{1i}}^2 \end{bmatrix} \right) \]

\[ \gamma_j \sim N(0, \sigma_\gamma^2) \]  
\[ \epsilon_{ij} \sim N(0, \sigma_{\epsilon_{0i}}^2) \text{ (for control)} \]
\[ \epsilon_{ij} \sim N(0, \sigma_{\epsilon_{1i}}^2) \text{ (for autism)} \]  

(5.16)

Without prior knowledge about model parameters, we assign flat, non-informative priors to these parameters.

\[ \sigma_{\beta_{0i}}^2 = 10000 \]
\[ \sigma_{\beta_{1i}}^2 = 10000 \]
\[ \sigma_{\delta_{0i}}^2 = 1000 \]
\[ \sigma_{\delta_{1i}}^2 = 1000 \]
\[ \sigma_\gamma^2 = 1000 \]  

(5.17)

5.3.2 Model Update and Diagnosis

The ultimate goal of WinBUGS analysis is to sample from posterior distribution. It is crucial to check whether WinBUGS has converged. However, since the posterior distribution is of unknown nonstandard form in most cases, convergence can not be proven. It is customary to run diagnostic tests to evaluate whether WinBUGS has converged to stationary distribution.
As these tests can not prove convergence, it make senses to run more than one tests. The most often used diagnostics tests to assess convergence include the following:

1. Brooks-Gelman-Rubin (bgr) test

   The Brooks-Gelman-Rubin diagnostics was first introduced by (81) and then expended by (82). The original Gelman-Rubin diagnostics test was a univariate statistic, referred to as the potential scale reduction factor, or PSRF which is defined as

   $$ PSRF = \sqrt{\frac{n-1}{n} + \frac{m+1}{mn} \frac{B}{W}} $$

   (5.18)

   where $m$ is the number of individual chains, $n$ is the the last $n$ samples used in the calculation, $\frac{B}{n}$ is the between chain variance and $W$ is the within chain variance. The idea behind this ANOVA-type diagnostics is that when chains converge to target distribution the between-chain variation should become small relative to the within chain variation, yielding a PSRF close to 1. The PSRF was later extended to a new test that can simultaneously assess the convergence of multiple parameters in the form of multivariate potential scale reduction factor (MPSRF). The relationship between PSRF and MPSRF can be expressed as

   $$ \max_i PSRF_i \leq MPSF $$

   (5.19)

   A rule of thumb for non-convergence is 0.975 quantile of MPSF larger than 1.2.
2. Geweke test

The convergence of each chain can be examined by viewing the set of values simulated by MCMC as a time series (83). The mean from early segment of the chain is compared to the mean in a later segment. A \( z \) test is applied to check whether these two means are equal. If the hypothesis that the means at the beginning and the end of the MCMC output are equal is rejected, then the convergence of the chain cannot be assumed.

As WinBUGS relies on Gibbs sampling to generate random numbers from posterior distribution, it samples sequentially from each parameter’s full conditional distribution. Due to the sequential nature of MCMC algorithm, WinBUGS produces correlated samples from the true joint posterior distribution. It becomes an important issue to monitor the correlation between neighboring draws to make sure that the algorithm is stop at the right iteration.

5.3.3 Analysis Results

We preprocessed fMRI data in R and then fed it into WinBUGS using R2WinBUGS package. The model specified in Chapter 5.3.1 is fitted in WinBUGS.

We chose to run 3 chains for 1100 iterations with a burn-in of first 100 iterations to obtain 3000 sets of samples. Convergence is accessed using Gelman test and Geweke test.

Part of the outputs from Geweke diagnostics test are listed in Table XI. The mean of first 10% samples are compared with last 50% sample in the \( Z \) score calculation. Since the \( Z \) statistic only apply to a single chain, the test is applied separately to each of the three chains. The \( z \) scores from all three chains are not significant at the 0.05 level, indicating no evidence of deviation from convergence.
The BGR diagnostic was also obtained via CODA package. First of all, PSRF does not provide evidence of deviation from convergence as 0.975 quantiles are all less than 1.2. Moreover, MPSRF is 1 which confirms that MCMC has converged.

Convergence can also be investigated graphically via trace-plots which display a time series plot of individual sampled for individual parameter in each chain. If all the values are within a zone without apparent periodicities, then we can assume convergence has been achieved. The trace-plots for several parameters in our model is presented in Figure 14(a). It seems that MCMC has converged based on these plots. In addition, the density plots have nice bell-shape which is consistent with the fact that the posterior has normal distribution.

After convergence was confirmed via multiple tests and graphical tools, we run analysis based on outputs from WinBUGS and applied FDR adjustments. Links with significant difference between autism and control subjects are presented in Table XII.
5.4 Result Comparison: Hierarchical Bayesian vs Empirical Bayesian

A comparison between Table XII and Table VII reveals that for the most part Bayesian hierarchical modeling in this Chapter is consistent with empirical Bayesian analysis in Chapter 3. In fact, 12 links that identified by EB are all confirmed by HB modeling. In addition, HB analysis found 6 extra links.

It seems that EB is a nice approximation to HB when computation is challenging. In fact in this analysis, HB took around 100 hours to converge while it only took 10 hours to implement EB analysis.

Another challenge for HB is to pick a sensible prior. In practice, to specify a prior can require a lot of investigation for statisticians. Although a flat or non-informative prior is
Figure 15: Brain network Bayesian analysis.
<table>
<thead>
<tr>
<th>Num</th>
<th>Link</th>
<th>Mean Diff</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63-66</td>
<td>0.12401059</td>
<td>3.66E-05</td>
</tr>
<tr>
<td>2</td>
<td>63-64</td>
<td>0.11482648</td>
<td>6.85E-05</td>
</tr>
<tr>
<td>3</td>
<td>08-35</td>
<td>0.09312393</td>
<td>6.87E-05</td>
</tr>
<tr>
<td>4</td>
<td>14-54</td>
<td>0.08030308</td>
<td>0.000151718</td>
</tr>
<tr>
<td>5</td>
<td>63-68</td>
<td>0.10010339</td>
<td>0.000283524</td>
</tr>
<tr>
<td>6</td>
<td>14-27</td>
<td>0.06500111</td>
<td>0.000316542</td>
</tr>
<tr>
<td>7</td>
<td>07-66</td>
<td>0.0973133</td>
<td>0.000327247</td>
</tr>
<tr>
<td>8</td>
<td>57-82</td>
<td>-0.08345572</td>
<td>0.000409923</td>
</tr>
<tr>
<td>9</td>
<td>61-66</td>
<td>0.09834726</td>
<td>0.000588685</td>
</tr>
<tr>
<td>10</td>
<td>08-74</td>
<td>-0.08531813</td>
<td>0.000740115</td>
</tr>
<tr>
<td>11</td>
<td>20-38</td>
<td>0.0666724</td>
<td>0.000944472</td>
</tr>
<tr>
<td>12</td>
<td>25-77</td>
<td>0.07971842</td>
<td>0.001273403</td>
</tr>
<tr>
<td>13</td>
<td>61-70</td>
<td>0.07243045</td>
<td>0.00130313</td>
</tr>
<tr>
<td>14</td>
<td>63-72</td>
<td>0.08507215</td>
<td>0.001332772</td>
</tr>
<tr>
<td>15</td>
<td>22-61</td>
<td>0.08128412</td>
<td>0.001361554</td>
</tr>
<tr>
<td>16</td>
<td>65-66</td>
<td>0.10029997</td>
<td>0.001408855</td>
</tr>
<tr>
<td>17</td>
<td>07-08</td>
<td>0.12526509</td>
<td>0.00146067</td>
</tr>
<tr>
<td>18</td>
<td>63-70</td>
<td>0.08580603</td>
<td>0.001479168</td>
</tr>
</tbody>
</table>

