Jialu Ran

July 16, 2024

Jialu Ran

Date

Advanced Statistical Methods for Brain Connectivity Analysis

By

Jialu Ran
Doctor of Philosophy

Department of Biostatistics and Bioinformatics

Ying Guo, PhD; David Benkeser, PhD
Advisor

Benjamin Risk, PhD
Committee Member

Sarah Shultz, PhD
Committee Member

Accepted:

Kimberly Jacob Arriola, PhD
Dean of the James T. Laney School of Graduate Studies

July 16, 2024
Date

Advanced Statistical Methods for Brain Connectivity Analysis

By

Jialu Ran
Bachelor of Science

Advisor: Ying Guo, PhD; David Benkeser, PhD

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 for the degree of
Doctor of Philosophy
in Department of Biostatistics and Bioinformatics
2024

Abstract

Advanced Statistical Methods for Brain Connectivity Analysis
By Jialu Ran

The analysis of brain network connectivity has become an important tool for investigating brain function and various disease characteristics. However, this type of analysis is challenging due to the complex structure and high dimensionality of brain imaging data. To extract meaningful information from these data, sophisticated analytical methods are required. This dissertation focuses on the development of statistical methods to enhance our understanding of the human brain.

In Chapter 1, we introduce the background of brain connectivity analysis, discuss the associated challenges, review existing solutions and their drawbacks, and highlight our contributions to the field.

In Chapter 2, we introduce dyna-LOCUS, a method for identifying neural circuits in dynamic brain functional connectivity. This method performs dynamic connectome source separation using low-rank modeling of connectivity traits with sparsity and temporal smoothness regularization. We validate the performance of dyna-LOCUS through simulations and apply it to the Philadelphia Neurodevelopmental Cohort (PNC) study. This application identifies latent neural circuits underlying observed dynamic connectivity, reveals key brain regions that drive each of these circuits, investigates their temporal expression profiles, and identifies neural circuits associated with gender and neurodevelopment.

In Chapter 3, we develop a longitudinal-LOCUS method to study changes in brain connectomes over time. This method decomposes longitudinal functional connectivity (FC) measures using blind source separation with low-rank structures and angle-based sparsity regularization. We present an efficient iterative node-rotation algorithm to solve the non-convex optimization problem for learning longitudinal-LOCUS. Simulations demonstrate superior accuracy in recovering latent sources and mixing coefficients compared to existing methods. We apply the method to the Adolescent Brain Cognitive Development (ABCD) data to investigate developmental changes in neural circuits and their differences between genders.

In Chapter 4, we address a challenge that arises in brain imaging data when participants move their heads during a brain scan. This motion can cause spurious associations when studying differences between functional connectivity across groups. We propose decomposing neural and motion-induced sources of group differences under a causal mediation framework. We establish the theoretical properties of our proposed estimators and validate the theory using simulation studies. The framework is applied to estimate the difference in functional connectivity between autism spectrum disorder (ASD) children and typically developing children in the ABIDE study. Our analyses indicate that some long-range connections between a seed region in the default mode network and frontal-parietal regions exhibit hyperconnectivity in ASD. Naively including high-motion children appears to cause spurious connectivity differences. Naively excluding high-motion children removed group differences.

In Chapter 5, we expand on the second project by developing MoCo, an R package for motion control in MRI studies. In addition to motion control, MoCo allows for missingness in MRI data due to collection or preprocessing issues. We demonstrate MoCo with examples related to ASD and attention-deficit/hyperactivity disorder (ADHD), showing it effectively mitigates motion artifacts, enhances data utilization, addresses selection biases, and is more robust to preprocessing pipelines.

Advanced Statistical Methods for Brain Connectivity Analysis

By

Jialu Ran
Bachelor of Science

Advisor: Ying Guo, PhD; David Benkeser, PhD

A dissertation submitted to the Faculty of the
Emory College of Arts and Sciences of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Department of Biostatistics and Bioinformatics
2024

Acknowledgments

First and foremost, I am deeply grateful to my advisors, Dr. Ying Guo and Dr. David Benkeser, and committee member, Dr. Ben Risk, for their invaluable advice, inspiration, encouragement, and support throughout my Ph.D. journey. Dr. Guo is a distinguished specialist in neuroimaging. Under her guidance, I have gained valuable insights into this field, and I deeply admire her meticulous attention to detail and profound analytical approach. Meanwhile, she has been more than just an academic advisor; she has guided me with her knowledge and experience in many aspects of my life. Dr. Benkeser's expertise in causal inference and his diligence have profoundly influenced and inspired me. He is responsive in fostering the growth of his students, always promptly addressing my questions whenever he is available. Two of the most important lessons I've learned from him are how to simplify complex concepts and the value of constant diligence. I'm consistently impressed by his excellent communication skills, whether in casual conversations or during lectures, and I aspire to be as humorous as him someday. Dr. Ben Risk has provided invaluable support from both neuroimaging and causal inference perspectives. His attention to detail, inquisitive nature, and commitment to continuous learning, despite his expertise, have greatly influenced and motivated my academic journey. Working under their supervision has been an invaluable experience.

I would like to express my sincere appreciation to my dissertation committee member, Dr. Sarah Shultz, and to other professors who inspired me along this journey: Dr. Razi Nabi and Dr. Liming Peng. Their insightful comments and invaluable suggestions have played a crucial role in shaping my dissertation.

I am grateful to have had the pleasure and fortune of having supportive and encouraging colleagues and friends during my Ph.D. I would like to extend my thanks to all my colleagues: Dr. Yikai Wang, Anna Guo, Zihang Wang, Dr. Yutong Jin, Guangming Yang, Dr. Yaotian Wang, Dr. Yanxi Tan, and Qi Zhang.

I would also like to extend my heartfelt appreciation to my dearest friends who provided unwavering support and encouragement during the challenging moments of my Ph.D.: Yuliang Zhang, Mengyu Di, Mingxi Yan, and Keyu Fang. Meeting you during my PhD has been a true blessing, and your friendship and belief in me have been invaluable.

Finally, I would like to thank my family for their unconditional love and support. I also want to express my enduring love for my grandmother; I will always cherish and miss you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent decades, there has been a growing interest among researchers in exploring the interconnections across the brain through brain connectivity analysis (Bullmore and Sporns, 2009; Van Den Heuvel and Pol, 2010; Deco et al., 2011; Lang et al., 2012; Shi and Guo, 2016; Wang et al., 2016; Wang and Guo, 2023). Research in this field not only sheds light on the human brain's functionality but also reveals neurodevelopment (DeCarli et al., 2012; Hoff et al., 2013; Satterthwaite et al., 2015; Iannilli et al., 2017; Chai et al., 2017) and provides insights into mental health disorders (Mayberg, 2003; Sorg et al., 2007; Seeley et al., 2009; Demirtaş et al., 2016; Miller et al., 2016). To measure brain connectivity, functional magnetic resonance imaging (fMRI), developed in the 1990s, has emerged as a cornerstone technique. In resting-state fMRI (rs-fMRI), individuals lie inside a scanner while their brain activity is recorded. rs-fMRI captures spontaneous neural activity in the absence of specific external stimuli (Biswal et al., 1995; Lv et al., 2018). The scanner records blood oxygenation level-dependent (BOLD) signals from hundreds of thousands of locations in the brain, producing a time series of brain images that reflect changes in activity over time for each brain region. Functional connectivity (FC), which is statistical association between the time series of different brain regions, is calculated subsequently (Friston et al., 1993). This association can be quantified using metrics such as correlation, partial correlation (Wang et al., 2016), mutual information (Hlinka et al., 2011), coherence (Chai et al., 2017), and others. By analyzing FC, researchers can comprehensively delineate and quantify the dynamic interactions between brain regions, gaining valuable insights into how brain networks function and evolve over time (Fox and Raichle, 2007).

Among various topics in brain connectivity analysis, network-oriented approaches have emerged as pivotal tools for investigating brain organization, neurodevelopmental processes, and mental disorders (Greicius et al., 2003; Bastos and Schoffelen, 2016; Van Den Heuvel and Pol, 2010). The analysis focuses on studying whole-brain connectivity, characterized by a symmetric functional connectivity matrix, where each element represents the functional connectivity between pairs of brain regions. However, analyzing FC matrices presents several major challenges. Firstly, the issue of high dimensionality. rs-fMRI captures a time series of brain images that depict changes in activity over time for each brain region, known as voxels. With the human brain containing nearly a million voxels, the FC matrix can expand to hundreds of thousands of dimensions (Chung, 2018). Previous studies have attempted to mitigate this dimensionality by grouping voxels into brain regions based on atlases or node systems (Tzourio-Mazoyer et al., 2002; Power et al., 2011; Glasser et al., 2013a; Schaefer et al., 2018). However, even after such grouping, the matrix size remains substantial, typically in the range of hundreds by hundreds (Wu et al., 2013; Solo et al., 2018). Secondly, these FC matrices capture complex underlying structures and aggregate information from diverse neural circuits (Cai et al., 2017; Wang and Guo, 2023). These underlying circuits can be considered latent sources or connectivity traits, representing a collection of connections between different brain regions that tend to occur together during neural processing. Understanding these latent connectivity traits is crucial for gaining valuable insights into the architecture and dynamics of brain organization. Advanced methods are needed to decompose FC matrices and recover the underlying neurocircuitry subsystems reliably. Lastly, the presence of mixed signals and noise during data collection, exacerbated by motion artifacts during scanning sessions, complicates and may introduce biases into brain connectivity analysis (Birn, 2012; Van Dijk et al., 2012; Satterthwaite et al., 2013c; Birn et al., 2014).

Regarding the first two challenges, various dimension reduction techniques have been developed to extract useful information from high-dimensional, complex datasets. Some of the most commonly used methods include principal component analysis (PCA) (Wold et al., 1987), support vector machines (SVM) (Cortes and Vapnik, 1995), and independent component analysis (ICA) (Hyvärinen and Oja, 2000). However, these methods need to be specifically tailored to accommodate unique characteristics and requirements of brain connectivity data, which are not only high-dimensional but also symmetric. Specifically, the self-connections represented by diagonal elements are not of interest in the FC matrix. Thus, the relevant information is effectively captured by either its lower or upper triangular form (Amico et al., 2017). Moreover, the FC matrix demonstrates a dependence structure across edges in the brain connectome and requires careful modeling. For example, ICA stands out as one of the most popular techniques among the various blind source separation methods developed for decomposing neuroimaging data. However, existing ICA methods have mainly focused on decomposing observed neural activity signals such as the blood oxygen level-dependent (BOLD) series from fMRI or the electrode signal series from EEG (Calhoun et al., 2001; Beckmann and Smith, 2004; Shi and Guo, 2016; Wang and Guo, 2019). The distinct properties of brain connectivity matrices, in contrast to activity data, limit the applicability of many existing methods. Recently, Amico et al. (2017) introduced connICA, a framework for connectivity-independent component analysis that vectorizes connectivity matrices and applies existing ICA algorithms for decomposition. While it offers a valuable tool for decomposing connectivity matrices, it treats each connection as an independent sample, which disregards the dependence structure across edges in the brain connectome. Meanwhile, this approach results in a large number of edge-wise parameters for estimation, leading to reduced accuracy. In the absence of sparsity regularization, connICA tends to produce densely connected components, thereby increasing the likelihood of spurious findings.

An extended follow-up challenge related to the first two challenges is the decomposition of more complex FC matrices, such as those in dynamic or longitudinal cases. Much of the research on brain connectome analysis has focused on stationary or static functional connectivity, which are association measures obtained using the whole fMRI BOLD series. However, studies have shown that brain connections dynamically reconfigure over seconds and are better captured dynamically (Chang and Glover, 2010; Cribben et al., 2012; Ekman et al., 2012). Various techniques have been developed to analyze dynamic functional connectivity (dFC), including the sliding window method (Chang and Glover, 2010; Sakoğlu et al., 2010; Kiviniemi et al., 2011; Allen et al., 2014), Hidden Markov Model (HMM) (Baker et al., 2014; Eavani et al., 2013), dynamic conditional correlations (Lindquist et al., 2014), dynamic connectivity regression (DCR) (Cribben et al., 2012, 2013), and time-frequency approaches (Chang and Glover, 2010; Demirtaş et al., 2016; Yaesoubi et al., 2015). These methods produce sequences of symmetric connectivity matrices, with each entry representing connectivity measures like correlations between brain regions, enabling the investigation of non-stationary changes in brain connectivity during imaging. There is a need to develop methods for decomposing these dFC matrices to study the temporal profiles of connectivity traits across time and investigate non-stationary changes in brain connectivity throughout the imaging acquisition. In the meantime, while most FC analyses consider data from a single time point, contemporary neuroimaging research is increasingly embracing longitudinal data collection, involving multiple scans over time (Weiner et al., 2017; Garavan et al., 2018). Notable examples include the Adolescent Brain Cognitive Development (ABCD) Study (Garavan et al., 2018), the largest longitudinal study on adolescent brain development, and the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Petersen et al., 2010), which collects longitudinal resting-state fMRI images to study Alzheimer's progression. To fully harness the potential of longitudinal brain connectivity data, there is a growing need to develop methodologies that effectively decompose longitudinal brain connectivity, enabling a deeper understanding of changes in connectivity patterns over time.

Regarding the third challenge, participant head motion during scanning sessions presents a pervasive issue in neuroimaging (Van Dijk et al., 2012; Baum et al., 2018), contributing to artifacts in brain connectivity analysis (Maknojia et al., 2019). Early studies found that brain connections change during neurodevelopment, with short-range connections weakening and long-range ones strengthening. However, these patterns can be confounded by the discovery that motion causes the same patterns (Van Dijk et al., 2012; Power et al., 2014a). Careful handling of motion artifacts is crucial, especially in studies involving children with conditions like Autism Spectrum Disorder (ASD) (Deen and Pelphrey, 2012) and Attention-deficit/hyperactivity disorder (ADHD) (Castellanos and Aoki, 2016; Aoki et al., 2018). Neglecting to address these artifacts can lead to spurious differences in brain connectivity measures between groups of interest (Power et al., 2012; Deen and Pelphrey, 2012). There is extensive literature on methods to mitigate motion artifacts in neuroimaging. In resting-state fMRI studies, multiple steps are typically applied. During preprocessing, confound regression is commonly applied to remove the effects of motion parameters, global signal, cerebral spinal fluid signal, and white matter from the fMRI time series (Ciric et al., 2017). Despite these efforts, residual relationships between motion and functional connectivity may persist due to potential nonlinear associations (Deen and Pelphrey, 2012). Consequently, motion quality control (QC) procedures are implemented, often involving the exclusion of scans with excessive motion based on various criteria outlined in the literature (Power et al., 2014a; Ciric et al., 2017). However, although removing scans may help mitigate motion artifacts across modalities, it can result in drastic data loss, some studies of more than 60% (Marek et al., 2022) and introduce selection bias due to the alteration of the study population (Cosgrove et al., 2022; Nebel et al., 2022). There is a need to develop a statistical method that makes more efficient usage of the data and avoids selection bias in order to draw unbiased inferences about brain development.

Topic I of my dissertation addresses challenges in analyzing brain functional connectivity matrices, focusing on dynamic and longitudinal FC analysis. For dynamic FC, we introduce dyna-LOCUS, a novel regularized blind source separation (BSS) method to decompose matrices and reveal latent connectivity traits and their temporal expression profiles. Using low-rank factorization and innovative regularizations, dyna-LOCUS efficiently maps dynamic brain connectome traits, characterizes temporal changes in observed dFC, and identifies whole-brain dFC states. We propose an iterative Node-Rotation algorithm for learning dyna-LOCUS, validated through simulations. Applying dyna-LOCUS to the Philadelphia Neurodevelopmental Cohort (PNC) uncovers latent traits, identifies key brain connections, and explores gender differences in executive function-related connectivity during neurodevelopment. For longitudinal FC matrices, we develop longitudinal-LOCUS, using blind source separation with low-rank structures and angle-based sparsity regularization. An iterative node-rotation algorithm solves its non-convex optimization problem, showing superior accuracy in recovering latent sources and coefficients. Applied to Adolescent Brain Cognitive Development (ABCD) data, longitudinal-LOCUS identifies time-specific neural circuits and captures neurodevelopmental differences between genders in brain connectivity traits.

Topics II and III of my dissertation address the challenge of head motion during brain scans, which can introduce spurious associations in studies comparing functional connectivity across groups. In Topic II, we propose a causal mediation framework to disentangle neural and motion-induced group differences. Our approach leverages machine learning to understand how motion influences brain connectivity, validated through theoretical analysis and simulations. We apply this framework to the Autism Brain Imaging Data Exchange (ABIDE) study, investigating connectivity differences between children with and without autism spectrum disorder (ASD). In Topic III, we expand on the second project by developing MoCo, an R package for motion control in MRI studies. In addition to motion control, MoCo allows for missingness in MRI data due to collection or preprocessing issues. We demonstrate MoCo with examples related to ASD and attention-deficit/hyperactivity disorder (ADHD), showing it effectively mitigates motion artifacts, enhances data utilization, addresses selection biases, and is more robust to preprocessing pipelines.

Chapter 2

Unveiling Hidden Sources of Dynamic Functional Connectome through a Novel Regularized Blind Source Separation Approach

2.1 Introduction

The human brain is a complex network consisting of a large number of functionally linked regions and their connections (Rubinov and Sporns, 2011; Van Den Heuvel and Pol, 2010). The analysis of brain connectomes has emerged as a key area of focus in neuroscience research, providing unprecedented insights into the organization of the brain and its crucial role in neurodevelopment, aging, behavior, as well as the progression and treatment of brain-related diseases. (Bullmore and Sporns, 2009; Deco et al., 2011; Satterthwaite et al., 2014; Kemmer et al., 2018; Wang et al., 2016; Wang and Guo, 2023). Functional connectivity (FC) derived from functional imaging such as fMRI has been measured and studied in most neuroimaging studies nowadays.

Much of the research on brain connectome analysis has focused on studying stationary or static functional connectivity. Studies have shown that the brain connections undergo dynamic reconfiguration over seconds (Chang and Glover, 2010; Cribben et al., 2012; Ekman et al., 2012), indicating that it is better characterized dynamically. Network reconfigurations may occur when the brain responds to external stimuli, and are potentially even more prominent in the resting state when mental activity is unconstrained (Leonardi et al., 2014; Allen et al., 2014). In recent years, various techniques have been developed to analyze dynamic functional connectivity (dFC). These include the widely used sliding window method (Chang and Glover, 2010; Sakoğlu et al., 2010; Kiviniemi et al., 2011; Allen et al., 2014), as well as other approaches such as Hidden Markov Model (HMM)(Baker et al., 2014; Eavani et al., 2013), dynamic conditional correlations (Lindquist et al., 2014), dynamic connectivity regression (DCR) (Cribben et al., 2012, 2013), time-frequency approaches (Chang and Glover, 2010; Demirtaş et al., 2016; Yaesoubi et al., 2015), and more. Generally, the dFC methods produce a sequence of symmetric connectivity matrices. Each entry in these matrices corresponds to a connectivity measure, such as the correlation between pairs of brain regions or nodes. These dFC matrices are analyzed to investigate non-stationary changes in brain connectivity over the course of imaging acquisition.

The organization of the brain is a highly intricate system that involves a vast network of neural circuits. The observed dFC matrices reflect collected connectivity patterns across the brain, representing aggregated information contributed by various underlying neural circuits. These underlying circuits can be considered as latent sources or connectivity traits, which represent a collection of connections between different brain regions that tend to occur together during neural processing. The observed dFC data are a combination of these latent connectivity traits. The network reconfigurations in dFC are a result of temporal changes in the expression levels of these latent traits. Gaining a reliable understanding of these latent connectivity traits is crucial for obtaining valuable insights into the architecture and dynamics of brain organization, both in healthy and diseased conditions. For instance, in the field of neurodevelopmental and aging research, studies have demonstrated that different neural circuits undergo maturation at distinct ages during early brain development (Hoff et al., 2013), and they also experience varying rates of deterioration as the brain ages in the elderly population (DeCarli et al., 2012; Iannilli et al., 2017). Additionally, research studies have revealed that demographic or disease-related alterations in the brain connectome typically manifest within specific neural circuits rather than affecting the entire brain connectome as a whole (Mayberg, 2003; Sorg et al., 2007; Williams, 2016). Furthermore, these connectivity traits demonstrate different characteristics. Investigations into dynamic connectivity have uncovered considerable variations in the temporal expressions of different neurocircuitry subsystems. Certain subsystems demonstrate a higher degree of persistence over time, while others exhibit more transient characteristics (Chai et al., 2017; Baker et al., 2014; Karahanoğlu and Van De Ville, 2015). To gain insights into the brain connectome, it is essential to comprehend the composition and temporal expression profiles of its latent connectivity traits, as well as the ways in which they synchronize and interact with each other. This understanding is critical for unraveling the complex dynamics and functional organization of the brain.

Existing methods for dFC primarily focus on identifying whole brain dFC states that deviate from stationary FC patterns, without explicitly revealing the underlying connectivity traits and their interplay that give rise to different dFC states (Allen et al., 2014; Leonardi et al., 2013; Karahanoğlu and Van De Ville, 2015). There is a critical need for reliable methods to uncover the latent connectivity traits within dFC and characterize their temporal expressions, which drive the reconfiguration of dFC patterns. Several significant challenges arise in addressing this need. Firstly, whole brain dFC matrices are high dimensional, often comprising hundreds of nodes and hundreds of thousands of connections (Chung, 2018; Solo et al., 2018; Wu et al., 2013; Wang et al., 2016). This enormous number of connections increases the likelihood of spurious findings when identifying significant edges within the connectivity traits and when examining brain-behavior associations. Moreover, there is a lack of suitable methods to decompose observed dFC matrices and uncover the underlying connectivity traits and their temporal expressions. Most existing source separation methods, such as Independent Component Analysis (ICA) (Calhoun et al., 2001; Beckmann and Smith, 2004; Shi and Guo, 2016; Wang and Guo, 2019) have mainly focused on decomposing observed neural activity signals such as the blood oxygen level-dependent (BOLD) series from fMRI or the electrodes signal series from EEG. The distinct properties of brain connectivity matrices, in contrast to activity data, limit the applicability of many existing methods. In recent years, Miller et al. (2016) applied tICA, sICA, and PCA methods to decompose dFC data to reveal underlying basis correlation patterns (CPs). Amico et al. (2017) introduced a connectivity-independent component analysis framework (connICA), which vectorizes connectivity matrices and utilizes existing ICA algorithms for decomposition. While the existing ICA, PCA methods, and connICA provide a valuable tool for decomposing connectivity matrices, it has certain limitations that affect the accuracy and reliability of extracting latent connectivity components. For instance, the method treats each connection as an independent sample, disregarding the dependence structure across edges in the brain connectome. This results in a large number of edge-wise parameters for estimation and loss of accuracy. Additionally, without sparsity regularization, the method generates densely connected traits, increasing the likelihood of spurious findings.

In this paper, we propose dyna-LOCUS which is a novel *lo*w-rank decomposition of brain *c*onnectivity with *u*niform *s*parsity for *dyna*mic FC. dyna-LOCUS is a fully data-driven blind source separation method that decomposes dFC matrices to extract latent connectivity traits and characterize their dynamic temporal expression profiles.

dyna-LOCUS is a general dFC source separation method that is applicable to connectivity measures derived from various aforementioned dFC techniques. Specifically, dyna-LOCUS decomposes temporally concatenated dFC measures into a combination of latent source signals weighted by temporal mixing coefficients. Each latent source signal corresponds to an underlying connectivity trait and can be mapped back to a connectivity matrix. The mixing coefficients characterize the dynamic temporal expression profile of the connectivity trait. To enhance the accuracy and reliability of recovering the connectivity traits and their temporal profiles, dyna-LOCUS incorporates several innovative and neurobiologically-motivated strategies. Firstly, we employ a low-rank structure to model the connectivity traits. This approach is motivated by the observation that brain connectivity trait patterns often exhibit specific structures, such as block-diagonal or banded structures, which can be effectively captured using low-rank matrix factorization techniques (Zhou et al., 2013). By incorporating a low-rank structure, we significantly reduce the number of parameters involved in the estimation process, leading to improved accuracy in extracting the underlying connectivity traits. We then propose a novel angle-based sparsity regularization on the recovered connectivity traits which further increases the reliability of the results by effectively reducing the presence of spurious edges and identifying connections that are genuinely relevant to a specific trait. Furthermore, to improve the accuracy in estimating the temporal expression profiles of the connectivity traits, we propose to include a temporal smoothness regularization in the optimization function. This is motivated by the observation that dynamic FC series typically exhibit a certain degree of temporal coherence, with general patterns displaying continuity across adjacent time windows.

Another major advantage of the dyna-LOCUS is that it provides a highly efficient and convenient approach for brain-behavior modeling and characterization of whole brain dFC states, which are two major research focuses in dynamic connectome studies.

In dyna-LOCUS, the extracted connectivity traits can be viewed as a set of basis connectivity matrices and the mixing coefficients are trait loadings by projecting the observed dFC matrices onto the connectivity trait basis matrices. These trait loadings are essentially low-dimensional representations of the original dFC matrices, characterizing the temporal expression profiles of the connectivity traits. By linking trait loadings with individual demographics, behavioral and clinical measures, we can perform brain-behavior modeling based on dFC to investigate age, gender, or disease-related distinctions in particular dynamic connectome traits. Furthermore, we introduce a novel approach for detecting whole-brain dFC states by employing techniques, such as clustering, on low-dimensional trait loadings. This method is not only computationally efficient but also generates dFC states that are sparser, more reliable, and easier to interpret.

To learn the dyna-LOCUS model, we formulate an optimization function that possesses the desirable property of block multi-convexity. This property ensures the existence of multiple convex subproblems within the optimization framework, facilitating more efficient and reliable parameter estimation. Furthermore, we introduce an efficient node-rotation algorithm, which enhances the effectiveness and computational efficiency of the estimation process. The proposed model and the estimation algorithm demonstrate superior performance in recovering the underlying connectivity traits through extensive simulation studies. We apply dyna-LOCUS to investigate latent connectivity traits underlying resting state fMRI of the Philadelphia Neurodevelopmental Cohort (PNC) study. dyna-LOCUS successfully identifies latent connectivity traits that exhibit sparse and easily interpretable patterns, many of which demonstrate high reproducibility. Using the temporal mixing coefficients from dyna-LOCUS, we are able to characterize the temporal expression profiles of each of the connectivity traits. This allows us to investigate the dynamic properties of the traits including their energy level and temporal variation and how connectivity traits are synchronized with each other in their dynamics. By employing our novel procedure based on dyna-LOCUS, we successfully identify seven whole-brain dFC states in the PNC study. Notably, our findings exhibit nice agreement with the results obtained using existing whole-brain dFC methods. However, our method and results offer several key advantages including significantly reducing computation time, generating more parsimonious and interpretable whole-brain dFC states, and providing deeper insights into the key latent connectivity traits that drive each state. Furthermore, leveraging results from dyna-LOCUS analysis of the PNC study, we find significant gender differences in the developmental changes in the temporal expression of a connectivity trait driven by the executive function network. This finding introduces new insights to previous research on developmental changes in executive function from childhood to adolescence.

The rest of the paper is organized as follows. In Section 2, the proposed model, estimation, and tuning parameter selection steps are introduced. Section 3 illustrates the strength of our method compared with other decomposition methods via simulation studies. In Section 4, we apply the proposed method to investigate dynamic connectivity for the PNC study. Finally, discussions and conclusions are presented in Section 5.

2.2 Materials and Methods

In this section, we introduce dyna-LOCUS, a regularized blind source separation method designed to decompose dynamic functional connectivity (dFC) matrices into a product of latent connectivity traits and temporal mixing matrices. It models the latent traits using a low-rank matrix factorization which is well-suited for connectivity matrices. To reduce spurious findings due to large number of edges in brain networks, we develop a novel angle-based element-wise sparsity regularization on the extracted connectivity traits. In addition, we propose a smooth regularization for the temporal mixing time series to account for the similarity in temporally adjacent dynamic connectivity matrices.

2.2.1 dyna-LOCUS model

dyna-LOCUS is applicable to decomposing dynamic FC measures derived from various methods such as sliding window method (Allen et al., 2014; Hutchison et al., 2013), jackknife correlation (Thompson et al., 2017a,b), temporal derivatives (Shine et al., 2015), wavelet coherence (Betzel et al., 2017), dynamic conditional correlation (Lindquist et al., 2014), and more. These methods generate connectivity matrices representing connections between different brain regions at a sequence of time points. For illustration purposes, we present dyna-LOCUS for decomposing connectivity matrices derived from the sliding window approach. The sliding window approach is a widely adopted strategy for investigating dynamic changes in resting-state FC. This technique involves partitioning the scanning time into small windows of fixed duration. These windows can be constructed using either a rectangular window with equal weights or a tapered window that gradually reduces the weights toward the edges. Within each time window, the fMRI BOLD series is utilized to compute the dFC matrix. The window is then shifted in time by a specified number of data points, with partial overlap between consecutive windows. By sliding these windows across the entire scanning session, the method calculates a series of dFC matrices that characterize the time-varying connectivity throughout the scan. Suppose we have N subjects, and each subject has fMRI BOLD signal series from V nodes or regions of interest at n_t time points. For subject i, we slide a tapered window (Allen et al., 2014; Rashid et al., 2014) to obtain a series of $V \times V$ connectivity matrices denoted as $\boldsymbol{Y}_{i1}, \boldsymbol{Y}_{i2}, ..., \boldsymbol{Y}_{iT}$ with T being the total number of windows. $\boldsymbol{Y}_{it}(t = 1, \ldots, T)$ is the connectivity matrix based on the tth sliding window, where $\boldsymbol{Y}_{it}(u, v) \in \mathcal{R}$ represents the connection between node u and v which is obtained by a proper transformation of the brain connectivity measure. Since the connectivity matrix $\boldsymbol{Y}_{it}$ is symmetric, and the diagonal, which represents self-relationships in the network, is typically not of interest, we define a vector $\boldsymbol{y}_{it}$ based on the upper triangular elements of $\boldsymbol{Y}_{it}$, i.e. $\boldsymbol{y}_{it} = \mathcal{L}(\boldsymbol{Y}_{it})$ where $\mathcal{L}(\boldsymbol{Y}_{it}) = [\boldsymbol{Y}_{it}(1,2), \boldsymbol{Y}_{it}(1,3), ..., \boldsymbol{Y}_{it}(V-1,V))]'$. Here, $\mathcal{L} : \mathcal{R}^{V \times V} \to \mathcal{R}^p$ with $p = \frac{V \times (V-1)}{2}$.

We propose the following dyna-LOCUS model to decompose the multi-subject dynamic connectivity matrices to extract latent connectivity traits. Specifically, dyna-LOCUS separates the observed connectivity data for the ith subject at the tth time window as combinations of q latent connectivity sources/traits, that is:

$$\boldsymbol{y}_{it} = \sum_{\ell=1}^{q} a_{it\ell} \boldsymbol{s}_\ell + \boldsymbol{e}_{it}, \tag{2.1}$$

where $\boldsymbol{s}_\ell \in \mathcal{R}^p (\ell = 1, ..., q)$ is the source signal of the ℓth latent connectivity source or trait, which is assumed to be independent across the q traits. A connectivity trait represents a set of brain connections that tend to occur together. The source signal $\boldsymbol{s}_\ell$ includes the weights of each of the p brain connections in the ℓth connectivity trait. By mapping connectivity trait $\boldsymbol{s}_\ell$ back to the $V \times V$ connectivity matrix form, we can recover the spatial composition of the underlying connectivity pattern. $\{a_{it\ell}\}$ are the mixing coefficients or trait loadings. They represent the presence or prominence of the ℓ's connectivity trait in ith subject at time point t. $\boldsymbol{e}_{it} \in \mathcal{R}^p$ is an error term independent of source signals. The number of latent sources, i.e. q, can be determined using methods such as the Laplace approximation (Minka, 2000) or based on the reproducibility and interpretability of the extracted latent sources. We can also rewrite the model (3.1) in the following matrix form aggregating across time windows and subjects:

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{E}, \tag{2.2}$$

where $\boldsymbol{Y} = \left[\boldsymbol{y}_{11}, \ldots, \boldsymbol{y}_{1T}, \ldots, \boldsymbol{y}_{N1}, \ldots, \boldsymbol{y}_{NT}\right]' \in \mathcal{R}^{NT \times p}$ is the multi-subject dynamic connectivity data, $\boldsymbol{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_q]' \in \mathcal{R}^{q \times p}$ is the connectivity trait matrix, $\boldsymbol{A} = \{a_{it\ell}\} \in \mathcal{R}^{NT \times q}$ is the mixing or trait loading matrix, and $\boldsymbol{E} = \Big[\boldsymbol{e}_{11}, \ldots, \boldsymbol{e}_{1T}, \ldots, \boldsymbol{e}_{N1}, \ldots, \boldsymbol{e}_{NT}\Big]' \in \mathcal{R}^{NT \times p}$.

