**Distribution Agreement**

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Tian Dai Date

# Agreement Methods for Complex Outcomes in Biomedical Studies

By

Tian Dai
Doctor of Philosophy

Biostatistics

---
Ying Guo, Ph.D.
Advisor

---
Amita K. Manatunga, Ph.D.
Committee Member

---
Limin Peng, Ph.D.
Committee Member

---
Tanja Jovanovic, Ph.D.
Committee Member

Accepted:

---
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

---
Date

# Agreement Methods for Complex Outcomes in Biomedical Studies

By

Tian Dai M.S., Emory University, 2014

Advisor: Ying Guo, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy
in Biostatistics
2016

**Abstract**

# Agreement Methods for Complex Outcomes in Biomedical Studies

By Tian Dai

In biomedical studies, researchers are often interested in evaluating the similarity of measurements produced by different methods on the same subjects. Given the importance of agreement studies, there has been extensive literature on measuring agreement for standard outcomes. However, very limited work has been done to assess agreement for complex biomedical outcomes. In this dissertation, we propose novel agreement methods for two complex outcomes often encountered in biomedical studies, survival data and high dimensional neuroimaging data.

The first objective of this dissertation is to study the local agreement pattern between two survival outcomes. Most classical agreement methods have been focused on global summary statistics, which cannot be used to describe various local agreement patterns. In this work, we propose a new agreement measure based on bivariate hazard functions to characterize the local agreement pattern between two correlated survival outcomes. The proposed measure naturally accommodates censored observations, fully captures the dependence structure between bivariate survival times and provides detailed information on how the strength of agreement evolves over time.

Next, we investigate statistical methods for assessing the reproducibility of neuroimaging data in multi-site studies. Considering the special features of imaging data, such as high dimensionality, we propose a two-stage network-based method to effectively and efficiently assess the similarity between the same subject's brain images acquired at different sites. In the first stage, we reduce the dimensionality of imaging data by extracting active functional networks under experimental conditions and estimating the corresponding temporal dynamics. In the second stage, we propose functional agreement indices to measure the agreement between the same subject's network-specific temporal dynamics estimated from brain images acquired across different sites.

Imaging-based brain connectivity measures based on resting-state fMRI data have become an important tool for investigating the pathophysiology, progression and treatment response of psychiatric disorders. In the last part of this dissertation, we propose a prediction method based on Bayesian hierarchical model that uses individual's earlier scans, coupled with relevant baseline characteristics, to predict the individual's future functional connectivity. The proposed prediction method could provide a useful tool to predict the changes in individual patient's brain connectivity with the progression of disease. It can also be used to predict a patient's brain connectivity after a specified treatment regimen which could potentially help guide individualized treatment plan. Another utility of the proposed method is that it could be applied to test-retest reproducibility imaging data to develop a more reliable estimator for individual functional connectivity.

# Agreement Methods for Complex Outcomes in Biomedical Studies

By

Tian Dai
M.S., Emory University, 2014

Advisor: Ying Guo, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy
in Biostatistics
2016

# Acknowledgments

First of all, I would like to express my deepest gratitude to my advisor Dr. Ying Guo for her guidance, support and patience throughout my dissertation research. Her systematic training, insightful suggestions, and continuous encouragement have helped me to a very great extent to accomplish this work.

My sincere gratitude also goes to Dr. Amita K. Manatunga, Dr. Limin Peng, and Dr. Tanja Jovanovic for their time and effort serving as my committee members and reviewing my dissertation. Their constructive comments and suggestions along the way have greatly improved this dissertation.

In addition, I would like to thank all the members in Center of Biomedical Imaging Statistics (CBIS) and agreement study group at Emory. I feel so fortunate to be a member of these two wonderful groups and have benefited a great deal from their dedicated help and inspiring discussions.

Moreover, I would like to extend my gratitude to all the faculties, staffs and fellow students in the Department of Biostatistics and Bioinformatics at Emory. I have had an enjoyable experience during the past years. Especially, I want to thank Mr. Kirk A. Easley for providing me with a lot of opportunities and guidance in statistical consulting.

Last but not least, I am indebted to my parents and my whole family for encouraging and supporting me to pursue this degree. Their unconditional love and consistent support is my most powerful motivation.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Overview

## 1.1 Motivation

In biomedical studies, researchers are often interested in evaluating the similarity of measurements produced by different raters or methods on the same subjects. For example, an outcome may be measured by a gold standard approach and an approximate method and it is of interest to decide whether the approximate method can act as a reasonable replacement for the gold standard (Laurent, 1998). In these cases, we need to assess the agreement between the correlated measurements. Given the importance of agreement studies in biomedical sciences, there has been extensive literature on measuring agreement for categorical and continuous outcomes (Cohen, 1960; Lin, 1989 etc.). However, very limited work has been done to assess agreement for complex biomedical outcomes. In my dissertation, we have proposed novel agreement methods for two types of biomedical data. In the first project, our goal is to assess agreement between correlated survival times with censored observations. In the second project, we are interested in measuring agreement for high dimensional neuroimaging data with replicated scans. In the third project, we propose a prediction method for brain functional connectivity in resting-state fMRI data. As will be elaborated later, considerable statistical challenges are involved in developing agreement methods that can appropriately address the data complications in biomedical studies.

The objective of our first project is to study the local agreement pattern between two continuous measurements subject to censoring (i.e. survival outcomes). Survival outcomes are frequently observed in biomedical studies and it may be of interest to assess the agreement between survival times measured on the same subjects using different methods. For example, in depression studies (Musselman et al., 2001), the time of onset of clinical depression is measured using both clinician-administered Hamilton depression rating scale (HAM-D) and a patient self-report dimensional instrument (Carroll-D). Evaluating agreement between the disease onset times based on the different instruments is useful in assessing whether the less time-consuming and easier to use patient self-report instrument

could be adopted as a reasonable replacement for the clinician-administered instrument. It is clear that the data setting considered here is more challenging than the typical case of complete observations of continuous measurements because the outcomes can be censored. Furthermore, as the measurements are observed over time, it is of interest to study how local agreement pattern changes over time.

The motivation for our second project is an increasing trend of conducting multi-site neuroimaging studies. The multi-site imaging studies have become popular due to a number of significant advantages (Friedman et al., 2006; Friedman and Glover, 2006). First of all, multi-site studies provide larger sample sizes and hence more statistical power for detecting the effect of interest when the effect size is relatively small in imaging studies. Secondly, they increase the generalizability of research conclusions by recruiting subjects from diverse geographic and demographic backgrounds. Additionally, they provide the capability to study rare psychiatric diseases. For example, the Functional Biomedical Informatics Research Network (fBIRN) has recently conducted a multi-site study sponsored by the NCCR/NIH to study regional brain dysfunction related to the progression and treatment of schizophrenia. Ideally, merging data across sites requires the interchangeability of data from different sites and is reasonable only if site differences in fMRI data can be minimized. However, one major challenge arises when combining data in multi-site imaging studies is that even the same subject' brain images can vary considerably across MRI sites since they are acquired using different scanners and protocols. Thus, it is crucial to effectively measure the reproducibility of brain images acquired from various sites before conducting any further analysis in order to draw reliable conclusions in multi-site studies. If the agreement between sites is low, sources causing the discrepancy need to be identified and addressed.

To assure the success of large-scale studies in the fBIRN project, researchers have conducted a pilot traveling-subject study prior to all other studies to evaluate the reproducibility of fMRI images acquired at different imaging sites. Fig 1.1 demonstrates a

general data structure for traveling-subject fMRI studies. In this fBIRN traveling-subject

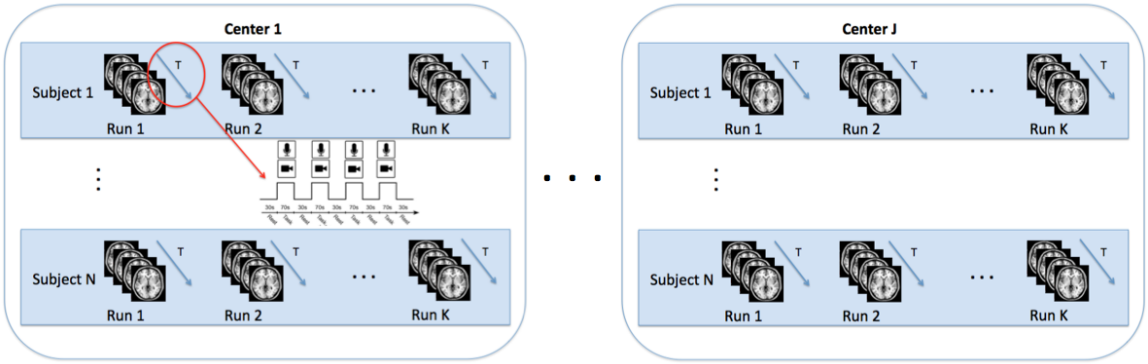



Figure 1.1: Data Structure of the Traveling Subjects Study

study, five healthy, English-speaking males in their 20s were imaged on two consecutive days on 10 different scanners located in nine geographically diverse sites within the United States. For each subject, a group of images, including T1 weighted scans, two cognitive fMRI tasks (auditory oddball (AO) and Serial Item Recognition Paradigm (SIRP) tasks), two calibration scans (breath hold (BH) and sensory motor (SM) tasks) and a resting state scan, were acquired during each visit under the identical tasks protocols across sites. Each site was instructed to design their own optimal imaging protocol. Therefore, scanning parameters, such as scanner vendor, field strength, k-space trajectory, echo time, head coil type, might vary across sites and introduced between-site variability. This study provides an excellent dataset to evaluate statistical tools for assessing reproducibility of fMRI images in multi-site studies (Zou et al., 2005; Friedman and Glover, 2006; Friedman et al., 2006, 2008).

Network-oriented analysis of fMRI, especially rs-fMRI, has revealed important associations between abnormal neural functional connectivity and brain disorders such as schizophrenia, major depression and Alzheimer's disease. Imaging-based brain connectivity measures have become important tools for investigating the pathophysiology, progression and treatment response of psychiatric disorders. Recent studies have begun to investigate the possibility of using functional neuroimaging to predict disease progression and guide treatment selection for individual patients. In the third project, we are interested in developing a prediction method based on Bayesian hierarchical model that uses individual's

earlier scans, coupled with relevant baseline characteristics, to predict the individual's future functional connectivity. The proposed prediction method is very general and can accommodate various study designs in the neuroimaging studies by modeling with different design matrices and distributional assumptions. In Chapter 4, we illustrate the practical utilities of the proposed prediction method using two real imaging studies– a longitudinal imaging study (ADNI2) and a test-retest reproducibility study (Kirby21).

In this dissertation, we develop new agreement methods, pursuing the advantages described above while appropriately handling the data complexities presented in real studies. In the next section, we present an outline of this dissertation.

## 1.2 Outline

In Chapter 2, we present a novel local agreement index for correlated bivariate continuous survival times based on the hazard functions. For estimation of the proposed agreement measure, we develop a non-parametric approach which does not impose any assumptions on marginal survival distributions or the dependence structure between the two survival times. Hence, the local agreement pattern reflected by our proposed measure is not restricted by specific dependence structures. We use kernel methods (Fermanian, 1997) to estimate the hazard function. Since survival outcomes are inherently bounded in the positive region, we propose to improve the estimation accuracy by adopting appropriate boundary kernel correction. The proposed estimators for the agreement measure have desirable asymptotic properties such as strong consistency and asymptotic normality. We perform simulation studies to evaluate the performance of the estimation method. We use a prostate cancer study to demonstrate the practical utility of our method.

In Chapter 3, we propose a two-stage network-based agreement framework to assess the agreement/reproducibility between images across sites. Our agreement method can effectively and efficiently characterize the similarity between functional networks extracted from

high dimensional neuroimaging data acquired at different sites. We present our method for task-related functional magnetic resonance imaging (task-fMRI), which is the most commonly used imaging modality. We develop nonparametric estimation methods for the proposed indices and establish their asymptotic properties. We evaluate the accuracy of the proposed estimation procedure via simulation studies. The proposed methods are applied to the fBIRN Traveling-Subject study to investigate site effect in multi-site imaging studies.

In Chapter 4, we present a general framework for predicting individual future resting-state functional connectivity (RSFC) based on his/her baseline rs-fMRI and relevant clinical and demographic characteristics, such as disease stage or treatment group. We illustrate the model specification for two types of imaging studies – longitudinal imaging studies which investigate disease progression and treatment-related changes in functional connectivity and test-retest studies which aim to investigate and improve the reliability of RSFC. To evaluate the accuracy of the proposed method, we conduct simulations studies using K-fold cross-validation approach. We illustrate the application of the proposed method using two real imaging datasets: a longitudinal ADNI2 data and a test-retest Kirby21 data. In both simulation studies and real data applications, we show that the proposed method provides a more reliable prediction over the alternative methods.

In Chapter 5, we provide a summary of our completed work and plans for future work.

# Chapter 2

# Agreement Methods for Survival Outcomes.

## 2.1 Introduction

In biomedical studies, researchers are often interested in assessing agreement on measurements taken on the same subjects using different methods or by different raters. There has been extensive literature on assessing agreement and making appropriate inference (Cohen, 1960, 1968; Fleiss, 1971; Kraemer, 1980; Agresti et al., 1995; Lin, 1989, 2000; Barnhart et al., 2007). For continuous data, various measures including both scaled measures, such as concordance correlation coefficient (CCC) (Lin, 1989) and coefficient of individual agreement (Barnhart et al., 2007), and unscaled ones, such as total deviation index (Lin, 2000), have been proposed. For categorical data, the kappa coefficient and its extensions (Cohen, 1960, 1968; Fleiss, 1971; Kraemer, 1980; Agresti et al., 1995) have been widely used.

All of the aforementioned methods take the strategy of quantifying the agreement of interest by a global summary measure. While being simple, they have been criticized for their limitations in fully capture agreement information (Tanner and Young, 1985; Darroch and McCloud, 1986). For example, with two categorical scales, Agresti (1989) showed that when a simple quasi-symmetry model holds for the contingency table, the kappa contains all relevant information about the structure of the agreement. However, when the quasi-symmetry model fails, various agreement patterns can produce the same kappa value and the kappa coefficient alone is not capable of distinguishing different agreement patterns. Given the limitation of kappa coefficient, Tanner and Young (1985), Agresti (1988, 1992) and others have promoted studying the structure of agreement instead of summarizing agreement into a single measure. To this end, they proposed alternative approaches based on log-linear modeling. For continuous data, researchers (Borg et al, 1995; Schild et al., 2000) have argued to use descriptive tools such as a scatter plot of two continuous readings with the 45 degree line as the concordance line or Bland and Altman (1986) plot of the difference against the average of paired measurements with the horizontal line of zero as the reference. These plots can be informative, in particular in visualizing the local patterns of agreement, and help capture the changes of measurement concordance along with the

magnitude of measurements. However, they are purely descriptive methods and cannot provide inference regarding agreement. A formal tool for assessing local agreement pattern would be desirable, but has been lacking in literature.

The objective of this chapter is to study the local agreement pattern between two continuous measurements subject to censoring (i.e. survival outcomes). Survival outcomes are frequently observed in biomedical studies and it may be of interest to assess the agreement between survival times measured on the same subjects using different methods. For example, in depression studies (Musselman et al., 2001), the time of onset of clinical depression is measured using both clinician-administered Hamilton depression rating scale (HAM-D) and a patient self-report dimensional instrument (Carroll-D). Evaluating agreement between the disease onset times based on the different instruments is useful in assessing whether the less time-consuming and easier to use patient self-report instrument could be adopted as a reasonable replacement for the clinician-administered instrument. It is clear that the data setting considered here is more challenging than the typical case of complete observations of continuous measurements because the outcomes can be censored. Furthermore, as the measurements are observed over time, it is of interest to study how local agreement pattern changes over time.

Most existing agreement measures with censored survival data have been focused on global measures (Liu et al., 2005; Guo and Manatunga, 2007, 2010; Guo et al., 2013) and cannot be used to characterize local agreement patterns. The descriptive tools for agreement patterns, such as scatter plot and Bland and Altman plot, are not applicable for survival data due to the presence of censoring. Recently, Zeng et al. (2015) proposed a time-dependent agreement method to assess temporal pattern of agreement between two time-to-event endpoints. Their approach is likelihood-based and thus requires parametric distribution assumptions and also relies on the assumption that the two event times have a positive probability of being identical.

In this chapter, we propose a local agreement measure for survival data to describe the local agreement pattern by considering the bivariate hazard function with an adjustment made to the expected chance agreement alone. It is sensible to consider hazard functions in this setting, due to its interpretability and analytical simplifications related to censoring. An appealing feature of our proposed local agreement measure is that it has a nice connection with the local kappa coefficient developed for assessing local agreement between two discrete survival times (Guo and Manatunga, 2005). Specifically, the proposed agreement measure can be viewed as an extension of the unscaled local kappa coefficient for continuous survival times. Our approach taken here is conceptually different from the modeling approaches suggested by Tanner and Young (1985), Agresti (1988, 1992), and Zeng et al. (2015), since we do not fit a series of models specified under specific types of dependence structures for describing the agreement pattern.

For estimation of the proposed agreement measure, we develop a non-parametric approach which does not impose any assumptions on marginal survival distributions or the dependence structure between the two survival times. Hence, the local agreement pattern reflected by our proposed measure is not restricted by specific dependence structures. We use kernel methods (Fermanian, 1997) to estimate the hazard functions. Since survival outcomes are inherently bounded in the positive region, we propose to improve the estimation accuracy by adopting appropriate boundary kernel correction. The proposed estimators for the agreement measure have desirable asymptotic properties such as strong consistency and asymptotic normality. We perform simulation studies to evaluate the performance of the estimation method. We use a prostate cancer study to demonstrate the practical utility of our method.

## 2.2 Methods

### 2.2.1 Proposed local agreement pattern measure $\varphi(t_1, t_2)$

In this section, we first present the proposed local agreement pattern measure and discuss its properties. To set notations, let $T_1$ and $T_2$ denote a pair of continuous survival times of the same individual based on different raters or methods. The joint survival function of the correlated survival times $(T_1, T_2)$ is denoted as $S(t_1, t_2) = Pr(T_1 \geq t_1, T_2 \geq t_2)$. To motivate our local agreement measure, consider a local contingency table for the conditional survival status of $T_1$ and $T_2$ at time $(t_1 + \Delta t_1, t_2 + \Delta t_2)$ given that $T_1 \geq t_1$ and $T_2 \geq t_1$, where $(t_1, t_2)$ is any time point within the support of the bivariate survival function, and $\Delta t_1$ and $\Delta t_2$ represent small time intervals around $t_1$ and $t_2$ respectively. Given $T_1 \geq t_1$ and $T_2 \geq t_2$, the joint survival status of $T_1$ and $T_2$ within the rectangular region $[t_1, t_1 + \Delta t_1) \times [t_2, t_2 + \Delta t_2)$ can be described with a $2 \times 2$ table with the four cells representing all possible situations: $T_1 \geq t_1 + \Delta t_1, T_2 \geq t_2 + \Delta t_2$; $T_1 \in [t_1, t_1 + \Delta t_1), T_2 \geq t_2 + \Delta t_2$; $T_1 \geq t_1 + \Delta t_1, T_2 \in [t_2, t_2 + \Delta t_2)$; and $T_1 \in [t_1, t_1 + \Delta t_1), T_2 \in [t_2, t_2 + \Delta t_2)$. The corresponding cell probabilities are defined as follows:

$$
\begin{aligned}
P_{00}(t_1, t_2) &= Pr\left(T_1 \geq t_1 + \Delta t_1, T_2 \geq t_2 + \Delta t_2 \mid T_1 \geq t_1, T_2 \geq t_2\right) \\
&= \frac{S(t_1 + \Delta t_1, t_2 + \Delta t_2)}{S(t_1, t_2)}, \\
P_{10}(t_1, t_2) &= Pr\left(T_1 \in [t_1, t_1 + \Delta t_1), T_2 \geq t_2 + \Delta t_2 \mid T_1 \geq t_1, T_2 \geq t_2\right) \\
&= \frac{S(t_1, t_2 + \Delta t_2) - S(t_1 + \Delta t_1, t_2 + \Delta t_2)}{S(t_1, t_2)}, \\
P_{01}(t_1, t_2) &= Pr\left(T_1 \geq t_1 + \Delta t_1, T_2 \in [t_2, t_2 + \Delta t_2), \mid T_1 \geq t_1, T_2 \geq t_2\right) \\
&= \frac{S(t_1 + \Delta t_1, t_2) - S(t_1 + \Delta t_1, t_2 + \Delta t_2)}{S(t_1, t_2)}, \\
P_{11}(t_1, t_2) &= Pr\left(T_1 \in [t_1, t_1 + \Delta t_1), T_2 \in [t_2, t_2 + \Delta t_2), \mid T_1 \geq t_1, T_2 \geq t_2\right) \\
&= \frac{S(t_1, t_2) - S(t_1 + \Delta t_1, t_2) - S(t_1, t_2 + \Delta t_2) + S(t_1 + \Delta t_1, t_2 + \Delta t_2)}{S(t_1, t_2)}.
\end{aligned}
$$

We define the following measure $k(t_1 + \Delta t_1, t_2 + \Delta t_2)$ to capture the chance-corrected local agreement between the two survival times within the rectangular

region $[t_1, t_1 + \Delta t_1) \times [t_2, t_2 + \Delta t_2)$ as,

$$k(t_1 + \Delta t_1, t_2 + \Delta t_2) = P_{00}(t_1, t_2) + P_{11}(t_1, t_2) - P_{1+}(t_1, t_2)P_{+1}(t_1, t_2) - P_{0+}(t_1, t_2)P_{+0}(t_1, t_2), \tag{2.1}$$

where $P_{0+}(t_1, t_2) = P_{00}(t_1, t_2) + P_{01}(t_1, t_2)$, $P_{+0}(t_1, t_2) = P_{00}(t_1, t_2) + P_{10}(t_1, t_2)$, $P_{1+}(t_1, t_2) = P_{11}(t_1, t_2) + P_{10}(t_1, t_2)$, $P_{+1}(t_1, t_2) = P_{11}(t_1, t_2) + P_{01}(t_1, t_2)$ are the marginal probabilities in the $2 \times 2$ table. From (2.1), we note that $k(t_1 + \Delta t_1, t_2 + \Delta t_2)$ is closely related to a local Kappa coefficient proposed by Guo and Manatunga (2005) for measuring local agreement between discrete survival times. Specifically, the observed local agreement probability for the two survival outcomes in the $2 \times 2$ table is $P_{00}(t_1, t_2) + P_{11}(t_1, t_2)$. The expected agreement probability when the two survival outcomes are locally independent is $P_{1+}(t_1, t_2)P_{+1}(t_1, t_2) + P_{0+}(t_1, t_2)P_{+0}(t_1, t_2)$. Therefore, $k(t_1 + \Delta t_1, t_2 + \Delta t_2)$ is the unscaled version of the local Kappa coefficient (Guo and Manatunga, 2005) defined based on the $2 \times 2$ table representing the local region $[t_1, t_1 + \Delta t_1) \times [t_2, t_2 + \Delta t_2)$.

Motivated by (2.1), we propose a local agreement measure $\varphi(t_1, t_2)$ by taking the limit of the regional chance-corrected agreement function $k(t_1 + \Delta t_1, t_2 + \Delta t_2)$ as the region area goes to 0. Specifically, $\varphi(t_1, t_2)$ is defined as follows,

$$\varphi(t_1, t_2) = \frac{1}{2} \lim_{\Delta t_1, \Delta t_2 \to 0} \frac{k(t_1 + \Delta t_1, t_2 + \Delta t_2)}{\Delta t_1 \Delta t_2} = \lambda_{11}(t_1, t_2) - \lambda_{10}(t_1, t_2)\lambda_{01}(t_1, t_2), \tag{2.2}$$

where $\lambda_{11}$,$\lambda_{10}$ and $\lambda_{01}$ are bivariate hazard functions defined as,

$$\lambda_{11}(t_1,t_2) = \lim_{\Delta t_1,\Delta t_2\to 0} \frac{Pr\left(T_1\in[t_1,t_1+\Delta t_1), T_2\in[t_2,t_2+\Delta t_2) \mid T_1\geq t_1, T_2\geq t_2\right)}{\Delta t_1 \Delta t_2},$$
$$\lambda_{10}(t_1,t_2) = \lim_{\Delta t_1\to 0} \frac{Pr\left(T_1\in[t_1,t_1+\Delta t_1) \mid T_1\geq t_1, T_2\geq t_2\right)}{\Delta t_1}, and$$
$$\lambda_{01}(t_1,t_2) = \lim_{\Delta t_2\to 0} \frac{Pr\left(T_2\in[t_2,t_2+\Delta t_2) \mid T_1\geq t_1, T_2\geq t_2\right)}{\Delta t_2}.$$

According to Equations (2.1) and (2.2), when there is no local agreement beyond expected by chance, our agreement measure $\varphi(t_1,t_2)$ equals 0. When the two survival outcomes have stronger agreement than expected by chance at $(t_1,t_2)$, $\varphi(t_1,t_2)$ has a positive value, with larger value indicating stronger local agreement. When the local agreement is weaker than the expected by chance, $\varphi(t_1,t_2)$ becomes negative. Since $\varphi(t_1,t_2)$ is defined based on the bivariate hazard functions which are not bounded above, $\varphi(t_1,t_2)$ is not restricted within a fixed range, which means it belongs to the category of unscaled agreement measures.

### 2.2.2 Properties of $\varphi(t_1,t_2)$ as a local agreement pattern measure

The proposed local agreement pattern measure $\varphi(t_1,t_2)$ has a connection with the cross ratio (Clayton, 1978; Oakes, 1982, 1986, 1989), a commonly used local dependence measure for continuous bivariate survival times. For any $(t_1,t_2)$ within the support of the bivariate survival function, the cross ratio $\theta(t_1,t_2)$ is defined as,

$$\theta(t_1,t_2) = \frac{S(t_1,t_2)\frac{d^2}{dt_1dt_2}S(t_1,t_2)}{\frac{d}{dt_1}S(t_1,t_2)\frac{d}{dt_2}S(t_1,t_2)}, \tag{2.3}$$

where $\theta(t_1,t_2)$ equals 1 when there is local independence between the two survival outcomes and is great than 1 when there is positive local dependence, with a higher

value indicating stronger positive dependence. From the definitions, we can show that

$$\varphi(t_1, t_2) = \lambda_{10}(t_1, t_2)\lambda_{01}(t_1, t_2)(\theta(t_1, t_2) - 1). \tag{2.4}$$

Equation (2.4) suggests that the local agreement pattern measure $\varphi(t_1, t_2)$ changes in the same direction as the local dependence measure $\theta(t_1, t_2)$. That is, as the dependence between $T_1$ and $T_2$ increases, the local agreement increases. Also, $\varphi(t_1, t_2)$ becomes zero when there is local independence between the two survival times. Another insight from Equation (2.4) is that when $\theta(t_1, t_2)$ and $\frac{\lambda_{10}(t_1,t_2)+\lambda_{01}(t_1,t_2)}{2}$ are fixed, $\varphi(t_1, t_2)$ increases as the absolute difference between $\lambda_{10}(t_1, t_2)$ and $\lambda_{01}(t_1, t_2)$ decreases. In other words, given constant strength of local dependence and average conditional marginal hazards, our local agreement measure increases when the conditional marginal hazards of $T_1$ and $T_2$ become more similar. This result shows that $\varphi(t_1, t_2)$ not only reflects the local dependence between the bivariate survival outcomes but also captures the homogeneity in the local marginal hazards of the two survival processes. This is a desirable feature for an agreement measure as described in Lin (1989) and Guo and Manatunga (2005).

To illustrate graphically, how the pattern of this local agreement measure captures both the strength of dependence and the marginal homogeneity of bivariate survival times, we first generate three random samples of bivariate survival times with the sample size of 500 from the Clayton model with the same marginal survival functions but with different cross ratios, which reflect different degrees of dependence. Next, we introduce 50% censoring to the bivariate survival times. In Fig 2.1(a) and 2.1(b), we plot the bivariate survival times with complete data and censored data, respectively. In Fig 2.1(a), when the two survival times have homogeneous marginal distributions, they are mostly distributed along the 45 degree line with more concentration along

this line when there is higher strength of dependence. However, when the survival times are subject to censoring (Fig 2.1(b)), this pattern is less obvious and it becomes harder to infer the strength of dependence between the two survival outcomes based on the scatter plot of the observed survival times. In Fig 2.1(c), we present the heatmaps representing the local agreement measure $\varphi(t_1, t_2)$ on the two dimensional time plane. The heatmaps of $\varphi(t_1, t_2)$ clearly demonstrate different patterns of the local agreement measure for survival times with different strengths of dependence. Specially, when the two survival times have the same marginal distributions and the highest dependence, the $\varphi(t_1, t_2)$ has the highest values along the 45 degree line, i.e. $t_1 = t_2$, and decreases fast as moving away from the 45 degree line, i.e. as $|t_1 - t_2|$ increases. When the two survival times are less dependent, $\varphi(t_1, t_2)$ is less elevated on the 45 degree line and decreases much slower as $(t_1, t_2)$ moving away from the 45 degree line. As a comparison, we also present heatmaps of the cross ratio in Fig 2.1(d), which remains constant across the whole time space in each scenario.

To demonstrate how the pattern of the $\varphi(t_1, t_2)$ can capture the marginal heterogeneity between two survival times, we generate three samples of bivariate survival times with the sample size of 500 from the Clayton model with the same cross ratio but with different marginal survival functions. Similarly, we introduce 50% censoring to the bivariate survival times. In Fig 2.2(a), with heterogeneous marginal distributions, the bivariate survival times are shifted away from the 45 degree line. This pattern, however, is not obvious from the scatter plots with censored observations in Fig 2.2(b). Fig 2.2(c) shows that the pattern of $\varphi(t_1, t_2)$ clearly captures this type of disagreement between the survival times with the highest local agreement area moving away from the 45 degree line as the two marginal survival functions become different. In comparison, the cross ratio in Fig 2.2(d) does not reflect the difference in the marginal distributions. Furthermore, we note that the

two different types of disagreement, i.e. marginal heterogeneity or low dependence, are not distinguishable when using global agreement measures such as Lin's CCC.

Another appealing feature of the proposed local agreement measure is that it can fully capture the dependence structure between the correlated survival times. As shown in the Proposition 1 of Appendix A1, the survival function $S(t_1, t_2)$ is jointly determined by the marginal survival functions $S_1(t_1)$, $S_2(t_2)$ and an integrable local agreement pattern measure function $\varphi(t_1, t_2)$ within the finite region $\boldsymbol{\tau}$.