usually recommended, in many cases flat prior can be problematic. (84) and (85) indicated that in several instances the seemingly non-informative prior could have undue impact on the posterior distribution.

As pointed out by (54), convergence of MCMC can be very difficult to diagnose as most of the usual diagnostic tools have drawbacks (86). Moreover, with increasing computation power, WinBUGS takes care of hierarchical Bayesian model implementation in the background with high efficiency. As it is becoming easy to fit a complicated hierarchical Bayesian model with
WinBUGS, statisticians tend to fit a hierarchical Bayesian model larger than the data can readily support.

Given the pros and cons of EB and HB, the choice depends on problem at hand. Fortunately in our analysis EB and HB lead to similar results, ensuring the validity of the list of links that are significantly different in autism as compared with control.
CHAPTER 6

META ANALYSIS

Recent years has seen several attempts to explore the brain regions affected by autism. Most studies have examined brain connectivity in autism using resting state fMRI data. The affected regions vary from study to study. For an excellent review, please refer to (87).

Part of the reasons that different studies lead to different results is that these studies are single-center studies and thus the sample size are relatively small. The goal of many fMRI studies is to identify regions with significant changes between patients with certain neurological conditions (experimental group) and normal subjects (control group). To achieve this goal, statistical test is perform to identify significant regions. The null hypothesis states that no difference exists between experimental group and control while the alternative hypothesis states that this difference exists. Results from single center study suffer from both type I and type II errors. Type I error corresponds to false positive. As the number of regions under study is large (in our case, 3864 different links), the chance of making a type I error can not be ignored.

In Chapter 3 and 4, we discussed how to control type I error when multiple regions are under study. Under the assumption that false positive occur randomly across different regions, the analysis that uses data from multiple studies can result in reduction in false positive. The reason is that a false positive result from one study is most likely not reproducible in other studies.
Another drawback associated with single-site study is false negative due to low power. This has been illustrated in many examples. For instance, a meta-analysis of streptokinase in the treatment of myocardial infarction was conducted from 33 studies (88). Patients were randomly assigned to either active treatment or placebo. The sample size varies from study to study with one trial having 20 subjects while two large studies having 12,000 to 17,000 patients. Among 33 studies, only 6 were statistically significant while 27 were not. The meta analysis that combined data from all studies yielded a p value of 0.0000008, which is highly significant. Further investigations revealed that the reason that 27 studies had non-significant results was not due to small treatment effect. In fact, the treatment effect of these studies were larger than 6 studies with significant results. Interestingly enough, the reason that these 27 studies had non-significant results was because these studies had small sample size and thus had low statistical power.

6.1 Vote Counting

In this dissertation, we have worked with fMRI data from 8 sites. As the data from each site is huge, a potential approach is to analyze data from each site separately to reduce computational burden. The significant links are then identified by finding significant links common to each site. The approach is called vote counting in the literature.

In a nutshell, vote counting can be carried out in three steps:

1. All available studies are classified into two categories: those that report significant results and those that do not.
2. The studies are counted up and the categories with most studies are declared as the winner.

3. The relative number of studies that vote for or against a category is considered as the evidence for magnitude of treatment effect.

Vote counting is easy to carried out and it has been widely used in many fields. For instance, a recent Science paper examined the relationship between net primary productivity and species richness (89) in several studies. The authors classified studies into five categories: non-significant, positive linear, negative linear, concave up and concave down. As the most studies were non-significant, they concluded that there was not relationship between productivity and species richness.

In recent years, vote counting has been mostly discarded in scientific research due to its serious pitfalls. Vote counting gave one vote to each study irregardless of its sample size (90). Thus a study with 10 subjects is treated the same as a study with 10000 subjects. Secondly, vote-counting only cares about the relative number of significant results. It tells nothing about magnitude of treatment effect which in many instances is the major interest of research. Thirdly, it has been demonstrated in (90) that the power of vote-counting procedure decrease with the number of studies included in the data integration, a particularly annoying feature.

6.2 **Overview of Meta Analysis**

As seen from previous discussions, pooling data from various studies can increase the reliability of findings and the power of statistical analysis. As such, we have performed several analyses using combined data in previous chapters. Due to above-mentioned reasons, these
analyses prove to be superior to single-site analysis and vote-counting procedure as they reduce both false positive and false negative results.

However these analyses do not explicitly address the difference between sites (studies). In fMRI studies, difference between sites can be caused by variations in scanner strength (91), study population (92) and analysis methods (93). It has been shown that inter-study variation can be a significant source of variation in fMRI studies (94; 91).

Many studies have shown that meta analysis is a technique that makes efficient use of multi-site fMRI data (95). Meta analysis, a term coined by (96), refers to “the statistical synthesis of results from individual studies for the purpose of integrating the findings”.

The goal of each study in a meta analysis is to compare the experiment group with the control group. The difference between experimental group and control group is the effect size. In term of statistical testing framework, this objective of each study can be expressed as

1. $H_0$: effect size=0.
2. $H_a$: effect size is not 0.

One of the central assumptions of meta analysis is that the true effect size is the same for each study in a meta analysis. This true effect size is denoted as $\theta$. The observed effect size $\hat{\theta}_i$ differ from study to study because of random error $\epsilon_i$ in each study.

$$\hat{\theta}_i = \theta + \epsilon_i$$ (6.1)
It follows that if each study has infinite number of subjects, then the sampling error would be zero and the observed effect size would be the same as the true effect size. In practice the sample size in each study is finite, hence the sampling error exists and the effect size is different from study to study.

The goal of meta analysis is to estimate the true effect size $\theta$ from observed effect size $\hat{\theta}_i$. This is accomplished by first compute the observed effect size $\hat{\theta}_i$ and variance $\hat{\sigma}_i^2$ of each study and then meta-analyze $\hat{\theta}_i$ across the studies.