Motivated by the observations that brain connectivity matrices often have block or banded structures, we model connectivity traits with a low-rank structure which can efficiently capture such kind of characteristics (Zhou et al., 2013) using a much smaller number of parameters. Specifically,

$$\boldsymbol{s}_\ell = \mathcal{L}(\boldsymbol{X}_\ell \boldsymbol{D}_\ell \boldsymbol{X}_\ell') = \mathcal{L}(\sum_{r=1}^{R_\ell} d_\ell^{(r)} \boldsymbol{x}_\ell^{(r)} \boldsymbol{x}_\ell^{(r)'}), \quad \ell \in 1, 2, ..., q. \tag{2.3}$$

Here, the source signal $\boldsymbol{s}_\ell$ is modeled via a low-rank factorization where $\boldsymbol{X}_\ell = [\boldsymbol{x}_\ell^{(1)}, \ldots, \boldsymbol{x}_\ell^{(R_\ell)}] \in \mathcal{R}^{V \times R_\ell}$ with rank $R_\ell < V$ and each tensor factor $\boldsymbol{x}_\ell^{(r)}$ $(r = 1, \ldots, R_\ell)$ is a $V \times 1$ vector with unit norm, i.e. $\|\boldsymbol{x}_\ell^{(r)}\|_2 = 1$ for identifiability purpose. $\boldsymbol{D}_\ell$ is a diagonal matrix with diagonal elements $\boldsymbol{d}_\ell = (d_\ell^{(1)}, .., d_\ell^{(R_\ell)})$. The low-rank factorization implies the V nodes reside in a reduced subspace with the dimension of R_ℓ, i.e. $\boldsymbol{s}_\ell = \mathcal{L}(\sum_{r=1}^{R_\ell} d_\ell^{(r)} \boldsymbol{x}_\ell^{(r)} \boldsymbol{x}_\ell^{(r)'})$ where the rth tensor factor $\boldsymbol{x}_\ell^{(r)} \in \mathcal{R}^{V \times 1}$ represents the latent coordinates of the V nodes in the rth dimension and each row of $\boldsymbol{X}_\ell$, i.e. $\boldsymbol{x}_\ell(v) \in \mathcal{R}^{R_\ell \times 1}$, represents the latent coordinates of the vth node in the R_ℓ-dimensional latent subspace. $d_\ell^{(r)}$ reflects the contribution of the rth dimension in generating $\boldsymbol{s}_\ell$. Note that we specify trait-specific rank R_ℓ in our low-rank model (2.3) which provides the flexibility to accommodate various connectivity traits with different network properties and topological structures.

The proposed low-rank model offers several advantages for modeling the brain connectome. Firstly, by incorporating a low-rank structure, the model achieves a substantial reduction in the number of parameters required to represent the brain connections. The number of parameters decreases from a quadratic complexity of $O(V^2)$ to a linear complexity of $O(V)$. This reduction in parameter space improves computational efficiency and alleviates the risk of overfitting, particularly in scenarios where the number of brain regions (V) is large. Secondly, the low-rank factorization offers appealing neuroscience interpretations in modeling brain connectivity. In the low-rank model, a node v is characterized by its latent coordinates in the R_ℓ-dimensional subspace, i.e. $\boldsymbol{x}_\ell(v)$, which potentially reflect the node's underlying neural activity. The connection between two nodes in the ℓth trait is modelled as $\boldsymbol{S}_\ell(u, v) = \boldsymbol{x}_\ell(u)'\boldsymbol{D}_\ell\boldsymbol{x}_\ell(v)$, which is the inner product between their latent coordinates. This implies the connection between the nodes depends on the similarity between the neural activity of the nodes characterized by their latent coordinates, which aligns with the understanding of brain connectivity in neuroscience (Friston et al., 1993; Friston, 2011). Another desirable property of this model is that all the brain connections involving a node v are based on the node's latent coordinate $\boldsymbol{x}_\ell(v)$ and hence are inherently related. This appropriately accounts for the dependence structure across edges in the brain connectome, which is disregarded by many existing methods. Finally, the latent coordinate $\boldsymbol{x}_\ell(v)$ can help identify key nodes driving each connectivity trait. To this end, we propose the following node contribution index to characterize a node's contribution to a connectivity trait,

$$g_\ell(v) = ||\tilde{\boldsymbol{D}}_\ell^{\frac{1}{2}}\boldsymbol{x}_\ell(v)||_2^2, \tag{2.4}$$

where $\tilde{\boldsymbol{D}}_\ell$ is a scaling matrix based on $\boldsymbol{D}_\ell$ that scales node v's R_ℓ-dimensional latent coordinate $\boldsymbol{x}_\ell(v)$ with the dimensions' contribution in generating the source signals. The index $g_\ell(v)$ measures the scaled magnitude of the node v in the latent subspace of the connectivity trait and reflects the contribution of this node to generating the connectivity trait. The node contribution index can be used to identify key nodes driving each connectivity trait.

We present a schematic plot of the proposed dyna-LOCUS model in Figure 2.1. The model decomposes observed dFC measures to generate two key results: (1) the source signals $\{\boldsymbol{s}_\ell\}$ that can be mapped back to the brain to depict the spatial composition of the connectivity traits, and (2) the trait loadings $\{a_{it\ell}\}$ that characterize the temporal expression profiles of the connectivity traits. The information provided by dyna-LOCUS model offers valuable information that can assist in understanding the intricate dynamics and underlying organization of the brain connectome.

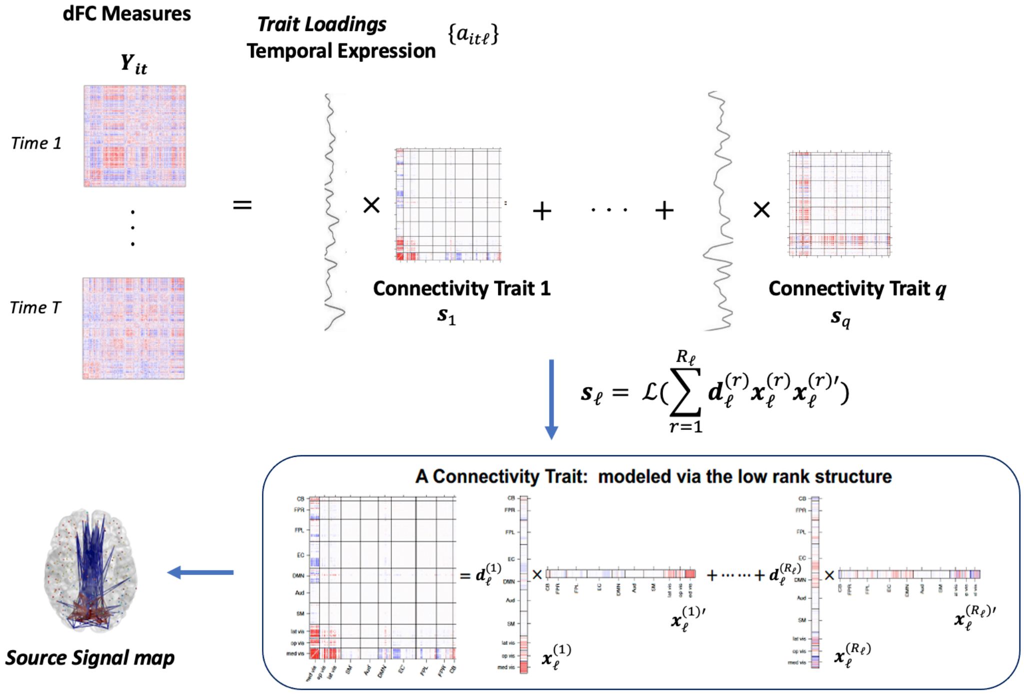


Figure 2.1: Schematic plot of dyna-LOCUS

2.2.2 Regularizations in dyna-LOCUS learning

To enhance the reliability in mapping dynamic functional connectivity traits, we integrate regularizations in dyna-LOCUS learning. These include a sparsity regularization for recovering the spatial source maps of the connectivity traits to reduce spurious connections and a temporal smoothness regularization for recovering the temporal expression profiles of the connectivity traits. The motivation for the spatial sparsity regularization derives from existing findings that functional connections in neural circuits exist only in a proportion of region pairs across the brain (Bullmore and Sporns, 2009; Lee et al., 2011; Huang et al., 2010). Given the enormous number of region pairs in a connectivity matrix, appropriate sparsity control is needed in order to obtain parsimonious results in recovering the spatial source maps of the connectivity traits. To this end, we propose a novel sparsity method that aims to achieve element-wise sparsity on the reconstructed connectivity traits based on the low-rank structure. Another observation from previous studies (Monti et al., 2017; Zhang et al., 2021) and our experiments is that the overall pattern of dynamic functional connectivity matrices generally demonstrates temporal similarity between adjacent time windows. To account for this, we include a temporal smoothness regularization on the trait loading series.

With the regularizations, the proposed dyna-LOCUS model is learned via the following optimization,

$$\min \sum_{i=1}^{N}\sum_{t=1}^{T}||\boldsymbol{y}_{it}-\sum_{\ell=1}^{q}a_{it\ell}\mathcal{L}(\boldsymbol{X}_\ell\boldsymbol{D}_\ell\boldsymbol{X}_\ell')||_2^2+\phi\sum_{\ell=1}^{q}\sum_{u<v}|\boldsymbol{x}_\ell(u)'\boldsymbol{D}_\ell\boldsymbol{x}_\ell(v)|+\lambda\sum_{i=1}^{N}\sum_{t=2}^{T}||\boldsymbol{a}_{it}-\boldsymbol{a}_{i(t-1)}||_2^2 \tag{2.5}$$

where ϕ is the tuning parameter for the sparsity control and λ is the tuning parameter for the temporal smoothness regularization, $\boldsymbol{a}_{it} \in \mathcal{R}^q$ is the tth row of the ith participant's trait loading matrix representing the ith participant's loadings on the q connectivity traits at time t. The first term in the optimization function (3.4) measures the closeness between the observed dFC and the reconstructed dFC based on the dyna-LOCUS model. The second term is the sparsity regularization aiming to achieve element-wise sparsity on the connectivity traits modeled via the low-rank structure, i.e. $\boldsymbol{S}_\ell(u,v) = \boldsymbol{x}_\ell(u)'\boldsymbol{D}_\ell\boldsymbol{x}_\ell(v)$. The sparsity penalty term in our model aims to minimize the sum of inner products of the latent coordinates for all pairs of nodes in the brain. This inner product corresponds to the angle between two nodes in the latent subspace. Consequently, our sparsity penalization can be understood as an angle-based regularization. By using an angle-based sparsity regularization, we are able to induce sparsity in the dependencies between pairs of nodes. This approach is both intuitive and theoretically sound, because our objective is to achieve sparsity in the connections between nodes. It is worth noting that in addition to the L1 norm specified in (3.4), alternative penalization functions such as SCAD (Fan and Li, 2001) can also be adopted for our sparsity regularization. The third term in (3.4) is the regularization for temporal smoothness in the mixing matrix and is based on the differences in the temporal mixing coefficients at adjacent time points.

Prior to the decomposition, several preprocessing steps, including centering, dimension reduction, and whitening, that are commonly adopted in blind source separation, are applied to the multi-subject connectivity data $\boldsymbol{Y}$. These preprocessing steps facilitate the subsequent decomposition by reducing the computational load and avoiding overfitting (Hyvärinen et al., 2001). Following the preprocessing procedures of the previous work (Beckmann and Smith, 2004; Shi and Guo, 2016; Wang and Guo, 2023), we first demean $\boldsymbol{Y}$ and then perform the dimension reduction and whitening as $\widetilde{\boldsymbol{Y}} = \boldsymbol{H}\boldsymbol{Y}$ with $\boldsymbol{H}$ being the whitening matrix. The detailed description of $\boldsymbol{H}$ is available in Appendix A. The preprocessed data $\widetilde{\boldsymbol{Y}}$ is of dimension $q \times p$ where the columns correspond to p connections in the brain. With the preprocessing, the dyna-LOCUS model in (2.2) can be rewritten on the reduced and sphered space as follows:

$$\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{A}}\boldsymbol{S} + \widetilde{\boldsymbol{E}}, \tag{2.6}$$

where $\widetilde{\boldsymbol{A}} = \boldsymbol{H}\boldsymbol{A} = \{\widetilde{a}_{i\ell}\} \in \mathcal{R}^{q\times q}$, $\widetilde{\boldsymbol{E}} = \boldsymbol{H}\boldsymbol{E}$. Note that the dimension reduction in the preprocessing is performed on the row space of $\boldsymbol{Y}$, which corresponds to the subject and time domain, and does not affect the column space of $\boldsymbol{Y}$ which corresponds to the connectivity domain. Therefore, the connectivity trait $\boldsymbol{S}$ is unaffected by the preprocessing. As in previous blind source separation methods (Hyvärinen and Oja, 2000; Beckmann and Smith, 2004), the whitening in the preprocessing leads to an orthogonal mixing matrix on the reduced space $\widetilde{\boldsymbol{A}}$ which facilitates the subsequent model estimation. With the preprocessed data and the orthogonal mixing matrix, the final optimization function for dyna-LOCUS is

$$\min_{\widetilde{\boldsymbol{A}},\{\boldsymbol{X}_\ell,\boldsymbol{D}_\ell\}} \sum_{l=1}^{q} ||\widetilde{\boldsymbol{Y}}'\widetilde{\boldsymbol{a}}_\ell - \mathcal{L}(\boldsymbol{X}_\ell \boldsymbol{D}_\ell \boldsymbol{X}_\ell')||_F^2 + \phi \sum_{\ell=1}^{q} \sum_{u<v} |\boldsymbol{x}_\ell(u)' \boldsymbol{D}_\ell \boldsymbol{x}_\ell(v)| + \lambda ||\boldsymbol{W}\widetilde{\boldsymbol{A}}||_F^2 \tag{2.7}$$

where $\widetilde{\boldsymbol{a}}_\ell$ is the ℓth column of $\widetilde{\boldsymbol{A}}$, $\boldsymbol{H}^*$ is the dewhitening matrix, and $\boldsymbol{W} = \boldsymbol{R}^*\boldsymbol{H}^*$ where $\boldsymbol{R}^*$ is a temporal contrasting matrix that measures the differences in the temporal mixing coefficients at adjacent time points. The closed forms of the whitening matrix $\boldsymbol{H}$ and the dewhitening matrix $\boldsymbol{H}^*$, as well as the derivation of the final optimization function (2.7), are provided in Appendix 2.5.1.

2.2.3 Estimation Algorithm and Tuning parameters selection

The objective function in (2.7) is non-convex but can be shown to be block multi-convex. A function is block multi-convex if there exists a partition of the set of parameters satisfying that the function is convex with respect to each of the individual arguments in the partition, while holding the others fixed. We can show the optimization function (2.7) is block multi-convex with respect to the partition of $\mathcal{P} = \{\boldsymbol{x}_1(1), .., \boldsymbol{x}_1(V), \ldots, \boldsymbol{x}_q(V), \boldsymbol{d}_1, .., \boldsymbol{d}_q, \widetilde{\boldsymbol{A}}\}$. Based on the block multi-convexity property, we derive an efficient node-rotation learning algorithm with closed-form solutions in each updating step. We initialize the algorithm with $\hat{\widetilde{\boldsymbol{A}}}^{(0)}, \{\hat{\boldsymbol{X}}_\ell^{(0)}, \hat{\boldsymbol{D}}_\ell^{(0)}\}$ derived from estimates based on existing methods such as connICA. The algorithm then estimates the parameters by iterating the following updating steps: *Step 1: Updating* $\boldsymbol{X}_\ell$. Following Wang and Guo (2023), we develop a node-rotation algorithm that updates $\boldsymbol{X}_\ell$ at one of the node v while conditioning on the rest of the nodes and then rotating across the nodes. Specifically, at the kth iteration, we update

$\hat{\boldsymbol{x}}_\ell^{(k)}(v)$, $v = 1, .., V$, conditioning on $\hat{\widetilde{\boldsymbol{A}}}^{(k-1)}$, $\hat{\boldsymbol{D}}_\ell^{(k-1)}$ and $\hat{\boldsymbol{X}}_\ell(-v)^{(k-1)}$. *Step 2: Updating* $\boldsymbol{D}_\ell$. The second step updates the diagonal estimates of the diagonal of $\boldsymbol{D}_\ell$, i.e. $\boldsymbol{d}_\ell, \ell = 1, ..., q$, conditioning on the estimates of $\hat{\boldsymbol{X}}_\ell^{(k)}$ and $\hat{\widetilde{\boldsymbol{A}}}^{(k-1)}$.*Step 3: Updating* $\widetilde{\boldsymbol{A}}$. An advantage of our proposed temporal smoothness regularization is that it admits an analytical form of solution to the temporal mixing matrix $\widetilde{\boldsymbol{A}}$. With that, we obtain the updated mixing matrix $\hat{\widetilde{\boldsymbol{A}}}^{(k)}$ based on the estimates of $\{\hat{\boldsymbol{X}_\ell}^{(k)}, \hat{\boldsymbol{D}_\ell}^{(k)}\}$. The details of the node-rotation estimation algorithm are provided in Appendix 2.5.2. Developed based on the block multi-convexity of the objective function, the node-rotation algorithm has the appealing theoretical property that updating each block of the parameters can be performed via convex optimization. Though being a highly efficient algorithm with analytic solutions, the node-rotation algorithm does involve rotating across the nodes to update the latent coordinates of each node. To further increase computation efficiency in learning dyna-LOCUS, we also develop an alternative estimation algorithm that simultaneously updates the latent coordinates of all the nodes using an eigenvalue decomposition. This alternative estimation algorithm could help reduce computation time for studies with large sample sizes and brain atlases involving a large number of nodes. Details on the alternative algorithm are presented in Section 7 of the Supplementary Material.

The optimization function in (2.7) involves three sets of tuning parameters: the rank parameters $\{R_\ell\}_{\ell=1}^q$ that control the dimension of the subspace of the connectivity traits, ϕ which regulates the influence of the sparsity penalization for the connectivity traits, and λ which regulates the influence of the temporal smoothness in the mixing matrix. In the dyna-LOCUS model, we specify trait-specific rank parameters $\{R_\ell\}_{\ell=1}^q$ in the low-rank factorization in order to accommodate the difference in the topology and structure across connectivity traits. However, selecting the appropriate rank parameter for each trait can be challenging using conventional approaches. To address this challenge, we propose an adaptive selection approach (Wang and Guo, 2023) to efficiently choose $\{R_\ell\}$ for the q latent sources. One of the main objectives of incorporating a low-rank structure is to utilize fewer parameters while effectively capturing the connectivity source signals. Therefore, our approach chooses the rank parameter R_ℓ to achieve a desired level of similarity between the estimated source signals with the low-rank structure, i.e. $\hat{\boldsymbol{s}}_\ell$, and the unconstrained source signal estimate obtained without assuming the low-rank structure, which is denoted as $\hat{\boldsymbol{s}}_\ell^*$. The unconstrained estimate $\hat{\boldsymbol{s}}_\ell^*$ can be conveniently obtained from our algorithm as an intermediate result before we project the estimates onto the reduced subspace spanned by the low-rank structure. Specifically, R_ℓ is selected to be the smallest integer value such that,

$$\|\hat{\boldsymbol{s}}_\ell - \hat{\boldsymbol{s}}_\ell^*\|_2^2 / \|\hat{\boldsymbol{s}}_\ell^*\|_2^2 \leq 1 - \rho, \tag{2.8}$$

where $\rho \in (0, 1)$ is a proportion parameter controlling the desired level of similarity between the unconstrained and low-rank structured estimates for the latent sources. Once the proportion parameter ρ is specified, the proposed approach adaptively selects the rank for each of the latent sources. The proposed method not only allows us to adaptively select an appropriate rank parameter to capture the varying patterns of each connectivity trait but also considerably simplifies the challenging task of selecting q rank parameters to only select a single parameter ρ. With the proposed adaptive selection approach, we propose to select ρ and ϕ, which are parameters related to learning the latent connectivity sources, via a BIC-type criterion,

$$\text{BIC} = -2\sum_{i=1}^{q} log(g(\widetilde{\boldsymbol{y}}_{\boldsymbol{i}}; \sum_{j=1}^{q} \hat{\widetilde{a}}_{ij}\hat{\boldsymbol{s}}_j; \hat{\widetilde{\sigma}}\boldsymbol{I}_p)) + log(N)\sum_{j=1}^{q} ||\hat{\boldsymbol{s}}_j||_0, \tag{2.9}$$

where $\widetilde{\boldsymbol{y}}_{\boldsymbol{i}}, i = 1, ..., q$ is the ith row in $\widetilde{\boldsymbol{Y}}$, $\{\hat{\widetilde{a}}_{ij}\}$ are the estimated mixing coefficients of preprocessed connectivity data, $\hat{\boldsymbol{s}}_j, j = 1, ..., \ell$, is the estimated connectivity trait, $\hat{\widetilde{\sigma}}^2 = \frac{1}{qp}\sum_{i=1}^{q} ||\widetilde{\boldsymbol{y}}_{\boldsymbol{i}} - \sum_{j=1}^{q} \hat{\widetilde{a}}_{ij}\hat{\boldsymbol{s}}_j||_2^2$, g is the pdf of a multivariate Gaussian distribution and $||.||_0$ denotes the L_0 norm which evaluate the number of non-zero elements in a vector. The BIC criterion seeks to achieve a balance by maximizing the likelihood of the observed dFC data while simultaneously minimizing the non-zero elements in the latent connectivity sources for the purpose of sparsity. The BIC criterion serves as a valuable guide in selecting the tuning the parameters ϕ and ρ. However, it is worth noting that the choice may not always be straightforward solely based on BIC in practice. Therefore, besides the BIC criterion, users can also employ supplementary selection strategies, such as specifying tuning parameters based on the desired sparsity level and the neuroscience interpretations they aim to achieve in the extracted connectivity traits. After obtaining ϕ and ρ, we select λ based on the goodness-of-fit of the model by computing the mean-squared reconstruction error, i.e. $||\boldsymbol{Y} - \hat{\boldsymbol{A}}\hat{\boldsymbol{S}}||_F^2$.

2.2.4 Reproducibility/Reliability of the extracted connectivity traits

One important criterion for evaluating the connectivity traits extracted from the imaging data is the reproducibility of the traits across replicated samples which reflects the reliability of the traits (Amico et al., 2017; Wang and Guo, 2023). To this end, we assess the reproducibility of the connectivity traits extracted by dyna-LOCUS from the PNC study across replicated bootstrap data samples using the following reliability index for blind source separation methods (Kemmer et al., 2018),

$$RI_\ell = \frac{\frac{1}{B}\sum_{b=1}^{B}\{h(\hat{\boldsymbol{s}}_\ell, \hat{\boldsymbol{s}}_\ell^{(b)})\} - \frac{1}{Bq}\sum_{b=1}^{B}\sum_{j=1}^{q}\{h(\hat{\boldsymbol{s}}_\ell, \hat{\boldsymbol{s}}_j^{(b)})\}}{1 - \frac{1}{B}\sum_{b=1}^{B}\{h(\hat{\boldsymbol{s}}_\ell, \hat{\boldsymbol{s}}_\ell^{(b)})\}}, \tag{2.10}$$

where h is a similarity measure such as the correlation coefficient or Jaccard Index, B is the total number of replicated samples (i.e. bootstrap data samples), $\hat{\boldsymbol{s}}_\ell$ is the latent sources extracted from the original data, and $\hat{\boldsymbol{s}}_\ell^{(b)}$ is the latent source estimated from the bth bootstrap data sample that is matched with $\hat{\boldsymbol{s}}_\ell$ from the original data.

Following previous work (Keeratimahat and Nichols, 2022; Wu et al., 2022), we employ a greedy matching algorithm to match the latent sources from the bootstrap sample and the original data.

The reliability index $RI_\ell (\ell = 1, \ldots, q)$, provides a scaled and chance-corrected measure to assess the reproducibility of each latent source. The index reflects the similarity of an extracted latent source from the original data and its matching estimates across the replication samples, removing by-chance similarity between the original latent source and any of the q extracted latent sources estimates. It is further scaled by its maximum possible value so that it typically ranges from 0 to 1, where RI_ℓ $= 0$ indicates the ℓth latent source is not reproducible across replication samples after correcting for by-chance similarity and RI_ℓ close to 1 indicates that the latent source is highly reproducible across replication samples. The reliability index is formulated in a similar way as the Cohen's kappa coefficient (Cohen, 1960). We can follow the kappa's guideline to interpret the reliability index. That is, we interpret $RI_\ell \leq 0$ as none reproducibility, 0.01–0.20 as slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial and 0.81–1.00 as almost perfect reproducibility. As a scaled and chance-corrected measure, RI_ℓ is comparable across different latent sources, i.e. connectivity traits, making it a desirable reliability measure for blind source separation methods (Kemmer et al., 2018; Wang and Guo, 2023).

2.2.5 Investigating brain dynamic connectome using dyna-LOCUS

In this section, we present several analysis strategies that leverage the results obtained from dyna-LOCUS to provide a detailed understanding of the complex dynamics of the brain's functional connectome. These include understanding the temporal expression variations and properties of the connectivity traits, exploring the interactions between different connectivity traits, and identifying and characterizing whole-brain dFC states.

Temporal expression of the latent connectivity traits

A valuable output from dyna-LOCUS is individual-level time-dependent loadings of each connectivity trait. These trait loadings contain information on the temporal expression of each of the latent connectivity traits in an individual's brain connectivity. We explore the dynamics of latent connectivity traits by examining two key aspects: the energy and variation of their temporal expression. Additionally, we investigate whether and how the connectivity traits synchronize with each other by studying the associations between their temporal loadings.

To identify highly expressed connectivity traits, we quantify the overall dynamic activation across the time windows of a connectivity trait expression by evaluating the energy of the temporal loading of the trait over the scanning session for each individual. Specifically, for individual i, the energy of connectivity trait ℓ is defined as $\sum_{t=1}^{T} a_{it\ell}^2$. Furthermore, to assess how stable or transient a connectivity trait is, we quantify the variation of the temporal expression of a connectivity trait by evaluating the mean of the absolute relative change of trait loadings throughout the scanning session. In particular, for individual i, the variation of connectivity trait ℓ is defined as $\sum_{t=1}^{T-1} \frac{1}{T-1} |\frac{a_{i(t+1)\ell} - a_{it\ell}}{a_{it\ell}}|$.

To investigate the interaction and synchronization between connectivity traits, we propose to evaluate the cross-correlation function (CCF) between trait loading time series. Denote $\{x_t\}, \{y_t\}, (t = 1, \ldots, T)$ as the trait loading time series of two connectivity traits of an individual. To facilitate notation, we order them by denoting the trait with higher reproducibility, i,e. higher RI index, as x_t and the trait with lower reproducibility as y_t. For a pair of traits, the CCF with a lag of k is $\text{CCF}_k(x_t, y_t) = \frac{\frac{1}{n}\sum_{t=1}^{n-k}(x_{t+k} - \bar{x})(y_t - \bar{y})}{SD_x SD_y}$, and CCF with a lag of $-k$ is $\text{CCF}_{-k}(x_t, y_t) = \frac{\frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(y_{t+k} - \bar{y})}{SD_x SD_y}$, where $\bar{x}$ and $\bar{y}$ are the mean value of $\{x_t\}$ and $\{y_t\}$ respectively, and SD_x and SD_y are their standard deviations. A large magnitude of CCF between a pair of connectivity traits indicates synchronized temporal expressions either in the same or opposite direction. For each pair of connectivity traits, we identify the lag on the population-level when they achieve the highest level of synchronization by taking the mode of the lags when the highest synchronization is achieved among individuals.

Identifying whole-brain dFC States

Another major contribution of dyna-LOCUS is that it provides a highly efficient and reliable approach for the identification and characterization of whole-brain dynamic connectivity states. In current neuroimaging studies, whole-brain dFC states are often obtained by clustering the observed dFC matrices (Allen et al., 2014; Damaraju et al., 2014; De Lacy et al., 2017; Geng et al., 2020). Using dyna-LOCUS, we propose a new approach by conducting the analysis on subjects' trait loadings, which are low dimensional representations of the observed dFC matrices, and then obtain the whole-brain dFC states via reconstruction. This method not only dramatically improves the computational efficiency but also generates dFC states that are sparser, more reliable, and easier to interpret.

We illustrate our new pipeline for investigating whole-brain dFC states in Figure 2.2. First, we apply dyna-LOCUS to decompose the observed dFC matrices to extract connectivity traits and individual-level trait loading time series. Here, connectivity traits can be viewed as a set of basis matrices for observed dFC data, and the trait loadings are low-dimensional representations of the observed dFC matrices obtained by projecting them onto the basis. Therefore, rather than clustering the dFC matrices to identify dFC states, we can cluster the low-dimensional trait loadings to identify cluster centroids, i.e. cluster medians. The whole-brain dFC states can then be reconstructed by multiplying the centroid trait loadings with the connectivity trait basis matrices.

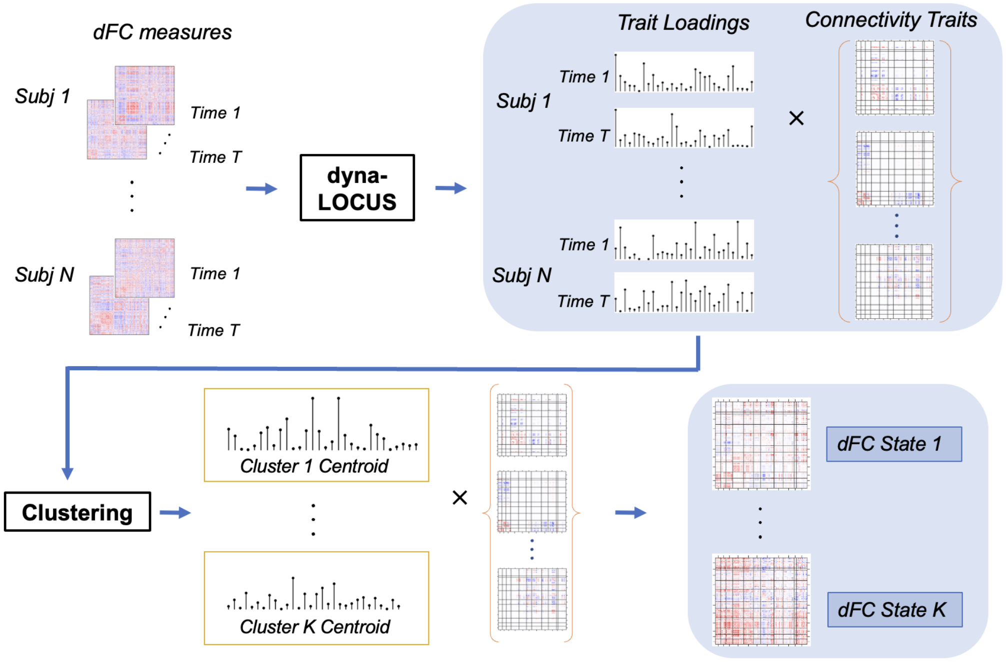


Figure 2.2: Schematic plot of the pipeline for identifying whole-brain dFC states based on dyna-LOCUS

2.3 Simulation

In this section, we investigate the performance of our model based on simulation studies. We compare the performance of dyna-LOCUS with two other source separation methods: connICA (Amico et al., 2017), which is a recently developed connectivity ICA method, and the dictionary learning (DL) method (Mairal et al., 2009), which is a popular sparse decomposition method. DL minimizes the L_1 norm of $\boldsymbol{s}_\ell$, and aims to achieve sparse estimates for the connectivity traits.

We specify $V = 50, q = 8$, consider two sample sizes $N = 20, 50$ and $T = 36$ windows. We generate the eight latent connectivity source signals based on specific connectivity patterns that we observe from the connectivity traits extracted from real imaging data including the PNC study. The mixing coefficients are also sampled from estimates from real imaging data. Furthermore, we add zero mean Gaussian noises to the mixture of signals where the variance is specified based on the signal-to-noise ratio observed from real data. Specifically, we consider three variance settings with

Term	N	Var.	**dyna-LOCUS**	connICA	DL
Latent Source Corr. (SD)	20	Low	0.955 (0.005)	0.805 (0.018)	0.947 (0.000)
		Mid	0.943 (0.005)	0.749 (0.014)	0.882 (0.017)
		High	0.828 (0.015)	0.626 (0.019)	0.772 (0.016)
	50	Low	0.956 (0.004)	0.784 (0.005)	0.955 (0.000)
		Mid	0.955 (0.005)	0.763 (0.007)	0.927 (0.001)
		High	0.932 (0.005)	0.716 (0.004)	0.884 (0.002)
Loading Matrix Corr. (SD)	20	Low	0.988 (0.002)	0.801 (0.011)	0.962 (0.000)
		Mid	0.950 (0.002)	0.771 (0.010)	0.930 (0.012)
		High	0.839 (0.011)	0.684 (0.018)	0.848 (0.020)
	50	Low	0.989 (0.001)	0.790 (0.005)	0.960 (0.000)
		Mid	0.956 (0.002)	0.764 (0.007)	0.928 (0.001)
		High	0.885 (0.003)	0.710 (0.004)	0.870 (0.002)

Table 2.1: Simulation results for comparing dyna-LOCUS and the existing connICA and DL methods based on 100 simulation runs conducted under three variance (Var.) settings. Values presented are mean and standard deviation of correlations between the true and estimated latent sources and loading/mixing matrices.

$\sigma^2 = 0.5^2, 1.5^2$, and 2.5^2, corresponding to low, medium, and high variance levels, respectively. In summary, we have 2×3 simulation settings with combinations of sample sizes and variance levels. For each setting, we generate 100 simulation runs to capture the variations in performance. Based on BIC and the goodness-of-fit criteria, we selected the following tuning parameters: $\rho = 0.95$, $\lambda = 0.01$, and $\phi = 1.5$.