### 2.2.3 Estimation and inference for $\varphi(t_1, t_2)$

We develop a nonparametric estimation method for the proposed local agreement pattern measure in the presence of censoring via kernel smoothing techniques. Let $(T_{i1}, T_{i2})$ $(i = 1, \cdots, n)$ be independent and identically distributed pairs of survival times observed from $n$ independent subjects, which are subject to bivariate censoring by a pair of independent random variables $\vec{C}_i = (C_{1i}, C_{2i})$. The observed data consist of random vectors $\left(\tilde{T}_{i1}, \tilde{T}_{i2}, \delta_{i1}, \delta_{i2}\right)$ $(i = 1, \cdots, n)$, where $\tilde{T}_{ij} = min\left(T_{ij}, C_{ij}\right)$ and $\delta_{ij} = I\left(T_{ij} \leq C_{ij}\right)$ for $j = 1, 2$. We propose to estimate $\varphi\left(t_1, t_2\right)$ by plugging in kernel estimators of the hazard functions. That is, the proposed estimator is given by

$$\hat{\varphi}(t_1, t_2) = \hat{\lambda}_{11}\left(t_1, t_2\right) - \hat{\lambda}_{10}\left(t_1, t_2\right)\hat{\lambda}_{01}\left(t_1, t_2\right), \tag{2.5}$$

where

$$\begin{aligned}
\hat{\lambda}_{10}(t_1, t_2) &= \int_0^{t_1} \frac{1}{h} K_1(\frac{v_1 - t_1}{h}) d\hat{\Lambda}_{10}(dv_1, t_2), \\
\hat{\lambda}_{01}(t_1, t_2) &= \int_0^{t_2} \frac{1}{h} K_1(\frac{v_2 - t_2}{h}) d\hat{\Lambda}_{01}(t_1, dv_2), and \\
\hat{\lambda}_{11}(t_1, t_2) &= \int_0^{t_1} \int_0^{t_2} \frac{1}{h^2} K_2(\frac{v_1 - t_1}{h}, \frac{v_2 - t_2}{h}) d\hat{\Lambda}_{11}(dv_1, dv_2).
\end{aligned}$$

Here, $\hat{\vec{\Lambda}} = \left(\hat{\Lambda}_{10}, \hat{\Lambda}_{01}, \hat{\Lambda}_{11}\right)^T$ are the empirical estimators of the corresponding cumulative hazard functions $\vec{\Lambda} = (\Lambda_{10}, \Lambda_{01}, \Lambda_{11})^T$ (Dabrowska, 1988), $K_1$ is a one-dimensional kernel function which has support on $[-1, 1]$ with integral one and has bounded variation and bounded first derivative, $K_2$ is a two-dimensional kernel function with similar properties, and $h$ is the smoothing parameter, which also known as the bandwidth. In Section 2.2.4, we will present more details on the choice of kernel functions and bandwidths.

Next, we establish the asymptotic properties of the proposed estimator. To facilitate the following derivation, we first define an alternative formulation for the proposed agreement pattern measure and its estimator. Denote $\vec{\lambda}(\cdot) = (\lambda_{10}(\cdot), \lambda_{01}(\cdot), \lambda_{11}(\cdot))^T$. Let $\tilde{\lambda}_0$ be the collection of bivariate hazard functions $\vec{\lambda}$ on $R^2$. We define the functional $g : \tilde{\lambda}_0 \mapsto R$ as follows:

$$g\left(\vec{\lambda}\left(\vec{t}\right)\right) = \vec{e_3}^T \vec{\lambda}\left(\vec{t}\right) - \vec{e_1}^T \vec{\lambda}\left(\vec{t}\right) \vec{e_2}^T \vec{\lambda}\left(\vec{t}\right), \tag{2.6}$$

where $\vec{t} = (t_1, t_2)$, $\vec{e}_1 = (1, 0, 0)^T$, $\vec{e}_2 = (0, 1, 0)^T$, $\vec{e}_3 = (0, 0, 1)^T$. It is straightforward to show the proposed local agreement pattern measure $\varphi\left(\vec{t}\right)$ can be expressed as a function of the bivariate hazard functions $\vec{\lambda}(\cdot)$ via $g(\cdot)$, i.e.

$$\varphi\left(\vec{t}\right) = g\left(\vec{\lambda}\left(\vec{t}\right)\right). \tag{2.7}$$

Our proposed estimator in (2.5) can be equivalently expressed as

$$\hat{\varphi}\left(\vec{t}\right) = g\left(\hat{\vec{\lambda}}\left(\vec{t}\right)\right), \tag{2.8}$$

where $\hat{\vec{\lambda}} = \left(\hat{\lambda}_{10}(\cdot), \hat{\lambda}_{01}(\cdot), \hat{\lambda}_{11}(\cdot)\right)^T$ are the kernel estimators of the hazard functions. In the following, we provide the asymptotic properties of $\hat{\varphi}$.

**Theorem 2.2.1.** *Consider a region* $\boldsymbol{\tau} = (0, \tau_1) \times (0, \tau_2)$ *where* $(\tau_1, \tau_2)$ *is in the support of the observed event times. Assuming the bivariate hazard function estimator* $\hat{\vec{\lambda}}$ *is uniformly strongly consistent within* $\boldsymbol{\tau}$*. For any* $\vec{t} = (t_1, t_2) \in \boldsymbol{\tau}$*, the proposed estimator* $\hat{\varphi}\left(\vec{t}\right)$ *has the following asymptotic properties as* $n \to \infty$*,*

*(i) The estimator* $\hat{\varphi}\left(\vec{t}\right)$ *is strongly consistent, i.e.,* $\left|\hat{\varphi}\left(\vec{t}\right) - \varphi\left(\vec{t}\right)\right| \to 0$ *with probability 1.*

*(ii) The proposed estimator* $\hat{\varphi}\left(\vec{t}\right)$ *has the following weak convergence result,*

$$r_n\left\{\hat{\varphi}\left(\vec{t}\right) - \varphi\left(\vec{t}\right)\right\} \xrightarrow{d} g'_{\vec{\lambda}}\left(\vec{W}\left(\vec{t}\right)\right),$$

*where* $r_n = (nh^2)^{1/2}$*,* $\vec{W}(\cdot)$ *is a multivariate zero-mean Gaussian process with the covariance function defined in equation (2.10) of Fermanian (1997), and* $g'_{\vec{\lambda}}$ *is the Hadamard derivative of* $g$ *at* $\vec{\lambda}$ *defined as*

$$g'_{\vec{\lambda}}\left(\vec{W}\right) = \vec{e}_3{}^T\vec{W} - \vec{e}_1{}^T\vec{W} \cdot \vec{e}_2{}^T\vec{\lambda} - \vec{e}_2{}^T\vec{W} \cdot \vec{e}_1{}^T\vec{\lambda}.$$

*Here,* $g'_{\vec{\lambda}}\left(\vec{W}\left(\vec{t}\right)\right)$ *follows a zero-mean normal distribution.*

*(iii) By randomly sampling with replacement from the observed data* $\left(\tilde{T}_{i1}, \tilde{T}_{i2}, \delta_{i1}, \delta_{i2}\right)$ $(i = 1, \cdots, n)$*, a bootstrap estimator* $\varphi^{\#}\left(\vec{t}\right)$ *can be obtained based on the bootstrap samples. Then* $r_n\left\{\varphi^{\#}\left(\vec{t}\right) - \hat{\varphi}\left(\vec{t}\right)\right\}$*, given the observed data, weakly converges to the same limiting distribution as* $r_n\left\{\hat{\varphi}\left(\vec{t}\right) - \varphi\left(\vec{t}\right)\right\}$ *in probability.*

We prove Theorem 2.2.1 based on the Hadamard differentiability of functional $g$ and the functional delta method. The detailed proof of this theorem is provided in Appendix A2. Due to the complexity of the covariance function of $\hat{\vec{\lambda}}$, an explicit expression for the asymptotic variance of $\hat{\varphi}(t_1, t_2)$ is analytically too complicated. Theorem 2.2.1(iii) suggests an alternative approach for estimating the variance of $\hat{\varphi}(t_1, t_2)$ via a bootstrap procedure.

### 2.2.4 Choice of the kernel functions and bandwidths for estimating the hazard functions

In this section, we provide some discussions on the specifications of the kernel functions and bandwidths for improving the performance of the proposed nonparametric estimator. We use the Epanechnikov kernels for estimating the hazard functions because they have been shown to be the most efficient kernel functions in minimizing the mean integrated squared error (Wang and Jones, 1995) and they also provide computational advantages over other kernel functions. For estimating $\lambda_{10}(t_1, t_2)$ and $\lambda_{01}(t_1, t_2)$, we adopt the univariate Epanechnikov kernel defined as $K_1(u) = 3/4(1 - u^2)I[u^2 \leq 1]$. For estimating $\lambda_{11}(t_1, t_2)$, we apply the bivariate product Epanechnikov kernel defined as $K_2(u_1, u_2) = K_1(u_1)K_1(u_2)$.

One important consideration in kernel hazard rate estimation is that the hazard functions have bounded support. In univariate hazard rate estimation, boundary bias occurs when the support of the kernel function at a time point within the interval $[0, h)$ exceeds the available range of the observed data, and thus leads to increased bias. For bivariate survival data, boundary effects are observed when either $T_1, T_2$ is close to zero. To reduce the boundary effects in hazard rate estimation, Müller and Wang (1994) proposed a class of boundary kernel functions which have shown numerical benefits in terms of smaller asymptotic mean squared error when estimating near the boundaries than other boundary kernels (Gasser and Müller, 1979, Keiding and Andersen, 1989, Gray, 1990). Specifically, for estimating univariate hazard function in the boundary region $B_L = \{t : 0 \leq t < h\}$, they proposed the following boundary kernel,

$$K_{1,t}(u) = \frac{12}{(1+q)^4}(u+1)\left[u(1-2q) + (3q^2 - 2q + 1)/2\right],$$

where $q = t/h$ and $u \in [-1, q]$. In this paper, we extend Müller and Wang

(1994)'s univariate boundary kernel to the bivariate case. For boundary regions $B_{L,I} = [t_1, t_2 : 0 \leq t_1 < h, t_2 \geq h]$, $B_{I,L} = [t_1, t_2 : t_1 > h, 0 \leq t_2 \leq h]$, and $B_{L,L} = [t_1, t_2 : 0 \leq t_1 < h, t_2 \geq h]$, the proposed bivariate boundary kernel is formulated as

$$
\begin{aligned}
K_{2,t_1,1}(u_1, u_2) &= K_{1,t_1}(u_1)K_1(u_2), && (t_1, t_2) \in B_{L,I} \\
K_{2,1,t_2}(u_1, u_2) &= K_1(u_1)K_{1,t_2}(u_2), && (t_1, t_2) \in B_{I,L} \\
K_{2,t_1,t_2}(u_1, u_2) &= K_{1,t_1}(u_1)K_{1,t_2}(u_2), && (t_1, t_2) \in B_{L,L}
\end{aligned}
$$

where $q_1 = t_1/h$, $q_2 = t_2/h$, $u_1 \in [-1, min(1, q_1)]$ and $u_2 \in [-1, min(1, q_2)]$.

Another consideration in our kernel estimation is the selection of bandwidths for estimating the bivariate hazard functions. Various bandwidth selection methods have been proposed in the literature for multivariate kernel density estimation (Wang and Jones, 1993, 1994, 1995; Duong and Hazelton, 2003, 2005; Duong, 2007), but very few have been proposed for multivariate hazard rate estimation. Fermanian (1997) proposed an asymptotically optimal plug-in bandwidth for estimating multivariate hazard functions. However, this bandwidth method typically requires a very large sample size and hence is not applicable in many studies with small to moderate sample sizes (less than several thousands). According to Fermanian (1997), a practical choice of bandwidth is to use Silverman's rule or Scott's rule (Silverman, 1986; Scott, 1992). In the simulation studies and data application, we choose the bandwidths for kernel hazard rate estimation based on Scott's rule (Scott, 1992) given by $h = n^{-1/(4+d)}\hat{\sigma}$, where $d$ is the number of dimensions of the hazard function and $\hat{\sigma}$ is the standard error estimator of the censored survival times (Steven, 2013).

## 2.3 Simulation Studies

We conducted simulation studies to assess the performance of the proposed estimation and inference procedure of the new local agreement pattern measure. In each simulation, we generated bivariate survival times $(T_1, T_2)$ from the Clayton model (Clayton, 1978) with a sample size of 500. We considered two sets of simulation studies with different setups for marginal distributions. In the first setup, we assumed $T_1$ and $T_2$ had identical marginal distributions which were standard exponential. In the second setup, the two survival times had different marginal distributions where $T_1$ and $T_2$ had exponential distributions with the means equal to 1 and 1.5, respectively. We specified a cross ratio of 3 in the Clayton model which indicated moderate dependence between bivariate survival times. The survival times generated from the Clayton model were subject to independent right censoring by two independent and exponentially distributed censoring variables. We considered three censoring rates of 17%, 33% and 50%, representing light, medium and heavy censoring, respectively.

Table 2.1 summarizes the results based on 500 simulation runs under various simulation scenarios. We selected nine time points $(t_1, t_2)$ on the two dimensional plane where the values of $t_j (j = 1, 2)$ were chosen as the $10th$, $30th$ and $50th$ percentiles of the standard exponential distribution. For each pair of time points, we presented the true value for the local agreement measure $\varphi(t_1, t_2)$, empirical mean and empirical standard error for $\hat{\varphi}(t_1, t_2)$, along with the average standard error estimate. We also presented coverage probability for the 95% confidence intervals based on the 200 bootstrap samples. The proposed nonparametric estimator for local agreement pattern measure demonstrated reasonable accuracy with the bias being less than 10% at most time points. The empirical standard error of the proposed estimator decreased as the censoring proportion decreased. The bootstrap standard error was close to the Monte Carlo standard error and the coverage probability

Table 2.1: Summary statistics of the proposed local agreement pattern measure estimator $\hat{\varphi}(t_1, t_2)$ for the Clayton family with identical or different marginal distributions.

| Marginal Distributions | $T_1$ | Censoring | $T_2$ | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 10% | | 30% | | 50% |
| Homogenerous | 10% | Light | | 1.550 (0.280,0.290,92.2%) | | 1.046 (0.169,0.165,90.0%) | | 0.596 (0.164,0.147,91.0%) |
| | | Medium | **1.395** | 1.552 (0.281,0.296,91.4%) | **0.967** | 1.050 (0.179,0.174,90.4%) | **0.556** | 0.604 (0.178,0.162,91.2%) |
| | | Heavy | | 1.549 (0.298,0.305,92.6%) | | 1.051 (0.187,0.191,92.2%) | | 0.605 (0.219,0.194,89.0%) |
| | 30% | Light | | 1.050 (0.152,0.166,94.0%) | | 0.840 (0.096,0.097,92.4%) | | 0.622 (0.096,0.097,93.4%) |
| | | Medium | **0.967** | 1.046 (0.163,0.175,93.6%) | **0.873** | 0.844 (0.106,0.104,92.4%) | **0.646** | 0.620 (0.106,0.107,92.4%) |
| | | Heavy | | 1.049 (0.174,0.190,95.4%) | | 0.842 (0.119,0.116,91.4%) | | 0.620 (0.123,0.127,92.6%) |
| | 50% | Light | | 0.602 (0.144,0.148,93.8%) | | 0.619 (0.095,0.096,92.4%) | | 0.609 (0.100,0.106,92.4%) |
| | | Medium | **0.556** | 0.601 (0.154,0.162,95.2%) | **0.646** | 0.621 (0.108,0.107,93.8%) | **0.654** | 0.608 (0.116,0.119,92.4%) |
| | | Heavy | | 0.603 (0.197,0.193,93.0%) | | 0.616 (0.128,0.128,93.0%) | | 0.609 (0.148,0.146,90.4%) |
| Heterogeneous | 10% | Light | | 1.227 (0.269,0.254,90.2%) | | 0.885 (0.158,0.149,91.2%) | | 0.591 (0.138,0.138,94.8%) |
| | | Medium | **1.085** | 1.225 (0.268,0.258,91.4%) | **0.833** | 0.883 (0.165,0.157,92.8%) | **0.562** | 0.586 (0.155,0.153,94.2%) |
| | | Heavy | | 1.224 (0.274,0.265,91.4%) | | 0.887 (0.181,0.172,92.0%) | | 0.601 (0.187,0.183,94.2%) |
| | 30% | Light | | 0.800 (0.137,0.144,94.6%) | | 0.668 (0.083,0.084,94.2%) | | 0.562 (0.087,0.087,92.2%) |
| | | Medium | **0.728** | 0.798 (0.144,0.151,94.4%) | **0.689** | 0.669 (0.090,0.090,92.8%) | **0.579** | 0.559 (0.099,0.097,91.8%) |
| | | Heavy | | 0.803 (0.162,0.165,94.0%) | | 0.670 (0.102,0.101,92.2%) | | 0.556 (0.116,0.114,92.0%) |
| | 50% | Light | | 0.442 (0.124,0.125,94.4%) | | 0.473 (0.087,0.083,93.8%) | | 0.481 (0.090,0.091,92.6%) |
| | | Medium | **0.407** | 0.445 (0.138,0.138,94.0%) | **0.465** | 0.469 (0.096,0.092,92.6%) | **0.501** | 0.486 (0.100,0.103,93.4%) |
| | | Heavy | | 0.437 (0.154,0.163,94.4%) | | 0.473 (0.110,0.110,93.4%) | | 0.477 (0.124,0.125,94.2%) |

of the estimated 95% confidence intervals was close to the nominal level in most cases.

To further illustrate the performance of the proposed estimator, we plotted the estimated local agreement pattern measures along with the true values over the two-dimensional time plane for identical (Fig 3a) and different (Fig 3b) marginal distributions. Specifically, the top panels of Figure 3a and 3b present the surfaces of the true local agreement pattern measure on the two-dimensional time plane. In the bottom panels, we first fixed $T_1$ at various values and then plotted the true and estimated profiles of $\varphi$ across $T_2$ ranging from 0.05 to 1.45 (which was approximately 5% to 75% quantile of the standard exponential distribution). We also plotted the 95% pointwise Monte Carlo confidence intervals based on the empirical variance of $\hat{\varphi}$. Then we fixed $T_2$ and plotted the profiles of $\varphi$ across $T_1$. Results from Figure 3a and 3b suggest that the proposed estimator provided fairly accurate estimation of the local agreement pattern measures across various time points in the two-dimensional time plane. The bias was in general small with the largest bias observed in the boundary region when either $T_1$ or $T_2$ was very close to 0. The Monte Carlo confidence bands for $\varphi$ provided a nice coverage of the profile of the local agreement measure. The confidence bands tended to be wider near the boundary regions as expected. The confidence bands for our local agreement measure were much narrower and tighter in the boundary regions as compared to the confidence bands previously developed for other local dependence measures (Hu et al., 2011).

## 2.4 Data Example

We illustrate the application of the proposed local agreement pattern measure using the data from a prostate cancer study. Prostate cancer is the most common cancer among US men. Various kinds of treatments are available for this disease

and it is of interest to compare the efficacy of different treatments (Critz et al., 1995, 1996; Jung et al., 2011). One major difficulty for comparing the efficacy across treatments is the lack of a standard definition of disease-free state after treatments. It's well known that post-treatment disease status is reflected in the level of prostate specific antigen (PSA) with higher-level PSA indicating cancer relapse. However, there is a lack of consensus regarding the exact pattern of PSA level that defines disease recurrence and physicians have been applying different definitions that have been traditionally used for specific treatments to define the disease-free state. The relapse-free survival rates are then obtained for different treatments based on the corresponding definitions and then used as important guidance for treatment selection. Since the relapse-free survival rates are derived based on different definitions of disease-free state, the potential discrepancies between the definitions may cause misleading conclusions on treatment efficacy. For example, radical prostatectomy and irradiation are two commonly used treatments for curing prostate cancer (Critz et al., 1995). Different definitions have been proposed for both treatments. For radical prostatectomy, post-treatment disease freedom is defined by reaching and maintaining an undetectable PSA nadir ranging between 0.2 ng/ml and 0.5 ng/ml (Critz et al., 1996). For irradiation, disease freedom is represented by a non-rising PSA with the increasing PSA defined as three consecutive PSA increases measured 6 months apart, according to the American Society of Therapeutic Radiation Oncology (ASTRO) consensus criteria (1997). There has been a debate on the relative effectiveness of the two treatments. Some researchers claim that two treatments have equivalent efficacy while others argued that radical prostatectomy is a better treatment than irradiation since it has a higher curing rate (Critz et al., 1996). In order to accurately understand and compare the relapse-free survival rates between the two treatments, it is important to first assess the agreement between the two disease-free definitions. In particular, we are interested in finding out how

the agreement between the two definitions evolves along the time after the treatments.

In a clinical study, 1369 men received simultaneous radiotherapy for prostate cancer followed by an external beam radiation. The disease-free status was evaluated frequently after radiation treatment. The relapse-free time was defined as the time from the end of the irradiation till the prostate cancer relapse based on two different definitions. $T_1$ was the observed relapse-free time with disease recurrence defined as post-treatment PSA level exceeding the nadir of 0.2 ng/ml. $T_2$ was defined based on ASTRO definition and represented the midpoint between the time when the lowest PSA was achieved after irradiation and the time when the first of three consecutive rises in the PSA level occurred. The relapse-free times for a patient were subject to independence censoring due to the end of the follow-up on this patient. Among the 1369 patients, 159 subjects were diagnosed with prostate cancer recurrence according to both definitions and 64 had relapses based on only one of the definitions, indicating approximate 80% of censoring. Figure 2.4 presents the distribution of patients' observed relapse-free times measured by the two definitions. The plot shows that the observed relapse-free times ranged from less than 1 year to 13.5 years and almost all of the observed cancer relapses happened within 8 years after the irradiation. Figure 2.4 also shows that most observed survival times were on the 45 degree line. For the remaining subjects, the observed relapse-free time by the ASTRO definition tended to be longer than that by the nadir definition.

We applied the proposed measure $\varphi$ to evaluate the local agreement pattern between $T_1$ and $T_2$ within 7.5 years after the irradiation. The estimated local agreement pattern measures are presented in Figure 2.5. On the top panel, we display the local agreement measure surface within the time space of $[0, 7.5] \times [0, 7.5]$. From the surface plot, we can see that within the first 4 years, the local agreement was high-

est along the 45 degree line and decreased dramatically when moving away from the 45 degree line, indicating there was strong agreement between the two relapse-free survival times. For year 4 and on, the highest local agreement region moved slightly away from the 45 degree line towards $T_2 > T_1$ area. This suggest that if a patient remained disease-free longer than 4 years after irradiation, he was more likely to have a disease relapse diagnosed earlier by the nadir definition. On the bottom panel, we present the estimated local agreement pattern measures on the 45 degree line within 7.5 years and the corresponding 95% bootstrap pointwise confidence bands based on 1000 bootstrap samples. It shows that the local agreement pattern measure on the 45 degree line was highest around year 2 after the irradiation, suggesting the nadir and ASTRO definitions agreed best for cancer recurrences that happened around year 2 after the treatment. The local agreement decreased after year 2 indicating more disagreement between the two definitions for recurrences that happened beyond 2 years after the irradiation. This result was supported by our observations from Figure 2.4 which shows that more bivariate survival outcomes were observed off the 45 degree line after year 2.

## 2.5 Remarks

In this paper, we propose a new framework for describing local agreement pattern between correlated survival outcomes. This new framework has three appealing features. First, it accommodates censored observations in time-to-event data. Second, it reveals how the local agreement pattern changes over time. Lastly, the local measure, as a function, fully captures the dependence structure between bivariate survival times.

The proposed local agreement pattern measure is not bounded by a fixed range

due to the nature of hazard functions being unbounded. As done with many other unbounded descriptive measures, such as cross ratio (Clayton, 1978) and local dependence function (Holland and Wang, 1987), the proposed measure can be interpreted with respect to the relative scale over the two-dimensional time space within the region of interest. It helps to address the nature of agreement when the relationship of two survival times is time dependent and can potentially be used for modeling such local relationship.

## 2.6 Appendix

### Appendix A1. Proposition 1

In the following proposition, we show that $\varphi(t_1, t_2)$ fully captures the dependence structure between correlated survival times by demonstrating how the survival function $S(t_1, t_2)$ is jointly determined by the marginal survival functions $S_1(t_1)$, $S_2(t_2)$ and an integrable local agreement pattern measure function $\varphi(t_1, t_2)$ within the finite region $\boldsymbol{\tau}$.

**Proposition 1.** *Let* $S_1(t) = S(t, 0)$, $S_2(t) = S(0, t)$, *and* $\boldsymbol{\tau} = (0, \tau_1) \times (0, \tau_2)$, *where* $\tau_1 = sup\{t : S_1(t) > 0\}$ *and* $\tau_2 = sup\{t : S_2(t) > 0\}$, *we have*

$$S(t_1, t_2) = S_1(t_1)S_2(t_2) \exp\left\{\int_0^{t_1} \int_0^{t_2} \varphi(u_1, u_2) du_1 du_2\right\}.$$

*Proof.* First, we show that $\varphi(t_1, t_2)$ equals to the second-order mixed partial derivative

of the log-transformed joint survival function, which follows from

$$\begin{aligned}\varphi(t_1,t_2) &= \lambda_{11}(t_1,t_2) - \lambda_{10}(t_1,t_2)\lambda_{01}(t_1,t_2) \\ &= \frac{f(t_1,t_2)}{S(t_1,t_2)} - \left\{\frac{d}{dt_1} logS(t_1,t_2)\right\}\left\{\frac{d}{dt_2} logS(t_1,t_2)\right\} \\ &= \frac{d^2}{dt_1 dt_2} logS(t_1,t_2).\end{aligned}$$

Then, we have

$$\begin{aligned}logS(t_1,t_2) &= [logS(t_1,t_2) - logS(t_1,0) - logS(0,t_2) + logS(0,0)] + logS(t_1,0) + logS(0,t_2) \\ &= \int_0^{t_1}\int_0^{t_2} \varphi(u_1,u_2)du_1du_2 + logS_1(t_1) + logS_2(t_2),\end{aligned}$$

or equivalently,

$$S(t_1,t_2) = S_1(t_1)S_2(t_2)\exp\left\{\int_0^{t_1}\int_0^{t_2}\varphi(u_1,u_2)du_1du_2\right\}.$$

□

## Appendix A2. Proof of Theorem 2.2.1

*Proof.* Let $D_g$ be a collection of bivariate and single hazard rate functions on $R^2$. For all converging sequences $t_n \to 0$, and $\vec{d_n} \to \vec{d} \in D_0 \subset D_g$ with $\vec{d_n} \in D_g$ and $\vec{\lambda} + t_n\vec{d_n} \in D_g$, we can prove that

$$\lim_{n\to\infty} \frac{g\left(\vec{\lambda} + t_n\vec{d_n}\right) - g\left(\vec{\lambda}\right)}{t_n} = \vec{e_3}^T\vec{d} - \vec{e_1}^T\vec{d}\cdot\vec{e_2}^T\vec{\lambda} - \vec{e_1}^T\vec{\lambda}\cdot\vec{e_2}^T\vec{d}. \tag{2.9}$$

From (2.9), $g$ is Hadamard differentiable at $\vec{\lambda}$ (van der Vaart and Wellner, 1996) with the derivative of

$$g'_{\vec{\lambda}}\left(\vec{d}\right) = \vec{e_3}^T\vec{d} - \vec{e_1}^T\vec{d}\cdot\vec{e_2}^T\vec{\lambda} - \vec{e_1}^T\vec{\lambda}\cdot\vec{e_2}^T\vec{d}.$$

The result of $\hat{\varphi}\left(\vec{t}\right) = g\left(\hat{\vec{\lambda}}\left(\vec{t}\right)\right) \xrightarrow{P} \varphi\left(\vec{t}\right) = g\left(\vec{\lambda}\left(\vec{t}\right)\right)$ follows from the continuity of the Hadamard differentiable function $g$ and the consistency of $\hat{\vec{\lambda}}$. Thus, the statement (i) of Theorem 2.2.1 is proven.

It has shown that $r_n\left\{\hat{\vec{\lambda}}\left(\vec{t}\right) - \vec{\lambda}\left(\vec{t}\right)\right\}$ weakly converges to a zero-mean Gaussian process $W\left(\vec{t}\right)$ (Fermanian, 1997) as the positive sequence $r_n \to \infty$. Specifically, Fermanian (1997) has showed that in $\boldsymbol{\tau}$, the kernel estimators $\hat{\lambda}_{10}, \hat{\lambda}_{01}, \hat{\lambda}_{11}$ are uniformly strongly consistent and have weak convergence results under some regularity conditions of the bandwidths. That is, for any $\vec{t} \in \boldsymbol{\tau}$, if $\lambda_{10}$, $\lambda_{01}$ and $\lambda_{11}$ are $C^2$ in the neighborhood of $\vec{t}$ and if there exists $\epsilon > 0$, such that the bandwidth sequence $\{h_n\}$ satisfies $h_n log_2 n \to 0$, $nh_n^{(1+\epsilon)} \to \infty$ as $n \to \infty$, we have the following asymptotic properties for $\hat{\vec{\lambda}}$ as below,

$$\begin{cases} \sqrt{nh_n}\left\{\hat{\lambda}_{10}(\vec{t}) - \lambda_{10}(\vec{t})\right\} \to W_{10}(\vec{t}) \\ \sqrt{nh_n}\left\{\hat{\lambda}_{01}(\vec{t}) - \lambda_{01}(\vec{t})\right\} \to W_{01}(\vec{t}) \end{cases} \Leftrightarrow \sqrt{nh_n^2}\left\{\hat{\vec{\lambda}}(\vec{t}) - \vec{\lambda}(\vec{t})\right\} \to \vec{W}(\vec{t})$$
$$\sqrt{nh_n^2}\left\{\hat{\lambda}_{11}(\vec{t}) - \lambda_{11}(\vec{t})\right\} \to W_{11}(\vec{t})$$

where $\vec{W}(\cdot) = (\sqrt{h_n}W_{10}(\cdot), \sqrt{h_n}W_{01}(\cdot), W_{11}(\cdot))^T$. The asymptotic covariance of the Gaussian process $W_{10}$, $W_{01}$, $W_{11}$ is defined in Equation (2.10) in Fermanian (1997). The function $g$ is proved to be Hadamard differentiable. Then, the following statement is true according to functional delta method, $r_n\left\{\hat{\varphi}\left(\vec{t}\right) - \varphi\left(\vec{t}\right)\right\} = r_n\left\{g\left(\hat{\vec{\lambda}}\left(\vec{t}\right)\right) - g\left(\vec{\lambda}\left(\vec{t}\right)\right)\right\}$ weakly converges to $g'_{\vec{\lambda}}\left(W\left(\vec{t}\right)\right)$. Thus, the statement (ii) is true.