More specifically, $\hat{\theta}_i$ is assumed to have the following distribution:

$$\hat{\theta}_i \sim N(\theta, \hat{\sigma}_i^2) \quad (6.2)$$

The overall effect size can be estimated by calculating the simple mean of $\theta_i$ if all studies in the analysis were equally precise. In many cases, however, some studies were more precise (due to larger sample size) than the others. It makes senses to assign more weights to these studies. In order to obtain an accurate estimate of overall effect size $\theta$ and to perform statistical testing on $\theta$, we calculate the weighted average of $\hat{\theta}_i$ and variance $\hat{\sigma}_i^2$. That is,

$$\hat{\theta} = \sum_{i=1}^{K} w_i \hat{\theta}_i \quad (6.3)$$

$$\hat{\sigma}^2 = \sum_{i=1}^{K} w_i^2 \hat{\sigma}_i^2 \quad (6.4)$$

where $w_i$ is the weight of study $i$ and $K$ is the number of studies included in the meta analysis.
The 95% confidence interval of $\theta$ is

$$\hat{\theta} \pm 1.96\hat{\sigma} \quad (6.5)$$

The null hypothesis of $\theta = 0$ can be tested by the z-test statistic,

$$z = \frac{\hat{\theta} - 0}{\hat{\sigma}} \quad (6.6)$$

There are several ways to assign weight to each study, which include the following:

1. Weighting by the number of studies in the meta analysis as $w_i = \frac{1}{K}$ where $K$ is the number of studies. Each study receives the same weight. This approach is the same as simple arithmetic mean.

2. Weighting by sample size in each study, or $w_i = \frac{n_i}{N}$ where $n_i$ and $N$ are the number of subjects in study $i$ and in all studies, respectively. Although this approach takes into account of difference in sample size among studies and assign more weights to studies with more subjects, it does not adjust for the variation in each study.

3. Weighting by number of subjects in experimental group and control group in each study.

$$w_i = \frac{n_{i0}n_{i1}}{n_{i0} + n_{i1}} \cdot \frac{1}{w} \quad (6.7)$$

$$w = \sum w_i \quad (6.8)$$
where $n_{i0}$ and $n_{i1}$ are numbers of subjects in the control group and in the experimental group in the $i^{th}$ study, respectively. This approach is more accurate than the previous one. However it still does not account for variations within each study.

4. Weighting by variation in each study, $w_i = \frac{1}{\hat{\sigma}_i^2} \frac{1}{w}$ and $w = \sum \frac{1}{\hat{\sigma}_i^2}$. This approach is referred to as the inverse variance approach. Research by (97) has proved that this is the optimal weight for meta-analysis. We will use this weighting in the following chapters.

6.3 Meta Analysis Applied to Study Data

Our fMRI datasets contains fMRI measurements from 361 subjects in 8 medical centers (sites). The first step of meta analysis is to run mixed-effects model described in Chapter 3.1 separately for each site.

Mixed effect analysis by site yields estimates of effect size and its variance from each site. These two estimates can be used for meta analysis. Let $\hat{\beta}_{ijk}$ denotes the mean fMRI measurement for $i^{th}$ site, $j^{th}$ link and $k^{th}$ sample where $k = 0$ and $k = 1$ correspond to control and autism, respectively. Similarly, denote the variance of $\hat{\beta}_{ij}$ by $V_{\hat{\beta}_{ij}}$.

The effect size for $i^{th}$ site $j^{th}$ link, denoted by $D_{ij}$, is the difference between $\hat{\beta}_{i0}$ and $\hat{\beta}_{i1}$,

$$D_{ij} = \hat{\beta}_{ij1} - \hat{\beta}_{ij0} \quad (6.9)$$
The variance of $D_{ij}$ is

$$V_{ij} = HV_{\beta_{ij}}H^t \quad (6.10)$$

$$H = [1, -1] \quad (6.11)$$

The next step of meta analysis is to compute the summary effect. The weight of each site is determined using the inverse variance method, namely,

$$w_{ij} = \frac{1}{V_{ij}} \quad (6.12)$$

The overall effect size for $j$th link, denoted by $\theta_j$, can be estimated by

$$\hat{\theta}_j = \frac{\sum_{i=1}^{k} w_{ij} D_{ij}}{\sum_{i=1}^{k} w_{ij}} \quad (6.13)$$

The estimated $\text{Var}(\hat{\theta}_j)$ is

$$\text{Var}(\hat{\theta}_j) = \frac{1}{\sum_{i=1}^{k} \frac{1}{V_{ij}}} \quad (6.14)$$

The null hypothesis of $\theta_j = 0$ can be tested using $z$ test as described in Chapter 6.2. The resulting p values are subject to FDR adjustment as described in section 3.5.
6.4 Meta Analysis Results

The links identified by meta-analysis as significant between autism subjects and control subjects are listed in Table XIII.