Following previous work (Beckmann et al., 2005; Wang and Guo, 2019, 2023), we evaluate the performance of each method based on the correlations between the truth and the model-based estimates on the source signals and mixing coefficients. We further examine the standard deviation of the correlations across 100 simulation runs to evaluate the robustness of the methods.

Results are summarized in Table 2.1 and Figure 2.3. Table 2.1 shows dyna-LOCUS consistently demonstrates better accuracy in recovering the latent sources and mixing coefficients as compared with connICA and DL. The standard deviation of dyna-LOCUS is generally lower than the other two methods, indicating our proposed method has better stability. Figure 2.3 shows that dyna-LOCUS demonstrates considerably better performance in recovering the spatial compositions of the connectivity traits. Compared with the results by the two existing methods, the source signal maps by dyna-LOCUS are more accurate, sparser with much fewer false positive findings, and show little crossing talking problems. Specifically, connICA, being a decomposition method without sparsity constraints and the low-rank structure, tends to yield noisy and inaccurate estimates. As a sparse decomposition method, DL minimizes the L_1 norm of $\boldsymbol{s}_\ell$ to achieve sparse estimates for the source signals. However, it doesn't model the source signals using the low-rank structure disregards the interdependence among brain connections. Instead, it treats connections as independent parameters, leading to a large number of parameters for DL to learn. As a result, DL estimations may lack accuracy compared to the proposed dyna-LOCUS approach, as evidenced by simulation results.

In addition to the simulation results presented in Table 2.1 and Figure 2.3, we further consider two additional scenarios featuring decreasing lower levels of sparsity and an increased number of connections in the true source signals. These additional simulations aim to evaluate the performance of dyna-LOCUS across varying degrees of source signal sparsity. Across the different sparsity levels, dyna-LOCUS consistently exhibits superior accuracy in recovering the underlying source signals and their respective temporal loadings compared to the other two methods. Please refer to Section 4 of the Supplementary Material for the additional simulation studies.

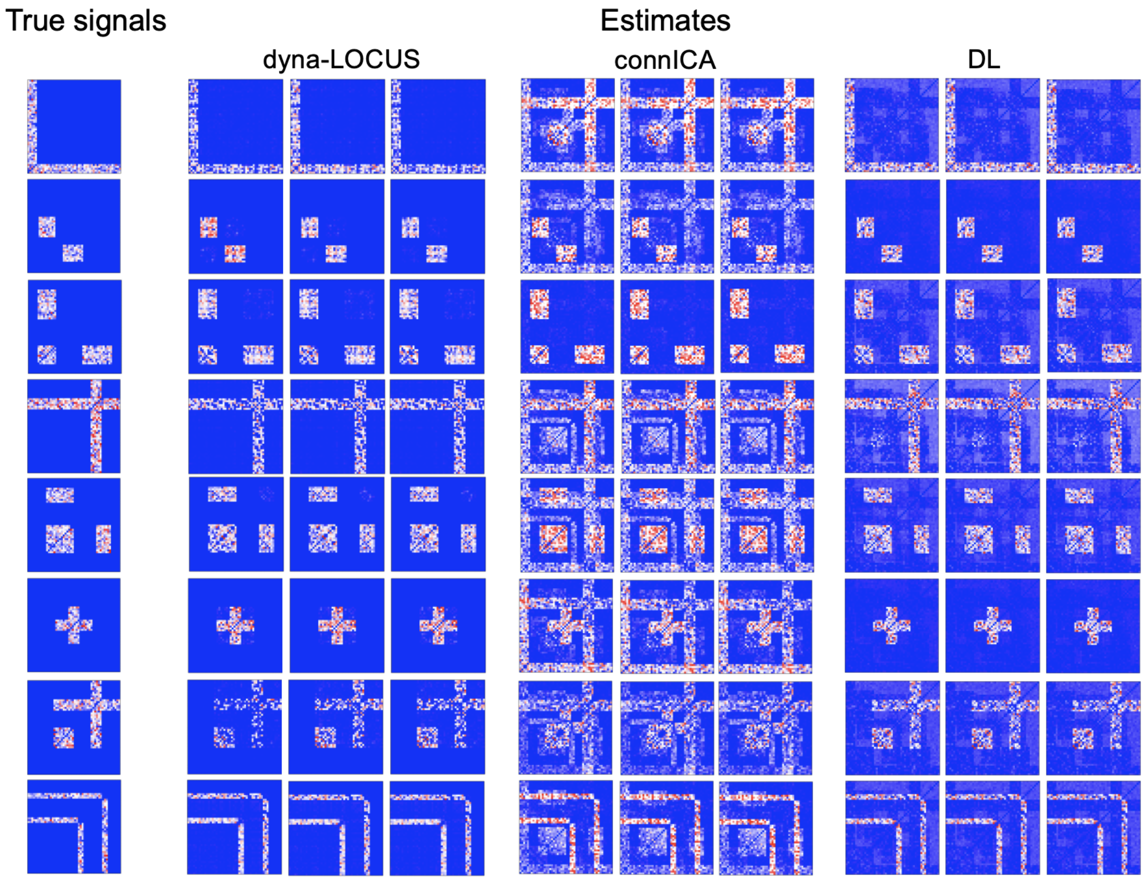


Figure 2.3: Results from the simulation study. The figures illustrate the true source signals and the estimated signals by dyna-LOCUS, connICA, and dictionary learning(DL) in three randomly selected simulation runs conducted under the low level variance setting

.

2.4 Investigating dynamic functional connectome for the Philadelphia Neurodevelopmental Cohort (PNC) study

2.4.1 Data acquisition and preprocessing

We apply dyna-LOCUS to analyze the resting state fMRI (rs-fMRI) data from the Philadelphia Neurodevelopmental Cohort (PNC) project (Satterthwaite et al., 2016, 2014). Children and adolescents between 8–21 at enrollment underwent multimodal neuroimaging which included rs-fMRI. Images were acquired using a single 3T Siemens TIM Trio whole-body scanner. Resting-state fMRI scans were acquired on a single-shot, interleaved multi-slice, echo planar imaging (GE-EPI) sequence. The nominal voxel size is $3 \times 3 \times 3$mm with full brain coverage achieved with parameters of TR/TE=3000/32ms, flip=90, and FOV=200x220 mm. Detailed descriptions of the inclusion and exclusion criteria, and the settings for the scanning session can be found in Satterthwaite et al. (2014)

Prior to analysis, we perform quality control procedures on the rs-fMRI. We remove subjects who had more than 20 volumes with relative displacement > 0.25mm to avoid images with excessive motion (Satterthwaite et al., 2015; Wang et al., 2016). 514 participants' rs-fMRI data meet the quality criterion and are used in our analysis. Among these subjects, 289 (56%) are female and the mean age is 15.28 years (SD = 3.11). For rs-fMRI preprocessing, skull stripping is performed on the T1 images to remove extra-cranial material. The first four volumes of the functional time series are removed to stabilize the signal, leaving 120 volumes for subsequent preprocessing. The anatomical image is registered to the 8th volume of the functional image and subsequently spatially normalized to the MNI standard brain space. The normalization parameters from MNI space are used for the functional images, which are smoothed with a 6 mm FWHM Gaussian kernel. Motion corrections are applied on the functional images. A validated confound regression procedure (Satterthwaite et al., 2015) is performed on each subject's time series data to remove confounding factors including motions, global effects, white matter (WM), and cerebrospinal fluid (CSF) nuisance signals. Furthermore, motion-related spike regressors are included to bind the observed displacement. Lastly, the functional time series data are band-pass filtered to retain frequencies between 0.01 and 0.1 Hz which is the relevant frequency range for rs-fMRI.

2.4.2 Dynamic connectivity analysis using dyna-LOCUS

In this study, we adopt Power's 264-node brain parcellation system (Power et al., 2011) for connectivity analysis. Each node in this node system is a 10 mm diameter sphere in the standard MNI space representing a putative functional area, and the collection of nodes provides good coverage of the whole brain. To facilitate interpretation, we assign the nodes to the functional networks of Smith's major resting-state network system (Smith et al., 2009). Specifically, the nodes are assigned to medial visual network ("Med Vis"), occipital pole visual network ("OP Vis"), lateral visual network ("LAT Vis"), default mode network ("DMN"), cerebellum ("CB"), sensorimotor network ("SM"), auditory network ("Aud"), executive control network ("EC"), and right and left frontoparietal networks ("FPR" and "FPL"). For the nodes whose assignments is uncertain in Smith's system, we interpret them using Power's resting-state network labels (Power et al., 2011).

For each subject, dynamic connectivity is assessed with the commonly used sliding window approach. We use a tapered window, created by convolving a rectangle (width = 15TRs = 45s) with a Gaussian kernel ($\sigma = 3$TRs) and slide in steps of 1TR, resulting in 106 windows. Our specification of the sliding window length of 45 seconds is based on the findings in the existing literature which recommend selecting a window length within the range of 30-60 seconds for fMRI (Hutchison et al., 2013; Shirer et al., 2012). This range of window length is shown to produce robust dynamic connectivity results by striking a balance between ensuring a sufficient number of time points in a window for reliable connectivity estimation and having an adequate number of windows across time to capture dynamic changes in brain connectivity. Within this recommended range, we select a window length of 45 seconds, similar to the one employed in numerous fMRI studies (Allen et al., 2014; Yang et al., 2014; Marusak et al., 2017). We extract the fMRI time series from each node and obtain 264×264 dynamic connectivity matrices for each subject by evaluating the pair-wise correlations between the node-specific fMRI series in each sliding window. Fisher's Z transformation is applied to the correlations to obtain the dynamic connectivity data for decomposition.

We apply dyna-LOCUS to decompose the dynamic connectivity data to reveal underlying connectivity traits. The choice of the number of latent sources q has been an open topic in blind source separation research. Typically, the number of sources is selected based on a specific objective function or the interpretability of the extracted latent sources. In our study, we select q based on the reproducibility and interpretability of the extracted sources. We evaluate the reproducibility of the extracted latent sources for a range of q values (Figure 4). The reproducibility of the extracted sources initially increases with q and then starts to level off at around $q = 30$. Although the reproducibility still rises slightly with $q > 30$, the interpretability of the extracted sources becomes less ideal when q becomes too large. Therefore, we choose $q = 30$ which achieves a good balance between the model size, reproducibility, and interpretability. Using the proposed tuning parameter selection method, we choose $\rho = 0.95$, $\phi = 2$, and $\lambda = \exp(-2)$ for the dyna-LOCUS optimization.

Regarding the computation time of the analysis, the most computationally intensive phase is the preprocessing step prior to the dyna-LOCUS decomposition, which involves dimension reduction and whitening of the dFC data through singular eigenvalue decomposition (SVD) of dFC across all subjects. The preprocessing of dFC data from the 514 subjects in the PNC study was conducted on the Emory Rollins School of Public Health High-Performance Computing (HPC) cluster, using a single compute node with 32 cores and 256GB of RAM. This process takes approximately 4 hours to complete. It's worth noting that the preprocessing step is a one-time procedure that only needs to be carried out once for a given dataset. Subsequently, the dyna-LOCUS decomposition of the preprocessed dFC data from the PNC study can be readily executed on a personal laptop. With a specific tuning parameter setting, the dyna-LOCUS decomposition of the PNC data took approximately 1 hour on a MacBook laptop equipped with an Apple M2 Chip and 16 GB of memory. For selecting tuning parameters, one can execute dyna-LOCUS decomposition in parallel for multiple parameter settings, requiring a computation time similar to that of a single decomposition.

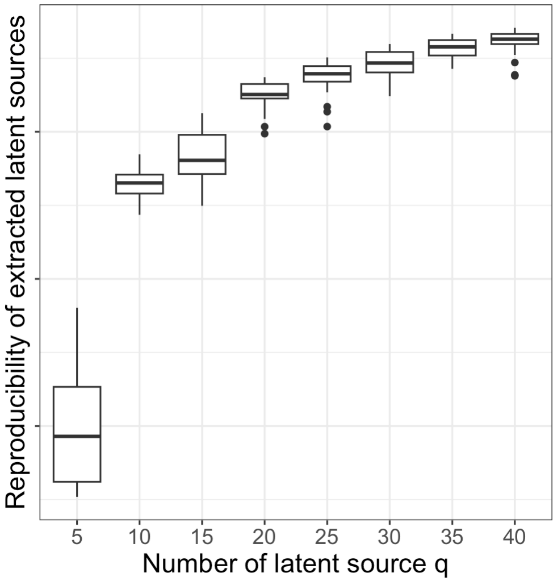


Figure 2.4: Boxplot of the reproducibility of the extracted latent sources from dyna-LOCUS under different choices of the number of latent sources q. The overall reliability across latent sources is close to 0.95 when q equals 30.

2.4.3 Results

Spatial composition of the latent connectivity traits

dyna-LOCUS uncovers 30 dynamic latent connectivity traits as well as the corresponding subject-level temporal trait loadings for each connectivity trait. The mean (SD) of the rank parameter R_ℓ across the latent traits is 4.2 (1.3), with the minimum, median, and maximum values being 2, 4, and 7, respectively. We first present source signal maps that reveal the key brain connections and brain nodes contributing to the connectivity traits. We label the traits in the order of their reproducibility based on the reliability index. Across these 30 traits, the reliability index ranges between 0.31

to 0.94, indicating all traits have at least fair reproducibility. In specific, 3 traits have fair reproducibility, 9 traits have moderate reproducibility, 11 traits have substantial reproducibility, and 7 traits have almost perfect reproducibility. We present the source signal maps for the most reproducible connectivity traits in Figure 3.1 (Part I and II). The results for all 30 extracted connectivity traits are presented in Figure S1 of the Supplementary Material. For a comprehensive analysis, we also visualize the 30 dynamic connectivity traits uncovered by connICA and DL in Section 2 of the Supplementary Material. To assess the effect of the window length selection on the results, we conduct a sensitivity analysis by considering alternative window lengths of 30 seconds (10TR) as well as 60 seconds (20TR). We calculate the correlations between the connectivity traits derived from our selected window size of 45 seconds and their matched traits obtained from the alternative window sizes. The median correlation is 0.75 for comparison with the window length of 30 seconds and 0.89 for the window length of 60 seconds, demonstrating a reasonable consistency in findings across varying window lengths.

Figure 3.1 presents the 12 most reproducible connectivity traits with a reliability index greater than 0.7. In the figure, the top 0.5% brain connections with the highest magnitude of source signal intensity in each of the connectivity traits are mapped onto the brain. Node contribution indices that help identify key brain nodes and networks that drive each connectivity trait are also shown in boxplots arranged by networks. An interesting discovery is that 8 out of the 12 most reproducible traits involve visual networks. This finding aligns well with earlier research (Zuo et al., 2010; Kang et al., 2011), which showed connections involving visual networks are highly reproducible and consistently observed over time across individuals. Results from dyna-LOCUS provide new insights and comprehensive information about the specific neural circuits involving the visual networks that exhibit a remarkably high level of reproducibility in brain connectome, shedding new light on their functional significance and potential

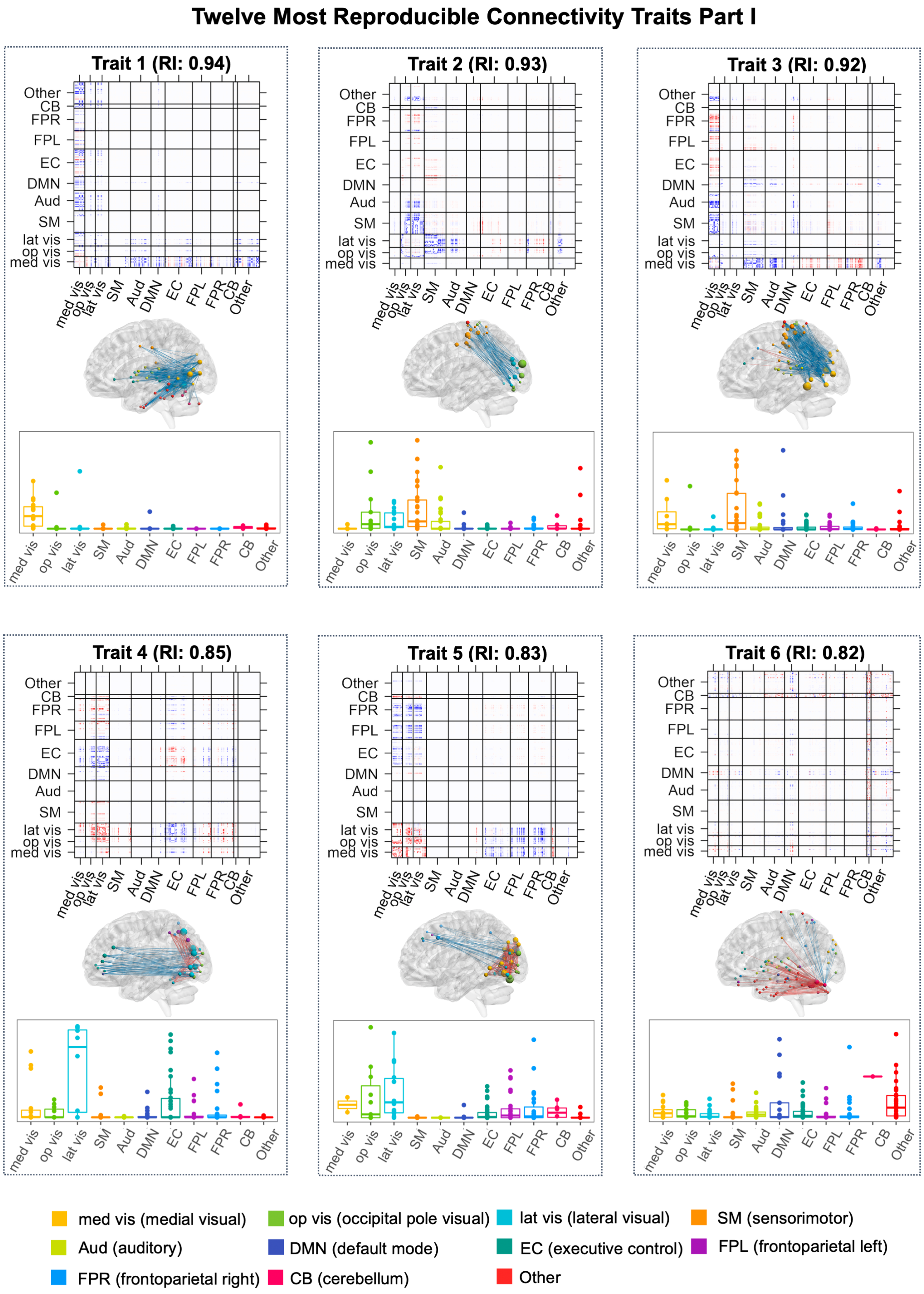


Figure 2.5: Part I: Twelve most reproducible connectivity traits extracted from the PNC study with reliability index greater than 0.7. The top 0.5% brain connections and significantly expressed nodes based on node contribution index are displayed in the brain maps. Node contribution index that help identify key brain nodes and networks that drive each connectivity trait are shown in the boxplot arranged by network.

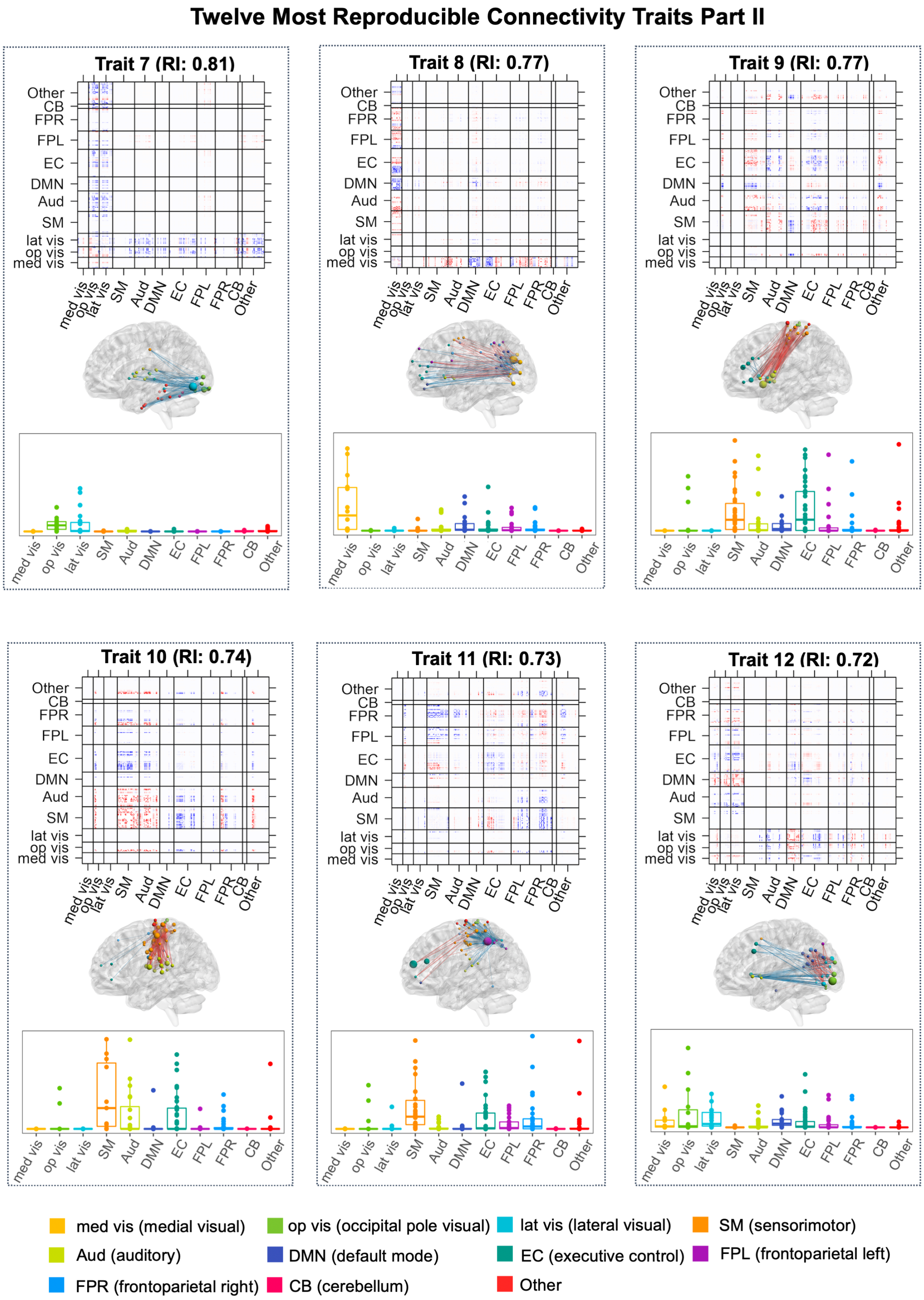


Figure 2.5: Part II: Twelve most reproducible connectivity traits extracted from the PNC study with reliability index greater than 0.7. The top 0.5% brain connections and significantly expressed nodes based on node contribution index are displayed in the brain maps. Node contribution index that help identify key brain nodes and networks that drive each connectivity trait are shown in the boxplot arranged by network.

implications. In specific, Trait 1 (Med vis-Aud-DMN-EC-CB), Trait 3 (Med vis-SM-Aud-FPR), and Trait 8 (Med vis-DMN-EC-FPL-FPR) are connectivity traits driven by the medial visual (Med vis) network including connections between Med Vis and other brain regions. Trait 2 (Vis-SM-Aud) Trait 4 (Vis-EC) and Trait 7 (Vis-Aud-DMN-EC) mainly feature connectivity between the other two visual networks, i.e. Op Vis/Lat Vis, and other brain regions. Specifically, Trait 2 (Vis-SM-Aud) features connectivity between Op Vis/Lat Vis and the sensory-motor and auditory networks. Trait 4 (Vis-EC) mainly consists of connections within and between the Op Vis and Lat Vis, connections within the executive control network, and connections between Op Vis/Lat Vis and the executive control networks. Trait 7 (Vis-Aud-DMN-EC) mainly consists of connections between nodes in Op Vis/Lat Vis and nodes in auditory, DMN, and executive control networks. Trait 5 (Vis-FPR-FPL) mainly consists of connections within the three visual networks and connections between the visual networks and other brain regions, i.e., executive control, left and right frontopartial, and cerebellum networks. Trait 12 (Vis-DMN-EC) involves connectivity between the three visual networks and cognitive networks, particularly DMN and executive control networks.

Among the 12 most reproducible traits, Trait 6 (CB) is a cerebellum-driven connectivity trait including connections within the cerebellum and also between the cerebellum and many other brain networks such as auditory, visual, DMN, executive control, etc.. Cerebellum is traditionally known for its role in movement coordination and motor learning, but recently increasing evidence suggests it may also be involved in higher-order functions, including emotional and cognitive processing (Schmahmann, 2004; Beuriat et al., 2020). Our finding provides exciting new evidence that the cerebellum has highly consistent neural connections with not only motor and sensory networks but also with higher-order emotional and cognitive networks in brain dynamic connectome. In addition, the source signal maps of dyna-LOCUS allow investigators to identify specific brain regions that have highly reliable connections with the cerebellum, which provide valuable information to advance the new understanding of this important part of the brain. For the rest of the 12 most reproducible traits, Trait 9 (SM-Aud-DMN-EC) involves connections within and between sensory-motor, auditory, DMN, and executive control networks. Trait 10 (SM-Aud-EC) is driven by the sensory-motor and auditory networks, and their connections with the executive control network. Trait 11 (FPR-FPL-SM-EC) mainly consists of the connections within the frontal partial networks and their connection with other brain networks.

Temporal expression of the latent connectivity traits

In addition to unveiling the spatial composition of connectivity traits, dyna-LOCUS also generates trait loadings that capture the temporal expression of these traits in brain dynamic connectome. These trait loadings offer new insights into the distinctive features exhibited by each connectivity trait throughout its dynamic profiles (Figure 2.6). Figure 2.6(A) plots the variation against the logarithm of energy of each connectivity trait. Our analysis reveals the presence of certain traits that exhibit distinctive characteristics in terms of one or both of these measures. We observe a diverse range of energy and variation patterns across these identified traits, encompassing distinct combinations of high-energy and high-variation, high-energy and medium-variation, medium-energy and high-variation, medium-energy and medium-variation, as well as low-energy and low-variation. These traits are denoted with different colors in Figure 2.6(A). In Figure 2.6(B), we display the trait loading series of these traits, and their source signal maps are presented in Figure 2.7. Among these traits, Trait 27 (SM-DMN-Aud) demonstrates particular high energy and high temporal variation. It mainly consists of connections between the sensorimotor network and other regions, especially the default mode network. Its temporal loading series in Figure 2.6(B) reveals that it is highly expressed during certain periods and then not expressed in others. On the other hand, Trait 6 (CB), the cerebellum-driven trait, demonstrates the lowest energy and also low variability across the scanning session, indicating it is a stable trait with a modest expression level across time. As another type of pattern, Trait 14 (EC-Aud-DMN-FPR), which includes connections within executive control and also between executive control and auditory, default mode, and frontal-parietal right networks, demonstrates high energy and only medium variation. The temporal loading series in Figure 2.6(B) reveals Trait 14 remains at a high level of expression for a good proportion of time during the scanning and is more stable in its temporal expression as compared to other high energy connectivity traits, indicating it is generally highly expressed across time. Traits 5 (Vis-FPR-FPL), which involve the frontal-parietal networks and their connections with other networks, exhibit medium energy and medium variation. It displays intermittent expression across scans, but not very frequently switching between the on and off status. Finally, Trait 8 (Med vis-DMN-EC-FPL-FPR), driven by the medial visual network, has medium-energy and high-variation. The loading pattern for this trait shows frequent switches between no expression and exhibiting expression.

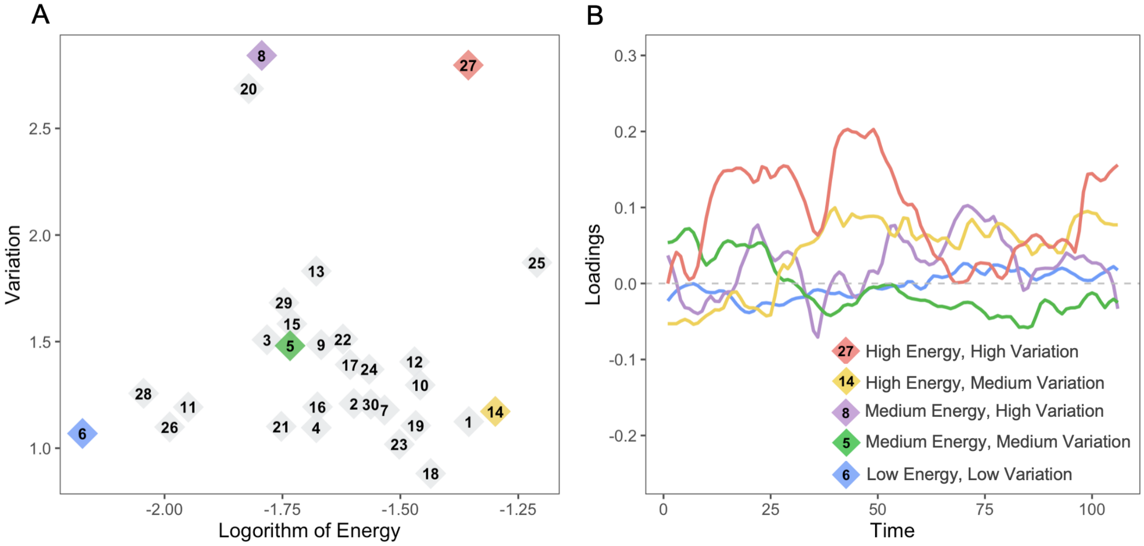


Figure 2.6: (A) Averaged variation against log(energy) of connectivity traits across subjects. Colored traits represent several types of energy and variation patterns including high-energy and high-variation(red), high-energy and medium-variation(yellow), medium-energy and high-variation(purple), medium-energy and medium-variation(green) and low-energy and low-variation(blue). (B) Temporal loading series of representative traits from sample subjects: high-energy and high-variation (Trait 27), high-energy and medium-variation signal (Trait 14), medium-energy and high-variation (Trait 8), medium-energy and medium-variation (Trait 5), and low-energy and low-variation (Trait 6).

In addition to examining the temporal expressions of each trait, we also investigate the interaction and synchronization between connectivity traits using the CCF measure. Every pair of connectivity traits exhibits their peak synchronization at a lag of either one or zero, indicating the rapid interactions and synchronization between these traits. This finding suggests that the expression of these connectivity traits are characterized by swift and efficient information exchange, facilitating seamless coordination and integration within the brain connectome. Among all pairs of connectivity traits, we identify six pairs that demonstrate a strong relationship with a median CCF magnitude greater than 0.35 across individuals. Figure 2.8 displays trait loading series of the trait pairs of example individuals to illustrate their synchronization. For example, Trait 13 (EC-FPL) and Trait 15 (Vis-SM-Aud) have a median CCF of -0.55 with a lag of zero, indicating that these two traits tend to be synchronized in the opposite

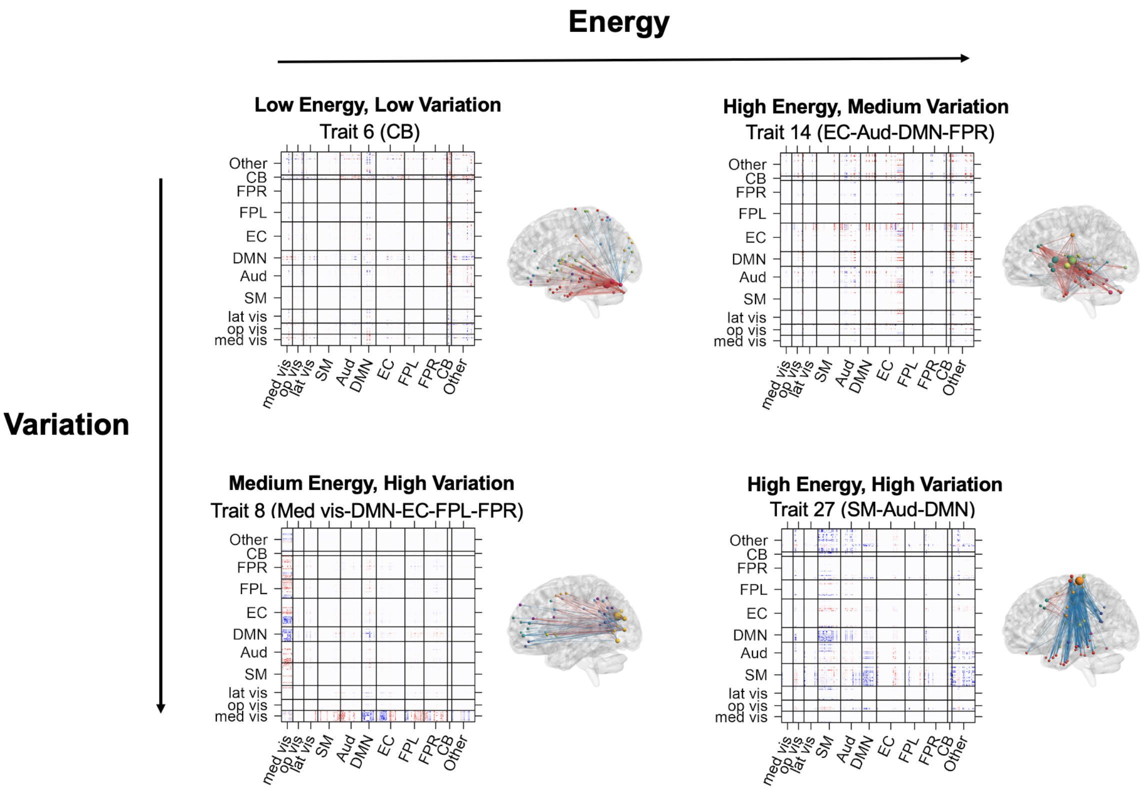


Figure 2.7: Connectivity traits with different types of energy and variation patterns. The top 0.5% brain connections and significantly expressed nodes based on the node contribution index are depicted in the brain maps.

direction, which is observed in their trait loading series in Figure 2.8. Trait 6 (CB) and Trait 14 (EC-Aud-DMN-FPR) have a median CCF of 0.40 with a lag of minus one, indicating they are synchronized in the same direction and Trait 14 is leading.