To show the asymptotic result for the bootstrap estimator in (iii), define $\varphi^{\#}\left(\vec{t}\right)$ as the nonparametric estimator of the local agreement pattern measure based on a bootstrap sample $\left(\tilde{T}_{i1}^{\#}, \tilde{T}_{i2}^{\#}, \delta_{i1}^{\#}, \delta_{i2}^{\#}\right)$ $(i = 1, \cdots, n)$ randomly selected with replacement from the observed data. Due to the Hadamard differentiability of the function $g$ and

the functional delta method for bootstrapping, $r_n \left\{ \varphi^{\#} \left( \vec{t} \right) - \hat{\varphi} \left( \vec{t} \right) \right\}$ conditional on the observed data weakly converges to the same limiting distribution as $r_n \left\{ \hat{\varphi} \left( \vec{t} \right) - \varphi \left( \vec{t} \right) \right\}$ in probability. $\square$

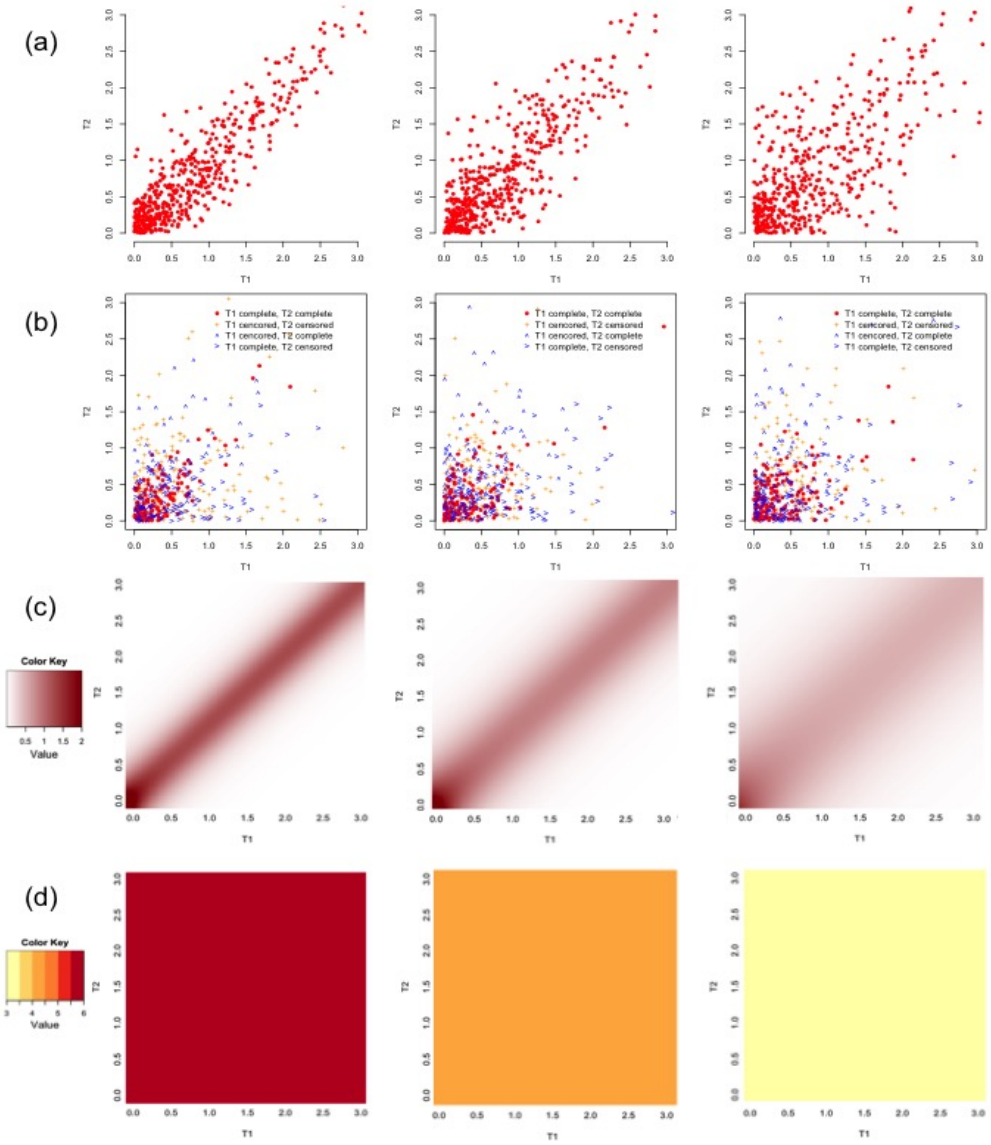



Figure 2.1: Clayton models with the same marginal exponential distributions for T1 and T2 but different dependence levels. The dependence between T1 and T2 decreases from left to right with cross ratio taking the values, 6, 4.3 and 3. From upper panel to lower panel, the figures represent the scatterplots with complete data, scatterplots with censored data, local agreement pattern measure heatmaps, and cross ratio heatmaps, respectively.

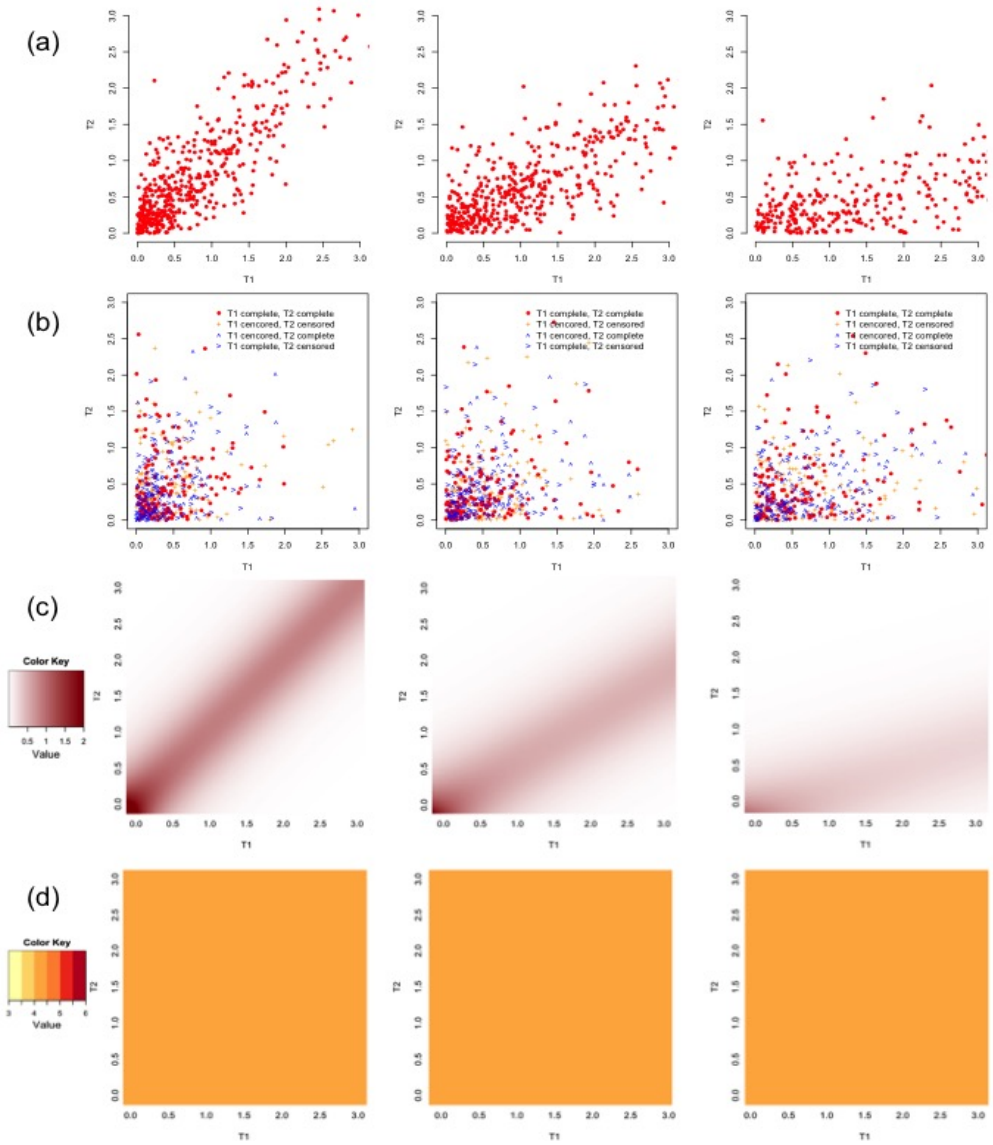



Figure 2.2: Clayton models when the marginal distribution of T2 changes while the marginal distribution of T1 and the association between T1 and T2 are fixed. The marginal distribution of T2 becomes more different from the distribution of T1 from left to right with rate parameter of T2 distribution taking values of 1, 0.6 and 0.3, respectively. From upper panel to lower panel, the figures represent the scatterplots with complete data, scatterplots with doubly censored data, local agreement pattern measure heatmaps, and cross ratio heatmaps, respectively.

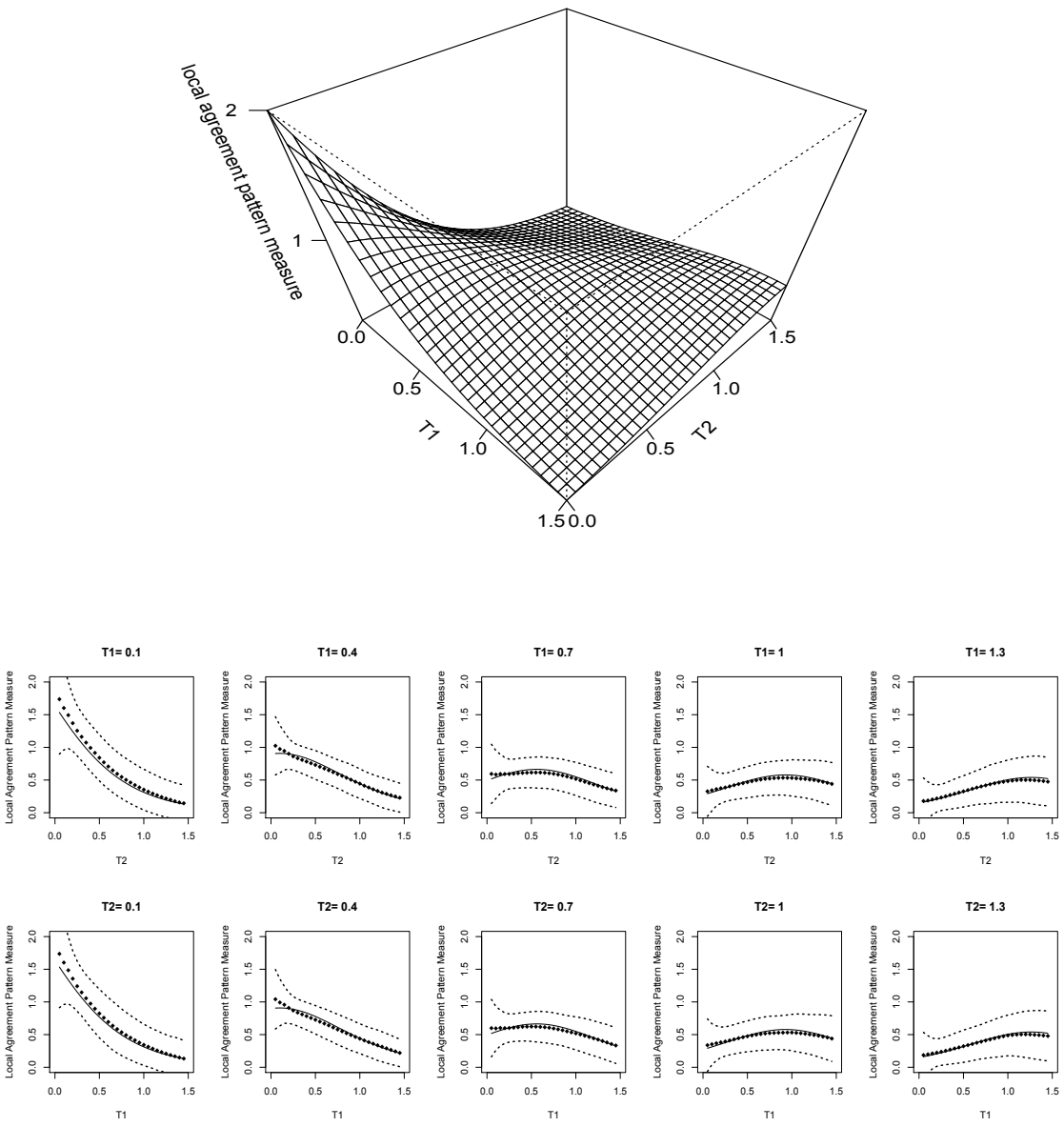



Figure 3a: Local agreement pattern measure for the Clayton model with identical marginal distributions. The surface plot in the top panel is the true local agreement pattern measure on the 2-D time space. Line curves in the bottom panels correspond to the lines in the surface plot when fixing one of the two survival times. Dot curves are the estimated local agreement pattern measures and the corresponding empirical pointwise 95% confidence bands.

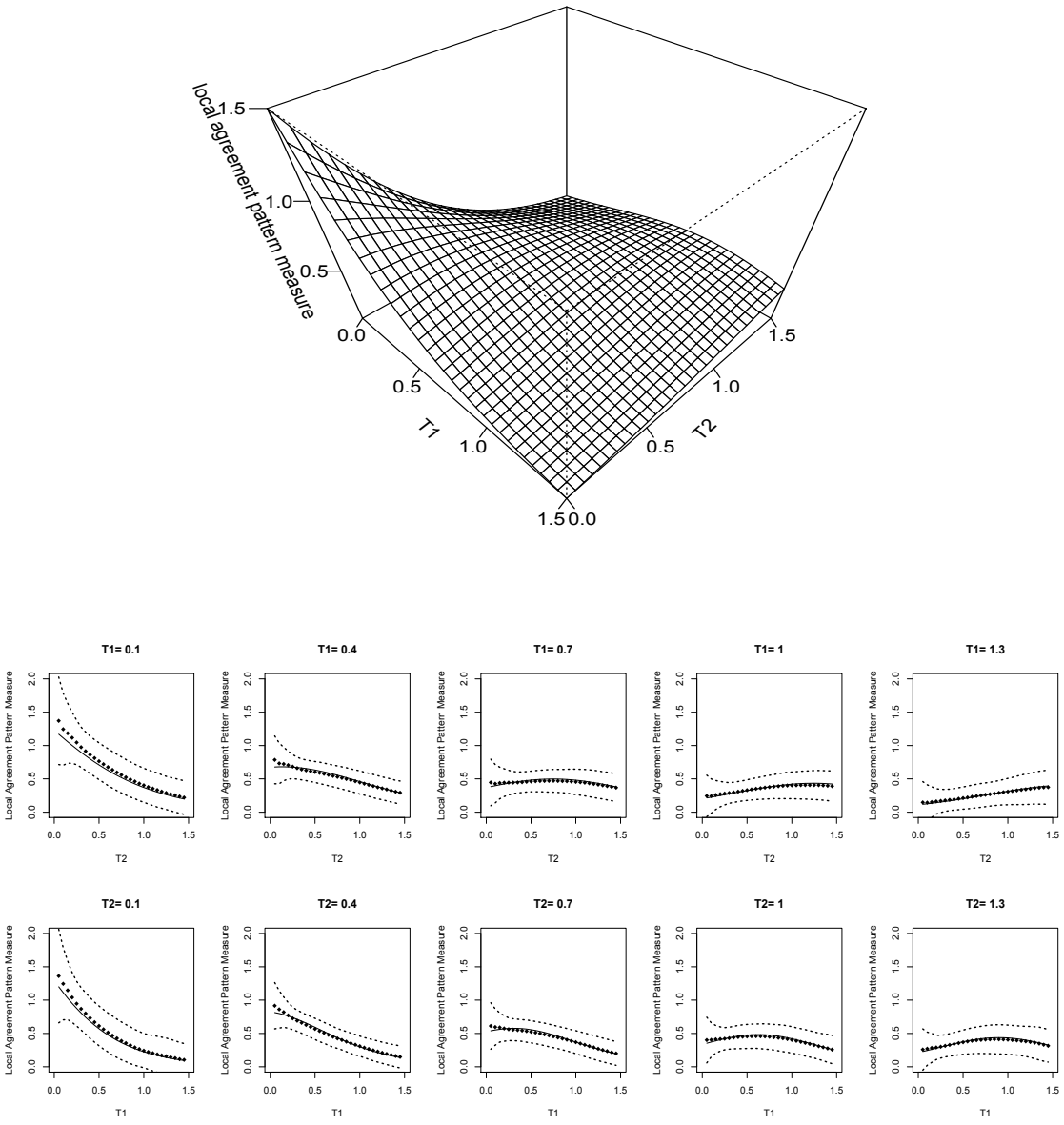



Figure 3b: Local agreement pattern measure for the Clayton model with different marginal distributions. The surface plot in the top panel is the true local agreement pattern measure on the 2-D time space. Line curves in the bottom panels correspond to the lines in the surface plot when fixing one of the two survival times. Dot curves are the estimated local agreement pattern measures and the corresponding empirical pointwise 95% confidence bands.

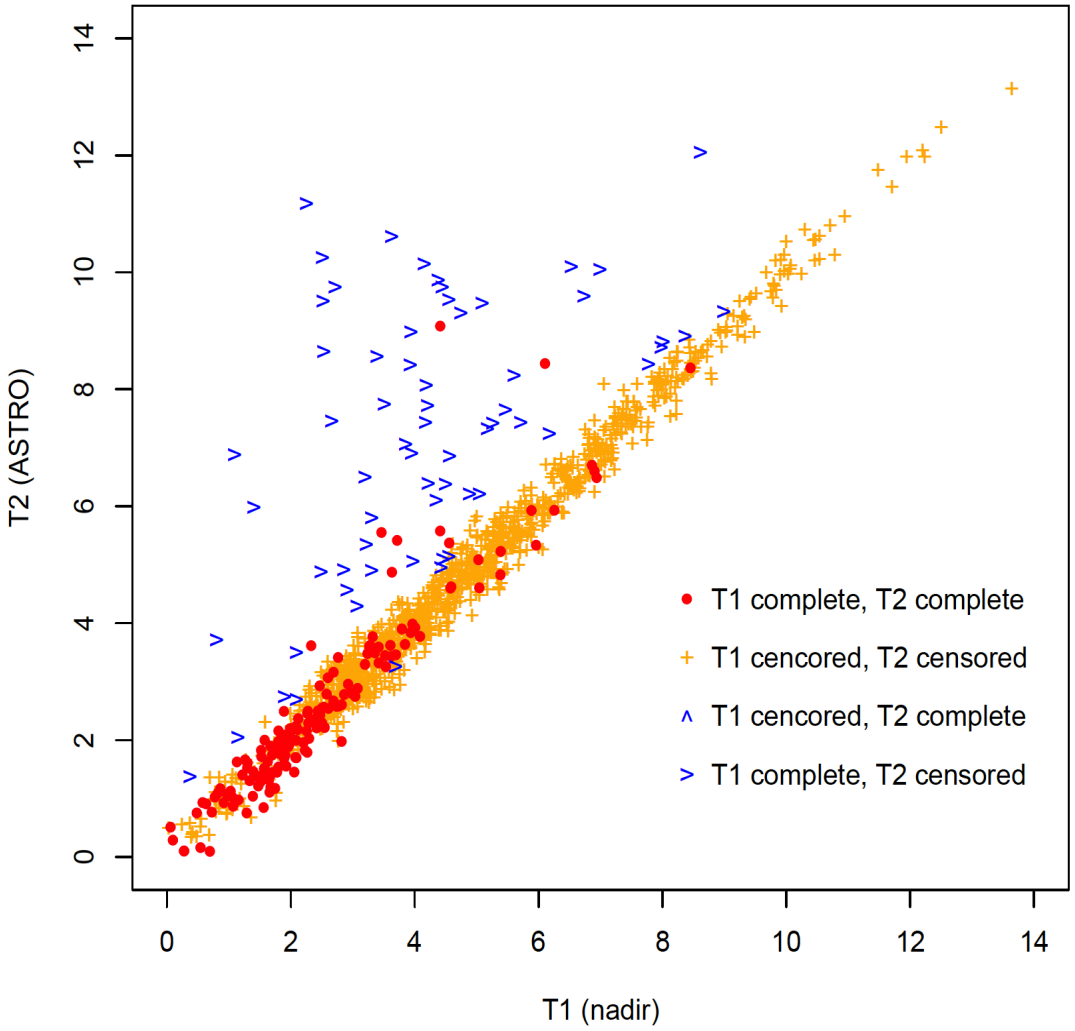



Figure 2.4: Prostate cancer relapse-free survival times after irradiation based on the two definitions of disease-free state. Red dots represent subjects with both survival times observed. The orange plus signs are corresponding to patients that did not experience disease relapse during the study according to both definitions. The blue arrow signs stand for patients that had been diagnosed with disease relapse with only one definition in the study.

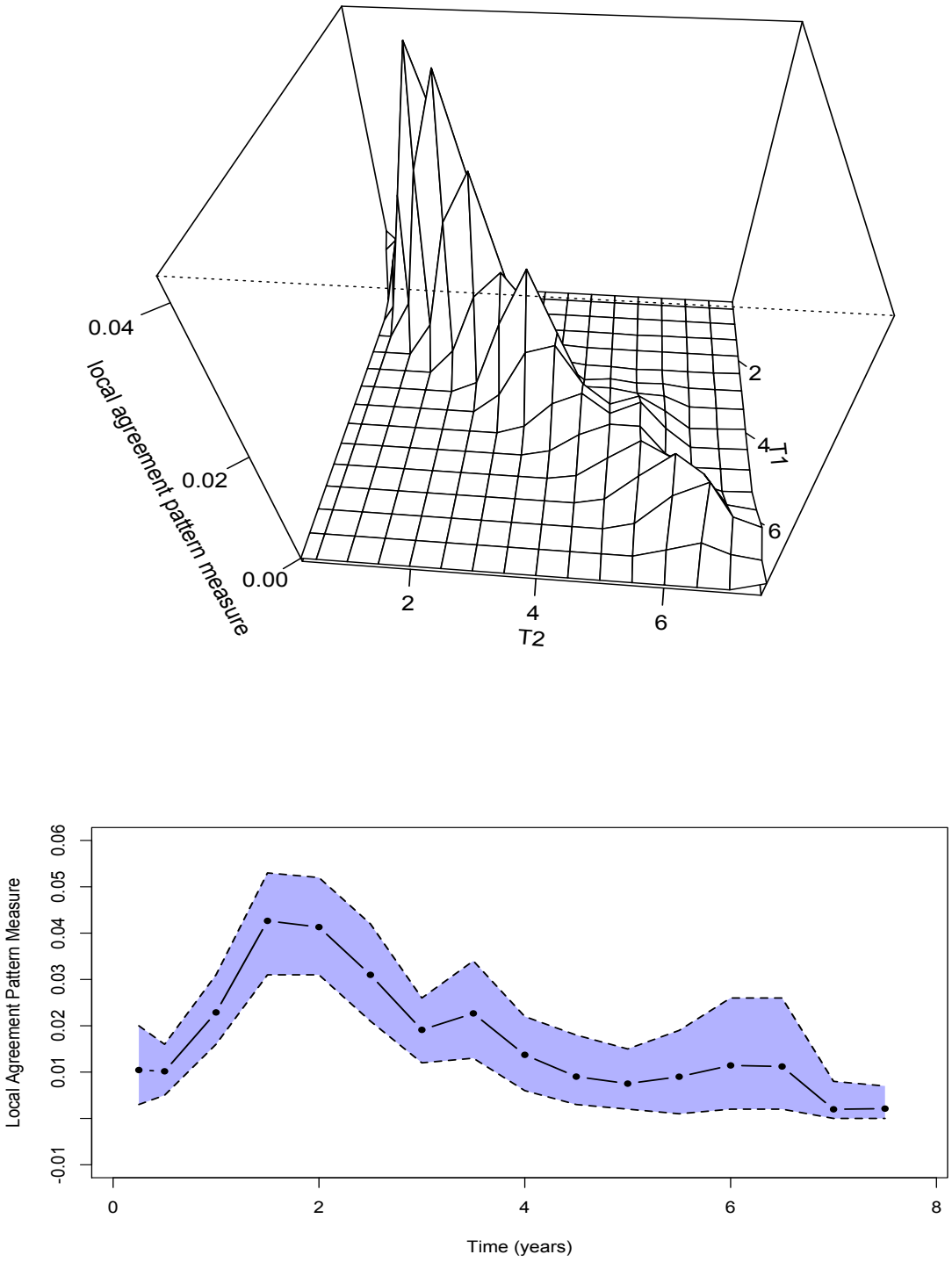



Figure 2.5: Estimated local agreement pattern measure surface (top) and diagonal-line curve with 95% bootstrap pointwise confidence bands (bottom).

# Chapter 3

# Agreement Methods for High Dimensional Neuroimaging Data

## 3.1 Introduction

Recently in the brain imaging community, there is an increasing trend of conducting multi-site studies. For example, the Functional Biomedical Informatics Research Network (fBIRN) has recently conducted a multi-site study sponsored by the NCCR/NIH to study regional brain dysfunction related to the progression and treatment of schizophrenia. Ideally, merging data across sites requires the interchangeability of data from different sites and is reasonable only if site differences in fMRI data can be minimized. However, one major challenge arises when combining data in multi-site imaging studies is that even the same subject' brain images can vary considerably across MRI sites since they are acquired using different scanners and protocols. Thus, it is crucial to effectively measure the reproducibility of brain images acquired from various sites before conducting any further analysis in order to draw reliable conclusions in multi-site studies. To assure the success of large-scale studies in the fBIRN project, researchers have conducted a pilot traveling-subject study prior to all other studies to evaluate the reproducibility of fMRI images acquired at different imaging sites. If the agreement between sites is low, sources causing the discrepancy need to be identified and addressed.

Agreement methodology provides a well-suited framework to address the needs for reproducibility assessment in multi-site imaging studies. Friedman et al. (2008) assessed the between-site reproducibility of brain images from the fBIRN Traveling Subject study by measuring the intra correlation coefficients (ICC) based on two summary statistics estimated from six regions of interest(ROI): percent signal change and contrast-to-noise ratio. By using scalar summary statistics, they directly applied existing statistical methods to assess the reliability for fMRI data. One major disadvantage of their approach was that they assessed the reliability only based on median- or maximum-based summary measures from selected regions

which represented very limited information from the observed images. Therefore, their conclusions could only be made with respect to these specific measures of brain activation. There were no clear justifications that the reproducibility in these measures could be generalized to the reproducibility in other features of fMRI images, such as reaction time. In addition, by using scalar summary statistics, their approach left out a lot of important information from the observed images, which would eventually result in the lost of statistical power for small-sample-sized studies.

Li and Chow (2005) proposed the concordance correlation coefficient for image data. Their approach was only applicable to measure the agreement between two single brain images. However, in fMRI study, we often have time series of brain images. Additionally, Li and Chow's (2005) approach for assessing agreement for image data used the whole brain images. As we know, a brain contains many functional networks. Observed brain imaging data represents the combination of signals from various underlying functional networks such as resting-state functional networks that control our physiological, sensory and cognitive activities and task-related networks that regulate our responses to experimental tasks. In the task-based brain imaging study, we are often interested in specific functional networks and their signals that correspond to experimental tasks. In these kinds of studies, it would be more efficient to only focus on those functional networks of interest and to eliminate the impacts of resting-state functional connectivity as well as background noises.

In the motivating data example, subjects were instructed to perform sensorimotor tasks for four replicated scans during each visit while their fMRI images were being acquired. In this case, we have four replications of time series of brain images for each subject at each site. In the literature, many agreement methods have been developed for assessing the agreement using replicated measurements. For example,

Barnhart et al. (2005, 2007) extended the classical CCC and proposed total-CCC and inter-CCC to accommodate data with replications. Barnhart et. al. (2007) introduced a new agreement measure for replicated data, i.e., coefficient of individual agreement (CIA), based on the concept of individual agreement. These agreement methods were all developed for standard scalar measurements and have very limited applications to the high dimensional and complex imaging outcomes.

Although the aforementioned statistical methods are all designed to quantify the agreement with replicated measurements, there are some underlying differences among them. Lin's CCC was originally designed to assess the agreement between continuous outcomes with single measurement and has later been modified to accommodate for replicated measurements. CIA differs from the CCC in its relationship with the within-subject variability ($\sigma_{Wj}$) and the between-subject variability ($\sigma_{Bj}$). CCC-type of measures use the independence model as the benchmark. Alternatively, CIA compares differences between measurements from different observers to the differences of replicated measurements of the same observer. Therefore, the estimation of CIA requires replications for the estimation of within-rater variability. Aforementioned agreement methods were all developed for standard scalar measurements and have very limited applications to the high dimensional and complex imaging outcomes.

In this chapter, we want to propose an agreement/reproducibility method that is suitable for high-dimensional brain imaging data. Specifically, we develop a functional-network-based agreement framework to assess the agreement between replicated images from different imaging sites. There are several reasons to consider a network-based agreement method. First of all, constructing a voxel-wise agreement measure is not efficient given the noisy imaging data, and does not summarize the

agreement very well. Secondly, hundreds of thousands of voxels in the brain can be grouped into some brain functional networks. Voxels in the same functional networks usually demonstrate similar temporal dynamics across the study period. These functional networks are often what researchers are interested in and studied in fMRI studies. The proposed network-based method can help us reduce the high dimensionality of imaging data and measure the agreement using the most relevant information from fMRI images. Our approach is consisted of two stages. In the first stage, we decompose the observed brain images into several functional networks and extract the temporal dynamics of these functional networks using blind source separation methods. This stage reduces the dimensionality of the data and extracts useful information from the noisy imaging data. In the second stage, we focus on comparing the temporal dynamics of the estimated functional networks across imaging sites. To takes full advantages of the informations contained in the replicated scans acquired at each site, we propose new agreement measures to assess how comparable these network-specific temporal dynamics are based on the replicated images.

The remainder of this chapter is organized as follows. In the method section, we first present the two-stage network-based agreement framework for replicated imaging data. We present our method for task-related functional magnetic resonance imaging (task-fMRI), which is the most commonly used imaging modality. We develop nonparametric estimation methods for the proposed indices and establish their asymptotic properties. We evaluate the accuracy of the proposed estimation procedure via simulation studies. The proposed method is applied to the fBIRN Traveling-Subject study to investigate site effect in multi-site imaging studies.

## 3.2 Method

In this section, we propose an efficient network-based agreement method for replicated task-fMRI data acquired from different imaging sites. The proposed method requires two-stage estimation. In the first stage, we propose to estimate the underlying brain functional networks based on the observed brain images using a blind source separation method. In the second stage, we propose a new agreement measure to assess the agreement between the temporal dynamics of the estimated brain functional networks across sites. The details are presented as follows.

### 3.2.1 Two-stage agreement methods for task-fMRI data

In the first stage, we propose to apply a group independent component analysis model (ICA) (Calhoun et al. 2001; Guo and Pagnoni, 2008; Guo et al., 2011) to identify and characterize the brain functional networks. Let $\mathbf{Y}_{ijk}$ be a $T \times V$ data matrix representing the observed fMRI data for subject $i$ from the $k$th scan at the $j$th site. Let $\mathbf{Y} = [\mathbf{Y}'_{111}, \cdots, \mathbf{Y}'_{11K}, \mathbf{Y}'_{121}, \cdots, \mathbf{Y}'_{NJK}]'$ be the $NJKT \times V$ temporal-concatenated group data matrix across scans, sites and subjects. The group ICA model is defined as

$$\mathbf{Y} = \mathbf{AS} + \mathbf{E},$$

where $\mathbf{S}$ is a $q \times V$ signal matrix with each row of $\mathbf{S}$ presents a common spatial map of a brain functional network, $\mathbf{A}$ is a $NJKT \times q$ mixing matrix with each column contains subject-site-scan-specific temporal information and $\mathbf{E}$ is the $NJKT \times V$ Gaussian noise term. We denote the $l$th column of $\mathbf{A}$ as $\mathbf{a}_l = [\mathbf{a}_{111l}, \cdots, \mathbf{a}_{NJKl}]'$, where $\mathbf{a}_{ijkl} = [a_{ijkl}(1), \cdots, a_{ijkl}(T)], (i = 1, \cdots, N; l = 1, \cdots, q; j = 1, \cdots, J; k = 1, \cdots, K)$, represents the temporal responses of the $l$th functional network for subject $i$ from scan $k$th at site $j$. The ICA decomposes the observed fMRI data as a linear combination of $q$ source signals where the $l$th column of $\mathbf{A}$ and $l$th row of $\mathbf{S}$ characterize

the spatial-temporal processes related to the $l$th source signal ($l = 1, \cdots, q$). In fMRI data, statistical independence are usually assumed among the spatial maps (Beckmann and Smith, 2005; Calhoun et al., 2001; Guo and Pagnoni, 2008; Guo et al., 2011), which are represented by the rows of $\mathbf{S}$ in the ICA model.