<table>
<thead>
<tr>
<th>Number</th>
<th>Link</th>
<th>p values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63-64</td>
<td>7.00E-07</td>
</tr>
<tr>
<td>2</td>
<td>08-35</td>
<td>7.87E-07</td>
</tr>
<tr>
<td>3</td>
<td>63-66</td>
<td>3.48E-06</td>
</tr>
<tr>
<td>4</td>
<td>14-54</td>
<td>4.33E-05</td>
</tr>
<tr>
<td>5</td>
<td>14-27</td>
<td>5.21E-05</td>
</tr>
<tr>
<td>6</td>
<td>61-66</td>
<td>8.28E-05</td>
</tr>
<tr>
<td>7</td>
<td>63-68</td>
<td>0.000148665</td>
</tr>
<tr>
<td>8</td>
<td>61-70</td>
<td>0.000165115</td>
</tr>
<tr>
<td>9</td>
<td>62-63</td>
<td>0.000247225</td>
</tr>
<tr>
<td>10</td>
<td>07-66</td>
<td>0.00027142</td>
</tr>
<tr>
<td>11</td>
<td>08-74</td>
<td>0.000288559</td>
</tr>
<tr>
<td>12</td>
<td>07-08</td>
<td>0.000330978</td>
</tr>
<tr>
<td>13</td>
<td>22-55</td>
<td>0.000341624</td>
</tr>
<tr>
<td>14</td>
<td>07-62</td>
<td>0.000393634</td>
</tr>
<tr>
<td>15</td>
<td>65-66</td>
<td>0.000397834</td>
</tr>
<tr>
<td>Number</td>
<td>Link</td>
<td>p values</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>----------------</td>
</tr>
<tr>
<td>16</td>
<td>65-76</td>
<td>0.000401205</td>
</tr>
<tr>
<td>17</td>
<td>34-60</td>
<td>0.000425462</td>
</tr>
<tr>
<td>18</td>
<td>63-72</td>
<td>0.000475461</td>
</tr>
<tr>
<td>19</td>
<td>59-62</td>
<td>0.00049853</td>
</tr>
<tr>
<td>20</td>
<td>20-38</td>
<td>0.000506677</td>
</tr>
<tr>
<td>21</td>
<td>63-70</td>
<td>0.000608131</td>
</tr>
<tr>
<td>22</td>
<td>27-52</td>
<td>0.000663006</td>
</tr>
<tr>
<td>23</td>
<td>07-48</td>
<td>0.000692082</td>
</tr>
<tr>
<td>24</td>
<td>38-54</td>
<td>0.000742584</td>
</tr>
<tr>
<td>25</td>
<td>62-65</td>
<td>0.000743391</td>
</tr>
<tr>
<td>26</td>
<td>22-61</td>
<td>0.000969964</td>
</tr>
<tr>
<td>27</td>
<td>57-82</td>
<td>0.001014205</td>
</tr>
<tr>
<td>28</td>
<td>28-47</td>
<td>0.001051393</td>
</tr>
<tr>
<td>29</td>
<td>17-29</td>
<td>0.00128563</td>
</tr>
<tr>
<td>30</td>
<td>15-44</td>
<td>0.001510052</td>
</tr>
<tr>
<td>31</td>
<td>61-62</td>
<td>0.001615111</td>
</tr>
<tr>
<td>32</td>
<td>33-38</td>
<td>0.001683454</td>
</tr>
<tr>
<td>33</td>
<td>21-25</td>
<td>0.001919102</td>
</tr>
<tr>
<td>34</td>
<td>38-74</td>
<td>0.001996688</td>
</tr>
<tr>
<td>Number</td>
<td>Link</td>
<td>p values</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>------------------</td>
</tr>
<tr>
<td>35</td>
<td>17-47</td>
<td>0.00207725</td>
</tr>
<tr>
<td>36</td>
<td>31-80</td>
<td>0.002078801</td>
</tr>
<tr>
<td>37</td>
<td>07-68</td>
<td>0.002301178</td>
</tr>
<tr>
<td>38</td>
<td>33-41</td>
<td>0.002684701</td>
</tr>
<tr>
<td>39</td>
<td>25-77</td>
<td>0.002774739</td>
</tr>
<tr>
<td>40</td>
<td>61-72</td>
<td>0.003072993</td>
</tr>
<tr>
<td>41</td>
<td>05-26</td>
<td>0.003111479</td>
</tr>
<tr>
<td>42</td>
<td>39-68</td>
<td>0.003279553</td>
</tr>
<tr>
<td>43</td>
<td>04-45</td>
<td>0.003314371</td>
</tr>
<tr>
<td>44</td>
<td>08-65</td>
<td>0.00357198</td>
</tr>
<tr>
<td>45</td>
<td>07-25</td>
<td>0.003576744</td>
</tr>
<tr>
<td>46</td>
<td>08-20</td>
<td>0.003856222</td>
</tr>
<tr>
<td>47</td>
<td>08-57</td>
<td>0.003881909</td>
</tr>
<tr>
<td>48</td>
<td>33-62</td>
<td>0.004113618</td>
</tr>
</tbody>
</table>

Comparing Table XIII with Table VII and Table XII, one can see that every link identified by mixed effect analysis or hierarchical analysis as significant are also detected by meta analysis.
Moreover, meta analysis uncovers additional significant links because it is able to account for random error at each site.

Figure 16 is a graphic representation of all the links that are identified as significantly different between autism and control by meta analysis at FDR of 0.3. The regions with more than 4 links are highlighted in red. Regions 7 (BA.13 L.Insular Cortex), 8 (BA.13R.Insular Cortex), 61 (BA.41 L.Primary Auditory Cortex), 62 (BA.41 R.Primary Auditory Cortex) and 63 (BA.42 L.Primary Auditory Cortex) are identified as hub regions as each of them has more than 4 links that are deemed to be significantly different between autism and control. In addition, shown in Figure 17 are the links identified by meta-analysis as exhibiting significant difference between autism and control. The clinical significance of these findings remains to be verified.

6.5 Meta Analysis Discussion

Figure 18 is a forest plot for mean difference at each site for one of the links (link 63-66). Each line in this plot represents a site. A box in the line indicates the mean difference for a site. The size of a box is proportional to weight given to that site in meta analysis. The width of the line represents the CI of mean difference. It is obvious that mean differences are not all the same among sites although they all negative or 0.

One may be tempted to include site as a covariate in the mixed effect model(Chapter 3) and hierarchical Bayesian model(Chapter 5). The risk of incorporating site as a covariate is that sometimes a model can be more complicated than the data can support. This is especially true in this study where, with data from 361 subjects, there are $3486 \times 9 = 31374$ parameters to
estimate even when study is not included in the mixed model or hierarchical model. Another complexity is that although including site as a covariate is conceptually straightforward, it is extremely difficult to implement due to large dataset and computational burden.

The meta analysis described in this chapter provides an sensible solution. In fact, it has been shown before that meta analysis can yield similar analysis results as the analysis that includes study as a covariate. The relative efficiency of analyzing original data with study as a covariate was compared with combining summary statistics(meta analysis) (98). The authors concluded that, “for all commonly used parametric and semiparametric models, there was no
Figure 17: Brain network from meta analysis.
Figure 18: Forest plot for mean difference at each site.

gain in efficiency by analyzing original data⁹. In addition, as indicated in the forest plot, meta
analysis assigns more weight to sites with smaller variation (in other words, data from these
sites are more accurate).
Autism is serious neurological disease characterized by poor social communication abilities and repetitive behaviors or restricted interests. As it is a behavior disease, so far no genetic or biochemical test is available for its diagnosis. Functional magnetic resonance imaging has emerged as an essential tool in autism research, due to its excellent contrast properties, spatial resolution, and temporal resolution.

As human brain is a complex network, fMRI is widely used in autism research to determine whether defects in neuroconnectivity contribute to autism. Early research indicated that neuroconnectivity was reduced in autism. These results led to the under-connectivity theory of autism. Several recent research, however, revealed that neuroconnectivity was actually over-connected in autism.

The apparent discrepancies among autism fMRI results can be partially attributed to small sample size. It is well-known that small sample size can result in increased type I error or reduced power. In order words, with small sample size, fMRI can not only pick up some links that are not truly different between autism and control but also miss some links that are truly different.

Another issue with previous autism research is the lack of FDR control. This issue renders many results to be not repeatable.
The last issue with previous research is inadequate statistical methods. As fMRI results are measured within the same brain, it is unrealistic to assume that all links are independent. Previous research employed simple t test which essentially treated each link independent from each other. Moreover, as fMRI results are measurement by different medical centers, the difference in practice in each center should also be taken into account.

Recent years has seen the establishment of large datasets for fMRI in autism. The ABIDE is the largest repository of autism fMRI data. In this work, we used fMRI results of 361 patient from 8 centers from ABIDE. The increased sample size in this study, compared with around 20-30 subjects in previous studies, alleviates the concern of small sample size.