Gender and age differences in the connectivity traits

dyna-LOCUS offers an efficient and convenient method for modeling brain-behavior relationships, enabling investigation of variations in connectivity traits linked to demographics, behavior, and disease. This is achieved through analyzing associations between individual trait loadings obtained from dyna-LOCUS and an individual's demographic attributes, behavior, or clinical symptoms.

In the PNC study, we are interested in exploring the maturation of neural circuits

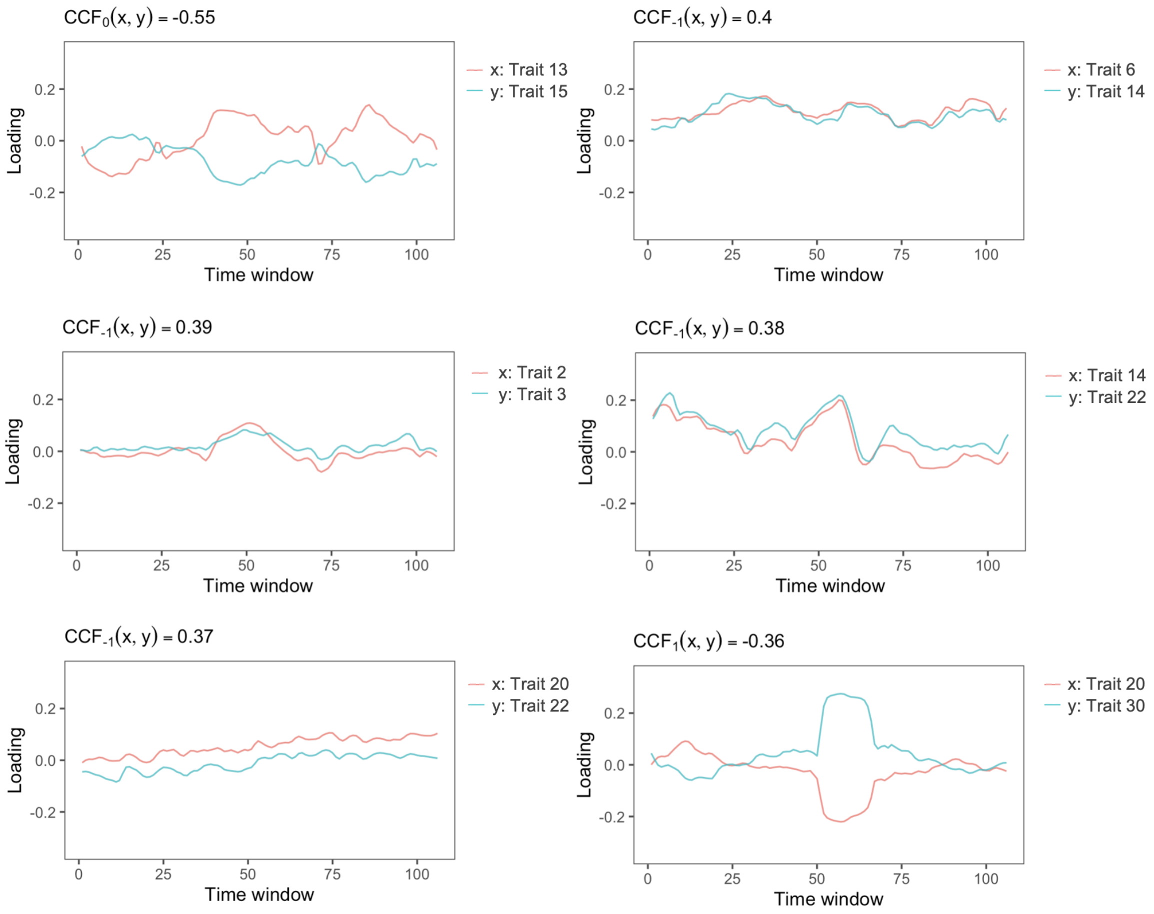


Figure 2.8: Temporal loading time series of six synchronized connectivity trait pairs from example individuals. Median CCF across individuals is displayed for each trait pair. Traits with higher reproducibility in each pair is shown in red and traits with lower reproducibility in a pair is in blue.

throughout adolescence and potential developmental differences between genders. Previous neurodevelopmental studies have reported developmental changes in the executive function and network from childhood to adolescence (Best and Miller, 2010; Chai et al., 2017). Here, we leverage results from dyna-LOCUS analysis of the PNC study to investigate age and gender differences in the connectivity traits involving executive function. To this end, we categorize participants into three age groups: middle and late childhood (ages 8 - 11), adolescence (ages 12 - 17), and early adulthood (ages 18 - 21), with 69, 291, and 143 participants in each group, respectively. We then employ regression models to investigate the relationship between the logarithm of energy of individual's trait loadings on Trait 14 (EC-Aud-DMN-FPR) (Figure 2.9), which is primarily driven by the executive function network, and the individual's gender and age group.

Trait 14 (EC-Aud-DMN-FPR) consists of connections within the executive control (EC) network and connections between the executive control network and the auditory, default mode, and frontal-parietal right networks. Our analysis of the traits' temporal expression in section 2.4.3 shows that Trait 14 (EC-Aud-DMN-FPR) is a stable and high energy connectivity trait that is often highly expressed across time (Figure 2.6), indicating the significance of this executive function related connectivity trait in the brain dynamic connectome. The association analysis between its trait loadings and individuals' age and gender reveals the energy, representing the expression level of the trait, increases with age in both gender groups (Figure 2.9). This observation suggests the presence and prominence of this particular trait become more pronounced during the process of neurodevelopment in childhood and adolescence. An interesting finding from our analysis is that males and females demonstrate distinctive developmental trajectories for this trait, with a significant gender×age interaction ($p = 0.03$). Specifically, females exhibit a notable increase in energy from late childhood to adolescence, and then a gradual upswing from adolescence to early adulthood. In contrast, males show relatively stable energy levels for Trait 14 between late childhood and adolescence, followed by a substantial increase upon entering early adulthood. This indicates the expression level of Trait 14 reaches higher levels earlier in females compared to males. To assess the reliability of this finding, we implement the data resampling validation method to evaluate the reproducibility of the result using replication samples generated through data resampling. The results provide supporting evidence for the robustness of the significant interaction effect between age and gender in the neurodevelopment of Trait 14 (Section 3 of the Supplementary Material). Notably, as young adults, males display significantly higher expression levels of Trait 14 compared to females

($p < 0.01$), indicating that Trait 14 is more prominent among young adult males. The gender difference in the expression level of Trait 14 is not significant during childhood ($p = 0.782$) and adolescence ($p = 0.113$).

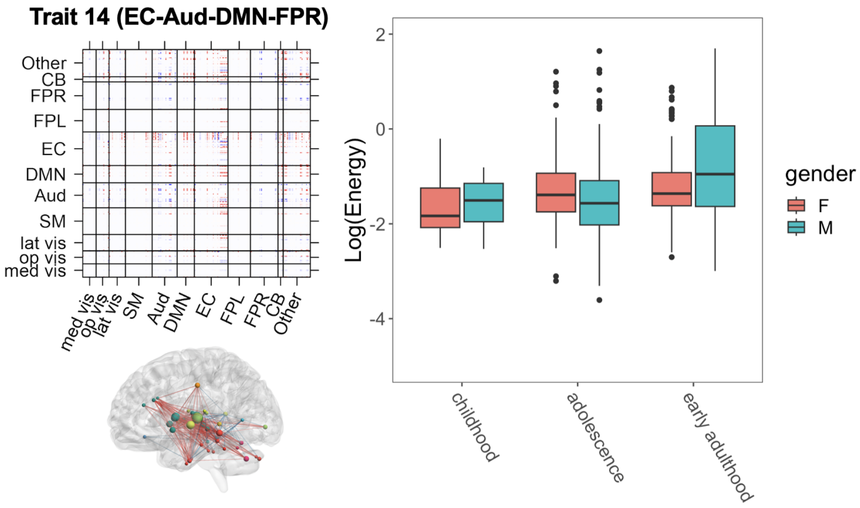


Figure 2.9: Gender differences in the development of an executive function related connectivity Trait 14 (EC-Aud-DMN-FPR). The left of the figure displays the connectivity source signals and its brain map (showing the top 0.5% brain connections). The right of the figure displays the logarithmic energy of the trait's temporal expression for males and females across three age groups: middle and late childhood (ages 8-11), adolescence (ages 12-17), and early adulthood (ages 18-21). Males and females demonstrate different developmental patterns for this trait.

Whole-brain dFC states

We implement the new dyna-LOCUS pipeline to identify whole-brain dFC states in the PNC study. After obtaining results from dyna-LOCUS, we perform clustering analysis on the trait loadings to identify brain state cluster centroids. To properly initialize the clustering analysis, we adopt an initializing strategy similar to the one in the previous dFC state analysis (Allen et al., 2014). Specifically, we randomly select 20 trait loading vector exemplars from each subject and perform clustering on the set of subject exemplars. This procedure is repeated 100 times with random initializations in order to escape local minima. The resulting centroids based on subject exemplars are then used to initialize the clustering of all trait loading data.

We determine the number of clusters by applying the elbow criterion to the ratio between the within-cluster distance and the between-cluster distance (Thorndike, 1953). After obtaining the trait loading centroids based on all data, we reconstruct whole-brain dFC states by multiplying the centroid trait loadings with the connectivity trait basis matrices derived from dyna-LOCUS. The top panel of Figure 2.10 displays the whole-brain dFC states, their percentage of occurrence, and their corresponding trait loading centroids for the PNC study. As a comparison, we derive the dFC states by directly clustering observed dFC matrices using the procedure in Allen et al. (2014). Following their paper, we randomly select 6 dFC matrices exemplars from each subject and perform clustering on all subjects exemplars. This procedure is repeated 100 times with random initializations to escape local minima. The resulting centroids are utilized to initialize the clustering of all dFC matrices. The number of clusters is chosen using the approach in Allen et al. (2014). The bottom panel of Figure 2.10 illustrates the clustering results using Allen's method.

Our findings demonstrate a notable level of concordance between the results obtained by dyna-LOCUS and Allen's method. Specifically, both methods identify seven whole-brain dFC states. We are able to readily match each of the dFC states identified by dyna-LOCUS with a corresponding dFC state identified by Allen's method, based on their connectivity patterns. Additionally, we observed a high degree of consistency in the percentage of occurrence of the matching dFC states between the two methods. The correspondence between the results obtained by dyna-LOCUS and Allen's method, which is a well-established approach for identifying dFC states, provides validation that the novel dyna-LOCUS pipeline is a reliable method for identifying whole-brain dFC states.

Our new dyna-LOCUS pipeline offers several key advantages over existing methods. Firstly, by clustering the low-dimensional trait loadings obtained from dyna-LOCUS, our method achieves a substantial improvement in computational efficiency compared to existing methods that perform clustering on the original dFC matrices. For example, for the PNC study, Allen's method takes 8 hours and 26 minutes to conduct clustering analysis on the 264×264 dFC matrices, whereas our dyna-LOCUS pipeline only uses 15.73 seconds to perform the analysis on the 30×1 trait loading vectors. Secondly, benefiting from dyna-LOCUS's sparsity regularization, our pipeline generates much sparser dFC states as compared to the existing method. Therefore, the new pipeline is able to reveal the most relevant edges in the dFC states, while effectively filtering out a large amount of less crucial connections. This results in more parsimonious and interpretable representations of the dFC states. Thirdly, our pipeline goes beyond solely identifying the dFC states; it also provides valuable information on the corresponding trait loadings. These trait loadings elucidate the contribution of the underlying connectivity traits to the whole-brain dFC states, revealing the key connectivity traits that drive each specific dFC state. Hence, researchers can gain deeper insights into the specific neural mechanisms and processes associated with different dFC states.

Comparison between the static and dynamic FC analysis

In this section, we compare the results between the static FC analysis using LOCUS (Wang and Guo, 2023) and the dynamic FC analysis using the proposed dyna-LOCUS. LOCUS assumes brain functional connectivity (FC) is stationary and extracts latent sources underlying static FC measures obtained using the whole fMRI BOLD series (Wang and Guo, 2023), ignoring changes in brain connectivity over time. In comparison, dyna-LOCUS models dynamic changes in functional connectivity and uncovers latent sources underlying the series of dynamic FC measures obtained using fMRI BOLD signals within short time windows that slide across the session.

We compare the latent connectivity sources extracted from the LOCUS and dyna-LOCUS from the PNC study. The static connectivity traits extracted from PNC study using LOCUS are presented in Section 6 of the Supplementary Material.

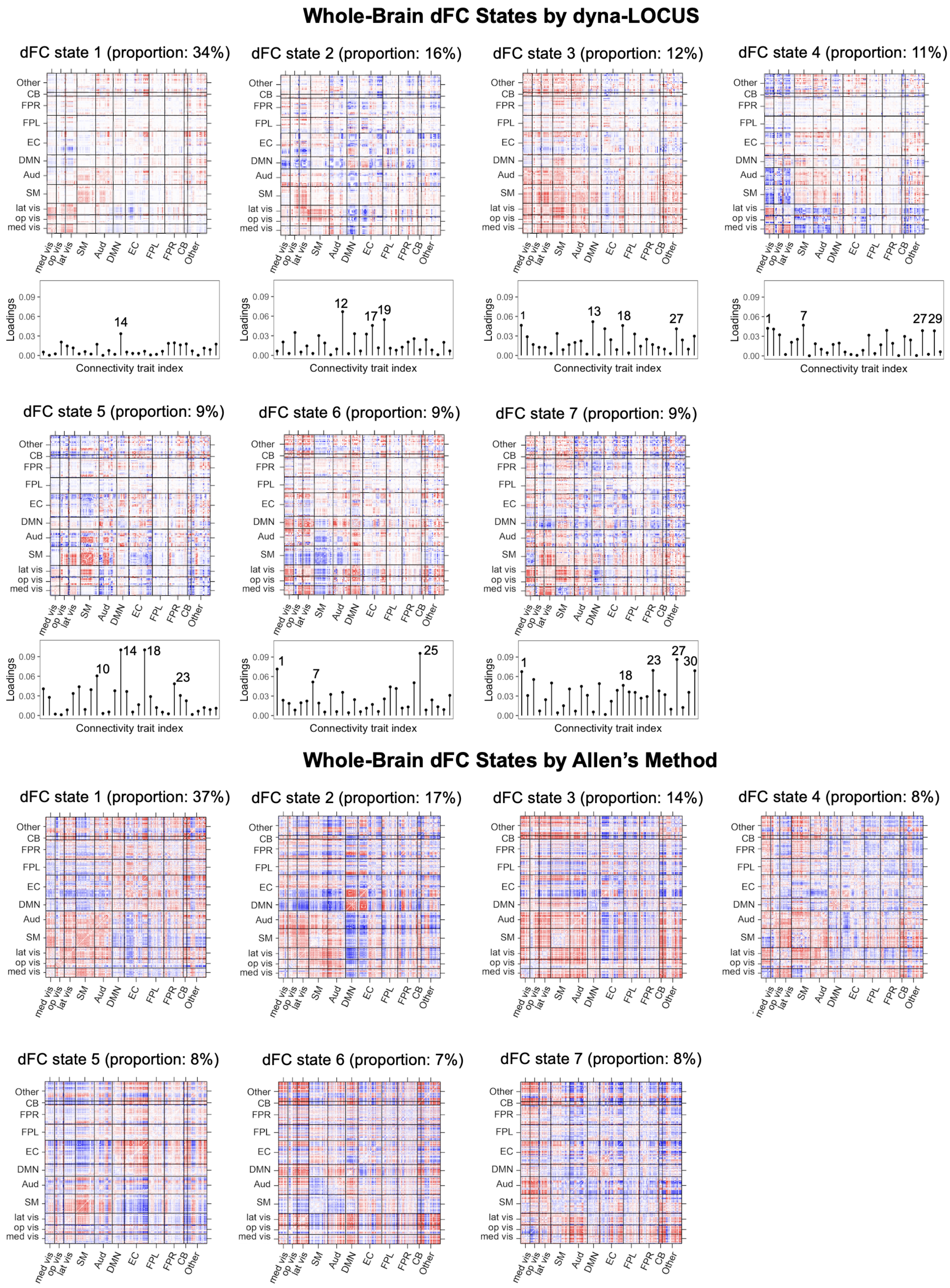
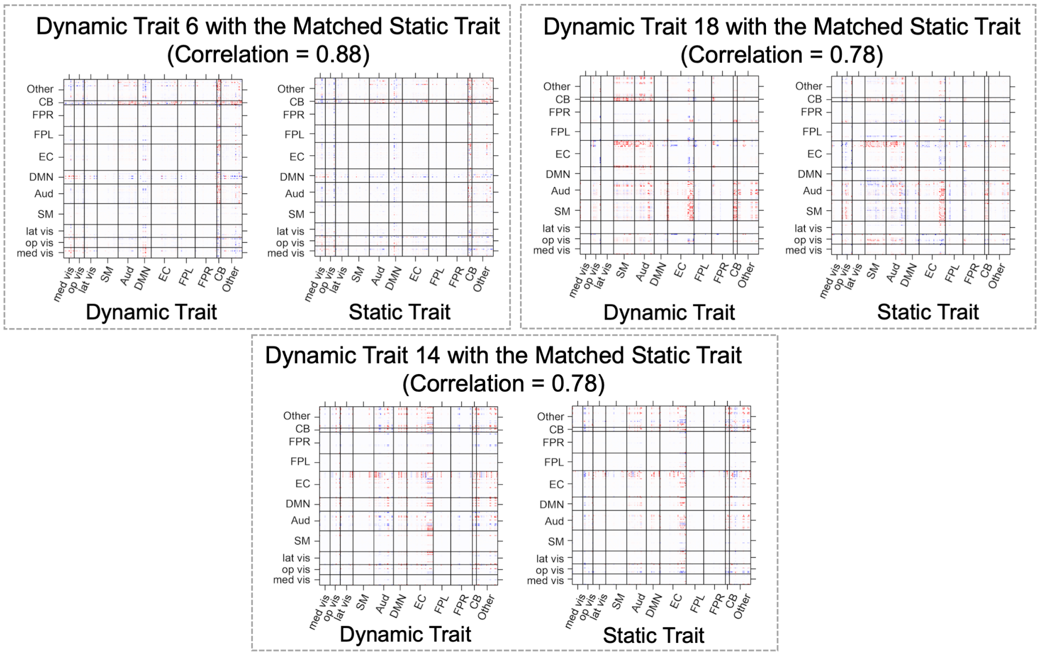
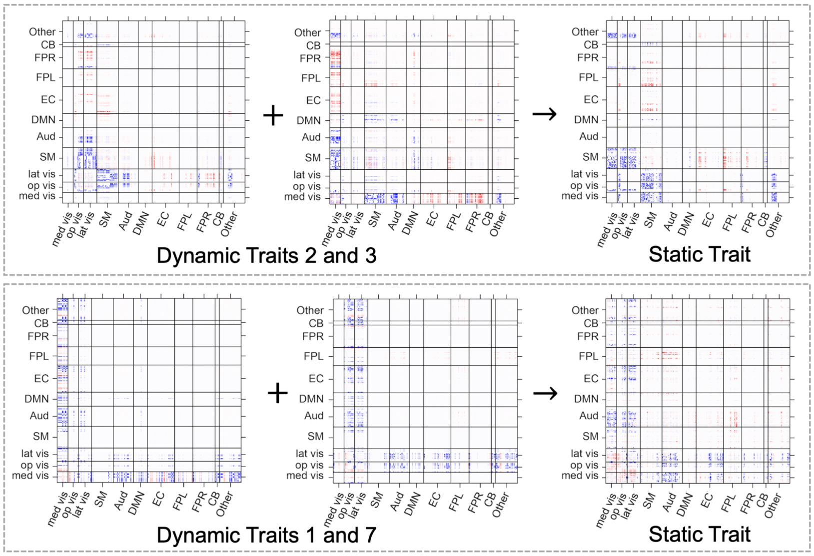


Figure 2.10: Whole-brain dynamic functional connectivity (dFC) states estimated by dyna-LOCUS and Allen's method, along with the corresponding proportions of their occurrence. dyna-LOCUS also provides the loadings of the underlying connectivity traits in each of the dFC states.

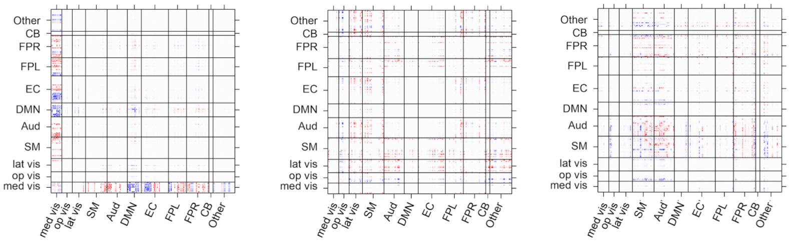
Our comparison reveals both similarities and differences in the results from the two methods. For certain dynamic latent sources extracted by dyna-LOCUS, we identify corresponding static latent sources extracted from LOCUS, which exhibit similar spatial compositions. Figure 2.11(A) illustrates three dynamic traits from dyna-LOCUS (Trait 6, 14 and 18) that have the most similar corresponding static traits from LOCUS. The temporal trait loading analysis of dyna-LOCUS (Figure 2.6) reveals that these dynamic traits display a low or medium level of variation in their temporal expression. This suggests a consistently stable presence over time in the brain connectome, elucidating why these dynamic traits closely resemble static traits identified by LOCUS. Furthermore, we notice situations where a combination of multiple dynamic traits is identified as a single static trait. For instance, the combination of dynamic Trait 2 and Trait 3 corresponds to a static trait identified by LOCUS (Figure 2.11(B)). The temporal trait loading analysis (Figure 2.8) reveals that Traits 2 and 3 exhibit strong synchronization in their temporal expression with a medium CCF of 0.39 across subjects. This synchronization explains their aggregation as a single static trait in the LOCUS analysis. This aggregation phenomenon also occurs for dynamic Trait 1 and 7 (Figure 2.11(B)). In addition to these consistent findings, dyna-LOCUS also unveils dynamic traits that are not distinctly identified among the static traits extracted by LOCUS (such as dynamic Trait 8, 15, and 29 displayed in Figure 2.11(C)). The temporal trait loading analysis of dyna-LOCUS (Figure 2.6) reveals that these dynamic traits exhibit high variation in their temporal expression. This variability suggests that their presence across time fluctuates significantly, potentially accounting for why these dynamic traits are not distinctly identified as static traits by LOCUS. The similarities and differences observed in the results from dyna-LOCUS and LOCUS highlight the potential of dyna-LOCUS to offer novel insights into brain functional connectivity beyond the static FC findings obtained by LOCUS.



(A) Dynamic traits with highly similar corresponding static traits



(B) The combination of multiple dynamic traits is identified as a single static trait



(C) Dynamic traits that are not distinctly identified among the static traits

Figure 2.11: (A) Three dynamic traits (Trait 6, 14, and 18) with highly similar static trait matches. (B) Examples of static traits representing combinations of multiple dynamic traits. (C) Examples of dynamic traits that are not clearly identified among static traits.

Ethics

This research utilized data from the PNC study. All study procedures of the PNC study were approved by the institutional review boards of the University of Pennsylvania and the Children's Hospital of Philadelphia.

Data and code availability

The PNC study data are publicly available to download from the database of Genotypes and Phenotypes (dbGaP) via Authorized Access. To request data access, investigators can login the dbGaP controlled-access portal at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000607.v3.p2 and submit a Data Access Request. The software package for dyna-LOCUS is available at `https://github.com/Emory-CBIS/dynaLOCUS`.

2.5 Appendix

2.5.1 The derivation of the final optimization function

The original objective function for dyna-LOCUS model with the sparsity and temporal smoothness regularization is,

$$\min \sum_{i=1}^{N} ||\boldsymbol{Y}_i - \boldsymbol{A}_i \boldsymbol{S}||_F^2 + \phi \sum_{\ell=1}^{q} \sum_{u<v} |\boldsymbol{x}_\ell(u)' \boldsymbol{D}_\ell \boldsymbol{x}_\ell(v)| + \lambda \sum_{i=1}^{N} \sum_{t=2}^{T} ||\boldsymbol{a}_{it} - \boldsymbol{a}_{i(t-1)}||_2^2 \quad (2.11)$$

where $\boldsymbol{S} = \begin{bmatrix} \boldsymbol{s}_1' \\ \dots \\ \boldsymbol{s}_q' \end{bmatrix} \in \mathcal{R}_{q \times p}$ and $\boldsymbol{s}_\ell = \mathcal{L}(\boldsymbol{X}_\ell \boldsymbol{D}_\ell \boldsymbol{X}_\ell')$.

Denote the loading matrix $\boldsymbol{A} = \{a_{it\ell}\} \in \mathcal{R}^{NT \times q}$ as $\boldsymbol{A} = [\boldsymbol{A}_1', \dots, \boldsymbol{A}_N']' \in \mathcal{R}^{NT \times q}$, and each $\boldsymbol{A}_i = [\boldsymbol{a}_{i1}, \dots, \boldsymbol{a}_{iT}]' \in \mathcal{R}^{T \times q}$, the temporal smoothness regularization term can be written in the matrix form:

$$\sum_{i=1}^{N}\sum_{t=2}^{T} ||\boldsymbol{a}_{it} - \boldsymbol{a}_{i(t-1)}||_2^2 = \sum_{i=1}^{N} ||\boldsymbol{R}\boldsymbol{A}_i||_F^2 \tag{2.12}$$

where $\boldsymbol{R} \in \mathcal{R}^{(T-1)\times T}$ is defined as: $\boldsymbol{R}_{ij} = 1$ if $i = j$, $\boldsymbol{R}_{ij} = -1$ if $i = j - 1$, else $\boldsymbol{R}_{ij} = 0$. Denote $\boldsymbol{R}^* = \boldsymbol{I} \bigotimes \boldsymbol{R} \in \mathcal{R}^{N(T-1)\times NT}$, where $\boldsymbol{I} \in \mathcal{R}^{N\times N}$, we can further write the temporal smoothness regularization in terms of $\boldsymbol{A}$. The original objective function can then be represented as,

$$\min \sum_{i=1}^{N}\sum_{t=1}^{T} ||\boldsymbol{y}_{it} - \sum_{\ell=1}^{q} a_{it\ell}\mathcal{L}(\boldsymbol{X}_\ell \boldsymbol{D}_\ell \boldsymbol{X}_\ell')||_2^2 + \phi \sum_{\ell=1}^{q}\sum_{u<v} |\boldsymbol{x}_\ell(u)'\boldsymbol{D}_\ell \boldsymbol{x}_\ell(v)| + \lambda ||\boldsymbol{R}^*\boldsymbol{A}||_F^2. \tag{2.13}$$

With the preprocessing, $\widetilde{\boldsymbol{Y}} = \boldsymbol{H}\boldsymbol{Y}$ with $\boldsymbol{H} = (\boldsymbol{\Lambda_q} - \widetilde{\sigma}_q^2 \boldsymbol{I})^{-1/2}\boldsymbol{U_q'}$. Here, $\boldsymbol{U_q}, \boldsymbol{\Lambda_q}$ contain the first q eigenvectors and eigenvalues based on singular value decomposition (SVD) of $\boldsymbol{Y}$. The residual variance, $\widetilde{\sigma}_q^2$, represents the variability in $\boldsymbol{Y}$ that are not explained by the extracted q latent sources and is estimated by the average of the smallest $NT - q$ eigenvalues in $\boldsymbol{\Lambda}_q$. The preprocessed data $\widetilde{\boldsymbol{Y}}$ is of dimension $q \times p$ where the columns correspond to p connections in the brain. The whitening in the preprocessing leads to an orthogonal mixing matrix on the reduced space $\widetilde{\boldsymbol{A}}$ which facilitates the subsequent model estimation. With the preprocessed data and the orthogonal mixing matrix, the final optimization function for dyna-LOCUS is

$$\min_{\widetilde{\boldsymbol{A}}, \{\boldsymbol{X}_\ell, \boldsymbol{D}_\ell\}} \sum_{l=1}^{q} ||\widetilde{\boldsymbol{Y}}'\widetilde{\boldsymbol{a}}_\ell - \mathcal{L}(\boldsymbol{X}_\ell \boldsymbol{D}_\ell \boldsymbol{X}_\ell')||_F^2 + \phi \sum_{\ell=1}^{q}\sum_{u<v} |\boldsymbol{x}_\ell(u)'\boldsymbol{D}_\ell \boldsymbol{x}_\ell(v)| + \lambda ||\boldsymbol{W}\widetilde{\boldsymbol{A}}||_F^2 \tag{2.14}$$

where $\widetilde{\boldsymbol{a}}_\ell$ is the ℓth column of $\widetilde{\boldsymbol{A}}$, and $\boldsymbol{W} = \boldsymbol{R}^*\boldsymbol{H}^*$, with $\boldsymbol{H}^* = \boldsymbol{U_q}(\boldsymbol{\Lambda_q} - \widetilde{\sigma}_q^2\boldsymbol{I})^{1/2}$ being the dewhitening matrix.

2.5.2 The Estimation Algorithm for dyna-LOCUS

Based on the block multi-convexity property, we derive an efficient algorithm with closed-form solutions in each updating step. The algorithm estimates the parameters by iterating three updating steps:

Step 1: Updating $\boldsymbol{X}_\ell$ We propose a node-rotation algorithm that updates $\boldsymbol{X}_\ell$ at a node v while conditioning on the other nodes and then rotating across the nodes. Since we have $||\widetilde{\boldsymbol{Y}}'\widetilde{\boldsymbol{a}}_\ell - \mathcal{L}(\boldsymbol{X}_\ell\boldsymbol{D}_\ell\boldsymbol{X}_\ell')||_2^2 = ||\widetilde{\boldsymbol{a}}_\ell'\widetilde{\boldsymbol{Y}} - \boldsymbol{s}_\ell'||_2^2 = \sum_{u<v}(\widetilde{\boldsymbol{a}}_\ell'\widetilde{\boldsymbol{Y}}(u,v) - \boldsymbol{x}_\ell(u)'\boldsymbol{D}_\ell\boldsymbol{x}_\ell(v))^2$, we can transform equation (3.6) to the edge-wise form.

$$\min_{\widetilde{\boldsymbol{A}},\{\boldsymbol{X}_\ell,\boldsymbol{D}_\ell\}} \sum_{\ell=1}^{q}\sum_{u<v}(\widetilde{\boldsymbol{a}}_\ell'\widetilde{\boldsymbol{Y}}(u,v) - \boldsymbol{x}_\ell(u)'\boldsymbol{D}_\ell\boldsymbol{x}_\ell(v))^2 + \phi\sum_{\ell=1}^{q}\sum_{u<v}|\boldsymbol{x}_\ell(u)'\boldsymbol{D}_\ell\boldsymbol{x}_\ell(v)| + \lambda||\boldsymbol{W}\widetilde{\boldsymbol{A}}||_F^2. \tag{2.15}$$

To update $\boldsymbol{X}_\ell$ at the kth iteration, we update the latent coordinate $\boldsymbol{x}_\ell(v)$ for node v, conditioning on $\widetilde{\boldsymbol{A}}$, $\boldsymbol{D}_\ell$ and $\boldsymbol{X}_\ell(-v)$ estimated from the $k-1$ iteration, where $\boldsymbol{X}_\ell(-v)$ is the sub-matrix of $\boldsymbol{X}_\ell$ which comprises the latent coordinates of nodes excluding v. Specifically, we update $\boldsymbol{x}_\ell(v)$ as follows,

$$\min_{\boldsymbol{x}_\ell(v)}||\widetilde{\boldsymbol{Y}}_{\{v\}}'\hat{\widetilde{\boldsymbol{a}}}_\ell - \hat{\boldsymbol{X}}_\ell(-v)\hat{\boldsymbol{D}}_\ell\boldsymbol{x}_\ell(v)||_2^2 + \phi\sum_{u=1,u\neq v}^{V}|\hat{\boldsymbol{x}}_\ell(u)'\hat{\boldsymbol{D}}_\ell\boldsymbol{x}_\ell(v)| \tag{2.16}$$

where $\widetilde{\boldsymbol{Y}}_{\{v\}}$ is a $q\times(V-1)$ sub-matrix of $\widetilde{\boldsymbol{Y}}$ which includes the subset of columns in $\widetilde{\boldsymbol{Y}}$ that correspond to connections involving node v.