In the second stage, we propose to assess the pairwise agreement for network-specific temporal responses measured on the same subjects but from two different imaging sites. Specifically, we extend the definition of scalar coefficient of individual agreement (CIA) to assess agreement for image data with replicated scans and denote the extended CIA definition as functional CIA (fCIA). The proposed agreement measure is defined based on a functional disagreement function $G(U, V)$ for two measurements $U$ and $V$ obtained on the same subject by two methods, which must satisfy (a) $G(U, V) \geq 0$, and (b) $G(U, V)$ increases as the disagreement between $U$ and $V$ increases for any $U(\cdot)$ and $V(\cdot)$ measured at $t \in \Gamma$, where $\Gamma$ is a finite closed real interval.

Consider that we measure the agreement separately for each functional network in the second stage, we thus omit the functional network index $l$ for the network-specific temporal responses when introducing the proposed measures. Let $\mathbf{a}_j(\cdot)$ represent the random network-specific temporal response estimated from a random scan at site $j$ and $\mathbf{a}_{ijk}(\cdot)$ be the estimated network-specific temporal response for subject $i$ acquired in $k$th replicated scan at site $j$. The disagreements for a particular subject $i$ are defined as

$$
\begin{aligned}
G_i(\mathbf{a}_1, \mathbf{a}_2) &= E\left[\int_\Gamma (\mathbf{a}_{i1k}(t) - \mathbf{a}_{i2k'}(t))^2 dt | i\right], \\
G_i(\mathbf{a}_1, \mathbf{a}_1') &= E\left[\int_\Gamma (\mathbf{a}_{i1k}(t) - \mathbf{a}_{i1k'}(t))^2 dt | i, k \neq k'\right], \\
G_i(\mathbf{a}_2, \mathbf{a}_2') &= E\left[\int_\Gamma (\mathbf{a}_{i2k}(t) - \mathbf{a}_{i2k'}(t))^2 dt | i, k \neq k'\right].
\end{aligned}
$$

And the overall disagreement functions are defined as

$$G(\mathbf{a}_1, \mathbf{a}_2) = E\left[G_i(\mathbf{a}_1, \mathbf{a}_2)\right],$$
$$G(\mathbf{a}_1, \mathbf{a}_1') = E\left[G_i(\mathbf{a}_1, \mathbf{a}_1')\right],$$
$$G(\mathbf{a}_2, \mathbf{a}_2') = E\left[G_i(\mathbf{a}_2, \mathbf{a}_2')\right].$$

The between-site disagreement $G(\mathbf{a}_1, \mathbf{a}_2)$ represents the expected squared difference of temporal responses for the same subjects between two imaging sites. In addition, the within-site disagreement $G(\mathbf{a}_j, \mathbf{a}_j')$ represents the expected squared difference of temporal responses between any pair of replicated scans for the same subjects at site $j$.

Motivated by CIA for scalar measurements (Barnhart et al., 2007), we propose a functional CIA (fCIA) for image data with replicated scans, which is specified as the ratio of the average within-site disagreement to between-site disagreement:

$$\psi_{fc} = \frac{G(\mathbf{a}_1, \mathbf{a}_1') + G(\mathbf{a}_2, \mathbf{a}_2')}{2G(\mathbf{a}_1, \mathbf{a}_2)}.$$

In the definition of fCIA, we adapt the form of original CIA for scalar measurements (Barnhart et al., 2007) and extend it to accommodate for high-dimensional image data. Similarly as the original CIA measure, the $\psi_{fc}$ is expected to lie between 0 to 1 and we claim satisfactory individual agreement with high value of $\psi_{fc}$ (Barnhart et al. 2007c). The recommendations for claiming good and excellent individual agreement is $\psi_{fc}$ greater than 0.445 and 0.8, respectively (Barnhart et al. 2007; Haber and Barnhart, 2007).

Motivated by total CCC for replicated scalar measurements (Barnhart et al. 2005, 2007), we extend Lin's CCC to acommondate high-dimensional fMRI data with replicated scans. The proposed functional total CCC (fTCCC) for replicated image data

is defined as

$$\rho_{fc} = 1 - \frac{G(\mathbf{a}_1, \mathbf{a}_2)}{G(\mathbf{a}_1, \mathbf{a}_2)_{\mathbf{a}_1 \perp \mathbf{a}_2}}.$$

The $\rho_{fc}$ ranges from $-1$ to 1 as Lin's CCC and higher value represents better agreement. As most agreement measures, $\psi_{fc}$ and $\rho_{fc}$ decrease as the location shift $|E(\mathbf{a}_1) - E(\mathbf{a}_2)|$ and scale shift $|\sigma(\mathbf{a}_1) - \sigma(\mathbf{a}_2)|$ between the same subjects' images from two sites increase and increase as the correlation $\rho(\mathbf{a}_1, \mathbf{a}_2)$ increases. One major difference between $\psi_{fc}$ and $\rho_{fc}$ lies in their relationships with respect to the magnitude of within- and between-subject variabilities. The $\psi_{fc}$ increases as the within-subject between-scan variability increases or as the between-subject within-site variability decreases, while $\rho_{fc}$ changes in the opposite direction. Other major differences between $\psi_{fc}$ and $\rho_{fc}$ are the use of the different benchmark models as well as the interpretations for the extreme values. For fTCCC, the squared distance between two temporal responses is scaled by the independence model. Therefore, the $\rho_{fc}$ equals 0 if and only if the correlation between two imaging sites equals 0. When $\rho_{fc}$ equals 1, we declare perfect agreement and suggest any pair of the temporal responses estimated from the same subjects and the two sites are identical. On the other hand, the fCIA measures the between-site disagreement relative to the within-site disagreement and we claim perfect individual agreement as long as the degree of between-site disagreement is equivalent to the degree of within-site disagreement.

The proposed fCIA and fTCCC share similar spirits as the functional extension for CCC (Li and Chow, 2005) in that, all these measures evaluate the scaled squared distance between the functional measures of two brain images acquired from the same subjects. To distinguish between the proposed measures and the functional CCC, we

note that the proposed measures evaluate the similarity between two fMRI images using summary temporal responses extracted from the most representative regions of the brain (i.e., functional networks) corresponding to the experimental task. These features of image data that we are comparing using fCIA and fTCCC have high signal-to-noise ratio, and reveal the most relevant image quality in an image study. Alternatively, functional CCC measures the voxel-level differences between the two images and averages across all voxels in the brain, which is inefficient for the noisy brain images.

### 3.2.2 Estimation and Inference

For estimation, we reorganized the proposed agreement measures as follows,

$$\psi_{fc} = \frac{C_1 + C_2}{B_1 + B_2 + D - 2A},$$
$$\rho_{fc} = \frac{2A}{B_1 + B_2 + D},$$

which consist of six components, i.e.,

$$A = E \int [E(\mathbf{a}_{i1k}(\nu)|i) - E(\mathbf{a}_1(\nu))][E(\mathbf{a}_{i2k}(\nu)|i) - E(\mathbf{a}_2(\nu))]d\nu,$$
$$B_1 = E \int [\mathbf{a}_{i1k}(\nu) - E(\mathbf{a}_1(\nu))]^2 d\nu, \qquad B_2 = E \int [\mathbf{a}_{i2k}(\nu) - E(\mathbf{a}_2(\nu))]^2 d\nu,$$
$$C_1 = E \int [\mathbf{a}_{i1k}(\nu) - E(\mathbf{a}_{i1k}(\nu)|i)]^2 d\nu, \quad C_2 = E \int [\mathbf{a}_{i2k}(\nu) - E(\mathbf{a}_{i2k}(\nu)|i)]^2 d\nu,$$
$$D = E \int [E(\mathbf{a}_1(\nu)) - E(\mathbf{a}_2(\nu))]^2 d\nu,$$

These components can be viewed as the covariance or variance structure for time series data. The corresponding estimators for the above components $A, B_j, C_j$ and $D$

$(j = 1, 2)$ are defined as follows,

$$
\begin{aligned}
A_N &= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\bar{\mathbf{a}}_{i1}^{K}(\nu_t) - \bar{\mathbf{a}}_{1}^{NK}(\nu_t)\right]\left[\bar{\mathbf{a}}_{i2}^{K}(\nu_t) - \bar{\mathbf{a}}_{2}^{NK}(\nu_t)\right]\Delta\nu_t, \\
B_{jN} &= \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbf{a}_{ijk}(\nu_t) - \bar{\mathbf{a}}_{j}^{NK}(\nu_t)\right]^2\Delta\nu_t, \quad (j = 1, 2) \\
C_{jN} &= \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbf{a}_{ijk}(\nu_t) - \bar{\mathbf{a}}_{ij}^{K}(\nu_t)\right]^2\Delta\nu_t, \quad (j = 1, 2) \\
D_N &= \sum_{t=1}^{T}\left[\bar{\mathbf{a}}_{1}^{NK}(\nu_t) - \bar{\mathbf{a}}_{2}^{NK}(\nu_t)\right]^2\Delta\nu_t,
\end{aligned}
$$

where $\bar{\mathbf{a}}_{ij}^{K}(\nu_t) = \frac{1}{K}\sum_{k=1}^{K}\mathbf{a}_{ijk}(\nu_t)$ and $\bar{\mathbf{a}}_{j}^{NK}(\nu_t) = \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\mathbf{a}_{ijk}(\nu_t)$ are the sample means of $\mathbf{a}_{ijk}(\nu_t)$. We propose the following sample estimators for $\psi_{fc}$ and $\rho_{fc}$ by plugging the above estimators.

$$
\begin{aligned}
\hat{\psi}_{fc} &= \frac{C_{1N} + C_{2N}}{B_{1N} + B_{2N} + D_N - 2A_N}, \\
\hat{\rho}_{fc} &= \frac{2A_N}{B_{1N} + B_{2N} + D_N}.
\end{aligned}
$$

In the following theorem, we establish the asymptotic properties of the proposed estimators. The proof is provided in the Appendix 3.6.

**Theorem 3.1.** *Suppose* $\{(\boldsymbol{a}_{i1k}(\nu_t), \boldsymbol{a}_{i2k}(\nu_t)), i = 1, \cdots, N, k = 1, \cdots, K, t = 1, \cdots, T\}$ *is a collection of random samples of the paired network-specific temporal responses for subject* $i$ *in the* $k$*th scan from sites* $j = 1, 2$*. If Conditions 1(A)-1(C) given in Appendix 3.6 hold, then*

*(i)* $\hat{\psi}_{fc}$ *is a consistent estimator for* $\psi_{fc}$*. The limiting distribution of* $\sqrt{N}(\hat{\psi}_{fc} - \psi_{fc})$ *as* $N \to \infty$*, is a zero mean normal distribution with variance* $\sigma^2_{\psi_{fc}} = H^T_{\psi_{fc}}\Sigma H_{\psi_{fc}}$*,*

*where $\Sigma$ is given in the Appendix and*

$$H_{\psi_{fc}} = \frac{\begin{pmatrix} 2C_1 + 2C_2 \\ -C_1 - C_2 \\ -C_1 - C_2 \\ B_1 + B_2 + D - 2A \\ B_1 + B_2 + D - 2A \\ -C_1 - C_2 \end{pmatrix}}{(B_1 + B_2 + D - 2A)^2}.$$

*(ii) $\hat{\rho}_{fc}$ is a consistent estimator for $\rho_{fc}$. The limiting distribution of $\sqrt{N}(\hat{\rho}_{fc} - \rho_{fc})$ as $N \to \infty$, is a zero mean normal distribution with variance $\sigma^2_{\rho_{fc}} = H^T_{\rho_{fc}} \Sigma H_{\rho_{fc}}$, where*

$$H_{\rho_{fc}} = \frac{(2B_1 + 2B_2 + 2D, -2A, -2A, 0, 0, -2A)^T}{(B_1 + B_2 + D)^2}.$$

According to Theorem 3.1, the standard errors of the proposed estimator $\hat{\psi}_{fc}$ and $\hat{\rho}_{fc}$ can be estimated by

$$SE(\hat{\psi}_{fc}) = \hat{\sigma}^2_{\psi_{fc}}/\sqrt{N-3},$$
$$SE(\hat{\rho}_{fc}) = \hat{\sigma}^2_{\rho_{fc}}/\sqrt{N-3},$$

where $\hat{\sigma}^2_{\psi_{fc}} = \hat{H}^T_{\psi_{fc}} \hat{\Sigma} \hat{H}_{\psi_{fc}}$, $\hat{\sigma}^2_{\rho_{fc}} = \hat{H}^T_{\rho_{fc}} \hat{\Sigma} \hat{H}_{\rho_{fc}}$ and $\hat{\Sigma}$, $\hat{H}_{\psi_{fc}}$ and $\hat{H}_{\rho_{fc}}$ are the corresponding sample counterparts. Here, $N-3$ is used instead of $N-2$ for small sample estimation (Li and Chow, 2005). Furthermore, the asymptotic confidence intervals for $\psi_{fc}$ and $\rho_{fc}$ can be constructed as

$$\hat{\psi}_{fc} \pm t_{N-3}(1-\alpha/2) SE(\hat{\psi}_{fc}),$$
$$\hat{\rho}_{fc} \pm t_{N-3}(1-\alpha/2) SE(\hat{\rho}_{fc}).$$

## 3.3 fBIRN Data Example

In this section, we applied the proposed agreement method to the fBIRN traveling-subject study, which was conducted in 2003 as the pilot study of fBIRN multi-site studies.

### 3.3.1 Preprocessing procedure

In this paper, we focused on the fMRI data acquired while subjects were performing sensorimotor (SM) tasks from the four imaging sites using 3T/4T scanners. The SM tasks employed a block design, with each block taking 10 TRs (30 sec) beginning with 5 TRs of (15 sec) of sensorimotor activity and followed by 5 TRs (15 sec) of rest. Each scan was consisted of eight complete on/off cycles and an initial 5-TR rest period for a total time of 85 TRs (255 sec). During the active phase, subjects were instructed to tap their fingers bilaterally as soon as they observed the signals given by simultaneous tone presentation and checkboard flash. The subjects were instructed to perform finger tapping in an alternating pattern as index, middle, ring, little, little, ring, middle, index, and repeated. In each imaging site, the same procedure was repeated for four times for each subject. Images were preprocessed according to the recommended procedure from the fBIRN website.

### 3.3.2 First Stage

We used GIFT software (GIFTv2.0a) to perform first-stage group ICA analysis and extracted 20 important components representing 20 potential brain functional networks. Four important ICs were identified out of these 20 by matching their locations with known functional networks according to Laird et al. (2011) and Smith et al. (2009). The identified ICs also overlapped with all six regions of interest (ROIs) identified in Friedman et al. (2008). Specifically, $IC_1$ covered left and right motor

cortices and bilateral supplementary motor area (MN), $IC_2$ was mainly consisted of left and right auditory cortices (AN). $IC_3$ and $IC_4$ were related to primary (VN1) and secondary (VN2) bilateral visual cortices, respectively (Figure 3.1). Figures 3.2-3.5 present the average time courses across scans corresponding to four ICs by site and subject. Most of the time series in these figures demonstrate nice and clear periodic patterns which were expected from the block task design. However, some time series appeared to be more variable than the others, for example, time series corresponding to the activities in the VN2 network at site 2. This maybe resulted from some malpractice in the imaging acquiring process and hence reduced the reliability in combine the data across sites.

### 3.3.3 Second Stage

In the second stage, we assessed the pairwise agreement among the four imaging sites using the proposed fCIA and fTCCC for the temporal responses corresponding to the four extracted functional networks (Tables 3.1). The estimated fCIA ranged between 0.5 to 0.75 indicating good individual agreement between the network-specific temporal dynamics. On the other hand, the estimated fTCCC was very small, indicating there was low agreement or no beyond chance agreement between the network-specific temporal dynamics. For VN2 network, Table 2 suggested that the agreement between site 2 and the others were consistently lower.

To test for the site effects on fMRI imaging reproducibility, permutation tests (Williamson et al, 2007) were carried out using the proposed measures (Table 3.2). The test hypothesis was specified as the average agreement between the site of interest and the other sites was equivalent to the average agreement among the other sites. Specifically, denote $\psi_{fc}(j_1, j_2), j_1, j_2 = 1, 2, 3, 4$ as the proposed fCIA between the network-specific temporal responses of fMRI images acquired in site $j_1$ and $j_2$.

Assuming site 1 is the site of interest, the test hypothesis was that

$$H_0 : \frac{\psi_{fc}(1,2) + \psi_{fc}(1,3) + \psi_{fc}(1,4)}{3} = \frac{\psi_{fc}(2,3) + \psi_{fc}(2,4) + \psi_{fc}(3,4)}{3}.$$

The test hypothesis for $\rho_{fc}$ was constructed along the same line. Site indices were permuted to generate the permutation datasets. For this data, there were in total 4*4*4*4=256 permutes resulting in the smallest possible p value of 0.004. In Table 3.2, all the p values were greater than 0.05 suggesting no differences were statistically significant for all four brain networks among all four sites. The small sample size of this data was likely to result in the lack of statistical power to detect the site effect. Despite none of the p values were lower than the 0.05 cut-point, site 1 and site 2 were shown to have marginally significant lower average agreement as compared to other sites for AN and VN2 networks, respectively (p value=0.082 and 0.081) using the proposed fCIA measure. These results were consistent with our previous findings.

Figure 3.1: Four identified ICs from group ICA analysis on the first stage.

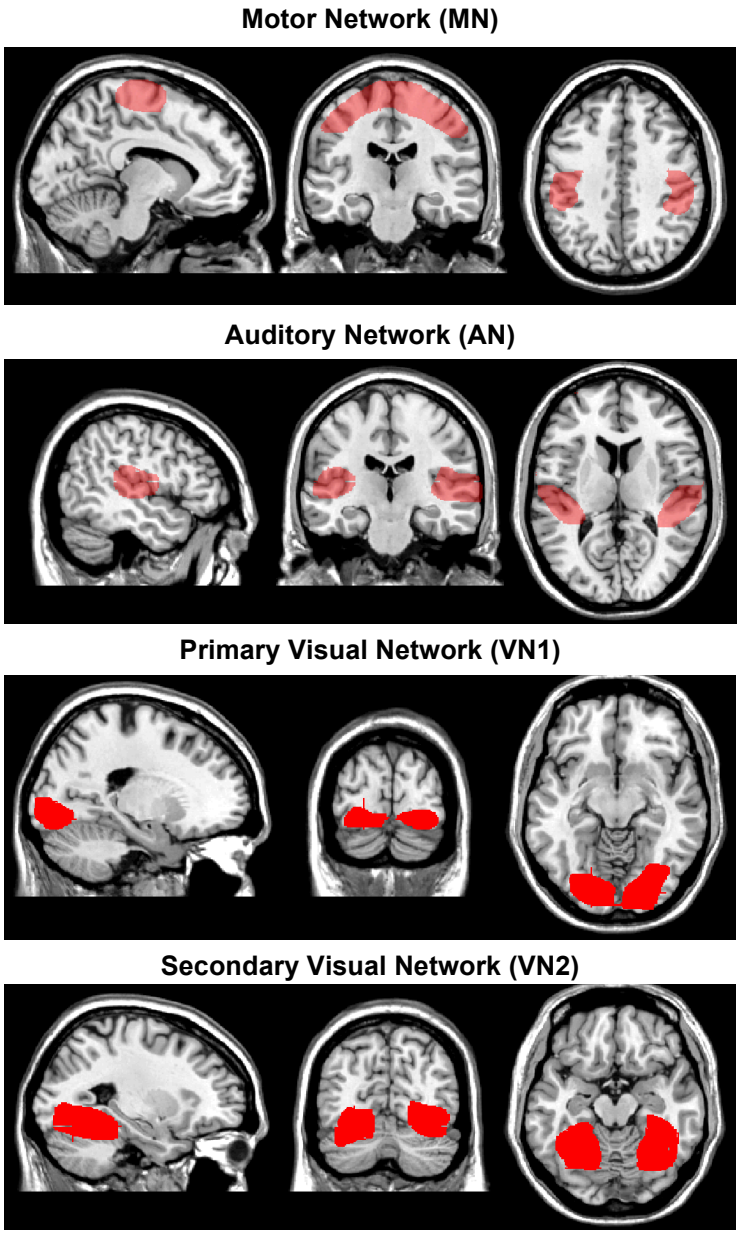


Figure 3.2: Average temporal responses of Motor Network (MN) across scan by site and subject.

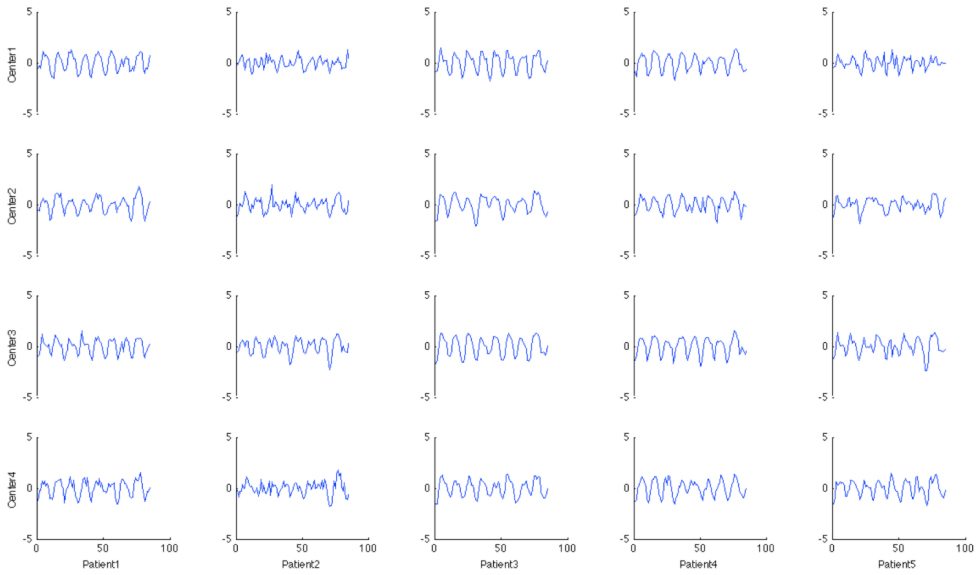



Figure 3.3: Average temporal responses of Auditory Network (AN) across scan by site and subject.

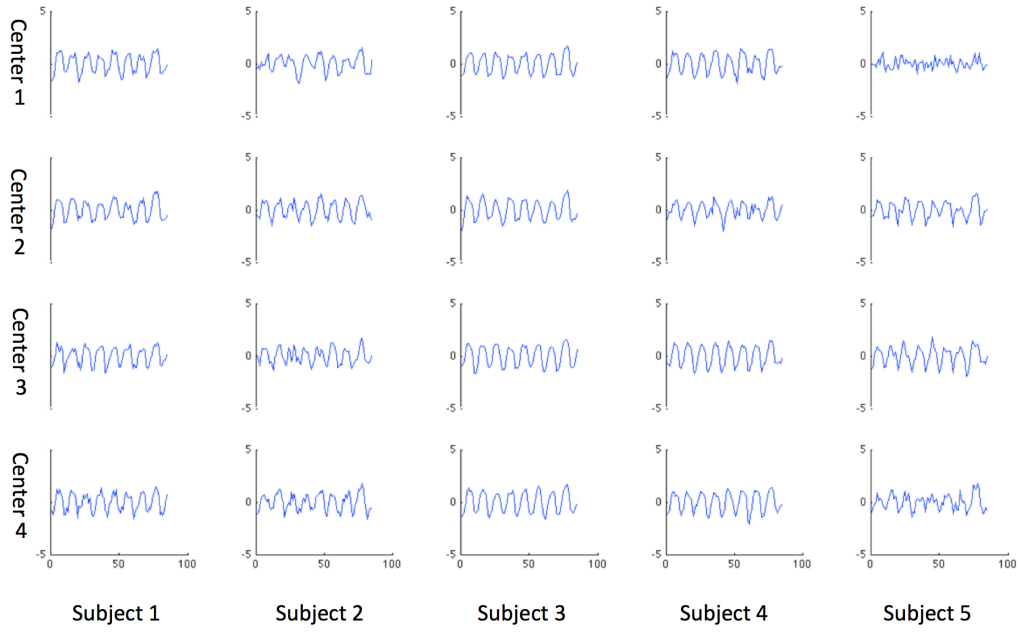


Figure 3.4: Average temporal responses of Primary Visual Network (VN1) across scan by site and subject.

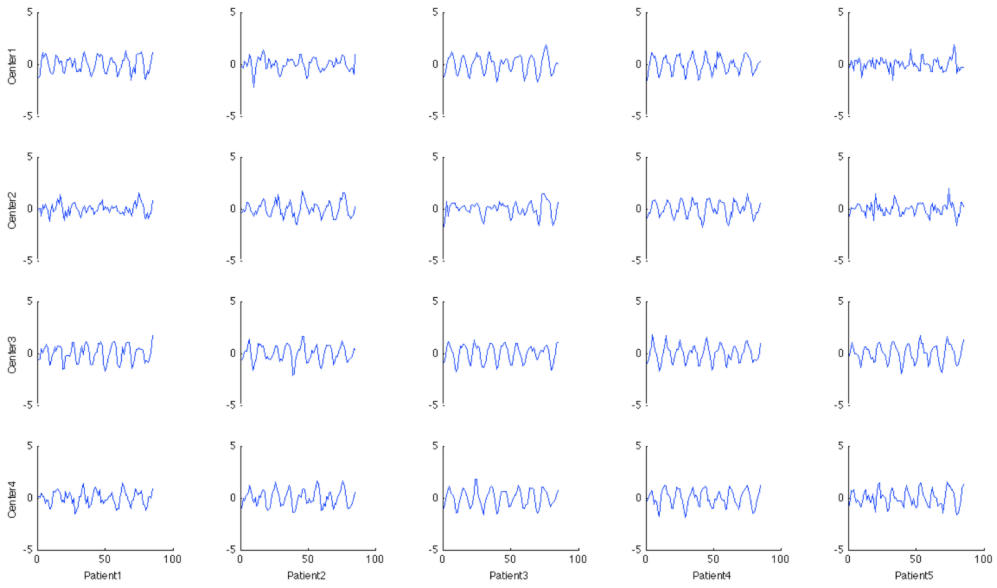



Figure 3.5: Average temporal responses of Secondary Visual Network (VN2) across scan by site and subject.

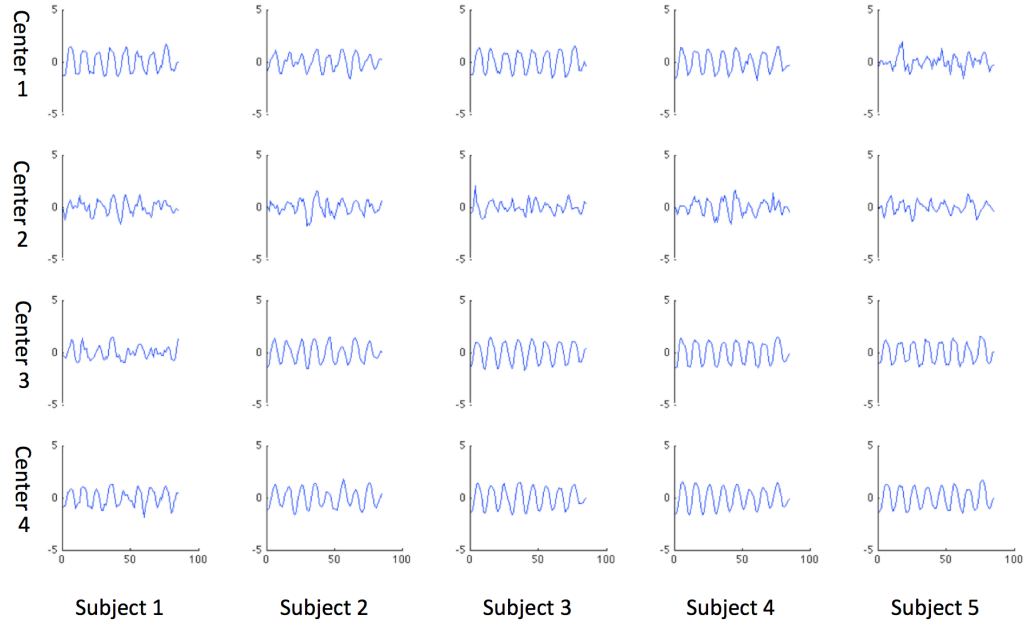


Table 3.1: Measuring agreement of the network-specific temporal responses between every pair of the four imaging sites using the proposed agreement indices $\psi_{fc}$ and $\rho_{fc}$.

| | | **Motor(MN)** | | | | **Auditory(AN)** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Site 1 | Site 2 | Site 3 | Site 4 | Site 1 | Site 2 | Site 3 | Site 4 |
| Site 1 | $\psi_{fc}$ | | 0.732 | 0.708 | 0.741 | | 0.684 | 0.670 | 0.708 |
| | $\rho_{fc}$ | | 0.050 | 0.046 | 0.062 | | 0.028 | 0.056 | 0.068 |
| Site 2 | $\psi_{fc}$ | | | 0.708 | 0.729 | | | 0.732 | 0.731 |
| | $\rho_{fc}$ | | | 0.052 | 0.049 | | | 0.074 | 0.044 |
| Site 3 | $\psi_{fc}$ | | | | 0.723 | | | | 0.736 |
| | $\rho_{fc}$ | | | | 0.072 | | | | 0.089 |
| Site 4 | $\psi_{fc}$ | | | | | | | | |
| | $\rho_{fc}$ | | | | | | | | |

| | | **Primary Visual(VN1)** | | | | **Secondary Visual(VN2)** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Site 1 | Site 2 | Site 3 | Site 4 | Site 1 | Site 2 | Site 3 | Site 4 |
| Site 1 | $\psi_{fc}$ | | 0.674 | 0.648 | 0.643 | | 0.619 | 0.611 | 0.653 |
| | $\rho_{fc}$ | | -0.023 | -0.007 | 0.015 | | 0.067 | -0.011 | 0.043 |
| Site 2 | $\psi_{fc}$ | | | 0.649 | 0.632 | | | 0.548 | 0.528 |
| | $\rho_{fc}$ | | | 0.058 | 0.029 | | | 0.013 | 0.028 |
| Site 3 | $\psi_{fc}$ | | | | 0.660 | | | | 0.682 |
| | $\rho_{fc}$ | | | | 0.076 | | | | 0.102 |
| Site 4 | $\psi_{fc}$ | | | | | | | | |
| | $\rho_{fc}$ | | | | | | | | |

Table 3.2: Permutation tests for the site effects on the task-fMRI imaging reproducibility. The test hypothesis was specified as the average agreement between the site of interest and the other three sites was equivalent to the average agreement among the other sites.

| Network | Index | Site 1 vs Others | | Site 2 vs Others | | Site 3 vs Others | | Site 4 vs Others | |
|---|---|---|---|---|---|---|---|---|---|
| | | Difference | P value | Difference | P value | Difference | P value | Difference | P value |
| **MN** | $\psi_{fc}$ | 0.007 | 0.717 | -0.001 | 0.954 | -0.021 | 0.266 | 0.015 | 0.441 |
| | $\rho_{fc}$ | -0.005 | 0.775 | -0.010 | 0.606 | 0.003 | 0.868 | 0.012 | 0.555 |
| **AN** | $\psi_{fc}$ | -0.046 | 0.082 | 0.011 | 0.741 | 0.005 | 0.871 | 0.030 | 0.346 |
| | $\rho_{fc}$ | -0.018 | 0.406 | -0.023 | 0.262 | 0.026 | 0.184 | 0.015 | 0.560 |
| **VN1** | $\psi_{fc}$ | 0.008 | 0.815 | 0.001 | 0.974 | 0.003 | 0.924 | -0.012 | 0.711 |
| | $\rho_{fc}$ | -0.059 | 0.156 | -0.007 | 0.872 | 0.035 | 0.425 | 0.031 | 0.485 |
| **VN2** | $\psi_{fc}$ | 0.042 | 0.356 | -0.084 | 0.081 | 0.014 | 0.744 | 0.029 | 0.531 |
| | $\rho_{fc}$ | -0.015 | 0.753 | -0.009 | 0.882 | -0.011 | 0.812 | 0.035 | 0.495 |

## 3.4 Simulation Studies

In this section, we performed simulation studies to evaluate the performance of the proposed two-stage estimation method and the inference procedure.