To account for difference among subjects, we introduced a random subject effect term in mixed effect analysis. Conceptually, this is an adequate analysis. To implement mixed effect model, however, turned out to be much complicated. The ABIDE dataset used in this work was hugh. As no commercially available statistical sofetware was able to fit mixed effect model using ABIDE dataset, we have adopted E-M algorithm. In the E step, we estimated random effect via empirical Bayesian; whereas in the M step we estimated fixed effect through MLE. The algorithm iterates between E step and M step until convergence. The mixed effect analysis identified 234 links that a p value of less than 0.05. After FDR adjustment(FDR=0.30), 12 links are deemed significantly different between autism and control.

The FDR control method (q value) in mixed effect analysis did not account for correlation among multiple tests. This is due to the lack of a FDR procedure that can incorporate dependence structure in FDR control. A groundbreaking paper (69) claimed to be able to control
FDP in the presence of arbitrary covariance structure. The method is called principal factor approximation, or PFA. This paper represents a breakthrough in FDR research. The original paper included a very brief proof. It turned out that the proof contained major mistakes which cast doubt on the validity of the PFA method. In this work, we took pain to go through each step in the derivation. Our theoretical derivation indicated that the result in (69) is still valid despite major missteps in the proof. We then applied PFA in the analysis of ABIDE dataset. We were able to identify 11 links with differential expression after incorporating the dependence structure.

The EB approach in mixed effect analysis is an approximation to full Bayesian method. EB suffers from some drawbacks. In recent years, with the advance in Bayesian computational tools such as MCMC, it becomes possible to directly sample from posterior means even in complicated models. In this work, we took advantage of WinBUG software and fitted fully Bayesian model to ABIDE data.

As above-mentioned three methods all identified links that are significantly different between autism and control, a natural question is that how do these methods compare. Did they pick up similar links? To address this question, Table XIV lists the number of links identified as significantly different at FDR of 0.3

Figure 19 is a Venn plot for significant links identified by mixed, Bayesian or PFA analysis. Among 3486 links analyzed in this work, 3468 links are not significant by any of the three analyses, indicating that autism probably is not caused by profound disruption of neuroconnectivity. 12 links identified by both mixed and Bayesian analysis. Fully Bayesian analysis identified 6
TABLE XIV: NUMBER OF SIGNIFICANT LINKS (FDR=0.3, 3 METHODS)

<table>
<thead>
<tr>
<th>Method: 1=mixed, 2=Bayesian, 3=PFA</th>
<th>Number of significant links</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>1,2</td>
<td>12</td>
</tr>
<tr>
<td>1,3</td>
<td>11</td>
</tr>
<tr>
<td>2,3</td>
<td>11</td>
</tr>
<tr>
<td>1,2,3</td>
<td>11</td>
</tr>
</tbody>
</table>

additional links as compared with mixed analysis which used empirical Bayesian. This result suggests that EB is a good approximation to fully Bayesian. Research have proved that these two analyses can be similar under certain conditions (99). A comparison between mixed analysis and PFA reveals that these two methods lead to almost identical results, indicating that the correlation structure between different links is probably not a determining factor in this dataset.

Another source of variation is difference between sites. The above-mentioned three method do not take into account site difference. One potential solution is to include a random effect term for site in mixed analysis. This approach can lead to a complicated model than the data can support. Moreover the computational burden is also huge. In this work we adopted a meta analysis approach. Briefly, we conducted mixed effect analysis by site. Meta analysis is then conducted based on parameters estimated from site-specific analysis. Meta analysis identified 48 significant links at FDR of 0.3.
To investigate the importance of site variation, Table XV lists number of significant links identified by meta analysis alone or by meta analysis and the other three analyses (mixed, Bayesian or PFA).

Figure 20 is a Venn plot for significant links identified by mixed, Bayesian, PFA or meta analysis. An important conclusion is that meta analysis is able to confirm all the significant links identified by mixed, Bayesian or PFA analysis. Moreover, meta analysis is able to identify 30 extra significant links, indicating that difference between site is an important source of variation that should be taken into account.
### TABLE XV: NUMBER OF SIGNIFICANT LINKS (FDR=0.3, 4 METHODS)

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of significant links</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>48</td>
</tr>
<tr>
<td>1,4</td>
<td>12</td>
</tr>
<tr>
<td>2,4</td>
<td>18</td>
</tr>
<tr>
<td>3,4</td>
<td>11</td>
</tr>
<tr>
<td>1,2,4</td>
<td>11</td>
</tr>
<tr>
<td>1,3,4</td>
<td>11</td>
</tr>
<tr>
<td>2,3,4</td>
<td>11</td>
</tr>
<tr>
<td>1,2,3,4</td>
<td>11</td>
</tr>
</tbody>
</table>

Figure 20: Venn plot for significant links (4 methods).
The 11 links that are identified by all four methods are listed in Table XVI. The hub plot (Figure 21) are also included. Interestingly enough, these 11 links are the same as the links picked up by PFA.

In this work, we have employed four statistical methods to analyze ABIDE dataset. In total, 11 links are considered as significantly different between autism and control by all four methods. Among them, 9 links show significantly decreased connectivity in autism compared with control, whereas only 2 links show slight increase in connectivity in autism (see Table XVI). Our analysis results seem to be more consistent with the under-connectivity theory of autism. The significance of findings in this work need to be verified by clinicians as statistical significance is not always the same as clinical significance.

<table>
<thead>
<tr>
<th>N</th>
<th>Region 1</th>
<th>Region 2</th>
<th>Ctl Mean</th>
<th>Aut.Mean</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BA.42 L.Primary Auditory Cortex</td>
<td>BA.43 R.Subcentral Area</td>
<td>0.12445</td>
<td>-0.00022</td>
<td>63-66</td>
</tr>
<tr>
<td>2</td>
<td>BA.42 L.Primary Auditory Cortex</td>
<td>BA.42 R.Primary Auditory Cortex</td>
<td>0.07344</td>
<td>-0.04197</td>
<td>63-64</td>
</tr>
<tr>
<td>3</td>
<td>BA.13 R.Insular Cortex</td>
<td>BA.3 L.Primary Somatosensory Cortex</td>
<td>0.02158</td>
<td>-0.07116</td>
<td>08-35</td>
</tr>
<tr>
<td>4</td>
<td>BA.19 R.Associative Visual Cortex</td>
<td>BA.38 R. Temporopolar Area</td>
<td>0.05311</td>
<td>-0.02674</td>
<td>14-54</td>
</tr>
<tr>
<td>5</td>
<td>BA.42 L.Primary Auditory Cortex</td>
<td>BA.44 R.IFC pars opercularis</td>
<td>0.0516</td>
<td>-0.04816</td>
<td>63-68</td>
</tr>
<tr>
<td>6</td>
<td>BA.13 L.Insular Cortex</td>
<td>BA.43 R.Subcentral Area</td>
<td>0.19518</td>
<td>0.09752</td>
<td>07-66</td>
</tr>
<tr>
<td>7</td>
<td>BA.19 R.Associative Visual Cortex</td>
<td>BA.25 L.Subgenual cortex</td>
<td>0.01138</td>
<td>-0.05342</td>
<td>14-27</td>
</tr>
<tr>
<td>8</td>
<td>BA.41 L.Primary Motor Cortex</td>
<td>BA.43 R.Dorsal Frontal Cortex</td>
<td>-0.06261</td>
<td>0.021</td>
<td>61-66</td>
</tr>
<tr>
<td>9</td>
<td>BA.41 L.Primary Auditory Cortex</td>
<td>BA.8 R.Subcentral Area</td>
<td>0.20255</td>
<td>0.10401</td>
<td>57-82</td>
</tr>
<tr>
<td>10</td>
<td>BA.21 R.Middle Temporal Gyrus</td>
<td>BA.30 R.Cingulate Cortex</td>
<td>0.0441</td>
<td>-0.02286</td>
<td>20-38</td>
</tr>
<tr>
<td>11</td>
<td>BA.13 R.Insular Cortex</td>
<td>BA.47 R.Inferior Prefrontal Gyrus</td>
<td>0.19128</td>
<td>0.27571</td>
<td>08-74</td>
</tr>
</tbody>
</table>
Figure 21: Network hub plot for significant links identified by all 4 methods.
Appendix A