Denote $\boldsymbol{b}_{\ell\{v\}} = \hat{\boldsymbol{X}}_\ell(-v)\hat{\boldsymbol{D}}_\ell\boldsymbol{x}_\ell(v) \in \mathcal{R}^{V-1}$, objective function (2.16) can be rewritten as

$$\min_{\boldsymbol{b}_{\ell\{v\}}} ||\widetilde{\boldsymbol{Y}}'_{\{v\}}\hat{\widetilde{\boldsymbol{a}}}_\ell - \boldsymbol{b}_{\ell\{v\}}||_2^2 + \phi||\boldsymbol{b}_{\ell\{v\}}||_1 \tag{2.17}$$

We can derive an analytical solution for $\boldsymbol{b}_{\ell\{v\}}$ fan2001variable and then obtain the estimate for $\boldsymbol{x}_\ell(v)$ as,

$$\hat{\boldsymbol{x}}_\ell^{(k)}(v) = \hat{\boldsymbol{D}}_\ell^{-1}(\hat{\boldsymbol{X}}_\ell(-v)'\hat{\boldsymbol{X}}_\ell(-v))^{-1}\hat{\boldsymbol{X}}_\ell(-v)'\hat{\boldsymbol{b}}_{\ell\{v\}}$$

After updating $\boldsymbol{x}_\ell(v)$, we rotate to the next node and repeat the procedure described above across all nodes to obtain an updated estimate for $\boldsymbol{X}_\ell$. An advantage of the proposed node-rotation algorithm is that it has analytic solutions and does not need gradient-based numerical approximation which makes it highly efficient and reliable.

Step 2: Updating $\boldsymbol{D}_\ell$

Our next step is to update the diagonal matrix $\boldsymbol{D}_\ell$ for $\ell = 1, ..., q$, given the estimate of $\boldsymbol{X}_\ell$ from the kth iteration and the estimate of $\widetilde{\boldsymbol{A}}$ from the $k-1$ iteration. Based on the objective function (3.6), we update the estimate of the diagonal of $\boldsymbol{D}_\ell$, i.e. $\boldsymbol{d}_\ell = \text{diag}(\boldsymbol{D}_\ell)$ via the following,

$$\min_{\boldsymbol{d}_\ell} ||\widetilde{\boldsymbol{Y}}'\hat{\widetilde{\boldsymbol{a}}}_\ell - \hat{\boldsymbol{Z}}_\ell\boldsymbol{d}_\ell||_2^2 + \phi||\hat{\boldsymbol{Z}}_\ell\boldsymbol{d}_\ell||_1, \tag{2.18}$$

where $\hat{\boldsymbol{Z}}_\ell \in \mathcal{R}^{p\times R_\ell}$ with the rth column of $\hat{\boldsymbol{Z}}_\ell$ being $\mathcal{L}(\hat{\boldsymbol{x}}_\ell^{(r)}\hat{\boldsymbol{x}}_\ell^{(r)'})$ $(r = 1, \dots, R_\ell)$. An analytical solution for $\boldsymbol{d}_\ell$ can be derived.

Step 3: Updating $\widetilde{\boldsymbol{A}}$

In this step, we update the mixing matrix $\widetilde{\boldsymbol{A}}$ given the estimates of $\{\boldsymbol{X}_\ell, \boldsymbol{D}_\ell\}$ from the kth iteration. The updated estimate for $\widetilde{\boldsymbol{A}}$ is derived as follows,

$$\min_{\widetilde{\boldsymbol{A}}} f = ||\widetilde{\boldsymbol{Y}} - \widetilde{\boldsymbol{A}}\hat{\boldsymbol{S}}||_F^2 + \lambda||\boldsymbol{W}\widetilde{\boldsymbol{A}}||_F^2, \tag{2.19}$$

where $\hat{\boldsymbol{S}}$ is based on the estimates of $\boldsymbol{X}_\ell$ and $\boldsymbol{D}_\ell$ from the kth iteration. By setting $\frac{\partial f}{\partial \widetilde{\boldsymbol{A}}} = 0$, we have

$$\widetilde{\boldsymbol{A}}\hat{\boldsymbol{S}}\hat{\boldsymbol{S}}' + \lambda\boldsymbol{W}'\boldsymbol{W}\widetilde{\boldsymbol{A}} = \widetilde{\boldsymbol{Y}}\hat{\boldsymbol{S}}' \tag{2.20}$$

We denote the eigen-decomposition of $\boldsymbol{S}\boldsymbol{S}'$ and $\boldsymbol{W}'\boldsymbol{W}$ as $\boldsymbol{Q}_1\boldsymbol{\Lambda}_1\boldsymbol{Q}_1'$ and $\boldsymbol{Q}_2\boldsymbol{\Lambda}_2\boldsymbol{Q}_2'$ respectively, where $\boldsymbol{\Lambda}_1 = \text{diag}(\lambda_1^{(1)}, \dots, \lambda_1^{(q)})$ and $\boldsymbol{\Lambda}_2 = \text{diag}(\lambda_2^{(1)}, \dots, \lambda_2^{(q)})$. The equation (2.20) can be rewritten as:

$$\widetilde{\boldsymbol{A}}^*\boldsymbol{\Lambda}_1 + \lambda\boldsymbol{\Lambda}_2\widetilde{\boldsymbol{A}}^* = \widetilde{\boldsymbol{Y}}^*$$

where $\widetilde{\boldsymbol{A}}^* = \boldsymbol{Q}_2'\widetilde{\boldsymbol{A}}\boldsymbol{Q}_1$, and $\widetilde{\boldsymbol{Y}}^* = \boldsymbol{Q}_2'\widetilde{\boldsymbol{Y}}\hat{\boldsymbol{S}}'\boldsymbol{Q}_1$. Thus, the solution for the (i, j)th element in $\widetilde{\boldsymbol{A}}^*$ is:

$$\widetilde{\boldsymbol{A}}^*_{i,j} = \frac{\widetilde{\boldsymbol{Y}}^*_{i,j}}{\lambda\lambda_2^{(i)} + \lambda_1^{(j)}} \tag{2.21}$$

And $\widetilde{\boldsymbol{A}}$ is updated as,

$$\hat{\widetilde{\boldsymbol{A}}}^{(k)} = \boldsymbol{Q}_2\widetilde{\boldsymbol{A}}^*\boldsymbol{Q}_1' \tag{2.22}$$

The proposed iterative estimation algorithm is summarized in Algorithm 1.

Algorithm 1: An Iterative Node-Rotation Algorithm

```
Initial: Initialize $\hat{\widetilde{\boldsymbol{A}}}^{(0)}, \{\hat{\boldsymbol{X}}_\ell^{(0)}, \hat{\boldsymbol{D}}_\ell^{(0)}\}$ based on estimates from existing methods such as connICA. For $\ell = 1...q$,    For $v = 1, \ldots, V$,
Step 1. Update $\boldsymbol{x}_\ell(v)$ :
    $\hat{\boldsymbol{b}}_{\ell\{v\}}^{(k)} = \underset{\boldsymbol{b}_{\ell\{v\}} \in \mathcal{R}^{V-1}}{\arg\min} \left\| \widetilde{\boldsymbol{Y}}'_{\{v\}} \hat{\widetilde{\boldsymbol{a}}}_\ell - \boldsymbol{b}_{\ell\{v\}} \right\|_2^2 + \phi \left\| \boldsymbol{b}_{\ell\{v\}} \right\|_1$
    $\hat{\boldsymbol{x}}_\ell^{(k)}(v) = \hat{\boldsymbol{D}}_\ell^{-1} (\hat{\boldsymbol{X}}_\ell(-v)' \hat{\boldsymbol{X}}_\ell(-v))^{-1} \hat{\boldsymbol{X}}_\ell(-v)' \hat{\boldsymbol{b}}_{\ell\{v\}}^{(k)}.$
    End for $v$  Step 2. Update $\boldsymbol{D}_\ell$:
    $\hat{\boldsymbol{d}}_\ell^{(k)} = \mathrm{diag}(\hat{\boldsymbol{D}}_\ell^{(k)}) = \underset{\boldsymbol{d}_\ell \in \mathcal{R}^{R_\ell}}{\arg\min} \left\| \widetilde{\boldsymbol{Y}}' \hat{\widetilde{\boldsymbol{a}}}_\ell - \hat{\boldsymbol{Z}}_\ell \boldsymbol{d}_\ell \right\|_2^2 + \phi \| \hat{\boldsymbol{Z}}_\ell \boldsymbol{d}_\ell \|_1,$
End for $\ell$
Step 3. Update $\widetilde{\boldsymbol{A}}$:
    $\hat{\widetilde{\boldsymbol{A}}}^{(k)} = \boldsymbol{Q}_2^{(k)} \hat{\widetilde{\boldsymbol{A}}}^{*(k)} \hat{\boldsymbol{Q}}_1^{(k)\prime}.$
    Perform an orthogonal transformation on $\hat{\widetilde{\boldsymbol{A}}}^{(k)}$
$\frac{\| \hat{\widetilde{\boldsymbol{A}}}^{(k)} - \hat{\widetilde{\boldsymbol{A}}}^{(k-1)} \|_F}{\| \hat{\widetilde{\boldsymbol{A}}}^{(k-1)} \|_F} < \epsilon_1$ and $\frac{\| \hat{\boldsymbol{S}}^{(k)} - \hat{\boldsymbol{S}}^{(k-1)} \|_F}{\| \hat{\boldsymbol{S}}^{(k-1)} \|_F} < \epsilon_2$
```

Chapter 3

Investigating latent neurocircuitry traits underlying longitudinal brain functional connectome

3.1 Introduction

The analysis of brain functional networks has become increasingly popular in neuroscience research, unveiling insights into the organization of the human brain, its role in neurodevelopment, aging, behavior, and the diagnosis or the understanding of the neurological procession of various brain diseases (Bullmore and Sporns, 2009; Deco et al., 2011; Kemmer et al., 2018; Satterthwaite et al., 2014; Wang et al., 2016; Wang and Guo, 2023). To investigate brain networks, functional connectivity (FC) obtained from functional magnetic resonance imaging (fMRI) or electroencephalogram (EEG) is widely used, where FC is defined as the temporal dependence between the time series of blood oxygen level-dependent signals originating from spatially distinct brain regions (Friston et al., 1993). Various measures of FC, including pearson correlation, partial correlation (Wang et al., 2016), mutual information (Hlinka et al., 2011), coherence (Srinivasan et al., 2007), and granger causality (Uddin et al., 2009) are commonly studied by researchers. The whole brain FC matrix aggregates connectivity patterns across the entire brain, representing the brain's functional networks. There is growing interest in decomposing this FC matrix to reveal underlying networks, known as latent sources or connectivity traits. These traits represent interconnected brain regions that tend to activate together during neural processing. Developing a reliable understanding of these latent connectivity traits is crucial for gaining valuable insights into the structure and dynamics of brain organization, as well as their association with neurological disease progression (Sorg et al., 2007; Williams, 2016), neurodevelopment (Hoff et al., 2013), and speeds of aging (DeCarli et al., 2012; Iannilli et al., 2017).

While most FC analyses typically consider data from a single time point, contemporary neuroimaging research is increasingly embracing longitudinal data collection, which involves multiple scans over time (Garavan et al., 2018; Weiner et al., 2017). For example, the Adolescent Brain Cognitive Development (ABCD) Study is the largest longitudinal study on adolescent brain development and behavior, and it has offered vast research possibilities to investigate developmental cognitive neuroscience (Garavan et al., 2018). The ADNI database contains longitudinal resting-state fMRI images collected at multiple time points, offering valuable data for studying the progression of early Alzheimer's disease (Petersen et al., 2010). To harness the full potential of longitudinal brain connectivity data for FC sub-system analysis, there is a growing need to develop methodologies capable of effectively decomposing longitudinal brain connectivity, unlocking its full potential for understanding brain longitudinal changes in connectivity patterns over time.

However, while methods for cross-sectional FC decomposition at a single time point exist, there are relatively few methodologies available for longitudinal FC decomposition. Several challenges arise in developing such models. Firstly, the decomposition of whole brain FC matrices is inherently challenging due to their high dimensionality.

FC matrices typically consist of hundreds of nodes and hundreds of thousands of connections, making reliable network estimation a significant challenge. Secondly, in terms of modeling, while Independent Component Analysis (ICA) (Beckmann and Smith, 2004; Calhoun et al., 2001) and its longitudinal extensions (Wang and Guo, 2019) are commonly employed for source separation, they primarily address the decomposition of observed series of neural activity signals, such as the blood oxygen level-dependent (BOLD) series from fMRI or electrode signal series from EEG. These methods are not ideally suited for decomposing brain connectivity matrices, which exhibit a symmetric structure and require consideration of the dependence or topological structure across edges in brain networks during decomposition. A recent proposal by Amico et al. (2017) introduced a connectivity-independent component analysis framework (connICA), which vectorizes connectivity matrices and utilizes existing ICA algorithms to decompose FC matrices. A natural extension of connICA stacks scans over time and applying the method for longitudinal FC decomposition. However, this approach treats each connection as an independent sample, disregarding the dependence structure across edges in the brain connectome. Additionally, it results in a large number of edge-wise parameters to estimate, which can lead to a loss of accuracy. Furthermore, it fails to capture differences in FC components across various time points. Previous research has found that functional connectivity (FC) exhibits aging effects that vary across different brain regions in cognitively normal (CN) individuals (Chen et al., 2016). Ren et al. (2016) demonstrated abnormal FC at various stages of Alzheimer's disease using longitudinal data. Ideally, we need methods that can model the changes in connectivity patterns over time, thereby providing a more accurate and comprehensive understanding of brain connectivity dynamics. Lastly, the absence of sparsity regularization in the aforementioned method leads to the generation of densely connected traits, which increases the likelihood of spurious findings.

In this paper, we introduce longitudinal-LOCUS, a novel low-rank decomposition method for brain connectivity that integrates uniform sparsity specifically for longitudinal functional connectivity (FC). Longitudinal-LOCUS is a fully data-driven blind source separation technique that decomposes longitudinal matrices to extract latent connectivity traits and characterize their dynamic changes in connectivity patterns over time. Specifically, longitudinal-LOCUS decomposes longitudinal FC measures into time-specific latent source signals, while accounting for their temporal dependence. To enhance accuracy and reliability, we employ a low-rank structure to model connectivity traits, capturing block-diagonal or banded patterns in brain connectivity matrices (Zhou et al., 2013) and reducing estimation parameters. For specific traits, our approach assumes shared eigenvectors across time while allowing eigenvalue changes, thereby effectively incorporating temporal correlation and modeling temporal differences. Additionally, we introduce a novel angle-based sparsity regularization on connectivity trait sets. This regularization further improves reliability by mitigating the presence of spurious connections and identifying connections that genuinely relevant to specific traits. The model is then applied to the ABCD dataset to capture the differences in FC related to neurodevelopment between females and males.

The rest of the paper is organized as follows. In Section 2, the proposed model is introduced. Section 3 discusses estimation and tuning parameter selection. Section 4 illustrates the superiority of our method compared to other decomposition methods through simulation studies. In Section 5, we apply the proposed method to investigate longitudinal connectivity in the ABCD study. Finally, discussions and conclusions are presented in Section 6.

3.2 Material and Methods

3.2.1 Longitudinal LOCUS Model

Let us consider a longitudinal study involving N participants, and each participant undergoes T visits. Let $\boldsymbol{Y}_{it} \in \mathbb{R}^{V\times V}$ denote the observed FC matrix for individual i at visit t, where $i \in 1,\dots,N$ and $t \in 1,\dots,T$. Here, V represents the number of nodes in the brain. Each element $\boldsymbol{Y}_{it}(u,v) \in \mathbb{R}$ quantifies the strength of the connection between nodes u and v. The matrix $\boldsymbol{Y}_{it}$ is symmetric, and the diagonal elements, which represent self-connections, are typically not of interest. To extract relevant and nonredundant information, we define a vector $\boldsymbol{y}_{it}$ comprising the upper triangular elements of $\boldsymbol{Y}_{it}$. Specifically, $\boldsymbol{y}_{it} = \mathcal{L}(\boldsymbol{Y}_{it})$, where $\mathcal{L}(\boldsymbol{Y}_{it}) = [\boldsymbol{Y}_{it}(1,2), \boldsymbol{Y}_{it}(1,3), \dots, \boldsymbol{Y}_{it}(V-1,V)]'$. Here, $\mathcal{L}$ is a mapping function from $\mathbb{R}^{V\times V}$ to $\mathbb{R}^p$, with $p = \frac{V(V-1)}{2}$. We model $\boldsymbol{y}_{it}$ as a linear combination of q group-visit-level latent connectivity sources or traits.

$$\boldsymbol{y}_{it} = \sum_{\ell=1}^{q} a_{it\ell}\boldsymbol{s}_{gt\ell} + \boldsymbol{e}_{it}, \tag{3.1}$$

where $a_{it\ell}$ are the individual-specific loadings on the ℓ-th latent connectivity trait, $\boldsymbol{s}_{gt\ell}$ are the ℓ-th latent connectivity traits at visit t for individuals belonging to group g, $\boldsymbol{e}_{it}$ is the error term for individual i at visit t independent of the connectivity traits. Here, $g \in \{1,2,\dots,G\}$ denotes the subgroup to which the individual belongs. These groups are mutually exclusive and could be defined by various factors such as sex (e.g., female or male) in studies investigating sex differences in neurodevelopment, different treatment conditions, or disease status in brain disorder analyses. If no grouping is considered, G can be specified as 1, and the group index g can be omitted. The model decomposes longitudinal connectivity matrices to extract group-visit-specific latent connectivity traits, denoted as $\boldsymbol{s}_{gt\ell} = \mathcal{L}(\boldsymbol{S}_{gt\ell}) \in \mathcal{R}^p$ for $\ell = 1,\dots,q$. These latent connectivity traits are assumed to be independent across the q traits. Each trait represents a set of brain connections that tend to occur together, with the elements indicating signal strength. The presence or prominence of the ℓ-th connectivity trait in the i-th individual at time point t is represented by $a_{it\ell}$.

We rewrite the model in matrix form by concatenating data from participants within the same group and visit. Let $\boldsymbol{Y}_{gt} = \left[\boldsymbol{y}_{1t}, \ldots, \boldsymbol{y}_{N_g t}\right]' \in \mathcal{R}^{N_g \times p}$ denote the multi-subject connectivity data in group g at visit t, where group g consists of N_g participants. The model can be expressed as:

$$\boldsymbol{Y}_{gt} = \boldsymbol{A}_{gt}\boldsymbol{S}_{gt} + \boldsymbol{E}_{gt}, \tag{3.2}$$

where $\boldsymbol{S}_{gt} = \left[\boldsymbol{s}_{gt1}, \ldots, \boldsymbol{s}_{gtq}\right]' \in \mathcal{R}^{q \times p}$ denotes the group-visit-level connectivity trait matrix. The model assumes that these connectivity traits are shared by a group at a specific time while varying across different groups. $\boldsymbol{A}_{gt} = \{a_{igt\ell}\} \in \mathcal{R}^{N_g \times q}$ represents the group-visit-level mixing or trait loading matrix. Additionally, $\boldsymbol{E}_{gt} = \left[\boldsymbol{e}_{1t}, \ldots, \boldsymbol{e}_{N_g t}\right]' \in \mathcal{R}^{N_g \times p}$ denotes the group-visit-level error matrix.

We model connectivity traits using a low-rank structure. This low-rank structure would effectively capture the block or banded patterns often observed in brain connectivity matrices (Zhou et al., 2013). By employing such structure, we significantly reduce the number of parameters required for estimation. Instead of estimating all the elements in the upper triangular of the connectivity matrix, which has a quadratic complexity of $O(V^2)$, we only need to estimate parameters with a linear complexity of $O(V)$. Specifically, the group-visit-level source signal $\boldsymbol{s}_{gt\ell}$ is modeled by:

$$\boldsymbol{s}_{gt\ell} = \mathcal{L}(S_{gt\ell}) = \mathcal{L}(\boldsymbol{X}_\ell \boldsymbol{D}_{gt\ell} \boldsymbol{X}_\ell') = \mathcal{L}(\sum_{r=1}^{R_\ell} d_{gt\ell}^{(r)} \boldsymbol{x}_\ell^{(r)} \boldsymbol{x}_\ell^{(r)'}), \quad \ell \in 1, 2, ..., q. \tag{3.3}$$

Here $\boldsymbol{X}_\ell = [\boldsymbol{x}_\ell^{(1)}, \ldots, \boldsymbol{x}_\ell^{(R_\ell)}] \in \mathcal{R}^{V \times R_\ell}$ is a matrix with rank $R_\ell < V$. We assume $\boldsymbol{X}_\ell$ is shared across groups and visits, in this way we can capture the correlation in latent sources among groups and visits. This shared structure facilitates the identification of common patterns and underlying similarities in the connectivity traits across the entire dataset. $\boldsymbol{D}_{gt\ell}$ is a group-visit-specific diagonal matrix with diagonal elements $\boldsymbol{d}_{gt\ell} = (d_{gt\ell}^{(1)}, .., d_{gt\ell}^{(R_\ell)})$, which reflects variations in latent sources across groups and visits. More specifically, in $\boldsymbol{X}_\ell$, each tensor factor $\boldsymbol{x}_\ell^{(r)}$ $(r = 1, \ldots, R_\ell)$ is a $V \times 1$ vector with unit norm, i.e. $\|\boldsymbol{x}_\ell^{(r)}\|_2 = 1$ for identifiability, and it represents the latent coordinates of the V nodes in the rth dimension. $d_{gt\ell}^{(r)}$ reflects the contribution of the rth dimension in generating $\boldsymbol{s}_{gt\ell}$. Each row of $\boldsymbol{X}_\ell$, i.e. $\boldsymbol{x}_\ell(v) \in \mathcal{R}^{R_\ell \times 1}$ represents the latent coordinates of the vth node in the a reduced R_ℓ-dimensional latent subspace, which potentially reflect the node's underlying neural activity. The connection between two nodes in the ℓth trait is modelled as $\boldsymbol{S}_{gt\ell}(u, v) = \boldsymbol{x}_\ell(u)' \boldsymbol{D}_{gt\ell} \boldsymbol{x}_\ell(v)$, which is the inner product between their latent coordinates. This implies the connection between the nodes depends on the similarity between the neural activity of the nodes characterized by their latent coordinates, which aligns with the understanding of brain connectivity in neuroscience (Friston et al., 1993; Friston, 2011). All the brain connections involving a node v are based on the node's latent coordinate $\boldsymbol{x}_\ell(v)$ and hence are inherently related. This appropriately accounts for the dependence structure across edges in the brain connectome, which is disregarded by many existing methods. We allow R_ℓ to vary across source signals, enabling us to accommodate diverse connectivity traits with different network properties and topological structures.

3.2.2 Regularization

As functional connections in neural circuits exist only in a proportion of region pairs across the brain (Bullmore and Sporns, 2009; Lee et al., 2011; Huang et al., 2010), we incorporate element-wise sparsity regularizations in Longitudinal LOCUS learning. This approach allows us to obtain parsimonious results and reduce spurious connections when recovering the spatial source maps of the connectivity traits. The optimization function with these regularizations is as follows:

$$\min \sum_{g=1}^{G} \sum_{t=1}^{T} ||\boldsymbol{Y}_{gt} - \boldsymbol{A}_{gt}\boldsymbol{S}_{gt}||_2^2 + \phi \sum_{g=1}^{G} \sum_{t=1}^{T} |\boldsymbol{S}_{gt}| \tag{3.4}$$

where ϕ is the tuning parameter for sparsity control. The first term in the optimization function (3.4) measures the closeness between the observed functional connectivity (FC) and the reconstructed FC, which stands for the goodness-of-fit of the model. The second term is the sparsity regularization, which penalizes each element of $\boldsymbol{S}_{gt\ell}$, i.e. $\boldsymbol{S}_{gt\ell}(u,v) = \boldsymbol{x}_\ell(u)'\boldsymbol{D}_{gt\ell}\boldsymbol{x}_\ell(v)$. This inner product corresponds to the angle between two nodes in the latent subspace. The sparsity penalty term in our model minimizes the sum of these inner products for all pairs of nodes, making it an angle-based regularization. This approach intuitively and effectively induces sparsity in the connections between nodes.

3.3 Estimation Algorithm and Selection of Tuning Parameters

3.3.1 Preprocessing

To optimize the subsequent decomposition process by reducing computational complexity and mitigating overfitting (Hyvärinen et al., 2001), we apply several common preprocessing procedures to the multi-subject connectivity data. These procedures include centering, dimension reduction, and whitening. This results in the transformed data, denoted as $\widetilde{\boldsymbol{Y}}_{gt}$, computed as $\widetilde{\boldsymbol{Y}}_{gt} = \boldsymbol{H}_{gt}\boldsymbol{Y}_{gt}$. The matrix $\boldsymbol{H}_{gt}$ is calculated as $\boldsymbol{H}_{gt} = (\boldsymbol{\Lambda}_{qgt} - \widetilde{\sigma}^2_{qgt}\boldsymbol{I})^{-1/2}\boldsymbol{U}'_{qgt}$. Here, $\boldsymbol{U}_{qgt}$ and $\boldsymbol{\Lambda}_{qgt}$ contain the top q eigenvectors and eigenvalues based on the singular value decomposition of $\boldsymbol{Y}_{gt}$. The residual variance, denoted as $\widetilde{\sigma}^2_{qgt}$, characterizes the unexplained variability in $\boldsymbol{Y}_{gt}$ that is not accounted for by the extracted q latent sources. This value is estimated by the average of the smallest $N_g - q$ eigenvalues in $\boldsymbol{\Lambda}_{qgt}$. The preprocessed data $\widetilde{\boldsymbol{Y}}_{gt}$ is of dimensions of $q \times p$ where the columns correspond to p connections in the brain. With the preprocessing, the Longitudinal LOCUS optimization function in (3.4) can be rewritten on the reduced and sphered space as following:

$$\min \sum_{g=1}^{G} \sum_{t=1}^{T} ||\widetilde{\boldsymbol{Y}}_{gt} - \widetilde{\boldsymbol{A}}_{gt} \boldsymbol{S}_{gt}||_F^2 + \phi \sum_{g=1}^{G} \sum_{t=1}^{T} |\boldsymbol{S}_{gt}| \tag{3.5}$$

where $\widetilde{\boldsymbol{A}}_{gt} = \boldsymbol{H}_{gt} \boldsymbol{A}_{gt} \in \mathcal{R}^{q \times q}$. It's important to note that the dimension reduction in the preprocessing occurs in the row space of $\boldsymbol{Y}_{gt}$. This space corresponds to the subject domain within group g and visit t, impacting how the data is organized over participants. However, it does not affect the column space of $\boldsymbol{Y}_{gt}$, which relates to the connectivity domain and represents the relationships between various brain regions. Therefore, the preprocessing does not impact the connectivity trait matrix $\boldsymbol{S}_{gt}$. Taking into account the orthogonal nature of the mixing matrix $\widetilde{\boldsymbol{A}}_{gt}$'s, the final optimization function can be expressed as follows:

$$\min_{\widetilde{\boldsymbol{A}}_{gt}, \{\boldsymbol{X}_\ell, \boldsymbol{D}_{gt\ell}\}} \sum_{g=1}^{G} \sum_{t=1}^{T} \sum_{l=1}^{q} ||\widetilde{\boldsymbol{Y}}_{gt}' \widetilde{\boldsymbol{a}}_{gt\ell} - \mathcal{L}(\boldsymbol{X}_\ell \boldsymbol{D}_{gt\ell} \boldsymbol{X}_\ell')||_2^2 + \phi \sum_{g=1}^{G} \sum_{t=1}^{T} \sum_{\ell=1}^{q} \sum_{u<v} |\boldsymbol{x}_\ell(u)' \boldsymbol{D}_{gt\ell} \boldsymbol{x}_\ell(v)| \tag{3.6}$$

where $\widetilde{\boldsymbol{a}}_{gt\ell}$ is the ℓth column of $\widetilde{\boldsymbol{A}}_{gt}$. The derivation of the final optimization function (3.6) is provided in Appendix 2.5.1.

3.3.2 A Node-rotation Estimation Algorithm

The objective function defined in (3.6) is non-convex, but it can be shown to be block multi-convex. This means the function can be divided into parameter subsets, such that the function is convex with respect to each parameter subset when the others are held constant. Using this property, we have developed an efficient node-rotation learning algorithm that provides closed-form solutions at each update step.

Initialization

We initialize the parameter values using LOCUS (Wang and Guo, 2023). We concatenate the FC matrices from different groups and visits into a single matrix, denoted as $\widetilde{\boldsymbol{Y}} = \left[\widetilde{\boldsymbol{Y}}_{11} \dots \widetilde{\boldsymbol{Y}}_{1T} \dots \widetilde{\boldsymbol{Y}}_{G1} \dots \widetilde{\boldsymbol{Y}}_{GT}\right]'$. Next, we decompose $\widetilde{\boldsymbol{Y}}$ using LOCUS to obtain a common set of latent sources $\boldsymbol{s}_1^*, \dots, \boldsymbol{s}_q^*$ and group-visit-specific loadings. For group g visit t, we have,

$$\widetilde{\boldsymbol{Y}}_{gt} = \widetilde{\boldsymbol{A}}_{gt}^* \boldsymbol{S}^* + \widetilde{\boldsymbol{E}}_{gt}^* \tag{3.7}$$

where $\boldsymbol{S}^* = \left[\boldsymbol{s}_1^*, \dots, \boldsymbol{s}_q^*\right]'$ and $\boldsymbol{s}_\ell^* = \mathcal{L}(\boldsymbol{X}_\ell^* \boldsymbol{D}_\ell^* \boldsymbol{X}_\ell^{*T})$. Then, we use a dual regression procedure to get group-visit-specific latent sources $\boldsymbol{S}_{gt}^{**} = \left[\boldsymbol{s}_{gt1}^{**}, \dots, \boldsymbol{s}_{gtq}^{**}\right]'$ by regressing group-visit-specific loadings $\widetilde{\boldsymbol{A}}_{gt}^*$ back into the group-visit-level data $\widetilde{\boldsymbol{Y}}_{gt}$. Finally, we map $\boldsymbol{s}_{gt\ell}^{**}$ to the common coordinate $\boldsymbol{X}_\ell^*$ to get $\boldsymbol{D}_{gt\ell}^*$. We initialize $\widetilde{\boldsymbol{A}}_{gt}$ using $\widetilde{\boldsymbol{A}}_{gt}^*$, $\boldsymbol{X}_\ell$ using $\boldsymbol{X}_\ell^*$, and $\boldsymbol{D}_{gt\ell}$ using $\boldsymbol{D}_{gt\ell}^*$.

Update $\boldsymbol{X}_\ell$

We develop a node-rotation algorithm that updates $\boldsymbol{X}_\ell$ at one of the node v while conditioning on the rest of the nodes and then rotating across the nodes. The objective function that are relevant to $\boldsymbol{x}_\ell(v)$ can be formatted as

$$\min \sum_{g=1}^{G} \sum_{t=1}^{T} ||\widetilde{\boldsymbol{Y}}_{gt\{v\}}' \widetilde{\boldsymbol{a}}_{gt\ell} - \boldsymbol{X}_\ell(-v) \boldsymbol{D}_{gt\ell} \boldsymbol{x}_\ell(v)||_2^2 + \phi \sum_{g=1}^{G} \sum_{t=1}^{T} \sum_{u<v} |\boldsymbol{x}_\ell(u)' \boldsymbol{D}_{gt\ell} \boldsymbol{x}_\ell(v)| \tag{3.8}$$

where $\widetilde{\boldsymbol{Y}}_{gt\{v\}}$ is a $q \times (V-1)$ matrix that represents $(V-1)$ edges that are relevant to node v, and $\boldsymbol{X}_\ell(-v)$ is a $(V-1) \times R_\ell$ matrix with the vth row removed. It is straightforward to show that the optimization in (3.8) is convex.