Given the sample size of the fBIRN dataset is fairly small (N=5), we performed Monte Carlo simulations for small (N=5) and moderate (N=20) sample sizes. We simulated three scans of 4-D fMRI data for each subject at each site assuming there were three independent components (IC) representing different brain functional networks. For each IC, we generated a 3D spatial map with 4 slices, where each slice consists of $25 \times 25$ voxels. These spatial maps were generated as linear combinations of the Gaussian random variability with a standard deviation of 0.5 at each voxel and the source signals that followed a distribution of $N(4, 0.5)$ at activated voxels. Figure 3.6 presents a set of the simulated IC spatial maps from one simulation run.

After creating the group spatial maps, we then simulated temporal responses that corresponded to these ICs $l = 1, 2, 3$ using the following procedure. To set notations, we denoted subject index $i = 1, \cdots, N$, site index $j = 1, 2$, scan index $k = 1, 2, 3$, signal index $l = 1, 2, 3$, and time index $\nu_1, \cdots, \nu_T$. First, we generated the mean temporal responses for subject $i$ and signal $l$ at two sites $(E'(\mathbf{a}_{i1kl}(t)|i, l), E'(\mathbf{a}_{i2kl}(t)|i, l))'$ jointly from Z-dependent gaussian processes with zero mean, unit variance and correlation of $\sigma_{\mathbf{a}_{1l}\mathbf{a}_{2l}}(t)$, for which we set Z=20 and $\{\nu_1, \cdots, \nu_T\}$ be 50 equally spaced time points over a period of time $[0, 1]$. Second, we sampled the between-scan random variability $\mathbf{e}_{ijkl}(t)$ from another zero-mean Z-dependent gaussian process with covariance $\sigma^2_{\mathbf{a}_{jKl}}(t)$, $j = 1, 2, l = 1, 2, 3$. Third, we simulated the site-subject-scan-specific temporal response for spatial signal $l = 1, 2, 3$ as $\mathbf{a}_{i1kl}(t) = E(\mathbf{a}_{i1kl}(t)|i, l) + \mathbf{e}_{i1kl}(t)$ and $\mathbf{a}_{i2kl}(t) = E(\mathbf{a}_{i2kl}(t)|i, l) + \mathbf{e}_{i2kl}(t)$, for $i = 1, \cdots, N$, $k = 1, 2, 3$. Parameters for simulating the temporal responses for the

three ICs were specified to create different levels of agreement as shown below:

IC 1: $\sigma_{\mathbf{a}_{11}\mathbf{a}_{21}}(t) = 0.5$, $\sigma^2_{\mathbf{a}_{1K1}}(t) = 2$, $\sigma^2_{\mathbf{a}_{2K1}}(t) = 1 \rightarrow \psi_{fc} = 0.75, \rho_{fc} = 0.2$.

IC 2: $\sigma_{\mathbf{a}_{12}\mathbf{a}_{22}}(t) = 0.7$, $\sigma^2_{\mathbf{a}_{1K2}}(t) = 2$, $\sigma^2_{\mathbf{a}_{2K2}}(t) = 1 \rightarrow \psi_{fc} = 0.833, \rho_{fc} = 0.28$.

IC 3: $\sigma_{\mathbf{a}_{13}\mathbf{a}_{23}}(t) = 0.7$, $\sigma^2_{\mathbf{a}_{1K3}}(t) = 2$, $\sigma^2_{\mathbf{a}_{2K3}}(t) = 4 \rightarrow \psi_{fc} = 0.857, \rho_{fc} = 0.125$.

Finally, 4-D fMRI data was simulated by adding a Gaussian background noise with a standard deviation of 1 to the mixed spatial sources based on the simulated group spatial maps and temporal responses.

The two-stage estimation method and the inference procedure were implemented for the simulated fMRI data. Given that there was no ordering algorithm for the ICs using group ICA method, we indexed the estimated ICs based on the highest spatial correlation between the estimated IC maps and the true spatial maps. At the second stage, we assessed the agreement on the matched temporal responses for each IC using the proposed agreement indices.

Simulation results based on 200 simulated data sets were summarized in Table 3.3. Under all simulation settings, the proposed estimators provided accurate estimation for both agreement indices. The mean standard deviation estimates based on the proposed inference procedure were very close to the empirical standard errors for both agreement indices. Furthermore, the coverage probabilities of the 95% confidence intervals were close to the nominal level. The simulation results implied that the proposed estimators and the corresponding inference procedure performed well for small to moderate small sizes.

Table 3.3: Simulation results of the two-stage estimation method and the inference procedure for $\psi_{fc}$ and $\rho_{fc}$. Bias, empirical standard deviation, mean of estimated standard deviation and coverage probability of the 95% confidence interval are presented along with true values of the proposed agreement indices for three spatial signals (ICs) with sample sizes of 5 and 20.

| N | Index | Truth | Bias | EmpSE | MeanSD | Cov Prob(%) |
|---|---|---|---|---|---|---|
| **IC 1** | | | | | | |
| 5 | $\psi_{fc}$ | 0.750 | 0.003 | 0.098 | 0.116 | 98.5 |
| | $\rho_{fc}$ | 0.200 | -0.039 | 0.091 | 0.092 | 99.0 |
| 20 | $\psi_{fc}$ | 0.750 | 0.005 | 0.045 | 0.048 | 92.5 |
| | $\rho_{fc}$ | 0.200 | -0.008 | 0.046 | 0.045 | 94.5 |
| **IC 2** | | | | | | |
| 5 | $\psi_{fc}$ | 0.833 | 0.011 | 0.086 | 0.115 | 98.0 |
| | $\rho_{fc}$ | 0.280 | -0.036 | 0.086 | 0.094 | 96.0 |
| 20 | $\psi_{fc}$ | 0.833 | 0.003 | 0.049 | 0.048 | 94.5 |
| | $\rho_{fc}$ | 0.280 | -0.017 | 0.043 | 0.043 | 90.5 |
| **IC 3** | | | | | | |
| 5 | $\psi_{fc}$ | 0.857 | 0.014 | 0.094 | 0.117 | 96.5 |
| | $\rho_{fc}$ | 0.125 | -0.018 | 0.067 | 0.077 | 98.0 |
| 20 | $\psi_{fc}$ | 0.857 | 0.000 | 0.047 | 0.049 | 93.5 |
| | $\rho_{fc}$ | 0.125 | -0.005 | 0.042 | 0.038 | 94.5 |

## 3.5 Remarks

In this paper, we propose a two-stage network-based agreement method for assessing the reproducibility of the same subjects' high dimensional brain imaging data acquired at different sites. We develop nonparametric estimation methods for the proposed indices and prove them to be consistent and asymptotically normal by theoretical derivation and simulation studies. The proposed methods are illustrated via fBIRN Phase I Traveling Subject study to investigate site effects in multi-site imaging studies. Nonparametric permutation test based on the fCIA index reveals some potential site effect between site 1, 2 and the other sites, which is consistent with visual examination of the data, while the fTCCC fails to detect the abnormality. Since our agreement method measures network-specific agreement, we are able to identify that the inconsistency of site 1 in task-fMRI data mostly comes from auditory network and the discrepancy in fMRI images at site 2 is mainly found in secondary visual network. Further examination into the task protocols and image acquiring procedure need be performed correspondingly to ensure better between-site agreement in the subsequent studies. Furthermore, the proposed agreement measures could potentially be used as an objective function for calibration purpose. Although detailed procedures will need further investigation.

## 3.6 Appendix

In this section, we present detailed proofs of Theorem 3.1. First, we present the regularity conditions.

**Definition 3.6.1.** Let $0 = \nu_0 < \cdots < \nu_T = 1$ be any partition $\mathscr{P}$ of [0,1], and $f(\nu)$ be a real function defined on [0,1]. $f$ is called a function of bounded variation if the variation $V_{\mathscr{P}} = \sum_{t=1}^{T} |f(\nu_{t+1}) - f(\nu_t)|$ has an upper bound, which is independent of the choice of $\mathscr{P}$. The least upper bound of $V_{\mathscr{P}}$ is called the total variation of $f$ and

is denoted by $V(f)$.

For any function of bounded variation, it has been shown that (Hua and Wang, 1996)

$$|\int_0^1 f(\nu)d\nu - \sum_{t=1}^{T} f(\nu_t)\Delta\nu_t| \leq V(f) \underset{0\leqslant t\leqslant T-1}{max}|\nu_{t+1} - \nu_t|.$$

Let $\mathscr{A}_1$ and $\mathscr{A}_2$ denote the space consisting of paths $\mathbf{a}_{i1k}(\cdot)$ and $\mathbf{a}_{i2k}(\cdot)$, $i = 1, \cdots, N; k = 1, \cdots, K$. We further define

$$\mathscr{B}_1 = \{\mathbf{a}_{i1k}(\cdot)\mathbf{a}_{i2k}(\cdot) : \mathbf{a}_{i1k}(\cdot) \in \mathscr{A}_1 \quad and \quad \mathbf{a}_{i2k}(\cdot) \in \mathscr{A}_2\}$$

$$\mathscr{B}_2 = \left\{\mathbf{a}_{i1k}^2(\cdot) : \mathbf{a}_{i1k}(\cdot) \in \mathscr{A}_1\right\}$$

$$\mathscr{B}_3 = \left\{\mathbf{a}_{i2k}^2(\cdot) : \mathbf{a}_{i2k}(\cdot) \in \mathscr{A}_2\right\}$$

**Condition 1.** *(A) For* $\mathscr{F} = \mathscr{A}_1, \mathscr{A}_2, \mathscr{B}_1, \mathscr{B}_2, \mathscr{B}_3$,

$$\underset{f\in F}{sup} V(f) < \infty \qquad a.s.$$

*(B) There exists a constant* $\Delta$ *such that*

$$\underset{0\leqslant t\leqslant T-1}{max}|\nu_{t+1} - \nu_t| \leqslant \frac{\Delta}{T}$$

*and* $\sqrt{N}/T \to 0$ *as* $N \to \infty$.
*(C)* $E\left[\int \boldsymbol{a}_{i1k}^2(\nu)\, d\nu\right]^2 < \infty$, $E\left[\int \boldsymbol{a}_{i2k}^2(\nu) d\nu\right]^2 < \infty$ *and* $E\left[\int \bar{\boldsymbol{a}}_{i1}^K(\nu)\bar{\boldsymbol{a}}_{i2}^K(\nu)\, d\nu\right]^2 < \infty$, *where* $\bar{\boldsymbol{a}}_{ij}^K = \frac{1}{K}\sum_{k=1}^K \boldsymbol{a}_{ijk}$.

To prove Theorem 3.1, we need the following lemma.

**Lemma 3.6.1.** *Under the above condition 1(A)-1(C), we have*

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\boldsymbol{a}}_{i1}^{K}(\nu_t)\bar{\boldsymbol{a}}_{i2}^{K}(\nu_t)\Delta\nu_t = \frac{1}{N}\sum_{i=1}^{N}\int \bar{\boldsymbol{a}}_{i1}^{K}(\nu)\bar{\boldsymbol{a}}_{i2}^{K}(\nu)d\nu + O_p(\frac{1}{T}) \tag{3.1}$$

$$\sum_{t=1}^{T}\bar{\boldsymbol{a}}_{1}^{NK}(\nu_t)\bar{\boldsymbol{a}}_{2}^{NK}(\nu_t)\Delta\nu_t = \int \bar{\boldsymbol{a}}_{1}^{NK}(\nu)\bar{\boldsymbol{a}}_{2}^{NK}(\nu)d\nu + O_p(\frac{1}{T}) \tag{3.2}$$

$$\frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{a}_{ijk}^{2}(\nu_t)\Delta\nu_t = \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\int \boldsymbol{a}_{ijk}^{2}(\nu)d\nu + O_p(\frac{1}{T}) \tag{3.3}$$

$$\sum_{t=1}^{T}\left[\bar{\boldsymbol{a}}_{j}^{NK}(\nu_t)\right]^2\Delta\nu_t = \int \left[\bar{\boldsymbol{a}}_{j}^{NK}(\nu)\right]^2 d\nu + O_p(\frac{1}{T}) \tag{3.4}$$

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\bar{\boldsymbol{a}}_{ij}^{K}(\nu_t)\right]^2\Delta\nu_t = \frac{1}{N}\sum_{i=1}^{N}\int \left[\bar{\boldsymbol{a}}_{ij}^{K}(\nu)\right]^2 d\nu + O_p(\frac{1}{T}) \tag{3.5}$$

*Proof.* By condition 1(B), it follows that

$$\begin{aligned}
&|\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\mathbf{a}}_{i1}^{K}(\nu_t)\bar{\mathbf{a}}_{i2}^{K}(\nu_t)\Delta\nu_t - \frac{1}{N}\sum_{i=1}^{N}\int \bar{\mathbf{a}}_{i1}^{K}(\nu)\bar{\mathbf{a}}_{i2}^{K}(\nu)d\nu| \\
&\leqslant \frac{1}{N}\sum_{i=1}^{N}|\sum_{t=1}^{T}\bar{\mathbf{a}}_{i1}^{K}(\nu_t)\bar{\mathbf{a}}_{i2}^{K}(\nu_t)\Delta\nu_t - \int \bar{\mathbf{a}}_{i1}^{K}(\nu)\bar{\mathbf{a}}_{i2}^{K}(\nu)d\nu| \\
&\leqslant \frac{1}{N}\sum_{i=1}^{N}V(\bar{\mathbf{a}}_{i1}^{K}\bar{\mathbf{a}}_{i2}^{K})\max_{0\leqslant t\leqslant T-1}|\nu_{t+1}-\nu_t| \\
&\leqslant \frac{1}{N}\sum_{i=1}^{N}V(\mathbf{a}_{i1k}\mathbf{a}_{i2k})\max_{0\leqslant t\leqslant T-1}|\nu_{t+1}-\nu_t| \\
&\leqslant \sup_{f\in F} V(f)\frac{\Delta}{T}
\end{aligned}$$

So equation (3.1) follows. Similarly, we can show that equations (3.2)-(3.5) also hold. □

**Proof of Theorem 3.1**

*Proof.* Using Lemma 3.6.1, we show

$$
\begin{aligned}
A_N &= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\bar{\mathbf{a}}_{i1}^{K}(\nu_t)-\bar{\mathbf{a}}_{1}^{NK}(\nu_t)\right]\left[\bar{\mathbf{a}}_{i2}^{K}(\nu_t)-\bar{\mathbf{a}}_{2}^{NK}(\nu_t)\right]\Delta\nu_t \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\mathbf{a}}_{i1}^{K}(\nu_t)\bar{\mathbf{a}}_{i2}^{K}(\nu_t)\Delta\nu_t-\sum_{t=1}^{T}\bar{\mathbf{a}}_{1}^{NK}(\nu_t)\bar{\mathbf{a}}_{2}^{NK}(\nu_t)\Delta\nu_t \\
&= \frac{1}{N}\sum_{i=1}^{N}\int\bar{\mathbf{a}}_{i1}^{K}(\nu)\bar{\mathbf{a}}_{i2}^{K}(\nu)d\nu-\int\bar{\mathbf{a}}_{1}^{NK}(\nu)\bar{\mathbf{a}}_{2}^{NK}(\nu)d\nu_t+O_p(\frac{1}{T}) \\
&= \frac{1}{N}\sum_{i=1}^{N}\int\left[\bar{\mathbf{a}}_{i1}^{K}(\nu)-E(\mathbf{a}_{i1k}(\nu)|i)\right]\left[\bar{\mathbf{a}}_{i2}^{K}(\nu)-E(\mathbf{a}_{i2k}(\nu)|i)\right]d\nu_t \\
&\quad+\frac{1}{N}\sum_{i=1}^{N}\int\left[E(\mathbf{a}_{i1k}(\nu)|i)-E(\mathbf{a}_1(\nu))\right]\left[E(\mathbf{a}_{i2k}(\nu)|i)-E(\mathbf{a}_2(\nu))\right]d\nu_t \\
&\quad+\frac{1}{N}\sum_{i=1}^{N}\int\left[\bar{\mathbf{a}}_{i1}^{K}(\nu)-E(\mathbf{a}_{i1k}(\nu)|i)\right]\left[E(\mathbf{a}_{i2k}(\nu)|i)-E(\mathbf{a}_2(\nu))\right]d\nu_t \\
&\quad+\frac{1}{N}\sum_{i=1}^{N}\int\left[\bar{\mathbf{a}}_{i2}^{K}(\nu)-E(\mathbf{a}_{i2k}(\nu)|i)\right]\left[E(\mathbf{a}_{i2k}(\nu)|i)-E(\mathbf{a}_2(\nu))\right]d\nu_t \\
&\quad-\int[\bar{\mathbf{a}}_{1}^{NK}(\nu)-E(\mathbf{a}_1(\nu))][\bar{\mathbf{a}}_{2}^{NK}(\nu)-E(\mathbf{a}_2(\nu))]d\nu_t+O_P(\frac{1}{T}) \\
&= \frac{1}{N}\sum_{i=1}^{N}\int\left[E(\mathbf{a}_{i1k}(\nu)|i)-E(\mathbf{a}_1(\nu))\right]\left[E(\mathbf{a}_{i2k}(\nu)|i)-E(\mathbf{a}_2(\nu))\right]d\nu_t \\
&\quad+\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=1}^{K}\int\left[\mathbf{a}_{i2k}(\nu)-E(\mathbf{a}_{i2k}(\nu)|i)\right]\left[E(\mathbf{a}_{i1k}(\nu)|i)-E(\mathbf{a}_1(\nu))\right]d\nu_t \\
&\quad+\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=1}^{K}\int\left[\mathbf{a}_{i1k}(\nu)-E(\mathbf{a}_{i1k}(\nu)|i)\right]\left[E(\mathbf{a}_{i2k}(\nu)|i)-E(\mathbf{a}_2(\nu))\right]d\nu_t+O_p(\frac{1}{NK})+O_p(\frac{1}{T}),
\end{aligned}
$$

which tends to $E\int[E(\mathbf{a}_{i1k}(\nu)|i)-E(\mathbf{a}_1(\nu))][E(\mathbf{a}_{i2k}(\nu)|i)-E(\mathbf{a}_2(\nu))]d\nu$ in probability by the week law of large numbers and $T\to\infty$.

Similarly, for $B_{jN}$, $C_{jN}$ and $D_N$, $j = 1, 2$, we have

$$\begin{aligned}
B_{jN} &= \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbf{a}_{ijk}(\nu_t) - \bar{\mathbf{a}}_j^{NK}(\nu_t)\right]^2 \Delta\nu_t \\
&= \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{a}_{ijk}^2(\nu_t)\,\Delta\nu_t - \sum_{t=1}^{T}\left[\bar{\mathbf{a}}_j^{NK}(\nu_t)\right]^2\Delta\nu_t \\
&= \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\int \mathbf{a}_{ijk}^2(\nu)\,d\nu - \int\left[\bar{\mathbf{a}}_j^{NK}(\nu)\right]^2 d\nu + O_p(\frac{1}{T}) \\
&= \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\int \left[\mathbf{a}_{ijk}(\nu) - E\left(\mathbf{a}_j(\nu)\right)\right]^2 d\nu - \int\left[\bar{\mathbf{a}}_j^{NK}(\nu) - E\left(\mathbf{a}_j(\nu)\right)\right]^2 d\nu + O_p(\frac{1}{T}) \\
&= \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\int \left[\mathbf{a}_{ijk}(\nu) - E\left(\mathbf{a}_j(\nu)\right)\right]^2 d\nu + O_p(\frac{1}{NK}) + O_p(\frac{1}{T}).
\end{aligned}$$

$$\begin{aligned}
C_{jN} &= \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbf{a}_{ijk}(\nu_t) - \bar{\mathbf{a}}_{ij}^{K}(\nu_t)\right]^2\Delta\nu_t, \quad (j=1,2) \\
&= \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{a}_{ijk}^2(\nu_t)\Delta\nu_t - \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\bar{\mathbf{a}}_{ij}^{K}(\nu_t)\right]^2\Delta\nu_t \\
&= \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\int \mathbf{a}_{ijk}^2(\nu)d\nu - \frac{1}{N}\sum_{i=1}^{N}\int\left[\bar{\mathbf{a}}_{ij}^{K}(\nu)\right]^2 d\nu + O_p(\frac{1}{T}) \\
&= \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\int \left[\mathbf{a}_{ijk}(\nu) - \bar{\mathbf{a}}_{ij}^{K}(\nu)\right]^2 d\nu + O_p(\frac{1}{T}) \\
&= \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\int \left[\mathbf{a}_{ijk}(\nu) - E(\mathbf{a}_{ijk}(\nu)|i) + E(\mathbf{a}_{ijk}(\nu)|i) - \bar{\mathbf{a}}_{ij}^{K}(\nu)\right]^2 d\nu + O_p(\frac{1}{T}) \\
&= \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\int \left[\mathbf{a}_{ijk}(\nu) - E(\mathbf{a}_{ijk}(\nu)|i)\right]^2 d\nu - \frac{1}{N}\sum_{i=1}^{N}\int\left[E(\mathbf{a}_{ijk}(\nu)|i) - \bar{\mathbf{a}}_{ij}^{K}(\nu)\right]^2 d\nu + O_p(\frac{1}{T}) \\
&= \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N}\int \left[\mathbf{a}_{ijk}(\nu) - E(\mathbf{a}_{ijk}(\nu)|i)\right]^2 d\nu + O_p(\frac{1}{N}) + O_p(\frac{1}{T})
\end{aligned}$$

$$
\begin{aligned}
D_N &= \sum_{t=1}^{T} \left[\bar{\mathbf{a}}_1^{NK}(\nu_t) - \bar{\mathbf{a}}_2^{NK}(\nu_t)\right]^2 \Delta\nu_t \\
&= \sum_{t=1}^{T} \left[\left[\bar{\mathbf{a}}_1^{NK}(\nu_t)\right]^2 + \left[\bar{\mathbf{a}}_2^{NK}(\nu_t)\right]^2 - 2\bar{\mathbf{a}}_1^{NK}(\nu_t)\bar{\mathbf{a}}_2^{NK}(\nu_t)\right] \Delta\nu_t \\
&= \int \left[\left[\bar{\mathbf{a}}_1^{NK}(\nu)\right]^2 + \left[\bar{\mathbf{a}}_2^{NK}(\nu)\right]^2 - 2\bar{\mathbf{a}}_1^{NK}(\nu)\bar{\mathbf{a}}_2^{NK}(\nu)\right] d\nu + O_p(\frac{1}{T}) \\
&= \int \left[\left[\bar{\mathbf{a}}_1^{NK}(\nu) - E(\mathbf{a}_1(\nu)) + E(\mathbf{a}_1(\nu))\right]^2 + \left[\bar{\mathbf{a}}_2^{NK}(\nu) - E(\mathbf{a}_2(\nu)) + E(\mathbf{a}_2(\nu))\right]^2\right] d\nu \\
&\quad - 2\int \left[\bar{\mathbf{a}}_1^{NK}(\nu) - E(\mathbf{a}_1(\nu))\right]\left[\bar{\mathbf{a}}_2^{NK}(\nu) - E(\mathbf{a}_2(\nu))\right] d\nu \\
&\quad - 2\int \left[\left[\bar{\mathbf{a}}_1^{NK}(\nu) - E(\mathbf{a}_1(\nu))\right] E(\mathbf{a}_2(\nu)) + E\left(\mathbf{a}_1(\nu)\right)\bar{\mathbf{a}}_2^{NK}(\nu)\right] d\nu + O_p(\frac{1}{T}) \\
&= \int E^2(\mathbf{a}_1(\nu))d\nu + \int 2E(\mathbf{a}_1(\nu))\left[\bar{\mathbf{a}}_1^{NK}(\nu) - E(\mathbf{a}_1(\nu))\right] d\nu + O_p(\frac{1}{NK}) \\
&\quad + \int E^2(\mathbf{a}_2(\nu))d\nu + \int 2E(\mathbf{a}_2(\nu))\left[\bar{\mathbf{a}}_2^{NK}(\nu) - E(\mathbf{a}_2(\nu))\right] d\nu + O_p(\frac{1}{NK}) + O_p(\frac{1}{NK}) \\
&\quad - 2\int \left\{\left[\bar{\mathbf{a}}_1^{NK}(\nu) - E(\mathbf{a}_1(\nu))\right] E(\mathbf{a}_2(\nu)) + E\left(\mathbf{a}_1(\nu)\right)\bar{\mathbf{a}}_2^{NK}(\nu)\right\} d\nu + O_p(\frac{1}{T})
\end{aligned}
$$

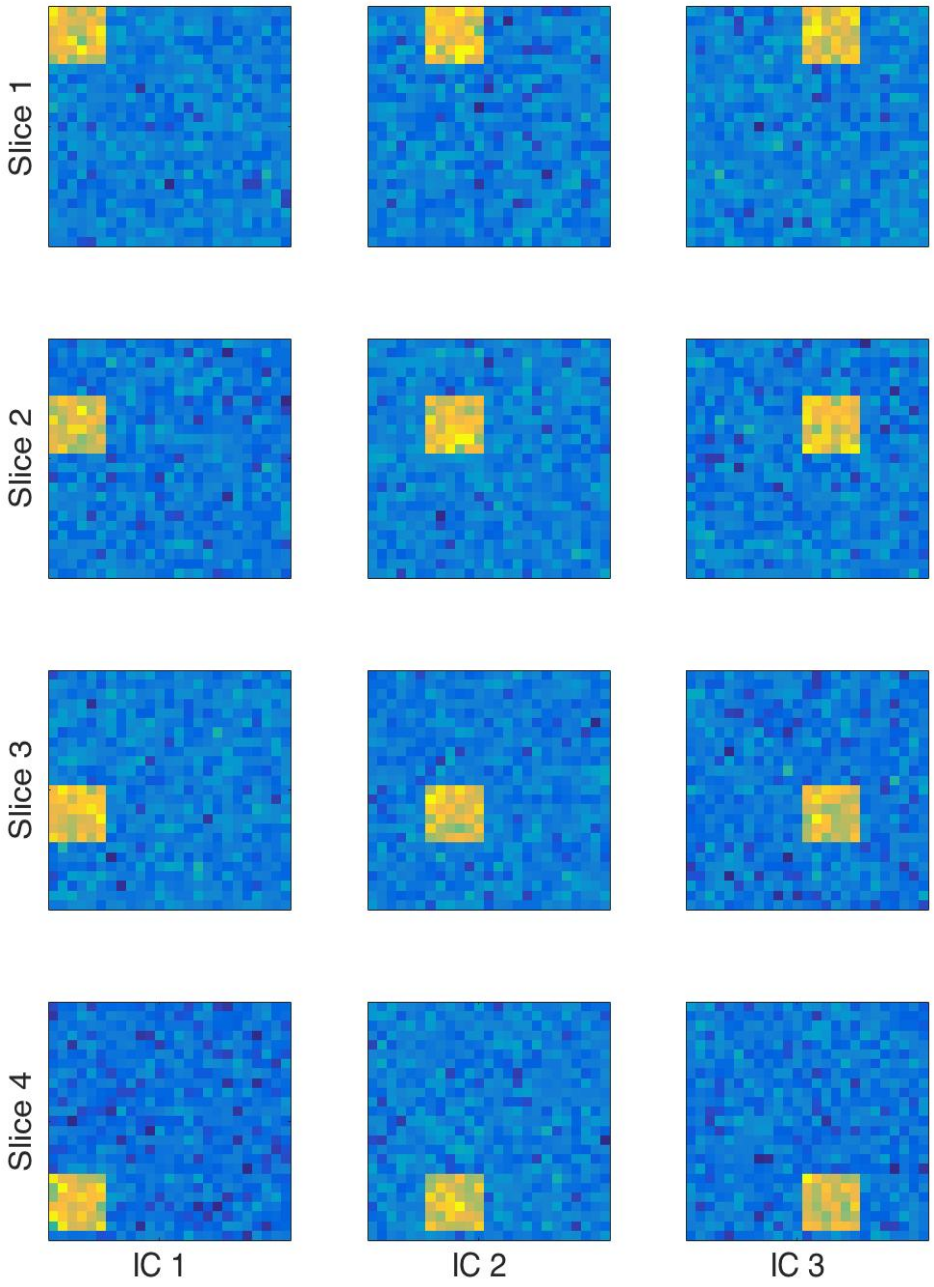
Define $V_N = (A_N, B_{1N}, B_{2N}, C_{1N}, C_{2N}, D_N)^T$ and $V = (A, B_1, B_2, C_1, C_2, D)^T$. Based on condition (B), the multivariate central limit theorem and the Slutsky theorem, $\sqrt{N}(V_N - V)$ has an asymptotic normal distribution with mean zero and the variance $\Sigma =$

$$
cov\left\{\begin{pmatrix}
\int \left[E(\mathbf{a}_{i1k}|i) - E(\mathbf{a}_1)\right]\left[E(\mathbf{a}_{i2k}|i) - E(\mathbf{a}_2)\right] + \left[\bar{\mathbf{a}}_{i2}^{K} - E(\mathbf{a}_{i2k}|i)\right]\left[E(\mathbf{a}_{i1k}|i) - E(\mathbf{a}_1)\right] \\
+ \left[\bar{\mathbf{a}}_{i1}^{K} - E(\mathbf{a}_{i1k}|i)\right]\left[E(\mathbf{a}_{i2k}|i) - E(\mathbf{a}_2)\right] d\nu \\
\int \left[\mathbf{a}_{1k} - E(\mathbf{a}_1)\right]^2 d\nu \\
\int \left[\mathbf{a}_{2k} - E(\mathbf{a}_2)\right]^2 d\nu \\
\int \left[\mathbf{a}_{1k} - E(\mathbf{a}_{i1k}|i)\right]^2 d\nu \\
\int \left[\mathbf{a}_{2k} - E(\mathbf{a}_{i2k}|i)\right]^2 d\nu \\
\int \{E(\mathbf{a}_1)^2 + E(\mathbf{a}_2)^2 - 2E(\mathbf{a}_1)\bar{\mathbf{a}}_2 + 2\left[\bar{\mathbf{a}}_1(\nu) - E(\mathbf{a}_1)\right] E(\mathbf{a}_1) \\
+2\left[\bar{\mathbf{a}}_2 - E(\mathbf{a}_2) - \bar{\mathbf{a}}_1 + E(\mathbf{a}_1)\right] E(\mathbf{a}_2)\} d\nu
\end{pmatrix}\right\}.
$$

Define $h_{\psi_{fc}}(v_1, v_2, v_3, v_4, v_5, v_6) = (v_4 + v_5)/(v_2 + v_3 + v_6 - 2v_1)$ and $h_{\rho_{fc}}(v_1, v_2, v_3, v_4, v_5, v_6) = 2v_1/(v_2 + v_3 + v_6)$. Thus, $\hat{\psi}_{fc} = h_{\psi_{fc}}(V_N')$, $\hat{\rho}_{fc} = h_{\rho_{fc}}(V_N')$.

Theorem 3.1 holds based on the functional delta method. □

Figure 3.6: Simulated IC spatial maps from one simulation run. Each column represents a simulated IC map, which consists of 4 slices and each slice consists of $25 \times 25$ voxels. The highlighted region in the each figure corresponds to the location of a source signal representing a functional network in the brain.