R CODE FOR ESTIMATING LINEAR MIXED-EFFECTS MODEL

```
setwd("C:/Users/fjie/Documents/Biostatistics/Dissertation/Fei Jie/combined")
ImagingData <- read.table("ImagingData.txt")
names(ImagingData) <- c("ID", "Group", "Region1", "Region2", "R1_R2", "Value")
numSubj0 <- length(unique(ImagingData[ImagingData$Group==0, 1]))
numSubj1 <- length(unique(ImagingData[ImagingData$Group==1, 1]))
GroupSubj <- ImagingData$Group[!duplicated(ImagingData$ID)]
numRegion <- length(unique(ImagingData[, 3])) + 1
numRegComb <- nrow(ImagingData)/(numSubj0 + numSubj1)
Y <- ImagingData$Value
X <- cbind(as.numeric(ImagingData$Group==0), as.numeric(ImagingData$Group==1))
Z <- rep(1, length(ImagingData$ID))
R1_R2 <- unique(ImagingData$R1_R2)

Beta <- matrix(0, numRegComb, 2)
Beta.var <- matrix(0, numRegComb, 4)
Beta.se <- matrix(0, numRegComb, 2)
Random <- matrix(0, (numSubj0 + numSubj1), 1)
Sigma <- matrix(0, numRegComb, 2)
Sigma.ran <- 0.005
sigma.ran.y <- array(0, c(1, 1, (numSubj0 + numSubj1)))
for (iSubj in 1:(numSubj0 + numSubj1)) sigma.ran.y[, , iSubj] <- 0.005
tol <- 1e-5
iter <- 1
diff <- rep(FALSE, numRegComb*2+(numSubj0 + numSubj1)+1+numRegComb*3+2)
```
while (iter < 100 && sum(diff) < length(diff)) {

  print(paste("Iteration", iter))

  # Update Estimates of Beta, Sigma, and Sigma.ran
  sigma.ran.mat <- t(t(Random)%*%Random / (numSubj0 + numSubj1) + apply(sigma.ran.y, 1, mean))
  Sigma.ran.new <- sigma.ran.mat
  Beta.new <- matrix(0, numRegComb, 2)
  Sigma.new <- matrix(0, numRegComb, 2)
  sigma0.inv <- matrix(0, numRegComb, numRegComb)
  sigma1.inv <- matrix(0, numRegComb, numRegComb)
  for (iRegComb in 1:numRegComb) {
    temp <- ImagingData[ImagingData$R1==R1[R2==R1][iRegComb], ]
    y <- Y[ImagingData$R1==R1[R2==R1][iRegComb]]
    x <- X[ImagingData$R1==R1[R2==R1][iRegComb]]
    z <- Z[ImagingData$R1==R1[R2==R1][iRegComb]]
    error0 <- matrix((y-x%*%Beta[iRegComb, ]-as.numeric(Random))[temp$Group==0], ncol = 1)
    error1 <- matrix((y-x%*%Beta[iRegComb, ]-as.numeric(Random))[temp$Group==1], ncol = 1)
    sigma.mat0 <- (t(error0)%*%error0 + sum(sigma.ran.y[, , GroupSubj==0])) / numSubj0
    sigma.mat1 <- (t(error1)%*%error1 + sum(sigma.ran.y[, , GroupSubj==1])) / numSubj1
    Sigma.new[iRegComb, ] <- c(sigma.mat0, sigma.mat1)
    sigma0 <- 1/Sigma.new[iRegComb, 1]
    sigma1 <- 1/Sigma.new[iRegComb, 2]
    # Define Error Covariance Matrix
    diag(sigma0.inv) <- sigma0
    diag(sigma1.inv) <- sigma1
    sigma.e <- matrix(0, numSubj0+numSubj1, numSubj0+numSubj1)
for (iSubj in 1:(numSubj0+numSubj1)) {
    sigma.e[iSubj, iSubj] <- Sigma.new[iRegComb, 1]*(1-GroupSubj[iSubj])
    + Sigma.new[iRegComb, 2]*GroupSubj[iSubj]
}
Beta.new[iRegComb,] <- solve(t(x)%*%solve(sigma.e)%*%x)%*%t(x)%*%solve(sigma.e)
Beta.var[iRegComb,] <- as.numeric(solve(t(x)%*%solve(sigma.e+Sigma.ran.mat%*%t(z))%*%x))
Beta.se[iRegComb,] <- sqrt(diag(solve(t(x)%*%solve(sigma.e+Sigma.ran.mat%*%t(z))%*%x)))
}

# Update EB Estimates and Variance of Random Effects
Random.new <- matrix(0, (numSubj0 + numSubj1), 1)
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 <- Sigma.ran.new%*%solve(Sigma.ran.new+solve(t(z)%*%sigma0.inv%*%z))
    else r <- Sigma.ran.new%*%solve(Sigma.ran.new+solve(t(z)%*%sigma1.inv%*%z))
    Random.new[iSubj,] <- r%*%solve(t(z)%*%x)%*%t(z)
    sigma.ran.y[, , iSubj] <- (1-r)%*%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
Appendix A (Continued)

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

write.table(round(Beta, 5), "Estimate/Beta.txt", quote = FALSE,
sep = "\t", row.names = FALSE, col.names = FALSE)
write.table(round(Beta.var, 5), "Estimate/Beta_var.txt", quote = FALSE,
sep = "\t", row.names = FALSE, col.names = FALSE)
write.table(round(Beta.se, 5), "Estimate/Beta_se.txt", quote = FALSE,
sep = "\t", row.names = FALSE, col.names = FALSE)
write.table(round(Sigma, 5), "Estimate/Sigma.txt", quote = FALSE,
sep = "\t", row.names = FALSE, col.names = FALSE)
write.table(round(Random, 5), "Estimate/Random.txt", quote = FALSE,
sep = "\t", row.names = FALSE, col.names = FALSE)
write.table(round(Sigma.ran, 5), "Estimate/Sigma.ran.txt", quote = FALSE,
sep = "\t", row.names = FALSE, col.names = FALSE)
Appendix B