We propose the following procedure for solving (3.8). First, we concatenate $\widetilde{\boldsymbol{Y}}'_{gt\{v\}}\widetilde{\boldsymbol{a}}_{gt\ell}$ and $\boldsymbol{X}_\ell(-v)\boldsymbol{D}_{gt\ell}\boldsymbol{x}_\ell(v)$ across groups and time, such that (3.8) can be reformulated as the following problem:

$$\min||\boldsymbol{z}_{\ell\{v\}} - \boldsymbol{b}_{\ell\{v\}}||_2^2 + \phi|\boldsymbol{b}_{\ell\{v\}}| \tag{3.9}$$

where $\boldsymbol{z}_{\ell\{v\}} = \begin{pmatrix} \widetilde{\boldsymbol{Y}}'_{11\{v\}}\widetilde{\boldsymbol{a}}_{11\ell} \\ \cdots \\ \widetilde{\boldsymbol{Y}}'_{1T\{v\}}\widetilde{\boldsymbol{a}}_{1T\ell} \\ \cdots \\ \widetilde{\boldsymbol{Y}}'_{G1\{v\}}\widetilde{\boldsymbol{a}}_{G1\ell} \\ \cdots \\ \widetilde{\boldsymbol{Y}}'_{GT\{v\}}\widetilde{\boldsymbol{a}}_{GT\ell} \end{pmatrix}$ and $\boldsymbol{b}_{\ell\{v\}} = \boldsymbol{R}_\ell(v)\boldsymbol{x}_\ell(v) = \begin{pmatrix} \boldsymbol{X}_\ell(-v)\boldsymbol{D}_{11\ell} \\ \cdots \\ \boldsymbol{X}_\ell(-v)\boldsymbol{D}_{1T\ell} \\ \cdots \\ \boldsymbol{X}_\ell(-v)\boldsymbol{D}_{G1\ell} \\ \cdots \\ \boldsymbol{X}_\ell(-v)\boldsymbol{D}_{GT\ell} \end{pmatrix} \boldsymbol{x}_\ell(v)$, and both are vectors of length $(V-1)GT$. We obtain analytical estimate $\boldsymbol{b}_{\ell\{v\}}$ following Fan and Li (2001),

$$\boldsymbol{b}_{\ell\{v\}} = \text{diag}(\text{sgn}(\boldsymbol{z}_{\ell\{v\}}))\delta(|\boldsymbol{z}_{\ell\{v\}}| - \frac{\phi}{2}\boldsymbol{1}_{(V-1)GT})$$

where sgn represents sign function for each element and δ denotes a rectifier function ($\delta(x) = x$ if $x > 0$ otherwise 0). In the kth iteration, we project $\hat{\boldsymbol{b}}^{(k)}_{\ell\{v\}}$ onto the low-rank space spanned by $\hat{\boldsymbol{R}}^{(k-1)}_\ell(v)$ to obtain the estimate of $\boldsymbol{x}_\ell(v)$, that is,

$$\hat{\boldsymbol{x}}^{(k)}_\ell(v) = (\hat{\boldsymbol{R}}'^{(k-1)}_\ell(v)\hat{\boldsymbol{R}}^{(k-1)}_\ell(v))^{-1}\hat{\boldsymbol{R}}'^{(k-1)}_\ell\hat{\boldsymbol{b}}^{(k)}_{\ell\{v\}}$$

After updating $\boldsymbol{x}_\ell(v)$, we proceed to the next node, systematically reiterating the aforementioned process across all nodes within the range of $v = 1, ..., V$. This iterative approach yields updated estimations for $\boldsymbol{X}_\ell$. A notable advantage of the node-rotation algorithm is its use of analytical solutions, which eliminates the need for gradient-based numerical approximations. This feature substantially enhances the algorithm's efficiency and reliability.

Update $\boldsymbol{D}_{gt\ell}$

Given the estimate of $\boldsymbol{X}_\ell$ from the kth iteration and the estimate of $\boldsymbol{A}_{gt\ell}$ from the $(k-1)$th iteration, we estimate the diagonal of $\boldsymbol{D}_{gt\ell}$, that is, $\boldsymbol{d}_{gt\ell} = \text{diag}(\boldsymbol{D}_{gt\ell})$ by minimizing the following objective function:

$$\min ||\widetilde{\boldsymbol{Y}}_{gt}^{\prime}\widetilde{\boldsymbol{a}}_{gt\ell} - \boldsymbol{Z}_\ell \boldsymbol{d}_{gt\ell}||_2^2 + \phi|\boldsymbol{Z}_\ell \boldsymbol{d}_{gt\ell}| \tag{3.10}$$

where $\boldsymbol{Z}_\ell \in \mathcal{R}^{p \times R_\ell}$ with the rth column being $\mathcal{L}(\boldsymbol{x}_\ell^{(r)} \boldsymbol{x}_\ell^{(r)\prime})$, $r = 1, \ldots, R_\ell$. The process for solving (3.10) is analogous to that employed for solving (3.9).

Update $\widetilde{\boldsymbol{A}}_{gt}$

The mixing matrix $\widetilde{\boldsymbol{A}}_{gt}$ is updated using the estimates of $\boldsymbol{X}_\ell$ and $\boldsymbol{D}_{gt\ell}$ obtained in the kth iteration as follows:

$$\hat{\widetilde{\boldsymbol{A}}}_{gt}^{(k)} = \widetilde{\boldsymbol{Y}}_{gt} \hat{\boldsymbol{S}}_{gt}^{\prime(k)} (\hat{\boldsymbol{S}}_{gt}^{(k)} \hat{\boldsymbol{S}}_{gt}^{\prime(k)})^{-1} \tag{3.11}$$

A summary of the algorithm is presented in Algorithm 2.

Algorithm 2: An Iterative Node-Rotation Algorithm

```
Initial: Initialize $\hat{\widetilde{\boldsymbol{A}}}_{gt}^{(0)}$, $\{\hat{\boldsymbol{X}}_{gt\ell}^{(0)}, \hat{\boldsymbol{D}}_{gt\ell}^{(0)}\}$ obtained from LOCUS and dual regression. For $\ell = 1...q$,     For $v = 1, \ldots, V$,
Step 1. Update $\boldsymbol{x}_\ell(v)$ :
    $\hat{\boldsymbol{b}}_{\ell\{v\}}^{(k)} = \underset{\boldsymbol{b}_{\ell\{v\}} \in \mathcal{R}^{(V-1)GT}}{\arg\min} \left\| \boldsymbol{z}_{\ell\{v\}} - \boldsymbol{b}_{\ell\{v\}} \right\|_2^2 + \phi \left\| \boldsymbol{b}_{\ell\{v\}} \right\|_1$
    $\hat{\boldsymbol{x}}_\ell^{(k)}(v) = (\hat{\boldsymbol{R}}_\ell^{\prime(k-1)}(v) \hat{\boldsymbol{R}}_\ell^{(k-1)}(v))^{-1} \hat{\boldsymbol{R}}_\ell^{\prime(k-1)} \hat{\boldsymbol{b}}_{\ell\{v\}}^{(k)}.$
    End for $v$     For $g = 1 \ldots G, t = 1 \ldots T$,  Step 2. Update $\boldsymbol{D}_{gt\ell}$:
        $\hat{\boldsymbol{d}}_{gt\ell}^{(k)} = \mathrm{diag}(\hat{\boldsymbol{D}}_{gt\ell}^{(k)}) = \underset{\boldsymbol{d}_\ell \in \mathcal{R}^{R_\ell}}{\arg\min} \left\| \widetilde{\boldsymbol{Y}}_{gt}' \hat{\widetilde{\boldsymbol{a}}}_{gt\ell} - \hat{\boldsymbol{Z}}_\ell \boldsymbol{d}_{gt\ell} \right\|_2^2 + \phi \| \hat{\boldsymbol{Z}}_\ell \boldsymbol{d}_{gt\ell} \|_1,$
    End for $g, t$
End for $\ell$
For $g = 1 \ldots G, t = 1 \ldots T$, Step 3. Update $\widetilde{\boldsymbol{A}}_{gt}$:
    $\hat{\widetilde{\boldsymbol{A}}}_{gt}^{(k)} = \widetilde{\boldsymbol{Y}}_{gt} \hat{\boldsymbol{S}}_{gt}^{\prime(k)} (\hat{\boldsymbol{S}}_{gt}^{(k)} \hat{\boldsymbol{S}}_{gt}^{\prime(k)})^{-1}$
    Perform an orthogonal transformation on $\hat{\widetilde{\boldsymbol{A}}}_{gt}^{(k)}$
End for $g, t$
$\frac{\|\hat{\widetilde{\boldsymbol{A}}}_{gt}^{(k)} - \hat{\widetilde{\boldsymbol{A}}}_{gt}^{(k-1)}\|_F}{\|\hat{\widetilde{\boldsymbol{A}}}_{gt}^{(k-1)}\|_F} < \epsilon_1$ and $\frac{\|\hat{\boldsymbol{S}}_{gt}^{(k)} - \hat{\boldsymbol{S}}_{gt}^{(k-1)}\|_F}{\|\hat{\boldsymbol{S}}_{gt}^{(k-1)}\|_F} < \epsilon_2$
```

3.3.3 Tuning parameter selection

The optimization function in (3.6) involves two sets of tuning parameters: the rank parameters $\{R_\ell\}_{\ell=1}^{q}$, which control the dimension of the subspace of the connectivity traits, and ϕ, which regulates the influence of the sparsity penalization for the connectivity traits. Following Wang and Guo (2023), we select R_ℓ for each set of $\{s_{gt\ell}\}$ during the initialization stage. Specifically, R_ℓ is chosen based on the desired level of similarity ρ between the estimated source signals initialized with the low-rank structure and the unconstrained source signal estimate obtained without the low-rank structure. Therefore, we determine the values of all sets of $\{R_\ell\}_{\ell=1}^{q}$ based on a single parameter ρ that captures the desired similarity.

We propose the following BIC-type criteria to select ρ and ϕ,

$$\mathrm{BIC} = -2\sum_{g=1}^{G}\sum_{t=1}^{T}\sum_{i=1}^{q} log(g(\widetilde{\boldsymbol{y}}_{gti}; \sum_{j=1}^{q}\hat{\tilde{a}}_{gtij}\hat{\boldsymbol{s}}_{gtj}; \hat{\tilde{\sigma}}\boldsymbol{I}_p)) + log(N)\sum_{g=1}^{G}\sum_{t=1}^{T}\sum_{j=1}^{q}||\hat{\boldsymbol{s}}_{gtj}||_0, \tag{3.12}$$

where $\widetilde{\boldsymbol{y}}_{gti}, g = 1, \ldots, G$ and $t = 1, \ldots, T$ is the ith row in $\widetilde{\boldsymbol{Y}}_{gt}$, $\{\hat{\tilde{a}}_{gtij}\}$ are the estimated mixing coefficients of preprocessed connectivity data, $\hat{\boldsymbol{s}}_{gtj}, j = 1, ..., \ell$, is the estimated connectivity trait, $\hat{\tilde{\sigma}}^2 = \frac{1}{qpGT}\sum_{g=1}^{G}\sum_{t=1}^{T}\sum_{i=1}^{q}||\widetilde{\boldsymbol{y}}_{gti} - \sum_{j=1}^{q}\hat{\tilde{a}}_{gtij}\hat{\boldsymbol{s}}_{gtj}||_2^2$, g is the pdf of a multivariate Gaussian distribution and $||.||_0$ denotes the L_0 norm.

3.4 Simulation

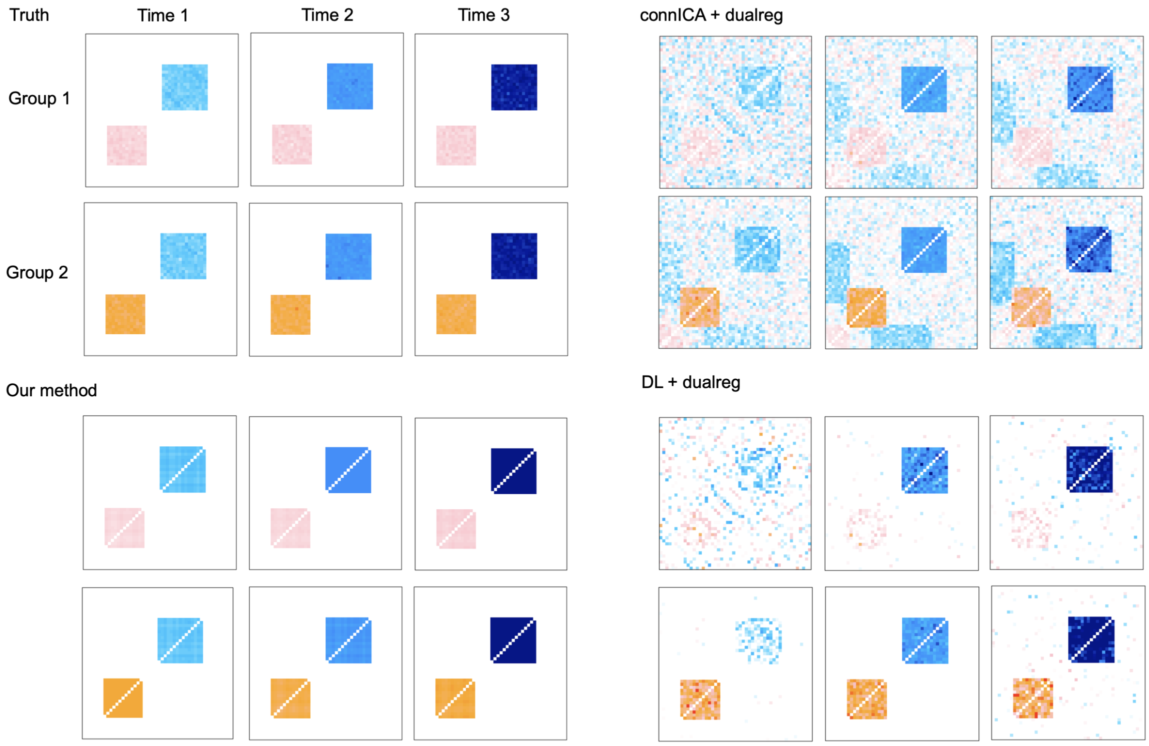
We conducted simulations with sample sizes of $N = 20, 50$, where each participant has three visits: baseline, visit 1, and visit 2 (at time points t = 0, 1, and 2). Participants were randomly assigned to either the treatment or control group with a probability of $p = 0.5$.

We generated FC from three underlying source signals ($q = 3$) with dimensions of 50×50. These source signals were designed to mimic patterns found in real fMRI data, featuring a low-rank structure. Latent source 1 has two active regions, one characterized by positive signals and the other by negative signals. The magnitude of the positive signals remains consistent over time, with group 1's magnitude lower than group 2, while the negative correlation areas exhibit a progressive increase in both groups. Latent source 2 consists of negative connections between modules (off-diagonal) and positive connections within modules (diagonal). The positive correlations increase over time, with group 2 displaying a lower magnitude compared to group 1, while the negative correlation area remains consistently unchanged over time. In latent source 3, the strength of the source signals within negatively activated regions increases over time, while within positively activated regions decreases over time. There are no observable differences between groups. The mixing coefficients were drawn from estimates obtained from real imaging data.

Furthermore, we incorporated zero-mean Gaussian noise into the signal mixture. The variance of this noise was set to three different levels: $\sigma^2 = 2^2, 3^2$, and 4^2, representing low, medium, and high variance levels, respectively. In summary, we have 2×3 simulation settings with various combinations of sample sizes and variance levels. For each setting, we generated 100 simulation runs to comprehensively evaluate performance variations.

We compared our method with connICA and Dictionary Learning (DL) in terms of the ability to recover group-visit-specific latent sources and loading matrix. Specifically, we applied each method to the concatenated data, yielding a population-level $\boldsymbol{S}$ matrix. Subsequently, we segregated the data into groups (treatment and control) and visits. Dual regression was then employed to derive group-visit-specific latent sources. We record these two method as 'connICA + dualreg' and 'DL + dualreg' respectively.

To assess the accuracy of each method, we evaluated the methods based on the correlations between the ground truth and the model-derived estimates of both source signals and mixing coefficients. Furthermore, we calculated the standard deviation

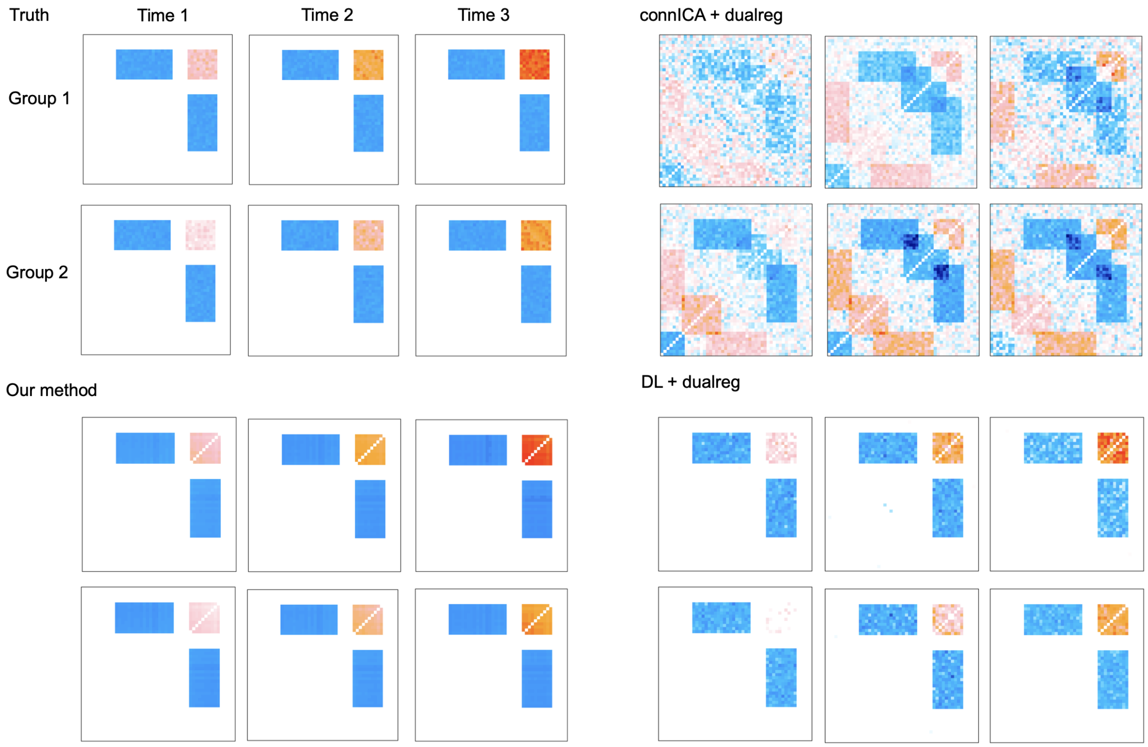


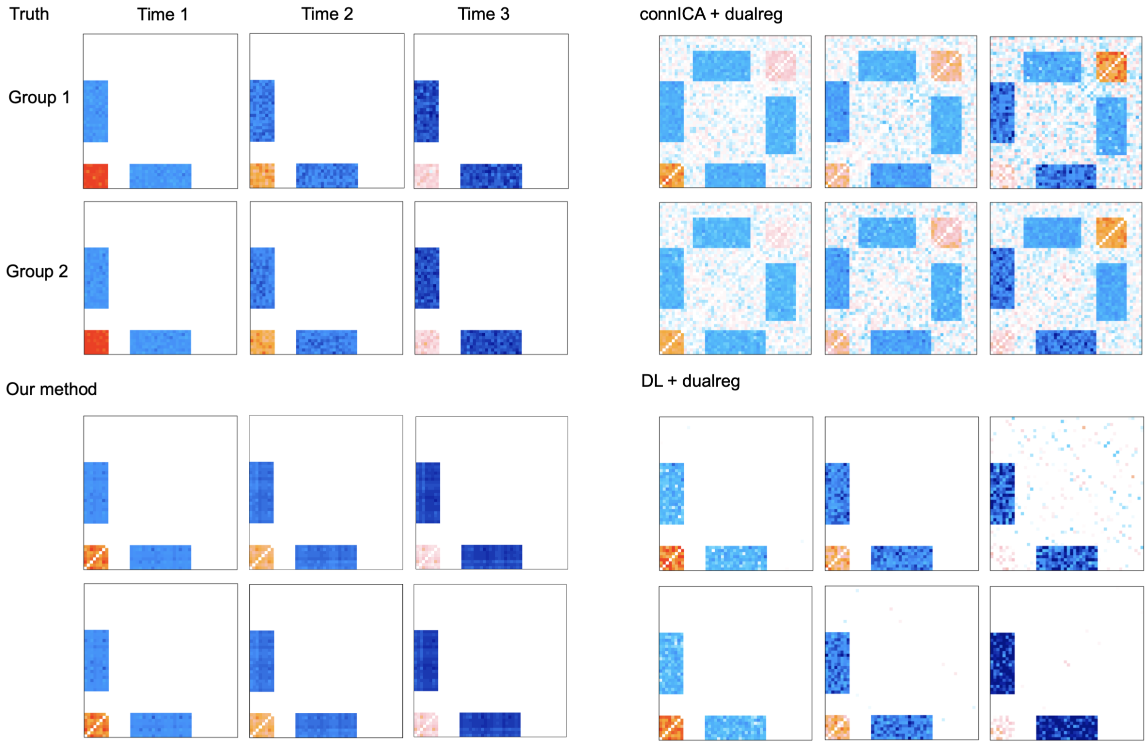
of these correlations across the 100 simulation runs to evaluate the variability of the estimations.

Results are summarized in Table 3.1. Our method consistently exhibits superior accuracy in recovering latent sources and mixing coefficients compared to connICA and DL. Meanwhile, the standard deviation of our method is notably lower than ‘connICA + dualreg’ and comparable to ‘DL + dualreg’, which underscores the enhanced stability of our proposed approach.

Term	N	Var.	llocus	connICA+dualreg	DL+dualreg
Latent Source Corr. (SD)	20	Low	0.9863 (0.0015)	0.7032 (0.0224)	0.9091 (0.0016)
		Mid	0.9588 (0.0083)	0.6453 (0.0230)	0.8466 (0.0023)
		High	0.9360 (0.0046)	0.5890 (0.0220)	0.7828 (0.0029)
	50	Low	0.9884 (0.0000)	0.8563 (0.0411)	0.9726 (0.0004)
		Mid	0.9880 (0.0000)	0.8266 (0.0391)	0.9437 (0.0008)
		High	0.9875 (0.0001)	0.7880 (0.0374)	0.9091 (0.0012)
Loading Matrix Corr. (SD)	20	Low	0.9958 (0.0014)	0.7726 (0.0435)	0.9862 (0.0004)
		Mid	0.9857 (0.0054)	0.7757 (0.0476)	0.9852 (0.0006)
		High	0.9781 (0.0040)	0.7784 (0.0492)	0.9837 (0.0008)
	50	Low	0.9989 (0.0001)	0.8341 (0.0840)	0.9739 (0.0003)
		Mid	0.9983 (0.0002)	0.8347 (0.0831)	0.9734 (0.0006)
		High	0.9973 (0.0003)	0.8276 (0.0839)	0.9725 (0.0008)

Table 3.1: Simulation results for comparing longitudinal-LOCUS and the existing connICA and DL methods with dual regression based on 100 simulation runs conducted under three variance (Var.) settings. Values presented are mean and standard deviation of correlations between the true and estimated latent sources and loading/mixing matrices.


Truth
Time 1
Time 2
Time 3
connICA + dualreg
Group 1
Group 2
Our method
DL + dualreg


Truth
Time 1
Time 2
Time 3
connICA + dualreg
Group 1
Group 2
Our method
DL + dualreg

3.5 Data analysis of longitudinal functional connectivity in ABCD study

3.5.1 Dataset and methods

The Adolescent Brain Cognitive Development (ABCD) study is an extensive longitudinal investigation using resting-state functional magnetic resonance imaging (fMRI) data collected at baseline and follow-up visits (Volkow et al., 2018). It's ideal for longitudinal connectivity studies due to its large sample size, detailed phenotypic data, standardized procedures, and data availability. We obtained minimally processed imaging data from ABCD 4.0 on the National Institute of Mental Health Data Archive (NDA), which includes standard registration and segmentation preprocessing steps to extract global, white matter (WM), and cerebrospinal fluid (CSF) signals. We then applied additional preprocessing steps adapted from the "DCANBOLDproc" stage in the ABCD-HCP BIDS fMRI Pipeline (Feczko et al., 2021; Marek et al., 2022). Framewise displacement (FD) (Power et al., 2012) was calculated using the motion parameters from the minimally preprocessed data, and frames with FD greater than 0.3 mm were censored. The remaining data were demeaned and detrended, and linear interpolation was used to replace the motion-censored frames with low-motion data. Next, the interpolated data were denoised using a general linear model (GLM) with Friston's 24 movement parameters (Friston et al., 1996), WM, CSF, and global signals as nuisance regressors. Finally, a second-order Butterworth bandpass filter between 0.009 and 0.08 Hz was applied to the residuals from the denoising step. We computed the average time series for regions of interest defined by Power's 264-node brain parcellation (Power et al., 2011). Nodes were assigned to functional networks based on Smith's major resting-state network system for interpretation (Smith et al., 2009). For nodes with ambiguous assignments in Smith's system, we interpret them using Power's resting-state network labels (Power et al., 2011).

We subsetted the preprocessed data to 3,869 children with complete 4 runs for both baseline and follow-up visits. After motion quality control with retaining data with FD < 0.2 mm for over 8 minutes, 2,672 children remained. The remaining children included 1,257 females (47.04%) and 1,415 males (52.96%), aged 8.92 to 11 years (mean 9.92, SD 0.63) in the baseline visit, and 10.75 to 13.83 years (mean 11.95, SD 0.65) in the follow-up visit. For each child, we extracted the fMRI time series from each node and obtained 264×264 longitudinal connectivity matrices by evaluating the pairwise correlations between the node-specific fMRI series for the baseline and follow-up visits. We then concatenated the upper triangular part of the matrices across time and children. Fisher's Z transformation was applied to the correlations to obtain the longitudinal connectivity data for decomposition. Site harmonization of functional connectivity (FC) matrices was performed using COVBAT (Chen et al., 2022) to account for site effects respectively for each year, with "site" as a factor and covariates including age, sex, and their interactions.

We applied longitudinal LOCUS to decompose the connectivity data and uncover underlying connectivity traits. We considered two groups based on gender and analyzed data from two time points: the baseline visit and the follow-up visit. The number of latent sources, q, was determined based on the reproducibility and interpretability of the extracted sources. The parameters ϕ and ρ were selected using the proposed BIC-type criteria. The reproducibility of the extracted sources increases with q, reaching around 0.9 at $q = 15$. While reproducibility continues to rise slightly beyond this point, the interpretability of the extracted sources diminishes as q becomes too large. Therefore, we select $q = 15$, which strikes a good balance between model size, reproducibility, and interpretability. Using the proposed tuning parameter selection method, we choose $\rho = 0.95$ and $\phi = 10$ for the longitudinal-LOCUS optimization.

After obtaining underlying longitudinal connectivity traits, we conducted statistical inference to compare differences in neurodevelopment as reflected in discrepancies in brain connectivity traits between genders. To quantify neurodevelopment within each gender, we subtracted the brain connectivity trait maps of the follow-up visit from those of the baseline visit. To determine the differences in neurodevelopment between genders, we further subtracted the difference trait maps between genders, resulting in "difference-in-difference" trait maps. Next, to identify significant connectivity differences, we constructed nonparametric statistical tests. For each bootstrap sample, we randomly shuffled the gender group index for each child while preserving their time order. We conducted a total of 5000 bootstrap iterations. In each iteration, we ran longitudinal LOCUS and computed the "difference-in-difference" map for each trait, determining the 99th percentile of the absolute values across all edges. For each trait, the 95th percentile of these values across all iterations was used to establish the 95% confidence band for the "difference-in-difference" maps.

3.5.2 Results

Spatial composition of the latent connectivity traits

Longitudinal-LOCUS uncovers 15 dynamic latent connectivity trait sets for both females and males at the baseline visit and follow-up visit. The figure below presents the retrieved source signal maps. An interesting finding is that 9 out of the 15 most reproducible traits involve visual networks. This finding aligns well with earlier research on single time point analysis (Wang and Guo, 2023), which shows that connections involving visual networks are highly reproducible and consistently observed in circuitry traits. Specifically, we categorize connectivity traits driven by lower-level sensory networks such as visual networks, auditory network ("Aud"), and sensorimotor network ("SM") as lower-level sensory connectivity traits. In contrast, traits driven by the default mode network ("DMN"), fronto-parietal networks ("FPL, FPR"), and cerebellum ("CB"), given its involvement in both motor and cognitive functions, are classified as higher-order cognitive connectivity traits. Traits driven by lower-level

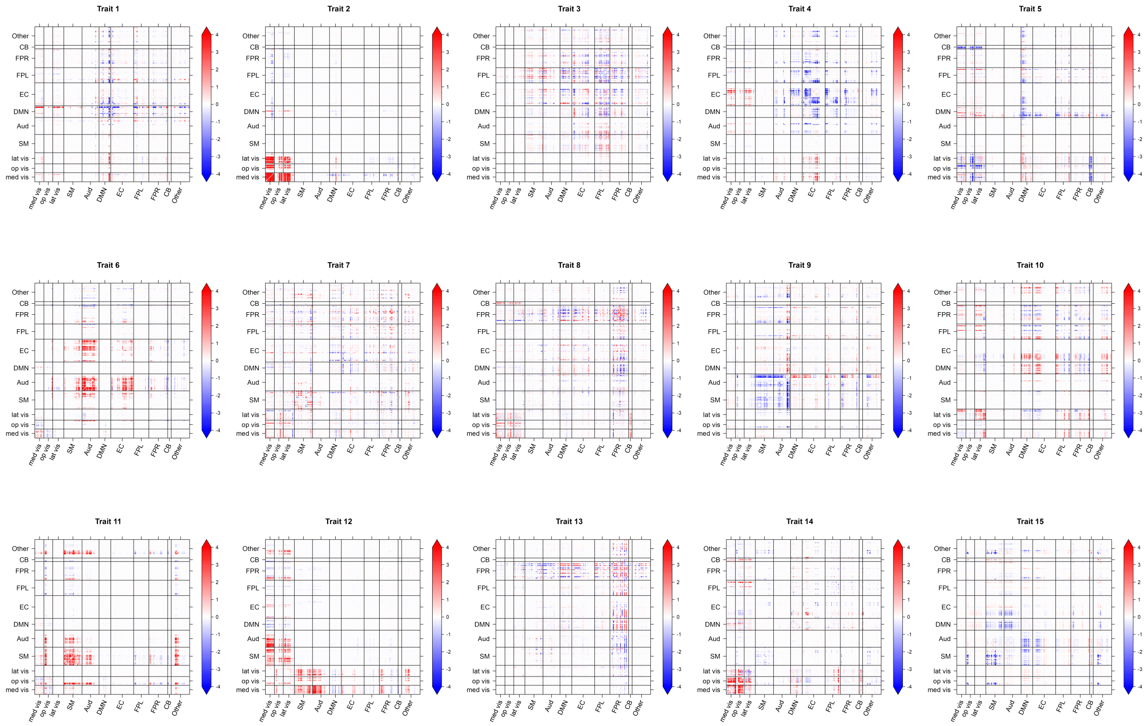

Trait 1
Trait 2
Trait 3
Trait 4
Trait 5
Trait 6
Trait 7
Trait 8
Trait 9
Trait 10
Trait 11
Trait 12
Trait 13
Trait 14
Trait 15

Figure 3.1: Part I: Fifteen connectivity traits extracted from the ABCD study for females at baseline.

sensory networks include Trait 2 (Vis), Trait 5 (Vis-DMN-FPL-CB), Trait 6 (Aud-EC), Trait 9 (Aud-SM), Trait 11 (Aud-SM-Vis), Trait 12 (Vis-Aud-SM), Trait 14 (Vis-FPR), and Trait 15 (SM). The identified higher-order cognitive traits include Trait 1 (DMN-Vis-EC-FPR-FRL), Trait 3 (EC-FPL), Trait 4 (EC-Vis), Trait 8 (FPR-Vis), Trait 10 (Vis-DMN), and Trait 13 (FPR-EC-DMN). Additionally, Trait 7 exhibits overall connectivity patterns.