# Chapter 4

# Predicting Brain Functional Connectivity in Resting-state fMRI Data Using a Bayesian Hierarchical Model

## 4.1 Introduction

Resting-state functional magnetic resonance imaging (rs-fMRI) has become a popular brain imaging technique to investigate intrinsic neural functional connection among spatially disjoint brain regions. One of the simple and most commonly used approaches for analyzing rs-fMRI in neuroimaging community is to perform correlation-type analysis between fMRI blood-oxygen-level dependent (BOLD) signals obtained from pairs of brain locations or regions (Biswal et al., 1995; Grecius et al., 2003; Fox et al., 2005, Fransson and Marrelec, 2008). These analyses produce subject-specific functional connectivity estimates that could provide useful information for exploring brain's functional organization and guiding brain parcellation for network analyses.

Functional connectivity has also shown great promises for studying pathophysiology of the development and progression of psychiatric diseases and their responses to treatments (Biswal et al., 1995; Greicius et al., 2003; Fox et al., 2005; Smith et al., 2009 among many others). Numerous studies identify distinct patterns of functional connectivity in brain networks among subjects with psychiatric disorders as compared to a healthy comparison population that support a neuropathophysiology of cognitive/behavioral problems associated with these disorders (Wang et al., 2007; Woodward and Cascio, 2015). Studies report mental disorders such as Alzheimer's disease affect brain functional connections across time with the disease progression (Damoiseaux et al., 2011; Yao et al. 2014). Furthermore, several studies have found significant pre- to post- treatment changes in brain connectivity following psychiatric treatments (Gay et al., 2014; Sarpal et al., 2015). These collective findings strongly implicate alterations in functional connectivity in the pathophysiology of psychiatric disorders and their responses to medical treatment. This provides motivation for the potential utility of forecasting disease-progression-related and treatment-related

brain functional connectivity reflected in fMRI.

There has been some work aimed to link baseline or pre-treatment brain connectivity to disease progression or the eventual clinical response to treatment (Damoiseaux et al., 2011; Yao et al. 2014). In a related manner, several authors (Brody et al., 1999; Goldapple et al., 2004; Kennedy et al., 2001; Lee et al., 2005; Mayberg et al., 2000, 2001; Moresco et al., 2000) establish associations between treatment response and pre- to post-treatment changes in brain activity. Important insights can be gained for disease progression or treatment response by evaluating baseline and follow-up scans or evaluating pre- and post-treatment scans. However, in clinical practice, these insights are offset due to the unavailability of follow-up scans or post-treatment scans at the time that a clinician makes treatment decision for a particular patient. This pragmatic shortcoming suggests the utility of developing a statistical framework to predict disease-related or treatment-related brain alterations, which could then be combined with baseline scans and patients' relevant risk factors to help inform clinical decision-making.

In this chapter, we present a general framework for predicting individual's future resting-state functional connectivity (RSFC) based on his/her baseline rs-fMRI and relevant clinical and demographic characteristics, such as disease stage or treatment group. The proposed prediction method provides a useful tool to predict the changes in an individual's brain connectivity with the progression of the disease or normal aging. It can also be applied to predict a patient's brain connectivity after a specified treatment regimen as the first step to potentially help guide individualized treatment plan. The proposed predictive model could potentially have important clinical applications by helping implement early intervention based on predicted disease progression trajectory to prevent severe outcomes from mental illnesses.

The method can also help guide the development of a more effective individualized treatment plan by taking into account both the population-level effects as well as a patient's distinctive neural connectivity characteristics. Another utility of the proposed method is to provide a more reliable estimator for individual RSFC than the noisy subject-specific estimator in test-retest rs-fMRI data. A reliable RSFC estimate can provide important information about the properties of brain networks and is the basis for many complex network-related analyses. For example, Shou et al. (2014) and Mejia et al. (2015) have shown that by improving the reliability of RSFC estimates, one can improve the reproducibility of individual brain parcellation, which is generated from the RSFC.

The remainder of this chapter is organized as follows. In the method section, we first present a general modeling and prediction framework for RSFC, and then illustrate the model specification for two types of imaging studies – longitudinal imaging studies which investigate disease progression and treatment-related changes in functional connectivity and test-retest studies which aim to investigate and improve the reliability of RSFC. To evaluate the accuracy of the proposed method, we conduct simulation studies using K-fold cross-validation approach. We illustrate the application of the proposed method using two real imaging datasets: a longitudinal ADNI2 data and a test-retest Kirby21 data. In both simulation studies and real data applications, we show that the proposed method provides a more reliable prediction over the alternative methods.

## 4.2 Method

We derive our predictive framework for brain functional connectivity map using a Bayesian hierarchical model (BHM), which is constructed using subjects' multiple

connectivity maps estimated at several time points over the study period. The first level of the proposed model models subject-specific connectivity measured at multiple scanning sessions associated with session-specific information, such as the timing of scans or different treatment periods in terms of subject-specific effects, and the second level models subject-specific effects in terms of population parameters. An expectation-maximization (EM) algorithm is developed to estimate the parameters in the model. The prediction algorithm is derived based on the estimated model. The proposed method is applicable to various study designs, such as test-retest studies and longitudinal studies. We start with presenting a general model structure and later we will provide detailed examples of using the proposed model for specific study designs.

### 4.2.1 Bayesian Hierarchical Model (BHM)

Let $i = 1, \cdots, N$ index subjects, $v = 1, \cdots, V$ index voxels, $k = 1, \cdots, K$ index scanning sessions. For any voxel pair $(v, v')$, we denote the observed connectivity estimated using subject $i$'s brain images from scanning sessions $1, \cdots, K$ as $\mathbf{r}_i(v, v') = (r_{i1}(v, v'), \cdots, r_{iK}(v, v'))^T$. The $K \times q$ matrix $\mathbf{X}_i^{(1)}(v, v')$ contains session-specific information such as the timing of scans, and usually takes the same value across all voxel pairs.

The first-stage of BHM models individual's connectivity for a voxel pair $(v, v')$ and is specified as follows

$$Level\ 1: \quad \mathbf{r}_i(v, v') = \mathbf{X}_i^{(1)}(v, v')\mathbf{R}_i(v, v') + \boldsymbol{\epsilon}_i^{(1)}(v, v'), \tag{4.1}$$

where $\mathbf{R}_i(v, v')$ is the $q \times 1$ vector of subject-specific effects of interest corresponding to $\mathbf{X}_i^{(1)}(v, v')$, such as the true unobserved connectivity. We assume that the random

error $\boldsymbol{\epsilon}_i^{(1)}(v,v') = (\epsilon_{i1}^{(1)}(v,v'), \cdots, \epsilon_{iK}^{(1)}(v,v'))^T$ follows a zero-mean multivariate normal distribution $N(0, \boldsymbol{\Sigma}^{(1)}(v,v'))$. We also assume that $\mathbf{R}_i(v,v')$ and $\boldsymbol{\epsilon}_i^{(1)}(v,v')$ are independent for all subjects. Since the connectivity maps for each subject $i$ are measured from independent scanning sessions and the subject-specific parameter $\mathbf{R}_i(v,v)$ already accounts for the correlations among within-subject images, it is reasonable to assume that scan-specific random errors in the first-stage model are independent. The random error $\boldsymbol{\epsilon}_i^{(1)}(v,v')$ in Equation (4.1) reflects the deviation of the estimated connectivity using individual's brain images from the true underlying connectivity, thus can also be considered as measurement errors, which mainly depend on scanner quality, the length of each scanning sessions, and other scanning conditions. If subjects' images are acquired under similar conditions (ex. following the same protocols), we can assume the variability of the random errors is the same across multiple sessions and specify the covariance matrix as $\boldsymbol{\Sigma}^{(1)}(v,v') = \lambda^{(1)}(v,v)\boldsymbol{I}_K$, where $\lambda^{(1)}$ is the covariance hyper-parameter and $\boldsymbol{I}_K$ is the identity matrix of order $K$. In a more complex setting when the scanning protocol is subject to change during a longitudinal study, we may accommodate these changes by specifying different variabilities in pre- and post-change periods, i.e. $\boldsymbol{\Sigma}^{(1)}(v,v') = \lambda_1^{(1)}(v,v)\begin{bmatrix} \boldsymbol{I}_{K_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \lambda_2^{(1)}(v,v)\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{K_2} \end{bmatrix}$, where $K_1(>1)$ and $K_2(>1)$ are the numbers of sessions before and after the change. We note that the specification of the covariance structures in the model only reflects the assumptions of sphericity/non-sphericity in repeated measures, the individual effects at different scanning times are modeled through individual parameters $R_i(v,v')$.

At the second stage, we focus on modeling the subject-specific effects with respect to population parameters of interest, such as disease status, treatment assignment or other relevant patient characteristics (e.g., gender, age). Therefore, the second-stage

model is specified as follows

$$Level\ 2 : \mathbf{R}_i(v, v') = \mathbf{X}_i^{(2)}(v, v')\boldsymbol{\beta}(v, v') + \boldsymbol{\epsilon}_i^{(2)}(v, v'), \tag{4.2}$$

where $\mathbf{X}_i^{(2)}(v, v')$ contains subject-specific covariates associated with brain functional connectivity and $\boldsymbol{\beta}(v, v')$ represents the population effects of interest corresponding to $\mathbf{X}_i^{(2)}(v, v')$. The random error $\boldsymbol{\epsilon}_i^{(2)}(v, v')$ represents the deviation of subject-specific effects from population mean $\mathbf{X}_i^{(2)}(v, v')\boldsymbol{\beta}(v, v')$ and is assumed to follow a zero-mean multivariate normal distribution with the covariance matrix $\boldsymbol{\Sigma}^{(2)}(v, v')$. The specification of $\boldsymbol{\Sigma}^{(2)}(v, v')$ depends on the individual effects that we are interested in modeling and we will illustrate it in the following examples.

### 4.2.2 Prediction Algorithm

Based on the proposed Bayesian hierarchical model, we develop prediction algorithm for predicting individual future brain connectivity. The proposed algorithm not only uses the same subject's brain connectivity information estimated from earlier scans, but also "borrows" strength from a larger population of subjects who share similar characteristics, such as subjects with the same disease status or in the same treatment group.

Let $K_1$ be the number of scanning sessions in which brain images have been collected from a subject whose brain connectivity is of interest for prediction and $K_2$ denote the number of sessions to be predicted. For this subject $i$, connectivity of a voxel pair $(v, v')$ across all scanning sessions can be expressed as a vector $\mathbf{r}_i(v, v') = (\mathbf{r}_{i1}^T(v, v'), \mathbf{r}_{i2}^T(v, v'))^T$, where $\mathbf{r}_{i1}(v, v')$ is a $K_1 \times 1$ vector representing the observed connectivity and $\mathbf{r}_{i2}(v, v')$ is a $K_2 \times 1$ vector representing the connectivity we are interested in predicting.

Based on the proposed two-stage hierarchical model, we derive the following marginal distribution for functional connectivity across all sessions,

$$\boldsymbol{r}_i(v,v') \sim N(\boldsymbol{\mu}_i^\star(v,v'), \boldsymbol{\Sigma}_i^\star(v,v')),$$

where $\boldsymbol{\mu}_i^\star(v,v') = \boldsymbol{X}_i^{(1)}\mathbf{X}_i^{(2)}\boldsymbol{\beta}(v,v')$ and $\boldsymbol{\Sigma}_i^\star(v,v') = \boldsymbol{\Sigma}^{(1)}(v,v') + \boldsymbol{X}_i^{(1)}\boldsymbol{\Sigma}^{(2)}(v,v')\boldsymbol{X}_i^{(1)T}$. The marginal covariance matrix $\boldsymbol{\Sigma}^\star(v,v')$ is the sum of the error covariance matrix $\boldsymbol{\Sigma}^{(1)}(v,v')$ from the first-stage model and the error covariance matrix $\boldsymbol{\Sigma}^{(2)}(v,v')$ from the second-stage model projected onto the measurement space by the first-stage design matrix $\boldsymbol{X}_i^{(1)}$. Therefore, $\boldsymbol{\Sigma}_i^\star(v,v')$ represents the total variability of $\boldsymbol{r}_i(v,v')$ from two levels of the hierarchical model.

Based on the joint marginal distribution of $\boldsymbol{r}_{i1}(v,v')$ and $\boldsymbol{r}_{i2}(v,v')$, we derive the predictive distribution of $\boldsymbol{r}_{i2}(v,v')$ given the earlier measurement $\boldsymbol{r}_{i1}(v,v')$,

$$[\boldsymbol{r}_{i2}(v,v')|\boldsymbol{r}_{i1}(v,v')] \sim N(\boldsymbol{\mu}_{2.1}(v,v'), \boldsymbol{\Sigma}_{2.1}(v,v')),$$

with the conditional mean

$$\boldsymbol{\mu}_{2.1}(v,v') = \boldsymbol{\mu}_{i2}^\star(v,v') + \boldsymbol{\Sigma}_{12}^\star(v,v')\boldsymbol{\Sigma}_{11}^{\star-1}(v,v')[\boldsymbol{r}_{i1}(v,v') - \boldsymbol{\mu}_{i1}^\star(v,v')] \tag{4.3}$$

and the conditional covariance matrix

$$\boldsymbol{\Sigma}_{2.1}(v,v') = \boldsymbol{\Sigma}_{22}^\star(v,v') - \boldsymbol{\Sigma}_{21}^\star(v,v')\boldsymbol{\Sigma}_{11}^{\star-1}(v,v')\boldsymbol{\Sigma}_{12}^\star(v,v'), \tag{4.4}$$

where $\boldsymbol{\mu}_{i1}^\star(v,v')$ and $\boldsymbol{\mu}_{i2}^\star(v,v')$ are the subvectors of $\boldsymbol{\mu}_i^\star(v,v')$ that correspond to the means of $\boldsymbol{r}_{i1}(v,v')$ and $\boldsymbol{r}_{i2}(v,v')$ and $\boldsymbol{\Sigma}_{11}^\star(v,v')$, $\boldsymbol{\Sigma}_{22}^\star(v,v')$, $\boldsymbol{\Sigma}_{12}^\star(v,v')$ and $\boldsymbol{\Sigma}_{21}^\star(v,v')$

are the submatrices of $\boldsymbol{\Sigma}^{\star}(v, v')$ that correspond to the variances and covariances of $\boldsymbol{r}_{i1}(v, v')$ and $\boldsymbol{r}_{i2}(v, v')$.

By plugging in the ML estimator of the variance parameters $\hat{\boldsymbol{\lambda}}(v, v')$ and the posterior mean of population mean parameters $\hat{\boldsymbol{\beta}}(v, v')$ calculated by the EM algorithm into Equations (4.3) and (4.4), we obtain the estimated conditional mean $\hat{\boldsymbol{\mu}}_{2.1}(v, v')$ and conditional variance $\hat{\boldsymbol{\Sigma}}_{2.1}(v, v')$. The connectivity $\boldsymbol{r}_{i2}(v, v')$ are predicted by the expectation of the estimated conditional distribution, i.e.,

$$\hat{r}_{i2}^{B}(v, v') = \hat{\boldsymbol{\mu}}_{2.1}(v, v') = \hat{\boldsymbol{\mu}}_{i2}^{\star}(v, v') + \hat{\boldsymbol{\Sigma}}_{12}^{\star}(v, v')\hat{\boldsymbol{\Sigma}}_{11}^{\star -1}(v, v')[\boldsymbol{r}_{i1}(v, v') - \hat{\boldsymbol{\mu}}_{i1}^{\star}(v, v')] \quad (4.5)$$

Additionally, we can construct the $100(1-\alpha)\%$ prediction interval for $\boldsymbol{r}_{i2}(v, v')$ based on the estimated conditional variance $\hat{\boldsymbol{\Sigma}}_{2.1}(v, v')$.

From Equation (4.5), one can see that the predicted value of subject $i$'s future connectivity consists of two components. The first component is the expected population mean of connectivity during $K_2$ sessions estimated from subjects who share similar characteristics as subject $i$. And the second component is added to for subject-specific adjustment based on the deviation of the same subject's previous connectivity estimate from the population mean of earlier connectivity. The amount of adjustment is jointly determined by the variability of earlier connectivity estimates and the association between earlier and later connectivity estimates. Specifically, when the within-subject variability is lower and the correlation between previous and future connectivity is higher, it is more reliable to use individual's previous images to adjust for the subject-specific prediction.

In the following, we will illustrate how to construct the hierarchical model for var-

ious study designs. Specifically, we consider two popular designs in imaging studies, which are longitudinal studies and test-retest studies.

### 4.2.3 Model Specification for Longitudinal Studies

In longitudinal imaging studies, subjects are scanned in multiple sessions over a considerate amount of time. Functional connectivity of certain brain regions may decrease over time as disease progresses or increase as a subject becomes more familiar with the experimental tasks. The goal of longitudinal imaging studies is to assess these changes in the neural networks as a result of disease progression or learning activities. Therefore in the hierarchical model, we model the subject-specific effects and population effects in terms of baseline connectivity and connectivity change rate. We denote the connectivity between voxel pair $(v, v')$ estimated for subject $i$ at scanning times $\mathbf{T}_K = (t_1, \cdots, t_K)^T$ as $\mathbf{r}_i(v, v') = (r_{i1}(v, v'), \cdots, r_{iK}(v, v'))^T$. Assuming connectivity changes at a constant rate over time and the baseline measurement is measured at $t_1 = 0$, we construct the hierarchical model as below,

$$
\begin{aligned}
Level\ 1: \quad & \begin{bmatrix} r_{i1}(v,v') \\ \vdots \\ r_{iK}(v,v') \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_K \end{bmatrix} \begin{bmatrix} R_{iB}(v,v') \\ R_{i\Delta}(v,v') \end{bmatrix} + \boldsymbol{\epsilon}_i^{(1)}(v,v'), \\
Level\ 2: \quad & \begin{bmatrix} R_{iB}(v,v') \\ R_{i\Delta}(v,v') \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{iB} & 0 \\ 0 & \mathbf{X}_{i\Delta} \end{bmatrix} \begin{bmatrix} \mathbf{R}_B(v,v') \\ \mathbf{R}_\Delta(v,v') \end{bmatrix} + \begin{bmatrix} \epsilon_{iB}^{(2)}(v,v') \\ \epsilon_{i\Delta}^{(2)}(v,v') \end{bmatrix},
\end{aligned}
$$

where $\mathbf{X}_{iB}$ and $\mathbf{X}_{i\Delta}$ contain subject-specific covariates that may influence the baseline and change rate of functional connectivity, respectively. In this hierarchical model for longitudinal studies, the first-stage models subject-scan-specific brain connectivity in terms of subject-specific baseline and change rate $(R_{iB}(v, v'), R_{i\Delta}(v, v'))^T$ and the second-stage models the subject-specific effects with respect to population

parameters. The random error $\boldsymbol{\epsilon}_i^{(1)}(v,v') = (\epsilon_{i1}^{(1)}(v,v'),\cdots,\epsilon_{iK}^{(1)}(v,v'))^T$ represents the deviation of the observed connectivity from the predicted connectivity based on the linear model with the subject-specific baseline and slope parameters. We assume $\{\epsilon_{ik}^{(1)}(v,v'), i = 1,\cdots,N; k = 1,\cdots,K\}$ to be independent across all subjects and all sessions, i.e., $\boldsymbol{\epsilon}_i^{(1)}(v,v') \sim N(0,\boldsymbol{\Sigma}^{(1)}(v,v'))$ with $\boldsymbol{\Sigma}^{(1)}(v,v') = \lambda^{(1)}(v,v')\mathbf{I}_K$. The random error $(\epsilon_{iB}^{(2)}(v,v'),\epsilon_{i\Delta}^{(2)}(v,v'))^T$ represents the deviation of the subject-specific baseline and slope from the population means of baseline connectivity and connectivity change and is assumed to follow a zero-mean multivariate normal distribution with the covariance matrix $\boldsymbol{\Sigma}^{(2)}(v,v') = (\lambda_B^{(2)}(v,v'),\lambda_\Delta^{(2)}(v,v'))\boldsymbol{I}_2$, where $\lambda_B^{(2)}(v,v')$ and $\lambda_\Delta^{(2)}(v,v')$ are the variabilities of subject-specific baseline and change rate, respectively.

### 4.2.4 Model Specification for Test-retest Studies

In test-retest studies, subjects are scanned for multiple sessions within a short study period. These types of studies often combine the information collected from multiple scanning sessions to improve the reliability of the connectivity estimates. Let $R_i(v,v')$ be the true underlying connectivity between a voxel pair $(v,v')$ for subject $i$ and let $r_{ik}(v,v')$ be the observed connectivity from session $k$. The classical measurement error model (Carroll et al., 2006) is specified as,

$$r_{ik}(v,v') = R_i(v,v') + \epsilon_{ik}(v,v'),$$

where $\epsilon_{ik}(v,v')$ is the measurement error for subject $i$ during session $k$ at voxel pair $(v,v')$, which can be affected by the length of time during a scanning session or the quality of scanners. As mentioned earlier, the main goal of conducting test-retest studies is to obtain a more stable estimate of functional connectivity by measuring the same underlying metric in multiple sessions. Specifically, by taking the aver-

age of measurements collected from the same subject during multiple sessions, we can reduce the variability of the connectivity estimate as compared to using a single measurement. In the case when we don't have multiple scans for a subject, the proposed prediction method helps to improve the subject-specific connectivity estimate by "borrowing" strength from a larger group of subjects over the raw estimate using only the individual's images. The hierarchical model is constructed as follows with the first-level model representing the measurement error model:

$$\begin{aligned} Level\ 1: &\quad \mathbf{r}_i(v,v') = \mathbf{1}_K R_i(v,v') + \boldsymbol{\epsilon}_i^{(1)}(v,v'), \\ Level\ 2: &\quad R_i(v,v') = \mathbf{X}_i\boldsymbol{\beta}(v,v') + \epsilon_i^{(2)}(v,v'), \end{aligned}$$

where the subject-specific covariate matrix $\boldsymbol{X}_i$ contains relevant patient characteristics such as medical or family history that may influence the functional connectivity in the brain and $\mathbf{1}_K$ is an all-ones vector of length $K$. We assume that the measurement error $\boldsymbol{\epsilon}_i^{(1)}(v,v')$ is independently and identically distributed with a zero-mean multivariate Gaussian distribution with covariance matrix $\boldsymbol{\Sigma}^{(1)}(v,v') = \lambda^{(1)}(v,v')\boldsymbol{I}_K$. The second-level model assumes that $R_i(v,v')$ are independently drawn from a population distribution with a mean of $\mathbf{X}_i\boldsymbol{\beta}(v,v')$, where $\mathbf{X}_i$ contains subject-specific information, such as age, gender or disease status, which may affect functional connectivity between the voxel pairs. The random error $\epsilon_i^{(2)}(v,v')$ represents between-subject variability and we assume $\epsilon_i^{(2)}(v,v')$ follows zero-mean normal distribution with a variance of $\lambda^{(2)}(v,v')$.

### 4.2.5 Connections and Differences with a Shrinkage Estimator in Test-retest Studies

In this section, we will show the connections and differences of our proposed BHM predictor with a recently proposed shrinkage estimator (Shou et al., 2014; Mejia et al., 2015) for predicting future brain connectivity in test-retest imaging studies.

Consider a simple test-retest study with two scanning sessions ($K = 2$), in which researchers are interested in predicting future connectivity matrices for a group of subjects based on their first brain scans as well as information collected from the other subjects in the study. We define $\gamma^B(v, v') = \frac{\lambda^{(2)}(v,v')}{\lambda^{(1)}(v,v')+\lambda^{(2)}(v,v')}$, which is often known as the intraclass correlation coefficient (ICC) and represents the relationship of between-subject variance $\lambda^{(2)}(v, v')$ and within-subject variance $\lambda^{(1)}(v, v')$. The BHM predictor of second-session connectivity based on the hierarchical model proposed for test-retest studies in Equation (4.5) can be written as,

$$\hat{r}^B_{i2}(v, v') = \hat{\gamma}^B(v, v')r_{i1}(v, v') + \left(1 - \hat{\gamma}^B(v, v')\right)\boldsymbol{X}_i\hat{\boldsymbol{\beta}}(v, v'). \tag{4.6}$$

Equation 4.6 shows that the BHM predictor is basically a weighted average of the same subject's first-session connectivity estimate and the population mean associated with covariates $X_i$. The weight $\gamma^B(v, v')$ ranges between 0 and 1 and represents the relative weight assigned to subject-level estimate compared to population mean. In many literatures, this kind of estimator is known as the "shrinkage estimator", and the weight $\gamma^B(v, v')$ is referred to as the shrinkage factor. As the within-subject variance $\boldsymbol{\lambda}^{(1)}(v, v')$ to between-subject variance $\boldsymbol{\lambda}^{(2)}(v, v')$ decreases, ICC increases, which means subject-level information is more reliable, more weight is assigned to the subject-level estimate $r_{i1}(v, v')$. On the contrary, as the within- to between-subject variance ratio increases, ICC decreases, which means subject-level information is less

reliable and we should benefit more by "borrowing" information from the population. The weight $\gamma^B(v, v')$ in BHM predictor is estimated by plugging in the model-based variance estimates $\hat{\boldsymbol{\lambda}}^B(v, v')$, i.e., $\hat{\gamma}^B(v, v') = \frac{\hat{\lambda}^{(2)B}(v,v')}{\hat{\lambda}^{(1)B}(v,v')+\hat{\lambda}^{(2)B}(v,v')}$.

Shou et al. (2014) and Mejia et al. (2015) recently proposed a shrinkage estimator for predicting future functional connectivity in test-retest studies, i.e.,

$$\hat{r}^S_{i2}(v, v) = \hat{\gamma}^S(v, v')r_{i1}(v, v) + (1 - \hat{\gamma}^S(v, v'))\bar{r}_1(v, v'), \tag{4.7}$$

where $\bar{r}_1(v, v')$ is the sample mean of first-session estimate of connectivity for voxel pair $(v, v')$ and $\gamma^S(v, v')$ is a shrinkage factor for voxel pair $(v, v')$ that takes any value in $[0, 1]$. Shou (2014) and Mejia (2015) proposed several methods for estimating $\gamma^S(v, v')$ based on the moment variance estimators $\hat{\boldsymbol{\lambda}}^S(v, v')$, i.e., $\hat{\gamma}^S(v, v') = \frac{\hat{\lambda}^{(2)S}(v,v')}{\hat{\lambda}^{(1)S}(v,v')+\hat{\lambda}^{(2)S}(v,v')}$. Numerical studies conducted in their papers have showed different performances across various shrinkage factor estimators in terms of the degree of shrinkage as well as prediction mean square error between the observed and predicted connectivities.

It is obvious that under this simple test-retest setting, the BHM predictor $\hat{r}^B_{i2}(v, v')$ is closely related to the shrinkage estimator $\hat{r}^S_{i2}(v, v')$, as both are defined as a weighted sum of subject-level estimate and population mean. The main differences lie in the calculation of shrinkage factor as well as the population mean. Shou's shrinkage factor is estimated empirically and separately from the population mean, while for BHM predictor, the shrinkage factor is estimated simultaneously with the sample mean in the EM iterations. For the estimation of population mean, Shou's method only uses the first-session data while in our proposed method, population mean is estimated using all available data from both sessions in the training dataset. Moreover, the

proposed BHM method utilizes covariate information to improve the population mean estimation, which is especially beneficial when there is large between-group difference and small between-subject variability within groups as will be shown in the following simulation studies.

## 4.3 Simulation Studies

We conducted simulation studies to evaluate the prediction performance of the proposed method under two settings. In the longitudinal setup, we compared the proposed predictor with a raw estimator and a General Linear Model(GLM) predictor. GLM predictor is derived based on the marginal model of the hierarchical structure with fixed covariate effects. In the test-retest setting, a mean estimator and a shrinkage estimator were also included for method comparison. Since all the methods are performed at voxel-pair level, our simulation studies were carried out for single voxel pairs, the results can be generalized to most connectivity metrics, such as voxel-by-voxel, node-by-node or seed-based correlation measure. Simulations were performed for 100 times with a sample size of 100 for each scenario. To compare the prediction accuracy of different methods, k-fold cross-validation was employed. Specifically, we split the data into K(= 5) subsets. In each iteration, one subset was used as the testing dataset, while the other subsets were combined as the training dataset. For BHM and GLM, we first fit the model using the training dataset, and used the estimated model to perform predictions for subjects in the testing dataset. For the other estimators, we adopted similar strategies in Shou et al. (2014) and Mejia et al. (2015) by using the first measurement of the same subject's data as the raw estimator, using the average first measurement in the training dataset as the mean estimator, and using the first two measurements of the training dataset to calculate the shrinkage factor and performing shrinkage with the raw and mean

estimators. Same procedure was repeated so that all the data were used as the testing dataset for once. We measured the prediction error using average prediction mean square error (PMSE) between the true underlying connectivity and predicted connectivity.

In the first setup, we simulated longitudinal data with three consecutive connectivity measurements for each subject. We assumed that, for subject $i$, the baseline connectivity was $R_{iB}$ and that the functional connectivity changed over time with a constant rate of $R_{i\Delta}$. We considered two covariate groups with the same baseline connectivity but different change rates. Subjects were randomly sampled from either group with equal probability. Given these assumptions, the fisher-transformed connectivity parameters in terms of baseline connectivity and change rate for subject $i$ were generated as follows,

$$\begin{bmatrix} z(R_{iB}) \\ z(R_{i\Delta}) \end{bmatrix} = \begin{bmatrix} 1 & I_1(i) & 0 \\ 1 & 0 & 1 - I_1(i) \end{bmatrix} \begin{bmatrix} z(\beta_B) \\ z(\beta_{\Delta 1}) \\ z(\beta_{\Delta 2}) \end{bmatrix} + \begin{bmatrix} \epsilon_{iB}^{(2)} \\ \epsilon_{i\Delta}^{(2)} \end{bmatrix}, \begin{bmatrix} \epsilon_{iB}^{(2)} \\ \epsilon_{i\Delta}^{(2)} \end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \lambda_B^{(2)} & 0 \\ 0 & \lambda_\Delta^{(2)} \end{bmatrix} \right),$$

where $I_1(i)$ was the indicator function of group assignment, which equals 1 if subject $i$ belongs to group 1 and equals 0 if the subject belongs to group 2, $\beta_B$ was the population baseline of brain connectivity, and $\beta_{\Delta 1}$ and $\beta_{\Delta 2}$ were the mean change rates for group 1 and group 2, and $z(\cdot)$ represents Fisher transformation. We assumed the between-subject variances of baseline connectivity $\lambda_B^{(2)}$ and connectivity change rate $\lambda_\Delta^{(2)}$ to be fixed at 0.03 and 0.01, respectively. The fisher-transformed connectivity $z(r_{ik})$ estimated at three time points $t_1, t_2, t_3$ was simulated by adding the within-

subject variance $\boldsymbol{\epsilon}_i^{(1)} = (\epsilon_{i1}^{(1)}, \epsilon_{i2}^{(1)}, \epsilon_{i3}^{(1)})^T$, i.e.,

$$\begin{bmatrix} z(r_{i1}) \\ z(r_{i2}) \\ z(r_{i3}) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & t_2 \\ 1 & t_3 \end{bmatrix} \begin{bmatrix} z(R_{iB}) \\ z(R_{i\Delta}) \end{bmatrix} + \begin{bmatrix} \epsilon_{i1}^{(1)} \\ \epsilon_{i2}^{(1)} \\ \epsilon_{i3}^{(1)} \end{bmatrix}, \quad \begin{bmatrix} \epsilon_{i1}^{(1)} \\ \epsilon_{i2}^{(1)} \\ \epsilon_{i3}^{(1)} \end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \lambda^{(1)} & 0 & 0 \\ 0 & \lambda^{(1)} & 0 \\ 0 & 0 & \lambda^{(1)} \end{bmatrix} \right),$$

where the within-subject variance $\lambda^{(1)}$ was chosen so that the variance ratio of between-subject variance and total variance at baseline, i.e. $\lambda_B^{(2)}/(\lambda^{(1)} + \lambda_B^{(2)})$, varied from 0.1 to 0.9. The simulated connectivity data was then back-transformed to the original scale and ranged between $[-1, 1]$. Method evaluation was performed under the original scale.