R CODE FOR SIMULATING NEUROIMAGING DATA

```r
setwd("M:/simulation")
rm(list=ls())
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 <- 25
numSubj1 <- 25
numRegion <- 50
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
  }
}
```

Appendix B (Continued)

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(1:numSubj, each = numRegComb),
                           rep(0:1, numRegComb*c(numSubj0, numSubj1)),
                           rep(R1, numSubj),
                           rep(R2, numSubj),
                           rep(0, numRegComb*numSubj),
                           rep("fMRI", each = numRegComb*numSubj),
                           rep(R1_R2, numSubj)), stringsAsFactors = FALSE)
names(DataStruc) <- c("ID", "Group", "R1", "R2", "Value", "Measure", "R1_R2")

mean.beta <- c(0, 0.05)
var.beta <- c(0.1, 0.2)
sig.gamma <- 0.1
SimResult <- vector("list", numSim)

set.seed(7)
# Simulate Data
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 <- rep(0.5, numRegComb)
  sig02 <- rep(0.5, numRegComb)
# sig11 <- rgamma(numRegComb, 0.3, 1)
# sig12 <- rgamma(numRegComb, 1.5, 1)
gamma <- rnorm((numSubj), 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)
    emat0 <- sig01
    emat1 <- sig02

    ImagingData[ImagingData$R1_R2==R1_R2[iRegComb] & ImagingData$Group==0, 5] <- as.
    numeric(rnorm(numSubj0, beta[iRegComb, 1], emat0) + gamma[1:numSubj0])
    ImagingData[ImagingData$R1_R2==R1_R2[iRegComb] & ImagingData$Group==1, 5] <- as.
    numeric(rnorm(numSubj1, beta[iRegComb, 2], emat1) + gamma[1:numSubj1+numSubj0])
}

# write.table(ImagingData, paste("ImagingData", iSim, ".txt", sep = ""),
# append = FALSE, quote = FALSE, row.names = FALSE, col.names = FALSE)
Y <- as.numeric(ImagingData$Value)
X <- cbind(as.numeric(ImagingData$Measure=="fMRI" & ImagingData$Group==0), as.numeric(
            ImagingData$Measure=="fMRI" & ImagingData$Group==1))

Z <- as.numeric(factor(ImagingData$Measure)) - 2
source("Model Estimate_temp.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, Beta.var=Beta.var)
Appendix B (Continued)

```r
save(SimResult, numSim, file = "Simulation Result28.Rdata")
```
Appendix C

R2WINBUGS CODE FOR BAYESIAN HIERARCHICAL MODEL

#this program performs Bayesian Mixed effect analysis using dataset of interest
(either site data or combined data)
rm(list=ls())
setwd("C:/Users/fjie/Documents/Biostatistics/Dissertation/Fei Jie/Winbug/combined")
library(R2WinBUGS)

# Write down the model in a text file and save it as a .bug type
# Define or generate the parameter in the models of WinBUGS
imaging.modle<function() {
  for (i in 1:m) { # RegComb
    for (j in 1:n) { # Subj

      Y[i, j] ~ dnorm(mu[i, j], tau[i, j])
      mu[i, j] <- Beta[i, 1]*(1-Group[j]) + Beta[i, 2]*Group[j] + gamma[j]
      tau[i, j] <- tau0[i]*(1-Group[j]) + tau1[i]*Group[j]
    }
  # Priors for fixed effects
  Beta[i, 1:2] ~ dnorm(mu.beta[ ], tau.beta[ , ])  
  # Measures of interest
  Beta.diff[i] <- Beta[i,1] - Beta[i, 2]
  tau0[i] ~ dgamma(0.001,0.001)
  sigma0[i] <- 1/tau0[i]
  tau1[i] ~ dgamma(0.001,0.001)
  sigma1[i] <- 1/tau1[i]
Appendix C (Continued)

```r
for (j in 1:n)
  { gamma[j] ~ dnorm(mu.gamma, tau.gamma) } 

tau.gamma ~ dgamma(0.001, 0.001) 

sigma.gamma <- 1/tau.gamma 

filename <- "imaging_new.bug" 

# write model file: 
write.model(imaging.modle, filename)

# dataset to be used by winbugs
ImagingData <- read.table("ImagingData.txt") 

names(ImagingData) <- c("ID", "Group", "R1", "R2", "R1_R2", "Value") 

ImagingData$Measure="fMRI" 

ImagingData <- ImagingData[,c(1,2,3,4,6,7,5)] 

numSubj <- length(unique(ImagingData[,1])) 

numRegion <- length(unique(ImagingData[,3])) + 1 

numRegComb <- nrow(ImagingData)/numSubj 

attach(ImagingData) 

Y <- matrix( Value, numRegComb, numSubj) 

GroupSubj <- Group[! duplicated(ID)] 

Y0 <- Y[GroupSubj==0, ] 

Y1 <- Y[GroupSubj==1, ] 

detach(ImagingData) 

n0 <- dim(Y0)[2]; n1 <- dim(Y1)[2] 

data <- list(m = numRegComb, n = numSubj, Y = Y, Group = GroupSubj, 

      mu.beta = rep(0, 2), tau.beta = diag(rep(0.0001, 2)), mu.gamma = 0) 

# initial value

inits <- function()

  list(Beta = matrix(0, numRegComb, 2), gamma = rep(0, numSubj),
```
Appendix C (Continued)

```r
tau0 = rep(1, numRegComb),
tau1 = rep(0.5, numRegComb),
tau.gamma = 1)
}

inits()

# call WinBUGS
imaging.sim <- bugs(data, inits, model.file = "C:/Users/fjie/Documents/Biostatistics/Dissertation/Fei Jie/Winbug/combined/imaging_new.bug",
parameters.to.save = c("Beta", "Beta.diff", "sigma0", "sigma1"), n.thin=1,
# n.thin=1, which means that we keep tracking the samples without dropping any of them
# if not setting n.thin, it will drop some terms
n.chains = 3, n.iter = 1100, n.burnin=100, debug=TRUE, codaPkg=TRUE,
# burning = 100, n.iter = MCMCsamples(each chain)+n.burn
bugs.directory = "C:/Users/fjie/Documents/WinBUGS14/"
# debug=TRUE will stop in WinBUGS rather than only shows the results

# posterior probability
attach.bugs(imaging.sim)
bdiff <- data.frame(t(rbind(apply(Beta.diff, 2, mean), apply(Beta.diff, 2, sd))))
names(bdiff) <- c("mean", "sd")
bdiff$pv <- 2*pnorm(0, abs(bdiff$mean), bdiff$sd)
detach.bugs()

# get link name
R1_R2 <- unique(ImagingData$R1_R2)
bdiff <- data.frame(R1_R2, bdiff)
# FDR
source("FDRfunction.R")
sink("WinBUGS - Mean diff FDR.txt")
```
for (q in 1:8*0.05) {
    print(paste("fMRI - Significant Regions FDR, q =", q))
    temp <- FDR(bdiff$pv, q)
    if (!any(is.na(temp))) {
        result <- bdiff[temp$Row,]
        rownames(result) <- 1:nrow(result)
        print(result)
    }
    cat("\n")
}
library(gdata)