Neurodevelopmental differences in connectivity traits between males and females

We identified three connectivity traits that demonstrate significant neurodevelopmental differences between males and females. Trait 10 (Vis-DMN) (Figure 3.2) is a cognitive-related connectivity trait primarily driven by the default mode and visual networks. Stronger connections are observed in the FPL-Visual, FRL-DMN, and within the

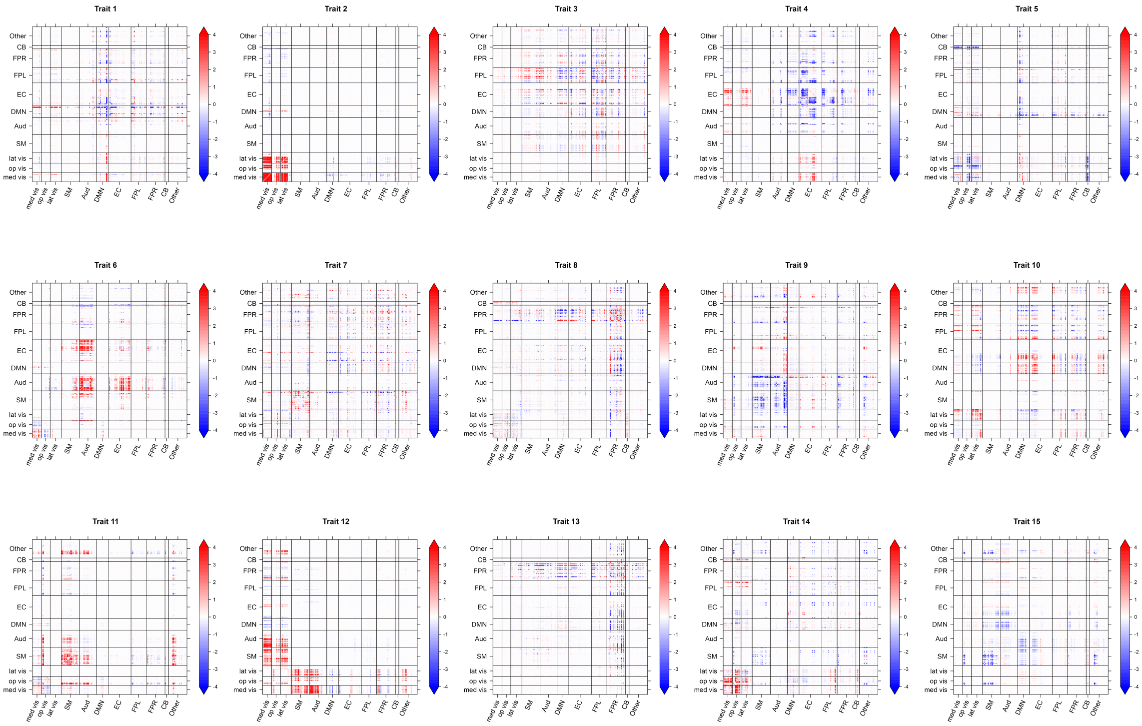
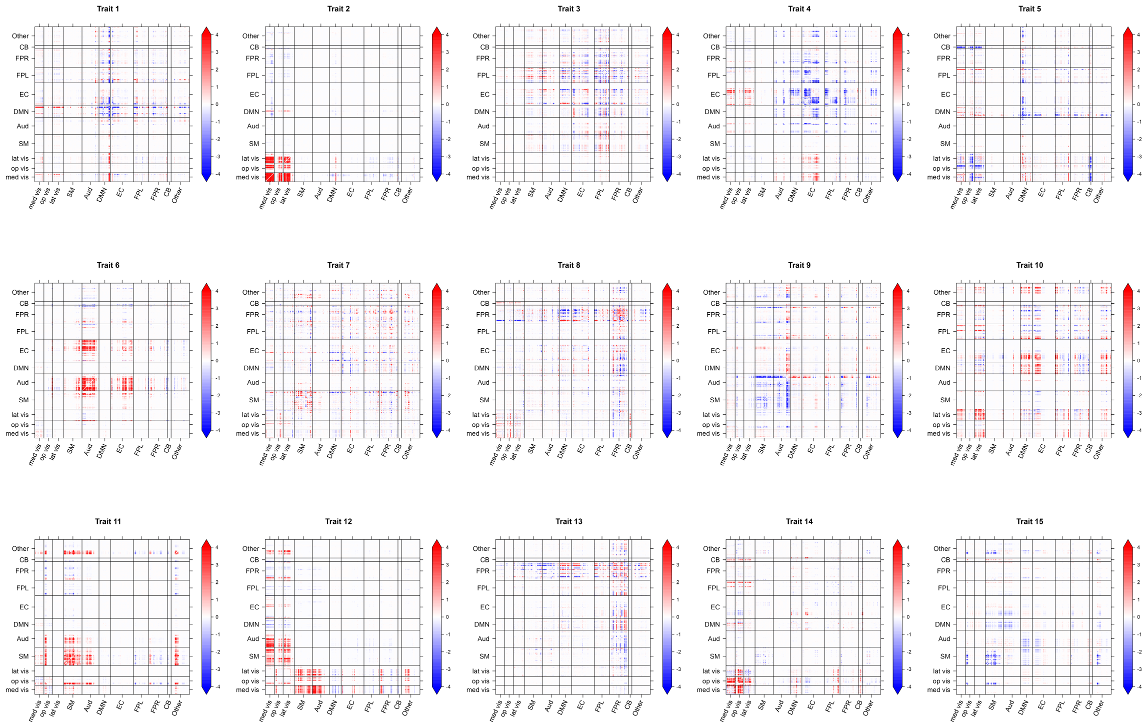


Figure 3.1: Part II: Fifteen connectivity traits extracted from the ABCD study for females at the follow-up visit.



Trait 1 Trait 2 Trait 3 Trait 4 Trait 5

Trait 6 Trait 7 Trait 8 Trait 9 Trait 10

Trait 11 Trait 12 Trait 13 Trait 14 Trait 15

Figure 3.1: Part III: Fifteen connectivity traits extracted from the ABCD study for males at baseline.

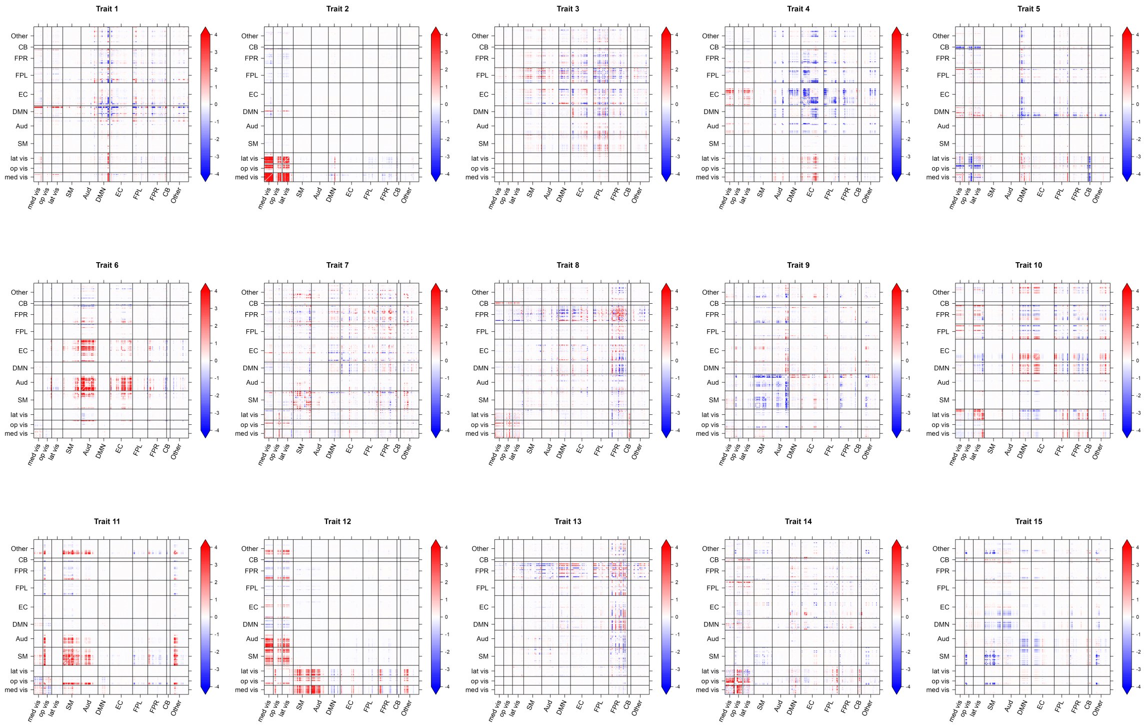


Figure 3.1: Part IV: Fifteen connectivity traits extracted from the ABCD study for males at the follow-up visit.

FPL regions as children mature. While males initially show weaker connections at baseline, their development progresses more rapidly compared to females. Overall, males exhibit greater developmental changes compared to females. Trait 14 (Vis-FPR) (Figure 3.3) is a connectivity trait primarily driven by the visual network and the frontal parietal network. Males show a decrease in connections within the visual and sensory-motor-related networks, while females generally exhibit either an increase or only a slight decrease in these connections. Trait 15 (SM) (Figure 3.4) is driven by the sensory-motor network. Males display stronger connections, while females exhibit slightly weaker connections during neural development. Significant differences are observed within the sensory-motor network, within the visual network, and between the sensory-motor and visual networks.

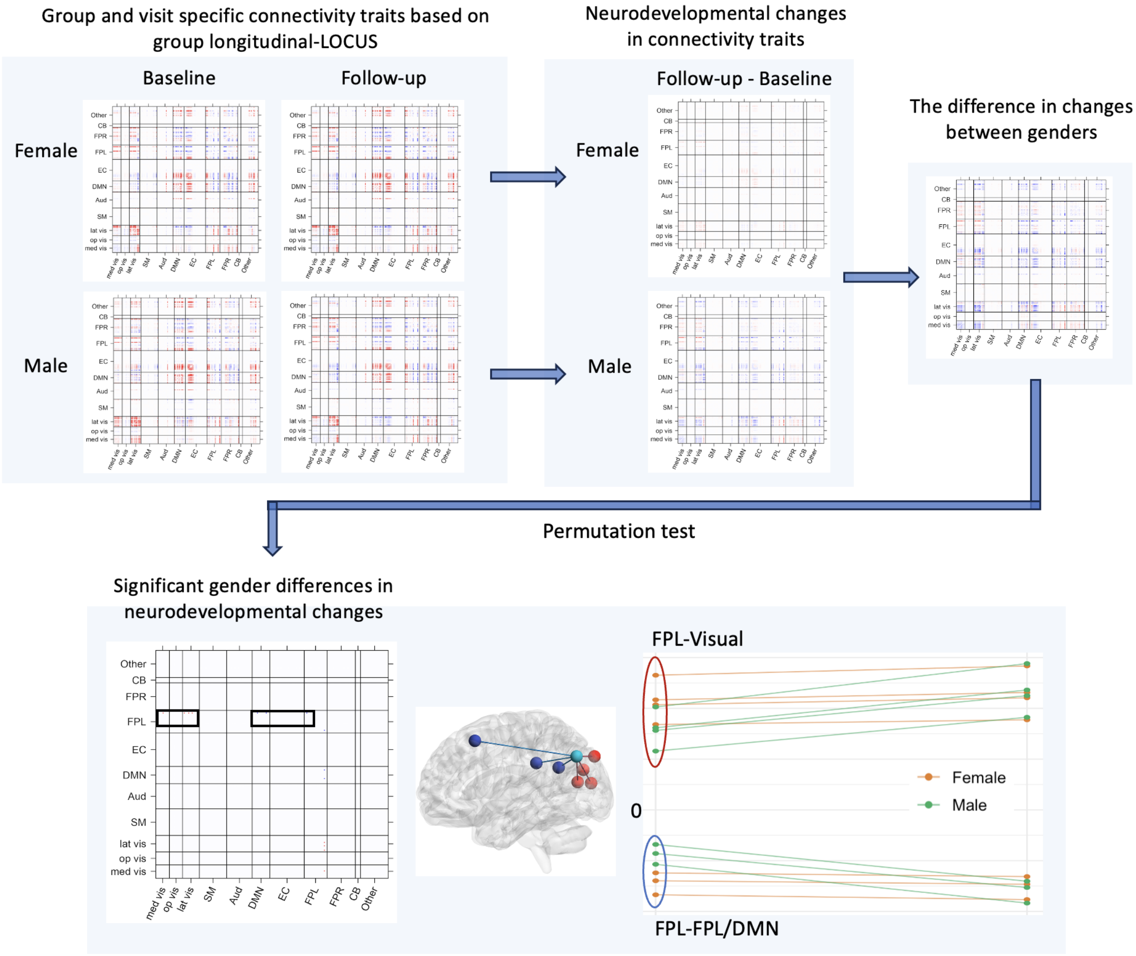


Figure 3.2: Trait 10 demonstrates significant neurodevelopmental differences between males and females

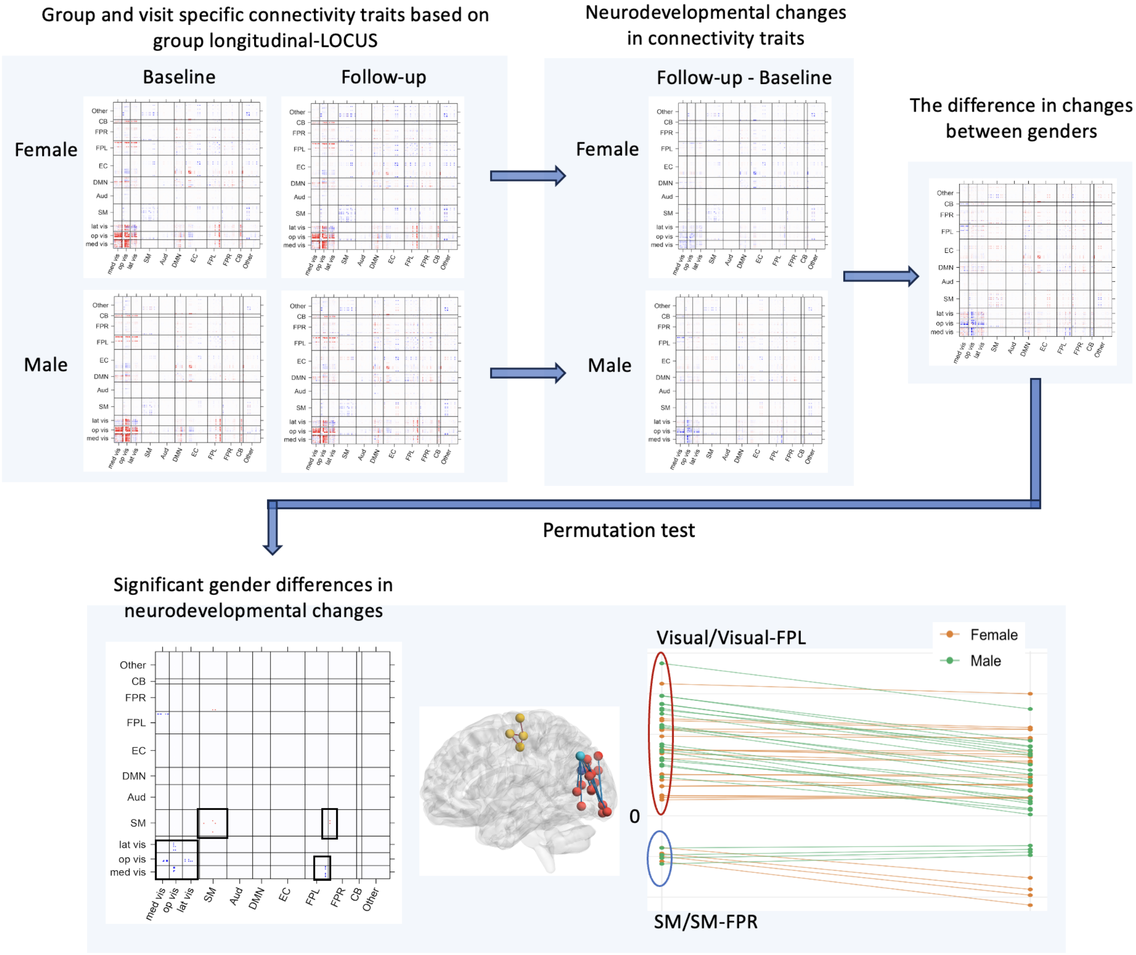


Figure 3.3: Trait 14 demonstrates significant neurodevelopmental differences between males and females

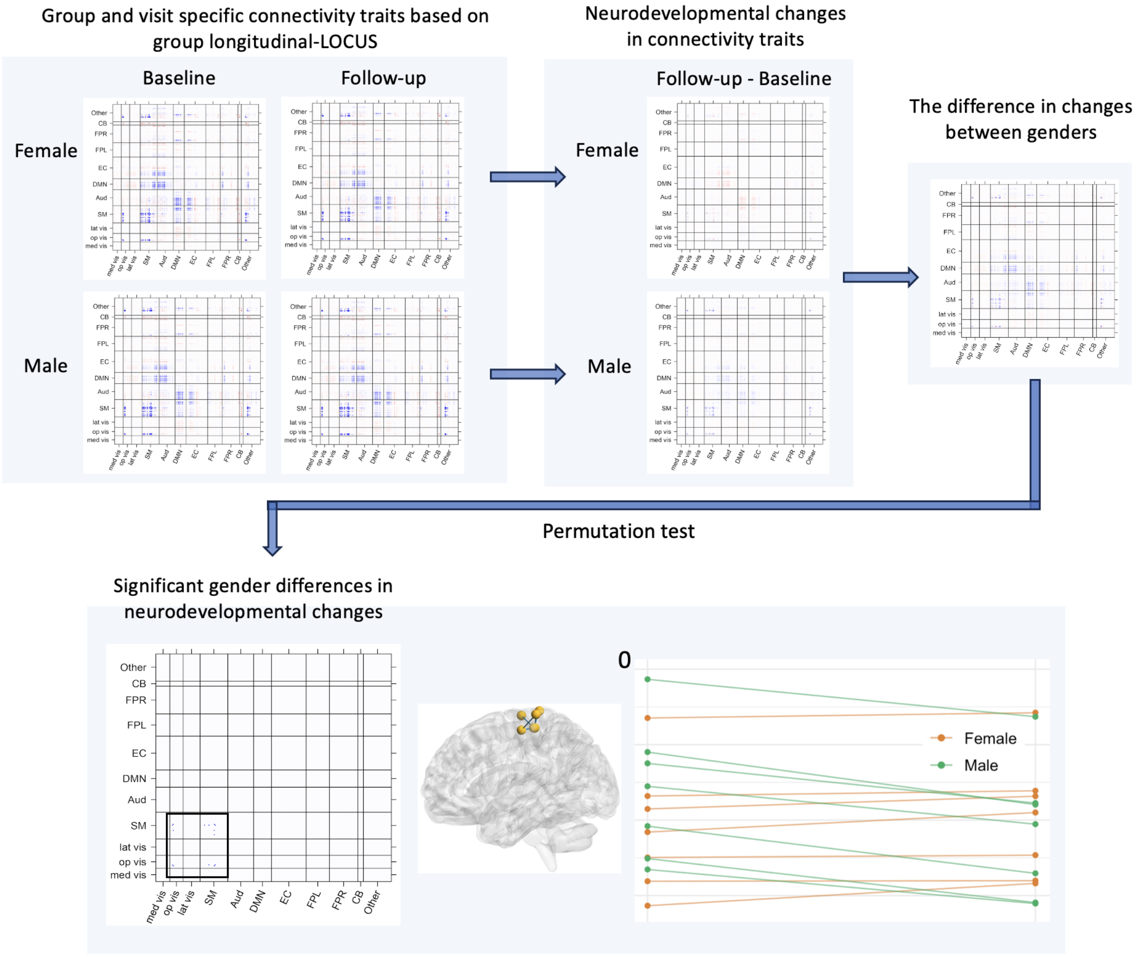


Figure 3.4: Trait 15 demonstrates significant neurodevelopmental differences between males and females

Chapter 4

Nonparametric Motion Control in Functional Connectivity Studies in Children with Autism Spectrum Disorder

4.1 Introduction

Early studies on neurodevelopment using functional magnetic resonance imaging found that short-range brain connections weakened and long-range brain connections strengthened during development (Fair et al., 2008), but these findings were undermined by the discovery that motion causes the same patterns (Van Dijk et al., 2012; Power et al., 2012; Colaço, 2024). This discovery led to the widespread adoption of motion quality control scan removal, which can result in drastic data loss. For example, multiple recent high-profile studies removed 60%-75% of approximately 11,500 children due to excessive motion (Marek et al., 2022; Nielsen et al., 2019). Removal of these children not only greatly decreases sample size, but also may introduce selection bias

(Cosgrove et al., 2022). There is a need to develop a statistical method that makes more efficient usage of the data and avoids selection bias in order to draw unbiased inferences about brain development.

Appropriate methods for motion quality control are especially warranted in studies of neurodevelopmental conditions. Autism spectrum disorder (ASD) is a neurodevelopmental condition with a rising prevalence that affects 1 in 36 children in the United States (Maenner, 2023). To study the neurobiology of ASD, investigators often use resting-state functional magnetic resonance imaging (rs-fMRI) to derive measures of *functional connectivity* between regions in the brain. Functional connectivity is commonly defined as the correlation between the blood oxygen level dependent signal of different brain regions across time, which provides an indirect measure of neuronal activity between those regions. Disruptions of functional connectivity have been proposed as an endophenotype in autistic children (Yerys et al., 2015), with decades of research reporting atypical patterns of functional connectivity in diverse brain systems (Hull et al., 2017). However, obtaining high-quality rs-fMRI data for functional connectivity analysis is challenging. Participants' head motion during the scanning session can induce *motion artifacts*, which can bias the analysis. The patterns of correlation induced by motion artifacts mimic the connectivity theory of autism, which predicts increased correlations between nearby brain regions and decreased correlations between distant brain areas (Deen and Pelphrey, 2012). Artifact-driven disruptions in brain networks can arise in comparisons of high and low motion rs-fMRI scans (Power et al., 2012).

Current guidelines for handling motion in rs-fMRI involve four steps, but the processing steps can introduce issues. First, rigid body motion correction is used to align fMRI volumes across time. Second, confound regression is applied. Several approaches to confound regression have been developed, which include regressing motion alignment parameters, global signal, cerebral spinal fluid signal, and white matter from the fMRI time courses, which may be combined with the removal of high-motion volumes or spike regression (Ciric et al., 2017). However, these approaches are still insufficient, and it is recommended to remove scans in which motion is deemed unacceptable (Power et al., 2014a). Nebel et al. (2022) found that 80% of autistic children compared to 60% non-ASD children were removed during quality control, and the removed autistic children had greater social deficits, worse motor control, and lower generalized ability index. Fourth, the effect of diagnosis is estimated using linear regression controlling for mean framewise displacement and demographic confounders (Yan et al., 2013), where the effect is estimated using only data that pass motion quality control.

Scan removal biases and inefficiencies have been largely unaddressed in the statistical and neuroimaging literature. Nebel et al. (2022) used an average-treatment effect that conditions on group in which the excluded scans are treated as missing data, which can address the selection bias issue. However, it does not utilize the outcome (resting-state correlations) in the excluded data, which may be inefficient. Additionally, it uses residuals from an initial linear regression of motion and demographic covariates restricted to the included scans, whereas a unified framework that flexibly models motion and demographic covariates has advantages. Sobel and Lindquist (2014) formulated a causal framework for task fMRI activation studies that considers the systematic error from motion, which is closely related to the nuisance regression step in resting-state fMRI preprocessing, but it does not address the problem of quality control in resting-state fMRI.

The objective of this study is to define and estimate an association between functional connectivity and ASD diagnosis that appropriately controls for the impact of motion. To this purpose, we use a causal mediation approach and consider motion as a mediator. To quantify the association between ASD and functional connectivity not related to motion, we conceptualize stochastic interventions (Muñoz and Van

Der Laan, 2012) on children's motion. Specifically, the stochastic intervention involves assigning the motion value based on a random sample from a specified distribution that corresponds to acceptable motion. By drawing children's motion from the same acceptable motion distribution across diagnostic groups, we are able to appropriately control for the impact of motion when evaluating associations of ASD diagnosis and functional connectivity. Our approach to motion control, which we call MoCo, is a novel solution to motion artifacts that uses all participants and avoids selection bias caused by motion quality control exclusion criteria. We propose an estimator and inference based on the efficient influence function that has multiple robustness properties. Our efficient estimator utilizes a data-adaptive ensemble of machine learning algorithms to flexibly model motion and other confounders (Van der Laan et al., 2007), while retaining $\sqrt{n}$-consistency (Van der Laan et al., 2011). Our analysis examines functional connectivity between a seed region in the default mode network and other brain regions in children in the Autism Brain Imaging Data Exchange (Di Martino et al., 2017).

4.2 Methods

Notation. Let $A \in \{0, 1\}$ denote the diagnosis group, which is equal to 1 if the participant has ASD and 0 otherwise. Let $M \in \mathcal{M}$ denote the motion variable. In our data application, we take M to be mean framewise displacement (FD), which is a commonly used measure of motion during the resting-state fMRI scan. Let $\Delta \in \{0, 1\}$ denote an inclusion indicator, which is equal to 1 if the participant meets a pre-specified set of criteria for inclusion in the study, related to the aggregate amount of movement during a child's scanning session. In our data application, we use the criteria from Power et al. (2014a), in which $\Delta = 1$ if a child has more than 5 minutes of data after removing frames with FD > 0.2 mm. Let $Y \in \mathcal{Y}$ denote the functional connectivity between two locations in the brain. For clarity, we initially define Y for a pair of regions, but in Section 7.4.2 we will extend it to the multivariate case with family-wise error control. Let $X \in \mathcal{X}$ denote covariates that are putatively related to functional connectivity and are possibly imbalanced across diagnosis groups. Such covariates could include age, sex, and handedness. Let $Z \in \mathcal{Z}$ be variables related to the diagnosis group and the pathophysiology of ASD. These are variables that have substantially different distributions with little or no overlapping support in ASD and non-ASD groups. Such variables could include scores on the autism diagnostic observation schedule (ADOS, a measure of social disability), full-scale intelligence quotient (FIQ) score, and current medication status. The distinction between X and Z is important: X are variables that would be balanced in an ideal experiment, while Z are variables that are related to diagnosis group, and thus have distributions that depend on A.

Let $O = (A, M, \Delta, X, Z, Y)$ represent a random variable with distribution P. Denote $O_1, \ldots, O_n$ as n i.i.d. observations of O, where $O_i = (A_i, M_i, \Delta_i, X_i, Z_i, Y_i)$. We assume $P \in \mathcal{P}$, where $\mathcal{P}$ is a statistical model for probability distributions on the support of O that is nonparametric up to certain positivity conditions that will be defined later in Section 7.3.1.

In our notation, an uppercase letter with no subscript denotes a random variable, an uppercase letter with an index, typically i, is an observed value of a random variable, and a lowercase letter indicates a typical realization of the random variable. For example, $E(Y \mid A)$ is a random variable, while $E(Y \mid A = a)$ is a scalar.

Let $P_{M|\Delta=1,A,X}(m \mid a, x)$ denote the probability distribution of M conditional on $\Delta = 1, A, X$ evaluated at value $(m, a, x) \in \mathcal{M} \times \{0, 1\} \times \mathcal{X}$. $P_{M|\Delta=1,A,X}$ is thus the probability distribution of motion given fixed diagnosis status A and covariates X, among children who meet the inclusion criteria. We use $p_{M|\Delta=1,A,X}(m \mid a, x)$ to denote a density with respect to some dominating measure. For simplicity, we write this density and all subsequent densities as being defined with respect to the Lebesgue measure. We follow similar conventions to define $p_{M|A,X,Z}(m \mid a, x, z)$ as the conditional density of M given A, X, Z and $p_{Z|A,X}(z \mid a, x)$ as the conditional density of Z given A, X.

Let $\mu_{Y|A,M,X,Z}(a, m, x, z)$ denote the conditional mean functional connectivity given diagnosis group $A = a$, motion level $M = m$, and covariates $(X, Z) = (x, z)$. Define $\pi_a(x)$ as the conditional probability that $A = a$ given $X = x$. We use an n-subscript to denote an estimate of the corresponding estimand, e.g., $\pi_{n,a}$ is an estimate of π_a.

4.2.1 Defining target parameter for group comparisons in fMRI studies

We propose a framework inspired by causal mediation analysis, whereby motion is considered a mediating variable. We aim to disentangle the impact of motion in the pathway between ASD and functional connectivity. To formalize our approach, we imagine a hypothetical experiment in which we are able to manipulate the motion of children during the scanning session. The hypothetical intervention is motivated by empirical studies in which children complete training in a mock scanner (De Bie et al., 2010). In the data set used in our analysis, children received at least one mock training, but oftentimes this training was insufficient to adequately control motion in the scanner. Therefore, we consider defining counterfactuals based on the hypothetical intervention that could be given to children prior to scanning that would successfully reduce motion during the scan. The hypothetical training program is applied to all children irrespective of diagnosis category. Our methods mathematically formalize the nature of the training program in terms of its impact on scanner motion, define the counterfactual data that would be observed under the hypothetical training program, describe assumptions under which inference can be made pertaining to the counterfactual data, and provide estimators of associations between diagnosis and functional connectivity that appropriately account for motion artifacts.

We first define the *total association* between ASD diagnosis and functional connectivity as $\theta_{\mathrm{O},1} - \theta_{\mathrm{O},0}$, where $\theta_{\mathrm{O},a} = E(E(Y \mid A = a, X))$, where the outer expectation is with respect to P_X, the marginal distribution of X. The total association can be interpreted as the mean difference functional connectivity between diagnosis groups after balancing measured confounders X and making *no effort* to control for motion. If X were balanced across diagnosis groups, i.e., if XA, then $\theta_{\mathrm{O},1} - \theta_{\mathrm{O},0} = E(Y \mid A = 1) - E(Y \mid A = 0)$. Therefore, the total association is adjusting for differences in the distribution of X across diagnosis groups, but nothing else. The subscript O denotes that this is an association in the *observed* data distribution that may include potentially inappropriate and imbalanced levels of motion.

Using a mediation-inspired approach, we then aim to isolate the associations of biological interest between ASD and functional connectivity from the total association by mitigating the influences of motion differences between diagnosis groups. This association of biological interest, which we term a *motion-controlled association* (MoCo), is an analogue to the classic direct effect in the causal mediation literature (Pearl, 2014). However, in this case, we do not view ASD as something that is inherently manipulable, so we restrict the interpretation of our analysis to describing direct *associations* of ASD with functional connectivity. On the other hand, we can view motion as something that *is* inherently manipulable.

We imagine a hypothetical training program where all children receive scanner training that results in a reduction of the motion of all children to a tolerable level, such that their data can be included in our analysis. The ideal motion distribution may be one where there is no motion whatsoever during a scanning session. However, such a choice is not practical since in the experimental data all children have at least *some* motion during a scanning session. Therefore, as a practical alternative, we suggest using an estimand inspired by a *stochastic intervention* on motion. In causal inference, *static interventions* imagine counterfactual scenarios wherein a variable is set to the same fixed value for everyone. On the other hand, a stochastic intervention is one in which the value of a variable is drawn at random from a user-specified distribution (Díaz et al., 2021). We define an *acceptable motion distribution* and imagine a counterfactual scenario where all children, irrespective of diagnosis category, receive sufficient training such that the distribution of motion that would be observed in the counterfactual scenario corresponds to this acceptable motion distribution. The acceptable motion distribution may be allowed to depend on covariates and should be selected to represent a distribution of motion that, if present in the experimental data, would still yield biologically meaningful readouts of functional connectivity.

One choice is to use $P_{M|\Delta=1,X}(\cdot \mid x)$ to generate the motion value for a child with covariates x. However, in our application, the distribution of motion in children that pass motion quality control differs between diagnostic groups (Figure 4.1). We observe that non-ASD children that pass quality control will generally only move within a tolerable level during a scanning session, such that we can still recover biologically meaningful signals from the resulting data. Thus, we suggest using $P_{M|\Delta=1,A,X}(\cdot \mid 0, x)$ as the acceptable distribution of motion. Let $M_0 \sim P_{M|\Delta=1,A,X}(m \mid 0, x)$ be the counterfactual motion that would be observed for a child in the counterfactual scenario under our stochastic intervention.

Let $Y(M_0)$ denote the counterfactual functional connectivity that would be measured under this training provided in our hypothetical experiment. The counterfactual data unit generated in this scenario is $O_{\mathrm{C}} = (A, M_0, X, Z, Y(M_0)) \sim P_{\mathrm{C}}$, where the probability distribution of evaluated at an observation $o_c = (a, m_0, x, z, y)$ can be written $P_{\mathrm{C}}(o_c) = P_{C,Y(M_0)|A,M,X,Z}(y \mid a, m_0, x, z) P_{M|\Delta=1,A,X}(m_0 \mid 0, x) P_{Z|A,X}(z \mid a, x) P_{A|X}(a \mid x) P_X(x)$. The scanner training program in our hypothetical experiment leads to a motion distribution under P_{C} that is described by $P_{M|\Delta=1,A,X}(\cdot \mid 0, x)$ as opposed to the motion distribution $P_{M|A,X,Z}(\cdot \mid a, x, z)$ observed in the real experiment.

The counterfactual motion is conditionally independent of diagnostic group (A) and diagnosis-specific variables (Z) given demographic covariates (X), which is consistent with our goal to control for the impact of motion in the two groups. Similarly, the conditional distribution of measured functional connectivity $Y(M_0)$ may be altered under P_{C} relative to the measured functional connectivity Y under P.

Using this counterfactual construction, we can then define a motion-controlled association $\theta_{\mathrm{C},1} - \theta_{\mathrm{C},0}$, where $\theta_{\mathrm{C},a} = E\{E_{\mathrm{C}}[Y(M_0) \mid A = a, X]\}$, and we use E_{C} to denote expectation under P_{C} and the subscript C to denote that this is an association defined with respect to a counterfactual distribution. The motion-controlled association provides a comparison of the ASD and non-ASD group that controls for (i) X differences between diagnosis groups (as with the total association); *and additionally* (ii) differences in motion between the two diagnosis groups by ensuring that the X-conditional motion distribution is the same across diagnosis groups.

4.2.2 Identifying the motion-controlled association of ASD and brain connectivity

Identifiability of $\theta_{\mathrm{C},a}$ for $a = 0, 1$ can be established under the following assumptions:

(A1) *Positivity*: (A1.1) for every x such that $p_X(x) > 0$, we also have $\pi_a(x) > 0$ for $a = 0, 1$; (A1.2) for every (x, z, m) such that $p_X(x)p_{Z|A,X}(z \mid a, x)p_{M|\Delta=1,A,X}(m \mid 0, x) > 0$, we also have that $p_{M|A,X,Z}(m \mid a, x, z) > 0$ for $a = 0, 1$.

(A2) *Mean exchangeability*: for all m such that $P\{p_{M|\Delta=1,A,X}(m \mid 0, X) > 0\} > 0$, $E_{\mathrm{C}}\{Y(m) \mid A = a, X, Z\} = E_{\mathrm{C}}\{Y(m) \mid A = a, M = m, X, Z\}$ a.e.-P.

(A3) *Causal Consistency*: for any child with observed motion value $M = m$, the observed functional connectivity measurement Y is equal to the counterfactual functional connectivity measurement $Y(m)$.

At the end of this section, we describe how our proposal yields biologically relevant inference even when (A2) and (A3) do not hold.