Figure 4.1 presents the average PMSEs of the three connectivity predictors against the baseline variance ratio. From left to right, $(\beta_{\Delta 1}, \beta_{\Delta 2})$ were specified as $(0.2, -0.2), (0.1, -0.1), (0, 0)$, sequentially, representing large, small to no group difference between the connectivity change rates. For all three scenarios, the baseline connectivity $\beta_B$ was 0.3 for both groups. Simulation results showed that the average PMSE of the BHM predictor was always the smallest as compared to GLM predictor and raw estimator. The mean PMSE of BHM and raw predictors decreased as the baseline variance ratio increased. The average PMSE of GLM predictor was very robust to the change in baseline variance ratio when fixing the other parameters. When large between-group difference was presented in the data, both BHM and GLM predictors had smaller PMSE compared to the raw estimator, since both BHM and GLM toke into account of this group information for prediction purpose. In addition, we presented the true connectivity distributions measured in three consecutive sessions with the predicted data distributions for session 3 connectivity by different methods in Figure 4.2 under the large between-group difference scenario. We specified a moderate within-subject variance of 0.03, which corresponded to a

baseline variance ratio of 0.5. Figure 4.2 shows that the box plot of the predicted data distribution by BHM predictor best resembled the true data distribution from the 3rd session. The GLM predictor also differentiated the population means of the two groups. However, since GLM predictor didn't account for individual variability in the prediction, the predicted distribution was homogenous across subjects from the same covariate group. The predicted distribution by the raw estimator didn't demonstrate any heterogeneity between the two groups, given the raw estimator was calculated based on the homogeneous baseline data and didn't account for longitudinal changes in the connectivity.

In the second setup, we simulated data with two scan-rescan connectivity measurements for each subject. Assume $r_{ik}$ was the simulated connectivity for subject $i(i = 1, \cdots, N)$ at session $k(k = 1, 2)$ with an underlying connectivity denoted as $R_i$, which was subject-specific and had the same value across two imaging sessions. We assumed that there were two covariate groups with different mean connectivities. Group assignment was simulated from a Bernoulli distribution with a probability of 0.5. Random variability among subjects $i = 1, \cdots, N$ was assumed to be the same across groups and was introduced by adding a Gaussian noise $\epsilon_i^{(2)}$ to the Fisher-transformed group means:

$$z(R_i) = \begin{bmatrix} I(i) & 1 - I(i) \end{bmatrix} \begin{bmatrix} z(\beta_1) \\ z(\beta_2) \end{bmatrix} + \epsilon_i^{(2)}, \;\; \epsilon_i^{(2)} \sim N(0, \lambda^{(2)}),$$

where $\beta_1$ and $\beta_2$ were the population means for group 1 and group 2, respectively. The session-specific connectivity $r_{ik}$ was simulated by including the within-subject variance $\epsilon_i^{(1)}$ as follows,

$$z(r_{ik}) = z(R_i) + \epsilon_i^{(1)}, \;\; \epsilon_i^{(1)} \sim N(0, \lambda^{(1)}).$$

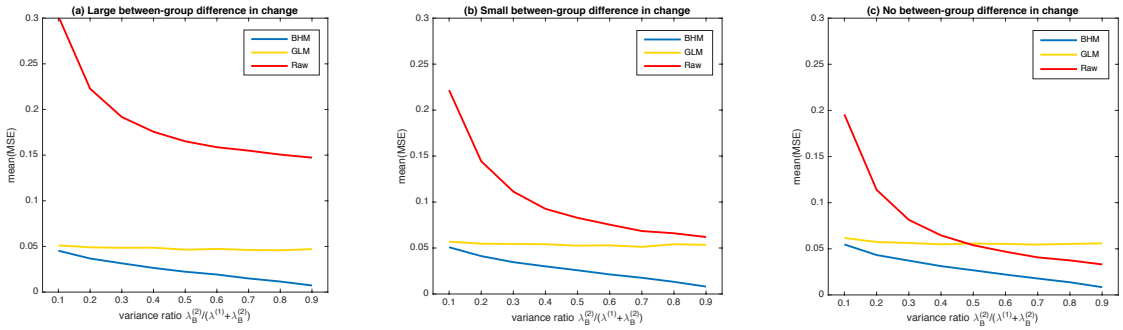



Figure 4.1: Effects of variance ratio on the average prediction mean square error (PMSE) for predicting future brain connectivity in longitudinal studies by BHM, GLM and raw estimator. We consider three scenarios: (a) large between-group difference in connectivity change rates: $(-0.2, 0.2)$; (b) small difference: $(-0.1, 0.1)$; (c) no difference: $(0.0, 0.0)$.

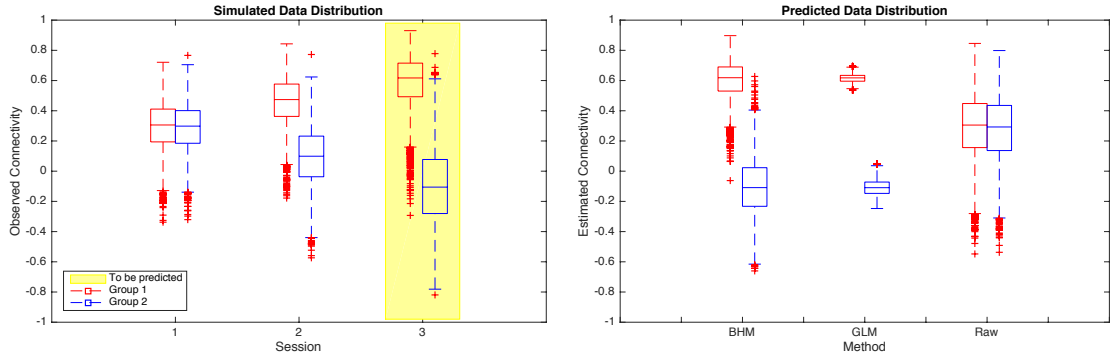



Figure 4.2: Consistency between the predicted connectivity distributions with the simulated data distributions in longitudinal studies with large between-group difference for connectivity change rate.

We specified the between-subject variance $\lambda^{(2)}$ as 0.03, and let the within-subject variance $\lambda^{(1)}$ vary so that the variance ratio of between-subject variance and total variance changed from 0.1 to 0.9.

In Figure 4.3, we plot the average PMSEs between the true connectivity $Ri$ and the predicted connectivity calculated by five different methods against the variance ratio. The results were presented for three scenarios with decreasing group differences from left to right with a constant group mean $\beta_2 = 0.2$ and a varying group mean $\beta_1 = 0.6, 0.4, 0.2$. As shown in Figure 4.3, the average PMSE of the proposed BHM predictor was always the smallest as compared to the other methods. As the variance ratio increased, the average PMSEs of BHM, shrinkage and raw predictors decreased, and difference among these three methods also decreased. On the other hand, the average PMSEs of GLM and mean predictors were relatively stable and did not change much as the variance ratio changed. When there was larger difference between two group means, the benefits of BHM and GLM predictors were more obvious, since these two methods took into accounts of the group effects. When there was no between-group difference, the results between BHM and shrinkage method are very close, hence it is hard to distinguish the two curves in the figure. Similarly, the results between GLM and mean method are very close, hence the corresponding curves appear to overlay. Figure 4.4 presents the simulated distributions of connectivity at two time points by group when there is large group difference and $\lambda^{(2)}/(\lambda^{(1)} + \lambda^{(2)}) = 0.5$. In the paralleled figure, we show the distributions of the predicted values for connectivity measured at session 2 using different predictors. The distribution of predicted values from BHM method best resembled that of the simulated data in terms of group mean and variance. GLM predictor successfully differentiated the population means for different groups. However, since GLM predictor didn't account for individual variability, the predicted distribution was

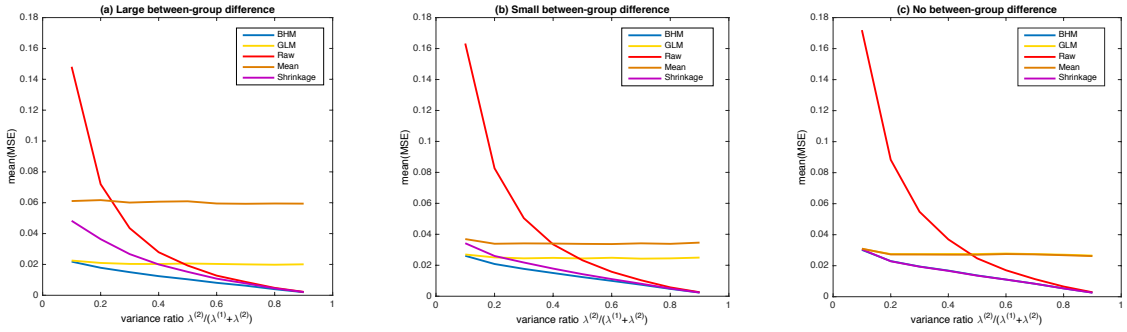



Figure 4.3: Effects of variance ratio on the average prediction mean square error (PMSE) for predicting brain connectivity in test-retest studies by BHM, GLM, raw, mean and shrinkage estimator. We consider three scenarios: (a) large between-group difference in mean connectivity: $(0.2, 0.6)$; (b) small difference: $(0.2, 0.4)$; (c) no difference: $(0.2, 0.2)$.

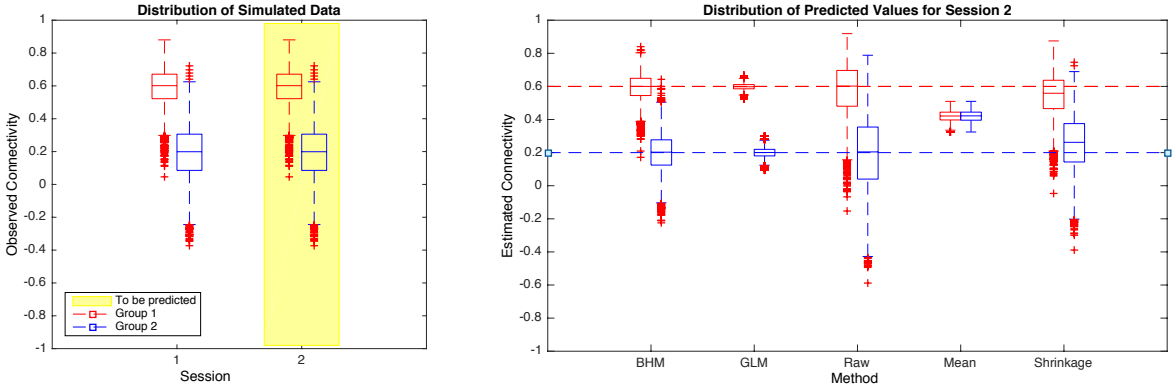



Figure 4.4: Consistency between the predicted connectivity distributions with the simulated data distributions in test-retest studies with large between-group difference. The horizontal reference lines correspond to the population means of connectivity for each group.

much more concentrated than the true distribution. Raw estimator also captured the difference in group means but was more variable than the true distribution because of the within-subject variability. Pooled mean predictor performed the worst by ignoring the large between-group difference. The group means of shrinkage estimator didn't align well with the true distribution since the shrinkage method shrank the mean towards the mean estimator. The variability of shrinkage estimator fell between that of raw and mean estimators.

## 4.4 Data Applications

### 4.4.1 Application to a Longitudinal ADNI2 Study

We applied the Bayesian hierarchical prediction approach to the Alzheimer's Disease Neuroimaging Initiative 2 (ADNI2) data. The purpose of ADNI2 project is to examine how the brain imaging and other biomarkers can be used to measure the progression of mild cognitive impairment(MCI) and Alzheimer's Disease(AD). Image acquisition parameters are described in details on `http://www.adni-info.org`. In short, T1-weighted high-resolution anatomical image (MPRAGE) and a series of resting state functional images were acquired with a 3.0 Tesla Siemens scanner during longitudinal visits.

In this analysis, eighty subjects with complete T1 and rs-fMRI images at screening, 6 months and 1 year published on ADNI website (`http://www.adni.loni.usc.edu`) before April 30, 2015 were included and have quality assured based on Mayo clinic quality control documentation (version 02-02-2015) and visual examination. Among these 80 subjects, 23 were normal, 47 had MCI, and 10 were diagnosed with AD at baseline. No significant age or gender differences were found across disease status groups (Age: one-way ANOVA, p = 0.2985; Gender: chi-square test, p=0.3134).

For the preprocessing procedure, the anatomical image was first registered to the mean of the corrected functional images and then spatially warped to the MNI standard brain space by using the segmentation routine of SPM5 (`http://fil.ion.ucl.ac.uk/spm/software/spm5`). The estimated warping parameters were subsequently applied to the slice and motion corrected functional images, which were finally smoothed with an 8 mm isotropic Gaussian kernel. Time series were detrended by regressing out a Legendre polynomial of order two,

demeaned and whitened.

We adopted the 264-node cortical parcellation system defined by Power et al. (2011), where each node is a 10 mm diameter sphere in MNI space. Compared to the full voxel-level analysis or 90 AAL regions segmentation, such parcellation has been shown to achieve a balance between spatial localization and dimension reduction (Power et al., 2011). Our connectivity analysis considered 258 of the 264 nodes that were within the boundary of data gray matter mask. To examine the prediction performance for the within- and between-network connectivity, we performed functional parcellation using these 258 nodes and classified them into nine resting state networks (RSNs) described by Smith et al. (2009). Network membership of each node was determined based on the largest value among RSN z-statistic maps at the core of that node. We excluded 44 nodes from the final analysis considering the largest values at these nodes were below a chosen threshold ($z > 3$) and excluded the cerebellum network which contained only 6 nodes.

To construct the connectivity maps, we performed singular value decomposition (SVD) on the times series for all the voxels within each node to extract the representative time series of that node. A $216 \times 216$ symmetric connectivity matrix was constructed for each subject by calculating Pearson correlation between the summary time series extracted from each node.

We built the hierarchical model by assuming a constant change rate of connectivity over time and potential group effects in both baseline connectivity and connectivity

change rate. The hierarchical model was specified as follows:

$$Level\ 1: \quad \begin{bmatrix} r_{i0}(v,v') \\ r_{i1}(v,v') \\ r_{i2}(v,v') \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & t_{i1} \\ 1 & t_{i2} \end{bmatrix} \begin{bmatrix} R_{iB}(v,v') \\ R_{i\Delta}(v,v') \end{bmatrix} + \begin{bmatrix} \epsilon_{i0}^{(1)}(v,v') \\ \epsilon_{i1}^{(1)}(v,v') \\ \epsilon_{i2}^{(1)}(v,v') \end{bmatrix},$$

$$Level\ 2: \quad \begin{bmatrix} R_{iB}(v,v') \\ R_{i\Delta}(v,v') \end{bmatrix} = \begin{bmatrix} I_{iN} & I_{iM} & I_{iA} & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{iN} & I_{iM} & I_{iA} \end{bmatrix} \begin{bmatrix} R_{B_N}(v,v') \\ R_{B_M}(v,v') \\ R_{B_A}(v,v') \\ R_{\Delta_N}(v,v') \\ R_{\Delta_M}(v,v') \\ R_{\Delta_A}(v,v') \end{bmatrix} + \begin{bmatrix} \epsilon_{iB}^{(2)}(v,v') \\ \epsilon_{i\Delta}^{(2)}(v,v') \end{bmatrix},$$

where $\epsilon_{ik}^{(1)}(v,v')$ followed a zero-mean normal distribution with the variance of $\lambda^{(1)}(v,v')$ for all $k$ and $(\epsilon_{iB}^{(2)}(v,v'), \epsilon_{i\Delta}^{(2)}(v,v'))^T$ followed a zero-mean bivariate normal distribution with the covariance matrix $\mathbf{\Sigma}^{(2)}(v,v') = (\lambda_B^{(2)}(v,v'), \lambda_\Delta^{(2)}(v,v'))\boldsymbol{I}_2$. The interval between measurements were the same for all subjects in this study. All subjects were scanned at baseline, 6 months and 1 year, i.e., $t_{i0} = 0, t_{i1} = 0.5$, $t_{i2} = 1$.

The BHM-based average baseline connectivity estimates across all 23220 unique node-pairs were 0.0290, 0.0259, 0.0266 for normal, MCI, and AD patients, respectively. And the average model-based connectivity change rates over one year were 0.0008, -0.0002, -0.0043, indicating a decreasing trend of functional connectivity in the brain for more advanced disease stages. In addition, BHM provided variance estimates for both within-subject variability and between-subject variability. The average within-subject variance across all voxel pair $\hat{\lambda}^{(1)}$ was 0.0373. The average between-subject variance for baseline connectivity $\hat{\lambda}_B^{(2)}$ and connectivity change rate $\hat{\lambda}_\Delta^{(2)}$ across all voxel pairs were 0.0136 and 0.0043, respectively. The results showed that within-subject variability was higher than between-subject variability.

Figure 4.5 illustrates the population effects on baseline connectivity estimated by BHM. Specifically, subfigures (a)-(c) present the estimated baseline connectivity maps in three status groups (NC: normal control; MCI: mild cognitive impaired; AD: Alzheimer's disease). Each point on the map corresponds to an edge connecting two nodes with red color indicating positive connectivity and blue color indicating negative connectivity. These connectivity maps demonstrate some similarities across three status groups. For example, we observed that most edges that connect nodes from the same functional network were positive. Edges connecting nodes among the three visual networks, between auditory and sensorimotor networks, and between default mode network and some part of executive control network were also positive. Most of the other inter-network connectivities were negative or showed weak correlation. To compare the edgewise baseline connectivity between two disease groups (MCI and AD) and the control group, we present the standardized mean differences (as measured by Cohen's d) in subfigures (d)-(g), and threshold their absolute values by 0.5 to identify group differences with at least medium effect size. The results are presented separately for positive and negative connectivities with brown color indicating stronger connectivity and green color indicating weaker connectivity in the disease group as compared with the control group. The figures show that connectivity within auditory network, between auditory and sensorimotor networks, between default mode and two frontoparietal networks, between medial visual and occipital pole visual networks were weaker in both disease groups, while connectivity within sensorimotor, executive control network and all three visual networks were stronger. In addition, the figures also show amplified connectivity between lateral visual network with auditory and executive control networks.

Figure 4.6 demonstrates the BHM model-based group effects on connectivity

change rate from baseline to one-year in the study. Subfigures (a)-(c) present the thresholded posterior probability maps for an absolute change rate greater than 0.2 by disease status. The results show that very few edges had large changes in NC and MCI subjects, while in AD group, some large alternations were observed for intra- and inter- lateral visual, auditory, executive control connectivity. To further illustrate the changing edges, we selected 50 edges with the largest posterior probabilities in AD group, and present their connectivity at baseline and one year in subfigures (d) and (e). Red lines connecting two nodes represent positive connectivity between the nodes, and blue lines represent negative connectivity. The widths of the lines reflect the magnitude in connectivity as estimated by BHM. For example, connectivity edges between some nodes in sensorimotor network and auditory network, default mode and auditory network decreased over the year, while connectivity between left and right frontoparietal nodes increased.

Figure 4.7 presents the prediction accuracy for BHM, GLM and raw estimator with prediction mean square error (PMSE) between the observed and predicted connectivity using 5-fold cross validation. Based on subfigures 4.7(a)-4.7(c), BHM performed the best across all voxel pairs with an average PMSE of $0.045(\pm 0.011)$. GLM predictor, which used group mean for prediction, also outperformed the raw estimator. Edges within and among executive control, frontoparietal left and right, and visual networks were the least predictable as compared to other edges. Individual connectivities within default mode, sensorimotor and auditory networks were the most predictable. Additionally, we compared the individual prediction accuracy by plotting the average PMSE across all node-pairs for each method in subfigures 4.7(d)-4.7(f). BHM predictor outperformed GLM predictor for most subjects (71/80) and outperformed raw predictor for all subjects.

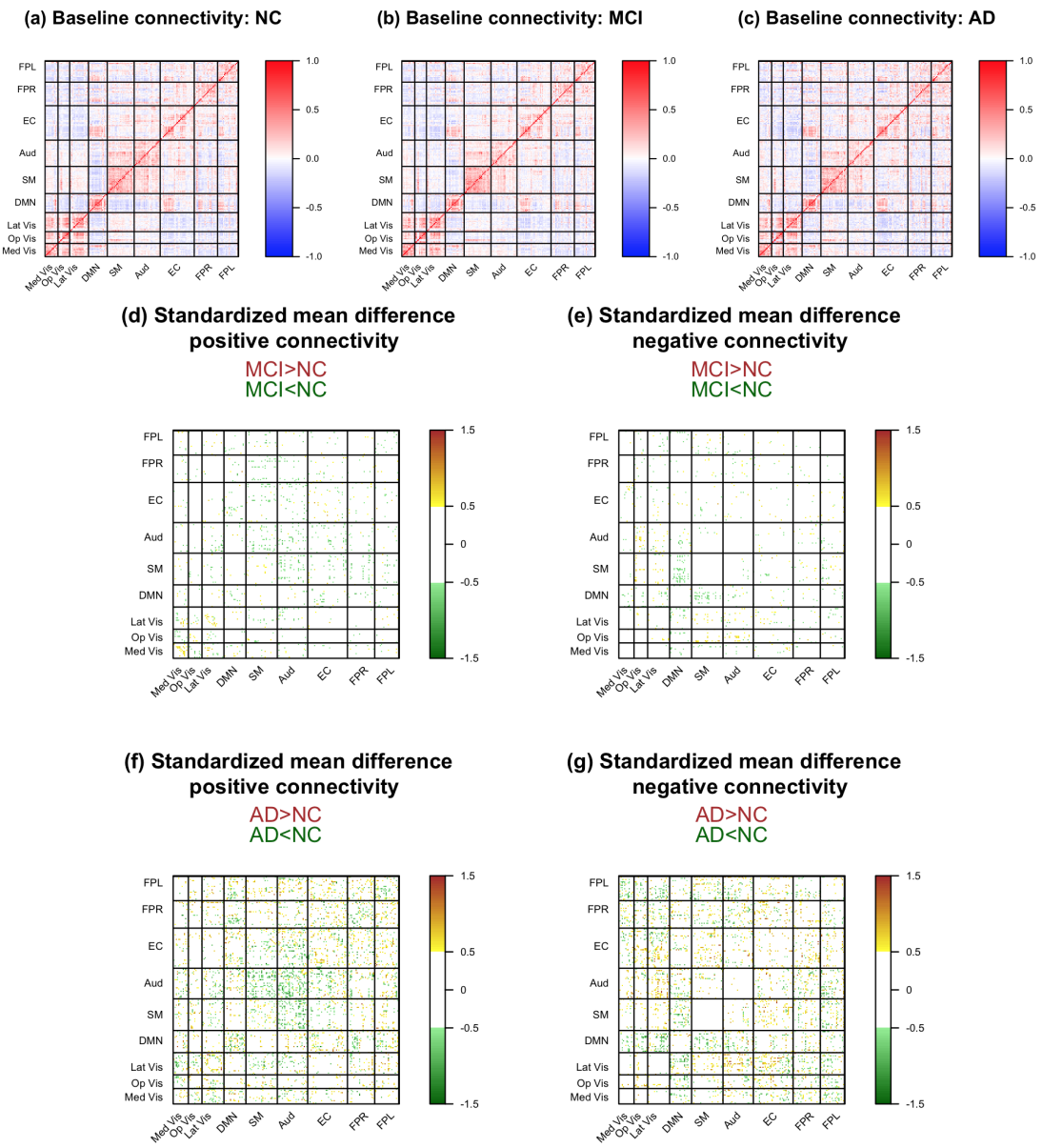



Figure 4.5: Population effects and group differences of baseline edgewise connectivity matrices estimated based on the bayesian hierarchical model (BHM). Subfigures (a)-(c) present the symmetric 216 $\times$ 216 baseline connectivity matrices (as measured by Pearson correlations and correspond to population parameters $[R_{BN}(v,v), R_{BM}(v,v), R_{BA}(v,v)]'$ in the BHM model) for normal controls, mild cognitive impaired and Alzheimer patients. Edges are grouped by their module membership, with red edges indicate positive connectivity and blue edges indicate negative connectivity. In subfigures (d)-(g), we present the standardized mean difference matrices (as measured by Cohen's d and thresholded by 0.5) for baseline connectivity between NC and two patient groups. The sign of each edge is determined by the sign of group average, and positive and negative edges are presented separately.

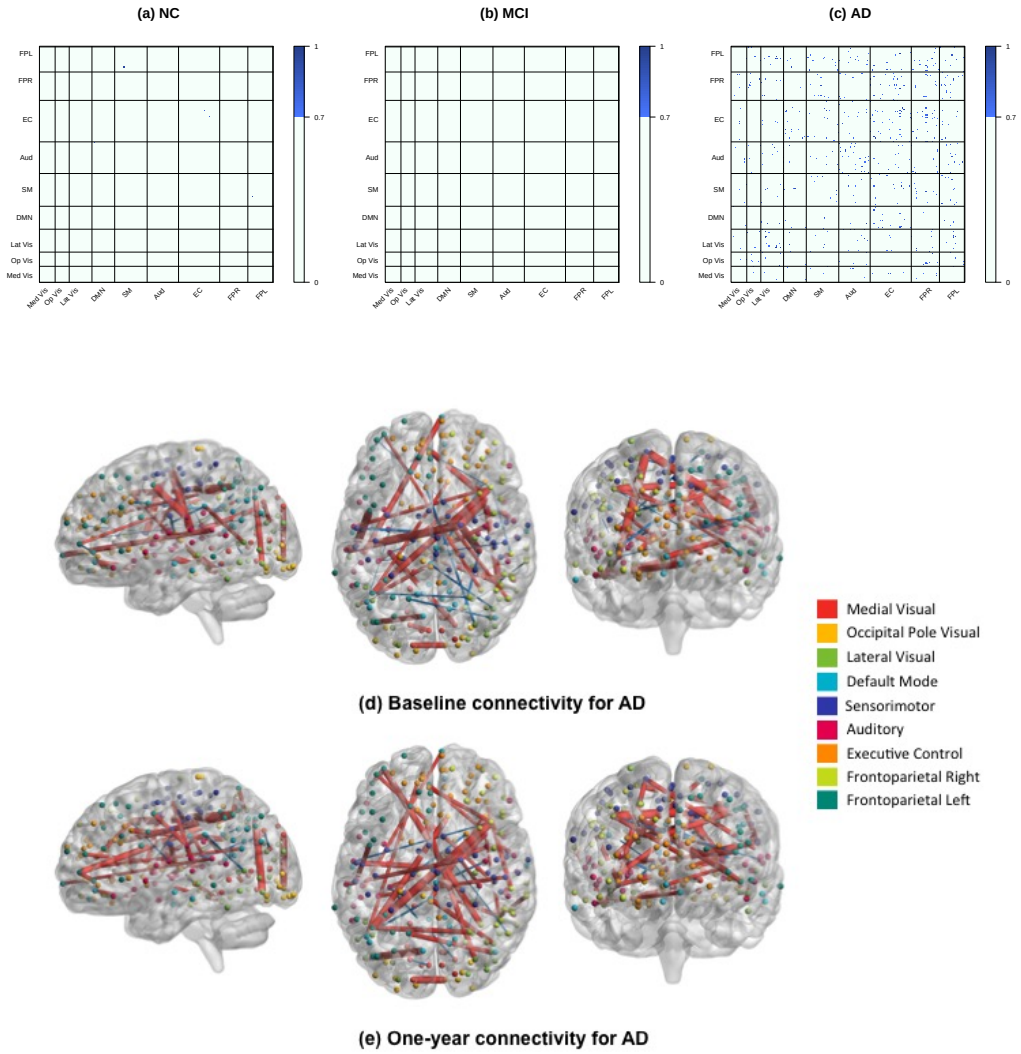



Figure 4.6: Population effects of connectivity change rate by disease status based on the bayesian hierarchical model (BHM). Subfigures (a)-(c) present the symmetric 216 $\times$ 216 thresholded posterior probability maps for absolute connectivity change over 1 year by disease status based on BHM for the ADNI2 data. Each point on the maps represents an edge between two nodes in the brain and edges are grouped by their module membership. Dark blue indicates a large posterior probability (probability greater than 0.7) of absolute change rate in connectivity per year greater than 0.2, i.e., $Pr[R_{\Delta J} > 0.2 \quad or \quad R_{\Delta J} < -0.2] > 0.7, J = N, M, A$. Several edges among nodes from EC, FPR, FPL, Aud and Lat Vis networks are identified to have higher posterior probabilities of a large change rate in AD group, while very few edges in NC and MCI are likely to change. In subfigures (d) and (e), we present the baseline and one-year connectivities for 50 edges with the largest posterior probabilities of a greater-than-0.2 absolute connectivity change rate in AD group. Red edges indicate positive connectivity and blue edges indicate negative connectivity. The size of the edges reflects the strength of connectivity.

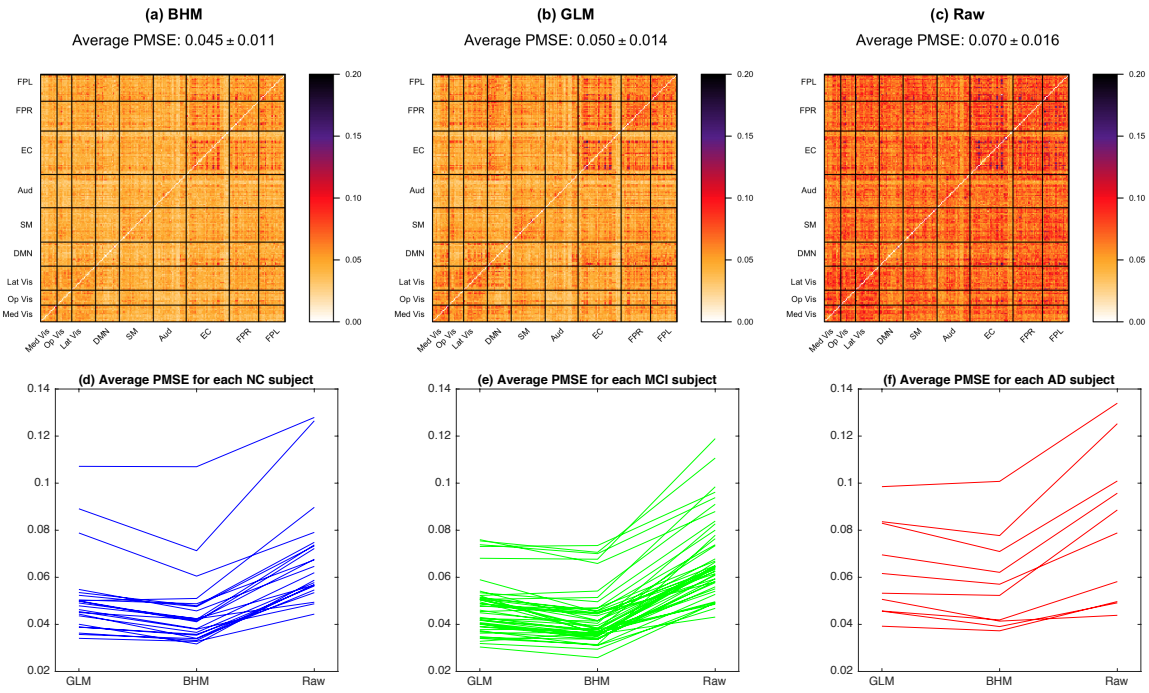



Figure 4.7: Comparison of prediction mean square error (PMSE) for predicting first-year functional connectivity in ADNI2 study using different prediction methods. Subfigures (a)-(c) present the average PMSE matrices across all subjects using the bayesian hierarchical model, general linear model, and raw estimator, respectively. Darker color represents larger PMSE, and lighter color represents smaller PMSE. In subfigures (d)-(f), we compare the performances of different prediction methods for each individual using average PMSEs across all node-pairs. Individual results are presented separately for normal controls, subjects with mild cognitive impairment and Alzheimer's disease.