read_beta <- function(indir, site)
{
  name <- paste("C:/Users/fjie/Documents/Biostatistics/Dissertation/Fei Jie/", indir, sep ="")
  setwd(name)
  vname <- paste("v", site, sep = "")
  bname <- paste("b", site, sep = "")
  ImagingData <- read.table("ImagingData.txt")
  names(ImagingData) <- c("ID", "Group", "Region1", "Region2", "R1_R2", "Value")
  numSubj0 <- length(unique(ImagingData[ImagingData$Group==0, 1]))
  numSubj1 <- length(unique(ImagingData[ImagingData$Group==1, 1]))
  GroupSubj <- ImagingData$Group[! duplicated(ImagingData$ID)]
  numRegion <- length(unique(ImagingData[, 3])) + 1
  numRegComb <- nrow(ImagingData)/(numSubj0 + numSubj1)
  Beta <- read.table("Estimate/Beta.txt", sep = "\t",
                     na.strings = ".", stringsAsFactors = FALSE)
  names(Beta) <- c("fMRI0", "fMRI1")

  Beta.var <- read.table("Estimate/Beta_var.txt", sep = "\t",
                         na.strings = ".", stringsAsFactors = FALSE)

  #rm(list=setdiff(ls(), c("region", "R1", "R2", "R1_R2", "FDR", "Beta", "Beta.var", "n0", "n1", "numRegComb", "vectomat")))
  R1_R2 <- unique(ImagingData$R1_R2)
  H1 <- matrix(c(1, -1), nrow = 1, byrow = TRUE)
  varb <- numeric(numRegComb)
Appendix D (Continued)

```r
wav <- numeric(numRegComb)
for (iRegComb in 1:numRegComb) {
  beta <- as.vector(Beta[iRegComb, ])
  beta.var <- matrix(as.numeric(Beta.var[iRegComb, ]), nrow = 2)
  varb[iRegComb] <- as.numeric(H1%*$beta.var*$t(H1))
  wav[iRegComb] <- (beta[,2] - beta[,1]) / varb[iRegComb]
}
Beta_out <- data.frame(R1, R2, varb, wav)
colnames(Beta_out) = c("R1", "R2", vname, bname)
return(Beta_out)
}

rm(list = setdiff(ls(), lsf.str()))
beta_nyu <- read_beta("nyu", 1)
keep(read_beta, beta_nyu)

beta_caltech <- read_beta("caltech", 2)
keep(read_beta, beta_nyu, beta_caltech)

beta_olin <- read_beta("olin", 3)
keep(read_beta, beta_nyu, beta_caltech, beta_olin)

beta_sbl <- read_beta("sbl", 4)
keep(read_beta, beta_nyu, beta_caltech, beta_olin, beta_sbl)

beta_sdsu <- read_beta("sdsu", 5)
keep(read_beta, beta_nyu, beta_caltech, beta_olin, beta_sbl, beta_sdsu)

beta_pitt <- read_beta("pitt", 6)
```
Appendix D (Continued)

```r
keep(read_beta,beta_nyu,beta_caltech,beta_olin,beta_sbl,beta_sdsu,beta_pitt)
beta_sjh <- read_beta("sjh",7)

keep(read_beta,beta_nyu,beta_caltech,beta_olin,beta_sbl,beta_sdsu,beta_pitt,beta_sjh)
beta_stanford <- read_beta("stanford",8)

#total

setwd("C:/Users/fjie/Documents/Biostatistics/Dissertation/Fei Jie/Meta")
list.of.data.frames = list(beta_nyu,beta_caltech,beta_olin,beta_sbl,beta_sdsu,beta_pitt,
beta_sjh,beta_stanford)
tot = Reduce(function(...merge(..., all=T), list.of.data.frames)
tot$v <- 1/tot$v1+1/tot$v2+1/tot$v3+1/tot$v4+1/tot$v5+1/tot$v6+1/tot$v7+1/tot$v8
tot$b <- tot$b1+tot$b2+tot$b3+tot$b4+tot$b5+tot$b6+tot$b7+tot$b8
tot$z <- tot$b/sqrt(tot$v)
tot$pv <- 2*pnorm(0, abs(tot$b), sqrt(tot$v))
tot <- tot[!is.na(tot$pv),]
sum(tot$pv < 0.05, na.rm = TRUE)
myvars <- c("R1,R2", "pv")
tot_n <- tot[myvars]

#FDR=0.3
source("FDRfunction.R")
q <- 0.3

temp <- FDR(tot_n$pv, q)
if (lany(is.na(temp))) {
  Num <- temp$Num
  write.table(tot_n[temp$Row, ],
```

Appendix D (Continued)

```r
region <- read.table("C:/Users/fjie/Documents/Biostatistics/Dissertation/Fei Jie/autism_region_names.csv")
sig.r <- tot.n[temp$Row, ]
sig.r$r1 <- substr(sig.r[,1],1,2)
sig.r$r1 <- ifelse(sig.r$r1 < 10, substr(sig.r$r1,2,2), sig.r$r1)
sig.r$r2 <- substr(sig.r[,1],4,5)
sig.r$r2 <- ifelse(sig.r$r2 < 10, substr(sig.r$r2,2,2), sig.r$r2)
Num <- 1:nrow(sig.r)
Sig.Regions.fMRI <- cbind(Num, region[sig.r[order(sig.r$pv), 3], ],
                          region[sig.r[order(sig.r$pv), 4], ],
                          sig.r[order(sig.r$pv), ])
names(Sig.Regions.fMRI) <- c("Num","Region 1","Region 2","R1_R2","Pval.fMRI","r1","r2")
write.csv(Sig.Regions.fMRI[,c(1,2,3,5)], paste("Meta.Sig.Regions.fMRI FDR = ", q, ".csv", sep = "", row.names = FALSE))
```

CITED LITERATURE


VITA

NAME: Fei Jie

EDUCATION: Bachelor of Medicine, Shanghai Medical University, 1990
PhD, Immunobiology, Iowa State University, 2005
MS, Statistics, Iowa State University, 2005
PhD, Biostatistics, University of Illinois at Chicago, 2016