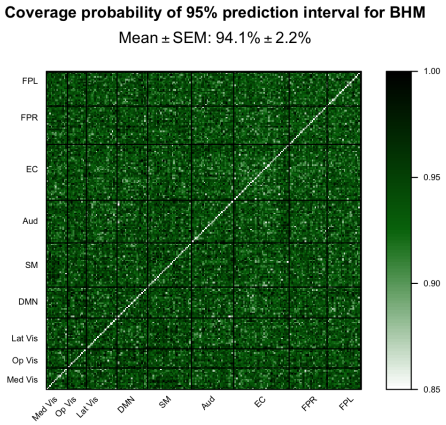



Figure 4.8: A symmetric 216 × 216 matrix showing coverage probabilities of the 95% prediction intervals for predicting first-year functional connectivity based on the Bayesian hierarchical model for the ADNI2 study. Each point on the map represents a node-pair with nodes defined based on the 264-node system and grouped by module membership. Across all node-pairs, 94.1% (with a standard error of 2.2%) of the observed connectivity measurements fall within the 95% prediction intervals.

### 4.4.2 Application to a Test-retest Kirby21 Study

We reanalyzed the Kirby 21 dataset (Landman et al., 2011) to demonstrate the utility of the proposed method in test-retest reliability studies. The data set contains two 7-minute scan-rescan resting-state fMRI images for 21 healthy adults. Imaging data was preprocessed using similar procedure as implemented for ADNI2 dataset, and details were omitted here. Given no significant covariates were reported for the 21 healthy adults, the hierarchical model was constructed as the example in section 2.3.2 with no covariate effects. General linear model was specified as the corresponding marginal model for BHM.

To evaluate the reliability of the connectivity predicts, we employed a K-fold (K=7) cross validation approach. The reliability of connectivity predictors in this study was measured by prediction mean square error (PMSE) between the observed connectivity from the second scan and the predicted connectivity based on previous connectivity estimates. Figure 4.9 compares the prediction accuracy of

BHM, GLM, raw, mean and shrinkage methods. Consistent with simulation results, BHM predictor and shrinkage estimator were the best among five methods for this test-retest dataset, with very similar PMSE across all edges. GLM predictor and mean estimator were the second best predictors with similar prediction accuracy (as measured by PMSE). We note that for healthy subjects, reliability of the functional connectivity estimates were very comparable across all voxel pairs, except for connectivity between some medial visual nodes and occipital pole visual and lateral visual nodes. In addition, we present the coverage probability map of the 95% prediction intervals based on BHM in Figure4.10. The map shows that, across all node-pairs, 92.3% of the observed connectivity measures fall within the 95% prediction intervals.

## 4.5 Remarks

In this paper, we present a general Bayesian hierarchical model for forecasting individual future functional connectivity based on the individual's earlier fMRI scans and relevant clinical and demographical characteristics. The utility of the proposed approach is demonstrated for both longitudinal and test-retest imaging studies. Through simulation studies and a longitudinal imaging study, we show that the proposed method provides improved prediction results for future RSFC compared to noisy individual estimates by modeling the longitudinal changes in RSFC. The proposed predictive model has many potential clinical applications. For example, it can help implement early intervention based on the predicted RSFC trajectory to prevent severe outcomes from mental illnesses. It can also provide guidance for developing a more effective individualized treatment plan by taking into account both the group-level effects as well as the patient's unique neural connectivity features.

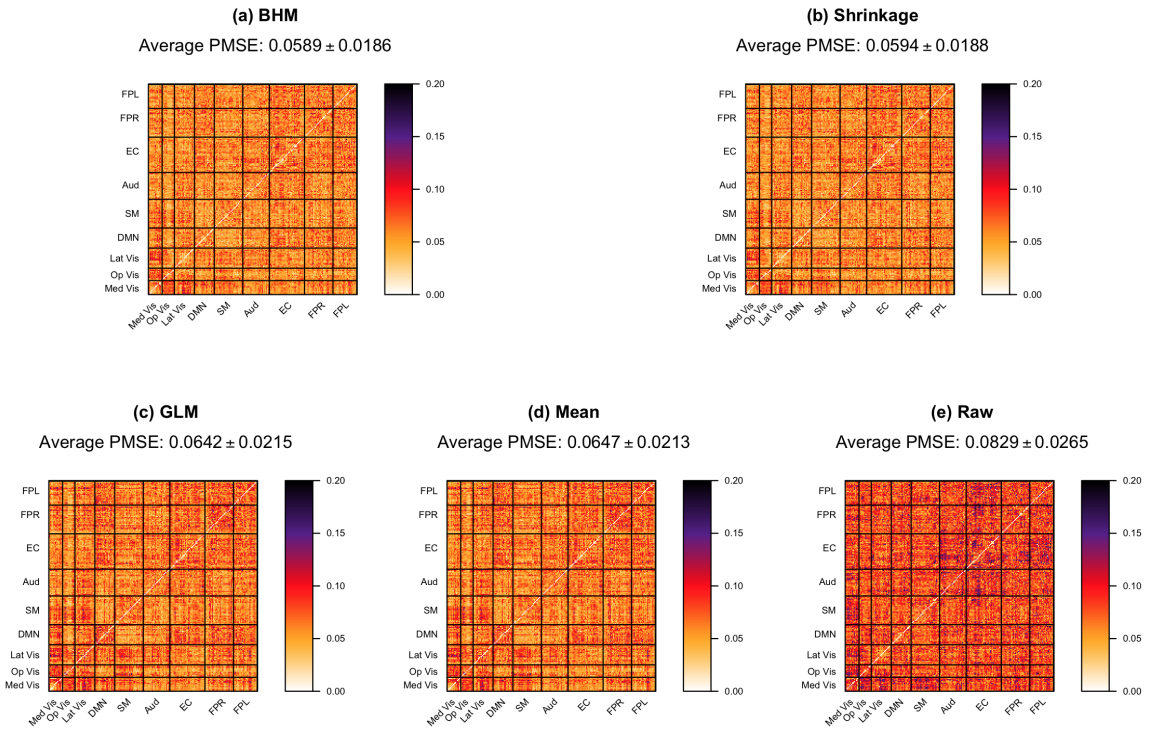



Figure 4.9: Average prediction mean square error (PMSE) matrices for predicting functional connectivity across all subjects in Kirby21 study using bayesian hierarchical model, shrinkage estimator, general linear model, mean estimator and raw estimator, respectively. Darker color represents larger PMSE, and lighter color represents smaller PMSE.

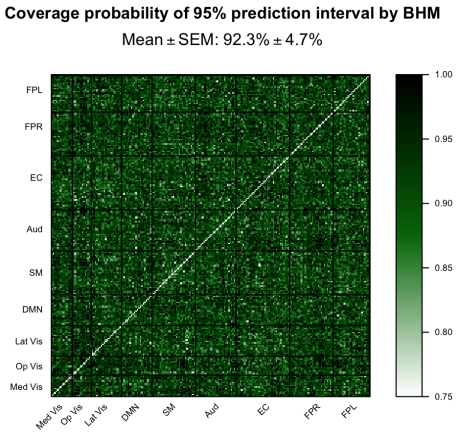



Figure 4.10: A symmetric 220 × 220 matrix showing coverage probabilities of the 95% prediction intervals for predicting first-year functional connectivity based on the Bayesian hierarchical model for the Kirby21 study. Each point on the map represents a node-pair with nodes defined based on the 264-node system and grouped by module membership. Across all node-pairs, 92.3% (with a standard error of 4.7%) of the observed connectivity measurements fall within the 95% prediction intervals.

For test-retest imaging studies, the proposed method can provide a more reliable estimator for individual RSFC than alternative methods. We show that there is a nice connection between our predictor of RSFC and the recently developed shrinkage estimator (Shou et al. 2014; Mejia et al. 2015). Both methods are essentially a weighting procedure that "borrow"s strength from a larger population to improve individual estimates. Compared with the shrinkage estimator, the proposed method based on Bayesian hierarchical model offers a number of methodological advances. First, our proposed method can account for covariate effects and thus can potentially provide more accurate estimates of RSFC for heterogeneous population, e.g., when the study cohort consists of both patients and healthy controls. Second, the proposed predictor provides a unified and model-based approach for determining the weighting factor which controls the relative contribution from the population-level and individual-level effects. Thirdly, we develop an inference procedure for the proposed predictor which allows for hypothesis testing and confidence interval construction for the RSFC.

In the data examples, we applied the prediction algorithm using a node-by-node connectivity map with nodes defined by a 264-node system (Power et al., 2014) and grouped by functional network membership (Smith et al., 2009). The advantage of using such a parcellation is that the node-level connectivity map achieves a balance between node localization and dimension reduction as compared to full voxel-by-voxel distance matrix. In addition, the grouping of the nodes allows us to examine the reliability and predictability of intra- and inter- network connectivity. Although we illustrated our method using node-based connectivity maps, the proposed method can be directly applied to any connectivity metrics, such as partial correlation or seed-based correlation maps, as long as the normality assumption is satisfied.

## 4.6 Appendix

This appendix describes an EM algorithm for the proposed Bayesian hierarchical model in Section 4.2.

For a given voxel pair, the two-stage model for all N subjects can be collapsed into the following representation:

$$\boldsymbol{r} = \boldsymbol{W}\boldsymbol{\theta} + \boldsymbol{\epsilon}^{(1)}$$

where $\boldsymbol{r} = [\boldsymbol{r}_1^T, \cdots, \boldsymbol{r}_N^T]^T$ is the connectivity from all $N$ subjects. The covariate matrix $\boldsymbol{W} = [\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(1)}\boldsymbol{X}^{(2)}]$, where $\boldsymbol{X}^{(j)} = diag(\boldsymbol{X}_1^{(j)}, \cdots, \boldsymbol{X}_N^{(j)})$ for $j = 1, 2$. The parameter $\boldsymbol{\theta} = [\boldsymbol{\epsilon}^{(2)^T}, \boldsymbol{\beta}^T]^T$ with $\boldsymbol{\epsilon}^{(2)} = [\boldsymbol{\epsilon}_1^{(2)T}, \cdots, \boldsymbol{\epsilon}_N^{(2)T}]^T$ and $\boldsymbol{\epsilon}^{(1)} = [\boldsymbol{\epsilon}_1^{(1)T}, \cdots, \boldsymbol{\epsilon}_N^{(1)T}]^T$. The parameter $\theta$ contains the population-level parameters, and the error terms in the secondary-level of the model are treated as parameters in Bayesian framework because they serve as priors for the subject-specific parameters on the first-level model. Based on the model specification, $\boldsymbol{\theta}$ is assumed to follow Gaussian distributions with the mean $\boldsymbol{\eta_\theta} = E(\boldsymbol{\theta}) = [\boldsymbol{0}^T, \boldsymbol{\eta_\beta}^T]^T$ and the covariance $\boldsymbol{\Sigma_\theta} = cov(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}^{(2)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma_\beta} \end{bmatrix}$. The likelihood and priors under the Gaussian distribution assumptions are

$$p(\boldsymbol{Y}|\boldsymbol{\theta}) \propto exp\{-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{W}\boldsymbol{\theta})^T\boldsymbol{\Sigma}^{(1)-1}(\boldsymbol{Y} - \boldsymbol{W}\boldsymbol{\theta})\},$$
$$p(\boldsymbol{\theta}) \propto exp\{-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\eta_\theta})^T\boldsymbol{\Sigma_\theta}^{-1}(\boldsymbol{\theta} - \boldsymbol{\eta_\theta})\},$$

which lead to a Gaussian posterior density for $\boldsymbol{\theta}$,

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) \propto exp\{-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\eta_{\theta|Y}})^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}|\boldsymbol{Y}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\eta_{\theta|Y}})\},$$

where $\boldsymbol{\Sigma_{\theta|Y}} = (\boldsymbol{W}^T\boldsymbol{\Sigma}^{(1)-1}\boldsymbol{W} + \boldsymbol{\Sigma_\theta}^{-1})^{-1}$ and $\boldsymbol{\eta_{\theta|Y}} = \boldsymbol{\Sigma_{\theta|Y}}(\boldsymbol{W}^T\boldsymbol{\Sigma}^{(1)-1}\boldsymbol{Y} + \boldsymbol{\Sigma_\theta}^{-1}\boldsymbol{\eta_\theta})$.

If we adopt the empirical Bayesian scheme with $\boldsymbol{\Sigma_\beta} = \infty$, $\boldsymbol{\Sigma_\theta}^{-1}\boldsymbol{\eta_\theta} = 0$ which means we do not need to specify the prior mean of $\boldsymbol{\beta}$ since they do not influence the posterior. Therefore, we only need to update the hyper-parameters $\boldsymbol{\lambda} = [\boldsymbol{\lambda}^{(1)T}, \boldsymbol{\lambda}^{(2)T}]^T$ in the error covariance matrices $\boldsymbol{\Sigma}^{(1)}$ and $\boldsymbol{\Sigma}^{(2)}$. To estimate the parameters and hyper-parameters in the above equations, we use the EM algorithm.

**E Step**

At the E step, we obtain the conditional distribution of the parameter $\boldsymbol{\theta}$ given the observed data $\boldsymbol{Y}$ and the current estimates of the hyper-parameters of the covariance components $\boldsymbol{\lambda}^{(m)}$, i.e., $p(\boldsymbol{\theta}|\boldsymbol{\lambda}^{(m)}, \boldsymbol{Y})$. Based on the model specification, the conditional distribution is Gaussian distribution with the mean and covariance of $\boldsymbol{\Sigma}^{(m)}_{\boldsymbol{\theta}|\boldsymbol{Y}} = (\boldsymbol{W}^T\hat{\boldsymbol{\Sigma}}^{(1)-1}\boldsymbol{W} + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}^{-1})^{-1}$ and $\boldsymbol{\eta}^{(m)}_{\boldsymbol{\theta}|\boldsymbol{Y}} = \boldsymbol{\Sigma}^{(m)}_{\boldsymbol{\theta}|\boldsymbol{Y}}(\boldsymbol{W}^T\hat{\boldsymbol{\Sigma}}^{(1)-1}\boldsymbol{Y} + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\eta_\theta})$, where $\hat{\boldsymbol{\Sigma}}^{(1)}$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$ are based on $\boldsymbol{\lambda}^{(m)}$.

**M Step**

At the M step, we use the estimated conditional distribution of $\boldsymbol{\theta}$ given $\boldsymbol{Y}$ from the E-step to update the maximum likelihood (ML) estimates for the covariance hyperparameters $\boldsymbol{\lambda}$ while keeping the parameters fixed. Specifically, we find the $\boldsymbol{\lambda}^{(m+1)}$ such that the log-likelihood $F = \ln p(\boldsymbol{Y}|\boldsymbol{\lambda})$ is maximized. This can be done using a Fisher-scoring iteration: $\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + (\frac{\partial^2 F}{\partial \boldsymbol{\lambda}^2})^{-1}\frac{\partial F}{\partial \boldsymbol{\lambda}}$.

# Chapter 5

# Summary and Future Work

Assessing agreement for measurements obtained from the same subjects by different methods has been an important topic in biomedical studies. Standard agreement methods are often challenged by the complexities in biomedical data. The main objective of our research is to explore novel agreement methods that accommodate two complex outcomes often encountered in biomedical studies, i.e., survival data and high dimensional neuroimaging data, to uncover the agreement structures in these data and to develop statistical methods that can improve the reproducibility of the data.

Our first proposed method is a local agreement pattern measure for correlated continuous survival times based on bivariate hazard functions. The proposed agreement index is shown to have several advantages over existing methods: (1) It easily accommodates for censored observations. (2) It fully captures the dependence structure between correlated survival outcomes. (3) It reveals how the strength of agreement changes along the time. We propose a nonparametric estimation method and show the desired asymptotic properties of the estimator. Our simulation studies show satisfactory performances of the proposed estimator and inference procedure. An application to the prostate cancer data is used to demonstrate the practical utilities of our proposals.

Our second proposed method is for assessing the reproducibility of the same subjects' high dimensional brain imaging data acquired at different imaging sites, which is a very important task to ensure the data quality in multi-site imaging studies. Specifically, we propose a two-stage network-based agreement method that first decomposes the whole brain into several functional networks, then identifies the important functional networks associated with the experimental tasks, and finally measures the agreement between imaging sites using our proposed

agreement indices with the identified network-specific temporal responses. We develop nonparametric estimation methods for the proposed indices, fCIA and fTCCC, and prove the consistency and asymptotical normality of the proposed estimators. The proposed methods are illustrated via fBIRN Phase I Traveling Subject study to investigate site effects in multi-site imaging studies. Nonparametric permutation test based on the fCIA index reveals some potential site effect between site 1, 2 and the other sites, which is consistent with visual examination of the data, while the fTCCC fails to detect the abnormality. Since our agreement method measures network-specific agreement, we are able to identify that the inconsistency of site 1 in task-fMRI data mostly comes from auditory network and the discrepancy in fMRI images at site 2 is mainly found in secondary visual network.

One of the future goals is to look into the task protocols and scanning procedures from these study sites with the goals of uncovering the scientific logistics behind data inconsistency. If found, revalidation and calibration should be made in order to ensure the interchangeability of data acquired from different imaging sites for the subsequent studies. One potential direction for future research is to develop imaging calibration tools by using the proposed indices as the objective functions. In this work, we develop our reproducibility method for task-fMRI data by taking advantages of the expected periodic pattern of the network-specific temporal responses corresponding to the experimental-task time series. For networks that are associated with resting-state functions, such as cognitive and emotional activities, we do not expect dominant periodic changes in their temporal responses. Therefore, different approaches should be considered for these resting-state networks. For example, Zou et al. (2004) evaluates the image segmentation quality based on a spatial overlap index, Dice similarity coefficient. Another potential direction for our future research is to investigate the reproducibility of individual parcellation for the same subjects

when their brain images are acquired from different imaging sites using resting-state fMRI data.

The third proposed method in this dissertation is a general Bayesian hierarchical model for forecasting individual's future resting-state functional connectivity (RSFC) based on the individual's earlier fMRI scans and relevant clinical and demographical characteristics. The utility of the proposed approach is demonstrated for both longitudinal and test-retest reproducibility imaging studies. Through simulation studies and a longitudinal imaging study, we show that the proposed method provides improved prediction results for future RSFC compared to noisy individual estimates by modeling the longitudinal changes in RSFC. The proposed prediction method could provide a useful tool to predict the changes in individual patient's brain connectivity with the progression of the disease. It can also be used to predict a patient's brain connectivity after a specified treatment regimen which could potentially help guide individualized treatment plan. Another utility of the proposed method, as demonstrated through the simulation studies and a test-retest data application, is that it could be applied to test-retest imaging data to develop a more reliable estimator for individual functional connectivity. In both data examples, we applied the proposed method using a node-by-node Pearson correlation map with nodes defined by a 264-node system (Power et al., 2014) and grouped by functional network membership (Smith et al., 2009). We note that the proposed method is not restricted to this specific connectivity measure and can be directly applied to any connectivity metrics, such as partial correlation or seed-based correlation maps, as long as the normality assumption is satisfied.

In the future, we plan to apply the proposed prediction method using the partial correlation map as the connectivity measure which has been shown to be an

effective method for minimizing the effects of global artifacts and artificial negative correlations, while accurately identifying the intrinsic functional networks (Chen et al, 2013). We expect our proposed prediction method will perform better using the partial correlation metric than the traditional Pearson correlation considering some artificial effects are removed from the imaging data. We also plan to apply the proposed method to the ADNI2 data when more data has been collected. Currently, to include as many subjects as possible for model estimation, we only utilize three longitudinal scans for each subject. We hope with more subjects and longer follow-up, the proposed method will provide better insights on the relationships between longitudinal alternation of the brain functional connectivity and psychiatric disorders, such as mild cognitive impairment and Alzheimer's disease, and further improve the prediction accuracy.

# Bibliography

Agresti, A. (1988). A model for agreement between ratings on an ordinal scale. *Biometrics* **44**, 539-548.

Agresti, A. (1989). An agreement model with kappa as parameter. *Statistics & Probability Letters* **7**, 271-273.

Agresti, A. (1992). Modelling patterns of agreement and disagreement. *Statistics in Medicine* **1**, 201-218.

Agresti, A., Ghosh, A. and Bini M. (1995). Raking Kappa: Describing Potential Impact of Marginal Distributions on Measures of Agreement. *Biometrics* **37**, 811-820.

Bagkavos, D. and Patil P. (2009). Variable Bandwidths for Nonparametric Hazard Rate Estimation. *Communications in Statistics–Theory and Methods* **38**, 1055-1078.

Banerjee, M., Capozzoli M., McSweeney L. and Sinha D. (1999). Beyond kappa: A review of interrater agreement measures. *Canadian Journal of Statistics* **27**, 3-23.

Barnhart, H.X., Lokhnygina Y., Kosinski A.S. and Haber M. (2007). Comparison of concordance correlation coefficient and coefficient of individual agreement in assessing agreement, *Journal of Biopharmaceutical Statistics* **17**, 721-738.

Biswal, B., Zerrin Yetkin, F., Haughton, V.M., Hyde, J.S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* **34(4)**, 537-541.

Bland, J.M. and Altman, D.G. (2003). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **i**, 307-310.

Borg, J., Mollgaard, A. and Riis, B.J. (1995). Single x-ray absorptiometry: performance characteristics and comparison with single photon absorptiometry. *Osteoporosis International* **5**, 377-381.

Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141-151.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20**, 37-46.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* **70**, 213-220.

Critz, F.A., Levinson, K., Williams, W.H., and Holladay, D.A. (1996). Prostate-specific antigen Nadir: the optimum level after irradiation for prostate cancer. *Journal of Clinical Oncology* **14**, 2893-2900.

Critz, F.A., Tarlton, R.S., and Holladay, D.A. (1995). Prostate specific antigen-monitored combination radiotherapy for patients with prostate cancer. *Cancer* **75**, 2383-2391.

Dabrowska, D.M. (1988). Kaplan-Meier Estimate on the Plane. *The Annals of Statistics* **16**, 1475-1489.

Damoiseaux, J.S., Prater, K., Miller, B.L., Greicius, M.D. (2012). Functional connectivity tracks clinical deterioration in Alzheimer's disease. *Neurobiology of Aging*, **33(4)**, 828.e19-828.e30.

Darroch J. N. and McCloud P. I. (1986). Category Distinguishability and Observer Agreement. *Australian Journal of Statistics* **28**, 371-388.

Duong, T. and Hazelton M.L. (2003). Plug-in bandwidth matrices for bivariate kernel density estimation. *Journal of Nonparametric Statistics* **15**, 17-30.

Duong, T. and Hazelton M.L. (2005). Convergence rates for unconstrained bandwidth matrix selectors in multivariate kernel density estimation. *Scandinavian Journal of Statistics* **32**, 485-506.

Duong, T. (2007). ks: Kernel density estimation and kernel discriminant analysis in R. *Journal of Statistical Software* **21(7)**, 1-16.

Fermanian, J.D. (1997). Multivariate Hazard Rates under Random Censorship. *Journal of Multivariate Analysis* **62**, 273-309.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin* **76**, 378-382.

Fransson, P., Marrelec, G. (2008). The precuneus/posterior cingulate cortex plays a pivotal role in the default mode network: evidence from a partial correlation network analysis. *NeuroImage* **42(3)**, 1178-1184.

Friston, K.J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: theory. *NeuroImage* **16(2)**, 465-483.

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E.

(2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci.* **102(27)**, 9673-9678.

Gay, C.W., Robinson, M.E., George, S.Z., Perlstein, W.M., Bishop, M.D. (2014). Immediate Changes After Manual Therapy in Resting-State Functional Connectivity as Measured by Functional Magnetic Resonance Imaging in Participants With Induced Low Back Pain. *J Manipulative Physiol Ther.* **37(9)**, 614-627.

Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci.* **100(1)**, 253-258.

Guo, Y. and Manatunga, A.K. (2005). Modeling the Agreement of Discrete Bivariate Survival Times using Kappa Coefficient. *Lifetime Data Analysis* **11**, 309-332.

Guo, Y. and Manatunga, A.K. (2007). Nonparametric estimation of the concordance correlation coefficient under univariate censoring. *Biometrics* **63**, 164-72.

Guo, Y., Bowman, F.D., Kilts, C., 2008. Predicting the Brain Response to Treatment Using a Bayesian Hierarchical Model With Application to a Study of Schizophrenia. *Hum. Brain Mapp.* **29(9)**, 1092-1109.

Guo, Y and Manatunga, A.K. (2010). A note on assessing agreement for frailty models. *Statistics and Probability Letters* **80**, 527-533.

Guo, Y, Li R, Peng L and Manatunga AK, 2013, A new agreement measures based on survival processes. *Biometrics* **69**, 874-882

Holland, P.W. and Wang, Y.J. (1987). Dependence function for continuous bivariate densities. *Communications in Statistics - Theory and Methods* **16**, 863-876.

Hu, T., Nan, B, Lin, X and Robins, J.M. (2011). Time-dependent cross ratio estimation for bivariate failure times. *Biometrika* **98**, 341-354.

Jones, M.C. (1996). The local dependence function. *Biometrika***83**, 899-904.

Jung, M., Dimtchev, A., Velena, A., Dritschilo, A., (2011). Combining radiation therapy with interstitial radiation-inducible TNF-[alpha] expression for locoregional cancer treatment. *Cancer Gene Therapy* **18(3)**, 189-195.

Kraemer, H.C. (1980). Extension of the kappa coefficient. *Biometrics* **36**, 207-216.

Landman, B.A., Huang, A.J., Gifford, A., Vikram, D.S., Lim, I.A., Farrell, J.A., Bogovic, J.A., Hua, J., Chen,M., Jarso, S., Smith, S.A., Joel, S., Mori, S., Pekar, J.J., Barker, P.B., Prince, J.L., van Zijl, P.C., 2011. Multi-parametric neuroimaging reproducibility: a 3-T resource study. NeuroImage 54(4), 2854-2866.

Lin, L.I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45**, 255-268.

Lin, L.I. (2000). Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence. *Statistics in Medicine* **19**, 255-270.

Lin, L.I., Hedayat, A.S., Sinha, B., Yang, M. (2002). Statistical methods in assessing agreement: Models, issues and tools. *Journal of American Statistical Association* **97**, 257-270.

Liu, X., Du, Y., Teresi, J. and Hasin, D. S. (2005). Concordance correlation in the measurements of time to event. *Statistics In Medicine* **24**, 1409-1420.

Mejia, A.F, Nebel, M.B., Shou, H., Crainiceanu, C.M., Pekar, J.J., Mostofsky, S., Caffo, B., Lindquist, M.A. (2015). Improving reliability of subject-level resting-state fMRI parcellation with shrinkage estimators. *NeuroImage* **112**, 14-29.

Musselman, D., Lawson, D., Gumnick, J., Manatunga, A., Penna, S., Goodkin, R., Greiner, K., Nemerof, C., and Miller, A. (2001). Paroxetine for the prevention

of depression induced by high-dose interferon alfa. *The New England Journal of Medicine* **344**, 961-966.

Müller, H.G. and Wang, J.L. (1994). Hazard Rate Estimation under Random Censoring with Varying Kernels and Bandwidths. *Biometrics* **50**, 61-76.

Oakes, D. (1982). A model for Association in Bivariate Survival Data. *Journal of the Royal Statistical Society: Series B* **44**, 414-422.

Oakes D. (1986). Semiparametric inference in a model for association in bivariate survival data. *Biometrika* **73**, 35361.

Oakes D. (1989). Bivariate survival models induced by frailties.*Journal of the American Statistical Association* **84**, 48793.

Power, J.D., Cohen, A.L., Nelson, S.M.,Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L. Petersen, S.E. (2011). Functional network organization of the human brain. *Neuron*, **72(4)**, 665-678.

Prentice, R.L. and Cai, J. (1992). Covariance and survival function estimation using censored multivariate failure time data. *Biometrika* **79**, 495-512.

Sarpal, D.K., Robinson, D.G., Lencz, T., Argyelan, M., Ikuta, T., Karlsgodt, K., Gallego, J.A., Kane, J.M., Szeszko, P.R., Malhotra, A.K. (2009). Antipsychotic treatment and functional connectivity of the striatum in first-episode schizophrenia. *JAMA Psychiatry* **72(1)**, 5-13.

Schild, RL, Fimmers, R, Hansmann, M. (2000). Fetal weight estimation by three-dimensional ultrasound. *Ultrasound in Obstetrics & Gynecology* **16**, 445-452.

Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, New York, Chichester.

Shou, H., Eloyan, A., Nebel, M.B., Mejia, A., Pekar, J.J., Mostofsky, S., Caffo, B., Lindquist, M.A., Crainiceanu, C.M. (2014). Shrinkage prediction of seed-voxel brain connectivity using resting state fMRI. *NeuroImage* **102(2)**, 938-944.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N.,Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., 2009. Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci.* **106(31)**, 13040-13045.

Steven P. M. (2013). *EnvStats: An R Package for Environmental Statistics.* Springer, New York. ISBN 978-1-4614-8455-4.

Tanner, M.A. and Young, M.A. (1985). Modeling ordinal scale disagreement. *Psychological Bulletin* **98(2)**, 408-415.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

Wang, M.P.; Jones, M.C. (1993). Comparison of smoothing parameterizations in bivariate kernel density estimation. *Journal of the American Statistical Association* **88**, 520-528.

Wang, M.P.; Jones, M.C. (1994). Multivariate plug-in bandwidth selection. *Computational Statistics* **9**, 97-177.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing* Chapman & Hall, London.

Wang, K., Liang, M., Wang, L., Tian, L., Zhang, X., Li, K., Jiang, T. (2007). Altered functional connectivity in early Alzheimer's disease: a resting-state fMRI study. *Hum. Brain Mapp.* **28(10)**, 967-78.

Woodward, N.D., Cascio, C.J. (2015). Resting-State Functional Connectivity in Psychiatric Disorders. *JAMA Psychiatry* **72(8)**:743-744.

Xia, M., Wang, J., He, Y. (2013). BrainNet Viewer: a Network Visualization Tool for Human Brain Connectomics. *PLoS ONE* **8(7)**: e68910.

Yao, H., Zhou, B., Zhang, Z., Wang, P., Guo, Y., Shang, Y., Wang, L., Zhang, X., An, N., Liu, Y., and Alzheimer's Disease Neuroimaging Initiative. (2014). Longitudinal Alteration of Amygdalar Functional Connectivity in Mild Cognitive Impairment Subjects Revealed by Resting-State fMRI. *Brain Connectivity* **4(5)**: 361-370.

Zeng D., Cornea E., Dong J., Pan J., and Ibrahim J. G. (2015). Assessing temporal agreement between central and local progression-free survival times. *Statistics in Medicine* **34**, 844-858.