Assumption (A1.1) states that, at a population level, there cannot be values of X that are observed exclusively in the ASD group or exclusively in the non-ASD group. In our application, X consists of age, sex, and handedness. These characteristics do not perfectly predict ASD and therefore assumption (A1.1) is plausible. (A1.2) stipulates that, for both ASD and non-ASD children with the same value x of X, it is possible to observe the *same range* of motion values *irrespective* of the value of Z. Recall that one of the components of Z is a measure of social disability, and we expect that children with higher support needs will move more in the scanner. This assumption requires that it is possible to obtain some low-motion data even in more challenging cases. This assumption could be scrutinized empirically by studying the distribution of an estimate of the ratio of motion distributions $r_a(M, X, Z)$ introduced in equation (7.10). We show results of such an analysis in Supplementary Material Section 7.2. The analysis indicates the positivity assumptions are plausible for the real data analysis.

Assumption (A2) implies that conditioned on X, Z, and diagnosis status A, there is no unmeasured confounding between $Y(m)$ and M. The plausibility of this assumption could be scrutinized, for example, by establishing conditional d-separation of Y and M in a graph (Richardson and Robins, 2013). However, because this assumption is fundamentally an assumption on the counterfactual distribution P_{C}, it cannot be fully verified empirically.

Assumption (A3) stipulates that the observed functional connectivity from children who naturally have motion level m is the same as the functional connectivity that would have been observed under our hypothetical experiment where children receive training in the scanner. This assumption would be violated, for example, if the hypothetical scanner training received by children had an impact on the underlying functional connectivity of the child's brain. This assumption seems plausible but is not verifiable empirically.

Under these assumptions, we have the following theorem.

Theorem 1. *Under (A1)-(A3), the counterfactual* $\theta_{C,a}$ *is identified by* θ_a*, where* $\theta_a = \iiint \mu_{Y|A,Z,M,X}(a,z,m,x) p_{Z|A,X}(z \mid a,x) p_{M|\Delta=1,A,X}(m \mid 0,x) p_X(x) dz dm dx$.

The proof is in the Supplementary Material Section 1. The parameter θ_a involves integrating the conditional mean functional connectivity $\mu_{Y|A,M,X,Z}$ over the distributions of Z, M, and X in a sequential manner. We note that the integration over Z is specific to diagnosis group a, while the integration over M and X is the same irrespective of diagnosis group. Thus, a comparison of θ_1 and θ_0 provides a marginal associative measure that describes the joint impact of diagnosis category A and diagnosis-specific variables Z on functional connectivity while controlling for both motion M and covariates X.

We conclude this section by noting that inference on $\theta_1 - \theta_0$ may be biologically relevant *even in settings* where the fundamentally untestable assumptions (A2) and (A3) do not hold. To make this argument, we define

$$\eta_{\mu|A,M,X}(a,m,x) = \int \mu_{Y|A,M,X,Z}(a,m,x,z) p_{Z|A,X}(z \mid a,x) dz \ , \tag{4.1}$$

and note that $\eta_{\mu|A,M,X}(1,m,x) - \eta_{\mu|A,M,X}(0,m,x)$ describes an m- and x-specific difference in functional connectivity between diagnosis groups. Thus,

$$\theta_1 - \theta_0 = \iint \{\eta_{\mu|A,M,X}(1,m,x) - \eta_{\mu|A,M,X}(0,m,x)\} p_{M|\Delta=1,A,X}(m \mid 0,x) p_X(x) dm dx \ ,$$

simply standardizes these m- and x-specific associations over the selected acceptable motion distribution and distribution of covariates, thereby controlling for motion and covariate differences between diagnostic groups. We argue that this is likely still a biologically relevant parameter for describing differences in functional connectivity between diagnosis groups even when (A2)-(A3) do not hold.

4.3 Estimation and Inference

4.3.1 Efficiency theory

In this section, we develop an efficient estimator of θ_a. A key step in developing our estimator is deriving the *efficient influence function* of regular, asymptotically linear estimators of θ_a. See Supplementary Material Section 2.1 for a short review of efficiency theory. To characterize this efficient influence function, we define $\pi_{\Delta=1|A,X}(0,x) = P(\Delta = 1 \mid A = 0, X = x)$ as the probability of a non-ASD child with covariate value x having usable data. We introduce the shorthand $\bar{\pi}_0(x) = \pi_0(x)\pi_{\Delta=1|A,X}(0,x)$ as the probability that $A = 0$ and $\Delta = 1$ conditional on $X = x$. We denote the indicator function ${}_a(A_i)$ equal to 1 if $A_i = a$ and zero otherwise; ${}_{0,1}(A_i, \Delta_i)$ equal to 1 if $A_i = 0$ and $\Delta_i = 1$ and equals zero otherwise. We also define for $a = 0, 1$

$$r_a(m,x,z) = \frac{p_{M|\Delta=1,A,X}(m \mid 0,x)}{p_{M|A,X,Z}(m \mid a,x,z)} \ , \tag{4.2}$$

$$\eta_{\mu|A,Z,X}(a,z,x) = \int \mu_{Y|A,M,X,Z}(a,m,x,z)p_{M|\Delta=1,A,X}(m \mid 0,x)dm \ , \tag{4.3}$$

$$\xi_{a,\eta|X}(x) = \iint \mu_{Y|A,M,X,Z}(a,m,x,z)p_{M|\Delta=1,A,X}(m \mid 0,x)p_{Z|A,X}(z \mid a,x)dmdz \ . \tag{4.4}$$

In these definitions, we use a subscript notation for the functional parameters η and ξ that attempts to make explicit both the integrand in the parameter's definition, as well as the random variables that are arguments of the function. For example, the definition of $\eta_{\mu|A,Z,X}$ (4.3) involves integrating $\mu_{Y|A,M,X,Z}$, while $\eta_{\mu|A,Z,X}$ is a function of the random variables appearing in the subscript, A, Z and X. Note that Fubini's theorem allows us to write $\xi_{a,\eta|X}$ equivalently in terms of either $\eta_{\mu|A,Z,X}$ or $\eta_{\mu|A,M,X}$,

$\xi_{a,\eta|X}(x) = \int \eta_{\mu|A,Z,X}(a,x,z) p_{Z|A,X}(z \mid a,x) dz = \int \eta_{\mu|A,M,X}(a,m,x) p_{M|\Delta=1,A,X}(m \mid 0,x) dm$.

Theorem 2. *(Efficient Influence Function). In a nonparametric model, the efficient influence function for* θ_a *evaluated on a typical observation* O_i *is*

$$\begin{aligned} D_{P,a}(O_i) = {} & \frac{{}_{a}(A_i)}{\pi_a(X_i)} r_a(M_i, X_i, Z_i) \left\{ Y_i - \mu_{Y|A,M,X,Z}(a, M_i, X_i, Z_i) \right\} \\ & + \frac{{}_{a}(A_i)}{\pi_a(X_i)} \left\{ \eta_{\mu|A,Z,X}(a, X_i, Z_i) - \xi_{a,\eta|X}(X_i) \right\} \\ & + \frac{{}_{0,1}(A_i, \Delta_i)}{\bar{\pi}_0(X_i)} \left\{ \eta_{\mu|A,M,X}(a, M_i, X_i) - \xi_{a,\eta|X}(X_i) \right\} \\ & + \xi_{a,\eta|X}(X_i) - \theta_a. \end{aligned} \tag{4.5}$$

A proof is included in the Supplementary Material Section 2.2.

We use the one-step estimation framework to define efficient estimators of θ_a (Bickel et al., 1993). Suppose we have an estimate of $\xi_{a,\eta|X}$ available, say $\xi_{n,a,\eta|X}$. An estimate of θ_a can be obtained by marginalizing $\xi_{n,a,\eta|X}$ over the empirical distribution of X, leading to an estimate of the form $\theta_{n,a} = n^{-1} \sum_{i=1}^{n} \xi_{n,a,\eta|X}(X_i)$. We refer to $\theta_{n,a}$ as a *plug-in estimate*. A *one-step estimator* of θ_a can be constructed as $\theta_{n,a}^{+} = \theta_{n,a} + n^{-1} \sum_{i=1}^{n} D_{n,a}(O_i)$, where $D_{n,a}$ is an estimate of $D_{P,a}$. Thus, to construct a one-step estimate of θ_a, we require as an intermediate step estimates of the various parameters of P that appear in $D_{P,a}$. We refer to these quantities as *nuisance parameters*; they are parameters that need to be estimated as an intermediate step in the estimation of θ_a.

Examining Theorem 2, we find that there are several nuisance parameters that appear in $D_{P,a}$ for which we will require estimates to construct our estimate $D_{n,a}$ of $D_{P,a}$. Estimation of several of these parameters is straightforward. For example, $\mu_{Y|A,M,X,Z}$ could be estimated using mean regression of Y on A, M, X, Z. On the other hand, the η and ξ parameters involve integration and conditional densities, which generally present practical challenges in implementation. Our approach for estimation outlined below emphasizes two key points: (i) wherever possible mean regression with pseudo-outcomes is used to avoid numeric integration and conditional density estimation and (ii) flexible estimation techniques are used.

We choose to emphasize the use of mean regression because it is a technique familiar to many applied statisticians and there are widely available tools. In our application, we focus on a flexible framework for regression, known as regression stacking or super learning (Van der Laan et al., 2007). Super learning is implemented by pre-specifying a so-called *library* of candidate regression estimators. Cross-validation is used to build a weighted combination of these estimators, with large sample theory indicating that the ensemble estimator will provide nuisance parameter estimates that are essentially as good or better than any of the individual candidate regression estimators considered.

Unfortunately, mean regression cannot be used exclusively in our estimation process for θ_a. We require estimates of certain conditional motion distributions described below. For this, we utilize the highly adaptive lasso, a flexible semiparametric conditional density estimator (Hejazi et al., 2022b).

To circumvent numerical integration, we make use of a technique proposed by Díaz et al. (2021) that re-casts these estimation problems that involve integrals and densities as an estimation problem that can be solved using mean regression with pseudo-outcomes. This technique is motivated by the fact that the definition of $\eta_{\mu|A,Z,X}(a,z,x)$ in (4.3) is equivalent to

$$\begin{aligned}\int \mu_{Y|A,M,X,Z}(a,m,x,z)\frac{p_{M|\Delta=1,A,X}(m \mid 0,x)}{p_{M|\Delta=1,A,X,Z}(m \mid a,x,z)}p_{M|\Delta=1,A,X,Z}(m \mid a,x,z)dm \\ = E\left[\mu_{Y|A,M,X,Z}(A,M,X,Z)\frac{p_{M|\Delta=1,A,X}(M \mid 0,X)}{p_{M|\Delta=1,A,X,Z}(M \mid A,X,Z)} \;\middle|\; \Delta=1, A=a, X=x, Z=z\right].\end{aligned}$$

This equivalence suggests that $\eta_{\mu|A,Z,X}$ could be estimated using mean regression, where a pseudo-outcome $\mu_{n,Y|A,M,X,Z}(A,M,X,Z)\frac{p_{n,M|\Delta=1,A,X}(M|0,X)}{p_{n,M|\Delta=1,A,X,Z}(M|A,X,Z)}$ is regressed onto A, X, and Z using only the observations with $\Delta = 1$. In this way, we can avoid both the challenges associated with numeric integration and instead utilize super learning-based mean regression for estimation. Similar techniques involving pseudo-outcomes can be applied for the estimation of $\eta_{\mu|A,M,X}$ and $\xi_{a,\eta|X}$, as described below.

4.3.2 Detailed implementation

Our estimator can be implemented in the following steps.

1. *Estimate mean functional connectivity* $\mu_{Y|A,M,X,Z}$. Fit a super learner regression using functional connectivity as the outcome and including diagnosis category A, mean FD M, demographic covariates X, and diagnosis-specific covariates Z as predictors in the super learner. Using this fit, evaluate the fitted value, $\mu_{n,Y|A,M,X,Z}(a, M_i, X_i, Z_i)$ for $i = 1, \ldots, n$ and for $a = 0, 1$. For a particular value of a, this can be achieved by predicting from the fitted super learner using the observed values of M, X, and Z, but replacing the observed value of A with the constant value a. Below we refer to this process as *evaluating the fitted value from the regression, setting A to a.*

2. *Estimate motion distributions* $p_{M|A,X}$, $p_{M|\Delta=1,A,X}$, $p_{M|A,X,Z}$, and $p_{M|\Delta=1,A,X,Z}$. For $p_{M|A,X}$, using the highly adaptive LASSO, estimate the conditional density of M given diagnosis A and demographic covariates X. The other three densities are estimated by further conditioning on diagnosis-specific covariates Z and/or subsetting to $\Delta = 1$. Using these fits, evaluate $p_{n,M|A,X}(M_i \mid a, X_i)$, $p_{n,M|\Delta=1,A,X}(M_i \mid a, X_i)$, $p_{n,M|A,X,Z}(M_i \mid a, X_i, Z_i)$, $p_{n,M|\Delta=1,A,X,Z}(M_i \mid a, X_i, Z_i)$ for $a = 0, 1$ and $i = 1, \ldots, n$.

3. *Estimate motion-standardized functional connectivity* $\eta_{\mu|A,Z,X}$ *using pseudo-outcome regression.* Using estimates obtained in steps 1 and 2, for $i = 1, \ldots, n$ create the pseudo-outcome $\hat{Y}_{M,i} = \mu_{n,Y|A,M,X,Z}(A_i, M_i, X_i, Z_i)\frac{p_{n,M|\Delta=1,A,X}(M_i|0,X_i)}{p_{n,M|\Delta=1,A,X,Z}(M_i|A_i,X_i,Z_i)}$. Using only observations with $\Delta_i = 1$, fit a super learner regression using $\hat{Y}_M$ as the outcome and including diagnosis category A, diagnosis-specific covariates Z, and demographic covariates X as predictors. Evaluate the fitted value from this regression setting A to

a to obtain $\eta_{n,\mu|A,Z,X}(a, Z_i, X_i)$ for $i = 1, \ldots, n$.

4. *Estimate Z-standardized functional connectivity $\eta_{\mu|A,M,X}$ using pseudo-outcome regression.* Use estimates obtained in steps 1 and 2, and for $i = 1, \ldots, n$ to create the pseudo-outcome $\hat{Y}_{Z,i} = \mu_{n,Y|A,M,X,Z}(A_i, M_i, X_i, Z_i)\frac{p_{n,M|A,X}(M_i|A_i,X_i)}{p_{n,M|A,X,Z}(M_i|A_i,X_i,Z_i)}$. Fit a super learner regression using $\hat{Y}_Z$ as the outcome and including M, demographic covariates X, and diagnosis category A as predictors. Evaluate the fitted value from this regression setting A to a to obtain $\eta_{n,\mu|A,M,X}(a, M_i, X_i)$ for $i = 1, \ldots, n$.

5. *Estimate motion- and Z-standardized functional connectivity $\xi_{a,\eta|X}$.* Fit a super learner regression using $\eta_{n,\mu|A,Z,X}$ as the outcome, include diagnosis category A and demographic-specific covariates X as predictors in the super learner. For $a = 0, 1$, evaluate the fitted value from this regression setting A to a to obtain $\xi_{n,a,\eta|X}(X_i)$ for $i = 1, \ldots, n$.

6. *Calculate plug-in estimate.* Define the plug-in estimate $\theta_{n,a} = n^{-1}\sum_{i=1}^{n}\xi_{n,a,\eta|X}(X_i)$.

7. *Estimate diagnosis distribution π_a and inclusion probability $\pi_{\Delta=1|A,X}$.* Fit a super learner regression using the diagnosis category as the outcome and including demographic covariates X as predictors. Evaluate the fitted value $\pi_{n,1}(X_i)$ for $i = 1, \ldots, n$ and set $\pi_{n,0}(X_i) = 1-\pi_{n,1}(X_i)$. Then fit an additional super learner using Δ as the outcome and including diagnosis category A and demographic covariates X as predictors in the regression. Evaluate the fitted value from this regression setting A to 0 to obtain $\pi_{n,\Delta=1|A,X}(0, X_i)$ for $i = 1, \ldots, n$. Compute $\bar{\pi}_{n,0}(X_i) = \pi_{n,0}(X_i)\pi_{n,\Delta=1|A,X}(0, X_i)$ for $i = 1, \ldots, n$.

8. *Evaluate estimated efficient influence function $D_{n,a}(O_i)$.* For $a = 0, 1$ and each $i = 1, \ldots, n$, evaluate $D_{n,a}(O_i)$ by substituting the fitted values based on the estimated nuisance parameters obtained in steps 1-7 into equation (7.13).

9. *Compute the one-step estimator.* For $a = 0, 1$, compute $\theta^{+}_{n,a} = \theta_{n,a}+n^{-1}\sum_{i=1}^{n}D_{n,a}(O_i)$.

4.3.3 Inference

Below we present two theorems establishing the consistency and asymptotic linearity, respectively, of the one-step estimator $\theta^+_{n,a}$. We define $\|\cdot\|$ to be the $L^2(P)$ norm of a given function f defined as $\|f\| = E[f(O)^2]^{1/2}$. We note that for the purposes of this definition, the function f is treated as given, even if it involves estimated quantities. Theorem 3 assumes the following:

(B1) *Boundedness:* $\pi_{n,a}$ is bounded below by some $\epsilon_1 > 0$, $\bar{\pi}_{n,0}$ is bounded below by some $\epsilon_2 > 0$, and $p_{n,M|A,X,Z}(m \mid a, x, z)$ is bounded below by some $\epsilon_3 > 0$.

(B2) $o_p(1)$*-convergence of certain combinations of nuisance parameters*: certain subsets of the nuisance parameters are consistently estimated, as described in Table 4.1.

	$\mu_{n,Y\|A,M,X,Z}$	$\eta_{n,\mu\|A,M,X}$	$\xi_{n,a,\eta\|X}$	$\bar{\pi}_{n,0}$	$\pi_{n,a}$	$p_{n,M\|\Delta=1,A,X}$	$p_{n,M\|A,X,Z}$
(B2.1)					✓	✓	✓
(B2.2)			✓			✓	✓
(B2.3)	✓	✓		✓	✓		
(B2.4)	✓				✓	✓	
(B2.5)	✓		✓			✓	

Table 4.1: Assumption (B2) of Theorem 3.2 (multiple robustness). Each row indicates a setting for consistency, where check marks indicate the nuisance parameters which, when they converge to true functions combined with assumptions (B1), (B3) and (B4), result in the consistency of $\theta^+_{n,a}$.

(B3) $L^2(P)$*-consistent influence function estimate*: $E[\{D_{P_\ell,a}(O) - D_{n,a}(O)\}^2] = o_P(1)$, where $D_{P_\ell,a}$ denotes the in-probability limit of $D_{n,a}$ as n approaches infinity and $D_{n,a}$ is treated as a fixed function of O in this expression.

(B4) *Glivenko Cantelli influence function estimate*: the probability that $D_{n,a}$ falls in a P-Glivenko Cantelli class tends to one as $n \to \infty$.

Assumption (B1) guarantees that estimated propensities and motion densities are appropriately bounded so that the one-step estimator is never ill-defined. Assumption

(B2) stipulates consistent estimations of the nuisance parameters. Assumptions (B3) and (B4) are necessary to ensure the negligibility of an empirical process term (Van Der Vaart et al., 1996).

Theorem 3. (Multiple robustness of the one-step estimator). *Under assumptions (B1) - (B4),* $\theta^{+}_{n,a} - \theta_a = o_p(1)$.

According to Theorem 3, our one-step estimators will only require *some* of the nuisance parameters to be consistently estimated to achieve consistency of our estimate of θ_a. For example, assumption (B2.1) implies that obtaining consistent estimates of the conditional motion densities, $p_{M|\Delta=1,A,X}(m \mid 0, x)$ and $p_{M|A,X,Z}(m \mid a, x, z)$, and the conditional probability of ASD as a function of covariates π_a is sufficient to ensure a consistent estimator of θ_a. A proof of the theorem is in the Supplementary Material Section 3.

The following theorem characterizes the large-sample behavior of the proposed estimator.

Theorem 4. (Asymptotic linearity of the one-step estimator). *Under (B1), (B3), and*

(C1) $n^{1/2}$-convergence of second order terms*:*

$$\|\xi_{n,a,\eta|X} - \xi_{a,\eta|X}\| \|\pi_{n,a} - \pi_a\| = o_P(n^{-1/2}),$$

$$\|\mu_{n,Y|A,M,X,Z} - \mu_{Y|A,M,X,Z}\| \left\{ \|p_{n,M|\Delta=1,A,X}(\cdot \mid 0, \cdot) - p_{M|\Delta=1,A,X}(\cdot \mid 0, \cdot)\| + \|p_{n,M|A,X,Z} - p_{M|A,X,Z}\| \right\} = o_P(n^{-1/2}),$$

and

$$\|p_{n,M|\Delta=1,A,X}(\cdot \mid 0, \cdot) - p_{M|\Delta=1,A,X}(\cdot \mid 0, \cdot)\| \left\{ \|\eta_{n,\mu|A,M,X} - \eta_{\mu|A,M,X}\| + \|\pi_{n,a} - \pi_a\| + \|\bar{\pi}_{n,0} - \bar{\pi}_0\| \right\} = o_P(n^{-1/2}).$$

(C2) Donsker influence function estimate*: the probability that* $D_{n,a}$ *falls in a* P*-Donsker class tends to one as* $n \to \infty$.

then $\theta^+_{n,a} - \theta_a = \frac{1}{n}\sum_{i=1}^n D_{P,a}(O_i) + o_P(n^{-1/2})$, *and*

$$n^{1/2}(\theta^+_{n,a} - \theta_a) \Rightarrow N(0, E[D_{P,a}(O)^2]) \ .$$

Assumption (C1) states that nuisance estimates converge to their true values at a sufficiently fast rate (so-called *quarter-rate conditions*). Assumption (C2) ensures large-sample negligibility of a certain second-order empirical process term (so-called *Donsker conditions*, Bickel et al. 1993). The Donsker conditions can be eliminated through the use of cross-fitting, as described in Section 7.4.2. A detailed discussion of the assumptions and the proof of the theorem are in the Supplementary Material Section 4.

When all nuisance regressions are consistently estimated, $\sigma_n^2 = (n-1)^{-1}\sum_{i=1}^n\{D_{n,a}(O_i) - n^{-1}\sum_{j=1}^n D_{n,a}(O_j)\}^2$ can be used as a consistent estimate of $E[D_{P,a}(O)^2]$. Thus, an asymptotically justified $1-\alpha$ confidence interval for θ_a may be constructed as $\theta^+_{n,a} \pm n^{-1/2} z_{1-\alpha/2}\sigma_n$, where $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of a standard Normal distribution.

By Theorem 4, we have $(\theta^+_{n,1} - \theta^+_{n,0}) - (\theta_1 - \theta_0) = \frac{1}{n}\sum_{i=1}^n\{D_{P,1}(O_i) - D_{P,0}(O_i)\} + o_P(n^{-1/2})$ and the limiting distribution of $n^{1/2}\{(\theta^+_{n,1} - \theta^+_{n,0}) - (\theta_1 - \theta_0)\}$ is $N(0, \tau^2)$, with $\tau^2 = \text{Var}(D_{P,1}(O) - D_{P,0}(O))$. The estimate $\tau_n^2 = (n-1)^{-1}\sum_{i=1}^n\{D_{n,1}(O_i) - D_{n,0}(O_i) - n^{-1}\sum_{j=1}^n(D_{n,1}(O_j) - D_{n,0}(O_j))\}^2$ will be consistent for τ^2 under the assumptions of Theorem 4. Consequently, an approximate $1-\alpha$ confidence interval for the direct association of ASD with brain connectivity in a single brain region is $(\theta^+_{n,1} - \theta^+_{n,0}) \pm n^{-1/2} z_{1-\alpha/2}\tau_n$.

4.3.4 Cross-fit one-step estimation

As mentioned above, cross-fitting avoids the necessity of Donsker conditions in the proof of Theorem 4, which may afford us the ability to utilize more aggressive machine learning techniques as part of the super learner, while still generating well-calibrated confidence intervals and hypothesis tests. Cross-fitting can also reduce finite-sample bias and improve confidence interval coverage in some settings (Zivich and Breskin, 2021).

The cross-fitting process involves randomly dividing the data set into K parts, followed by separate cross-validation routines on each part. $K-1$ parts of the data are used to estimate the nuisance parameters appearing in the efficient influence function using super learner with K'-fold cross-validation. In practice, we use $K = 5$ and $K' = 10$. Cross-fitting is used for all nuisance regressions and conditional density estimates. Consider the example of $\xi_{a,\eta|X}$. We denote by $\xi_{n,k,a,\eta|X}$ the estimate of $\xi_{a,\eta|X}$ obtained when the k-th part of the data is withheld from the nuisance estimation stage. Similarly, we denote by $D_{a,n,k}$ the efficient influence function evaluated at the nuisance parameters estimated without using the k-th part of the data. Denote by $\mathcal{I}_k$ the indices of observations in the k-th part of the data and denote the number of observations in this set by n_k. The cross-fit estimate of θ_a is $\theta^{\text{cf}}_{n,a} = \frac{1}{K}\sum_{k=1}^{K}\left[\frac{1}{n_k}\sum_{i\in\mathcal{I}_k}\xi_{n,k,a,\eta|X}(X_i) + \frac{1}{n_k}\sum_{i\in\mathcal{I}_k}D_{a,n,k}(O_i)\right]$. The asymptotic linearity of the cross-fit one-step estimator follows using the same arguments as in Theorem 4, where nuisance estimates are replaced by their k-specific counterparts and assumption (C2) is removed (van der Laan et al., 2011; Chernozhukov et al., 2018).

4.3.5 Simultaneous inference for associations

In order to control the family-wise error rate for tests of direct associations between ASD and brain connectivity across hundreds of regions, we conduct hypothesis testing using simultaneous confidence bands (Ruppert et al., 2003). Let $j = 1, \dots, J$ index the region. In our application, $J = 399$, and we use $\theta_{a,j}$ to denote the motion-controlled average functional connectivity in diagnosis group a between a seed region and region j, based on a 400-region parcellation. Similarly, we denote by $D_{P,a,j}$ the efficient influence function for diagnosis group a and region j, and by $\tau_{n,j}^2$ the region-specific estimate of the asymptotic variance. By Theorem 4,

$$n^{1/2}\left\{\begin{pmatrix}\theta_{n,1,1}^{+}-\theta_{n,0,1}^{+}\\ \vdots \\ \theta_{n,1,J}^{+}-\theta_{n,0,J}^{+}\end{pmatrix}-\begin{pmatrix}\theta_{1,1}-\theta_{0,1}\\ \vdots \\ \theta_{1,J}-\theta_{0,J}\end{pmatrix}\right\}\Rightarrow N\left\{\begin{pmatrix}0\\ \vdots \\ 0\end{pmatrix},\mathrm{Cov}\begin{pmatrix}D_{P,1,1}(O)-D_{P,0,1}(O)\\ \vdots \\ D_{P,1,J}(O)-D_{P,0,J}(O)\end{pmatrix}\right\}, \tag{4.6}$$

where $\theta_{n,a,j}^{+}$ is the estimator at $A=a$ at location $j\in\{1,...,J\}$. An approximate $1-\alpha$ simultaneous confidence interval is $(\theta_{n,1,1}^{+}-\theta_{n,0,1}^{+},\dots,\theta_{n,1,J}^{+}-\theta_{n,0,J}^{+})^{\top}\pm z_{\max,1-\alpha}(\tau_{n,1},\dots,\tau_{n,J})^{\top}$, where $z_{\max,1-\alpha}$ is the $1-\alpha$ quantile of the random variable $\max_{1\le j\le J}\{n^{1/2}|(\theta_{n,1,j}^{+}-\theta_{n,0,j}^{+})-(\theta_{1,j}-\theta_{0,j})|/\tau_{n,j}\}$, which depends on the covariance matrix in (7.14).

To approximate $z_{\max,1-\alpha}$, Monte-Carlo integration is performed by taking 10^5 independent draws of a J-dimensional mean-zero multivariate normal random variable with covariance matrix equal to an empirical estimate of the correlation matrix derived from the covariance matrix on the right-hand side of (7.14). This correlation matrix can be estimated via the empirical correlation of the vector $(D_{n,1,1}(O)-D_{n,0,1}(O),\dots,D_{n,1,J}(O)-D_{n,0,J}(O))^{\top}$. For each of the 10^5 random draws, the maximal absolute value of the components of the vector is calculated. The critical value $z_{\max,1-\alpha}$ is approximated by calculating the empirical $(1-\alpha)$-quantile of these maximum values. Wald hypothesis tests controlling family-wise error rate at level α are conducted by rejecting the null hypothesis of no association between diagnosis group and functional connectivity in the j-th region whenever $n^{1/2}|\theta_{n,1,j}^{+}-\theta_{n,0,j}^{+}|/\tau_{n,j}$ is larger than the estimated value of $z_{\max,1-\alpha}$.

4.4 Simulation study

We compared MoCo to two naïve approaches: one that mimics the removal of high-motion participants and one that does not remove any participants. We fixed the sample size to $n = 400$ and simulated covariates that are similar in distribution to the covariates in the observed data. We summarize our simulation design here, with details provided in Supplementary Material Section 5. The simulated demographic covariates X had three dimensions corresponding to sex, age, and handedness. Given covariates $X = x$, diagnosis A was drawn from a Bernoulli distribution with success probabilities defined by coefficients from logistic regression of the real data. Next, four diagnosis-specific covariates Z were generated to mimic the autism diagnostic observation schedule (ADOS), full-scale IQ (FIQ), a binary variable indicating the use of stimulant medication, and a binary variable for non-stimulants. The natural logarithm of mean FD M was generated from a normal distribution with mean defined from estimated coefficients of $A = a, X = x, Z = z$. Tolerable motion Δ was equal to 1 for $M \leq 0.2$.

We simulated the functional connectivity between a seed region equal to the default mode network and the six other parcels defined in the Yeo 7 parcellation (Yeo et al., 2011). We denote by $Y_1, \ldots, Y_6$ the simulated functional connectivity for these six parcels. We simulated 1000 data sets such that the true associations between the diagnosis group and functional connectivity for parcels $Y_1, \ldots, Y_4$ were set to zero, while Y_5 and Y_6 were simulated to have non-zero associations with diagnosis. This allowed us to evaluate both the type I error and the power. We simulated functional connectivity such that large and small negative associations existed between the diagnosis group and Y_5 and Y_6, respectively. The data generating process also included quadratic associations between motion and observed functional connectivity to examine the ability of super learner to account for possible non-linear relationships in the data. The covariance matrix for the multivariate normal was set equal to the covariance matrix of the errors obtained when fitting the functional connectivity using Super Learner. See Supplementary Material Section 5 for details.

We included the mean of the outcome, multivariate adaptive regression splines, LASSO, ridge regression, generalized additive models, generalized linear models (with and without interactions, and with and without forward stepwise covariate selection), random forest, and xgboost as candidate regressions in super learner (Polley et al., 2023). We compared MoCo to two naïve approaches. First, we calculated the sample mean difference in average functional connectivity between the ASD group and the non-ASD group and conducted a Welch's two-sample t-test. Second, we excluded high-motion participants ($\Delta = 0$).

MoCo demonstrated advantages in terms of bias, MSE, type I error, and power (Table 4.2). For regions with zero associations, the bias and type I error were lower in MoCo compared to the two naïve methods for all four regions, and the MSE was lowest in three of four regions. In regions with associations, MoCo exhibited greater power and lower bias in detecting these differences compared to the naïve methods. Figure 4.2 illustrates the results of MoCo with cross-fitting on one simulated dataset. The true association was that children with ASD had a decreased correlation between the default mode network and two large brain regions, marked in dark green and purple, with an association of zero for other brain regions. MoCo successfully recovered one of the two true associations. However, using the naïve method with participant removal resulted in missing both regions, while using the naïve method with all data picked up false positives and failed to recover any of the regions with true associations.

In the Supplementary Material Section 6, we include an additional simulation study examining the multiple robustness property of our estimators. The simulation demonstrates that the theoretical multiple robustness result has the practical impact that bias of the estimators may be well controlled even in settings in which some of the nuisance parameters are inconsistently estimated.

4.5 Data analysis of functional connectivity in ASD

4.5.1 Data and methods

We conducted a functional connectivity analysis using a seed region in the default mode network to investigate selection bias and motion impacts in ASD. The default mode network is a collection of brain regions that tend to co-activate during wakeful rest, including daydreaming or mind wandering. Hypoconnectivity between anterior and posterior parts of the default mode network was previously found in the Autism Brain Imaging Data Exchange (ABIDE) dataset (Di Martino et al., 2014). The default mode network has been identified as a possible endophenotype of ASD (Yerys et al., 2015). However, hypoconnectivity in the default mode network also arises from motion artifacts (Power et al., 2012). We applied our method to resting-state fMRI data from school-age children in the ABIDE dataset (Di Martino et al., 2014, 2017). We subset to 377 scans corresponding to 8 to 13-year-old children from the Kennedy Krieger Institute (KKI) and New York University (NYU) from ABIDE I and ABIDE II (Supplement Table 5). Imaging data were preprocessed using fMRIprep (Esteban et al., 2019) as described in Supplement Section 7. We included the following covariates: diagnosis (A); age, sex, and handedness (X); Autism Diagnostic Observation Schedule (ADOS) score, Full-scale Intelligence Quotient (FIQ) score, stimulant medication status, and non-stimulant medication status (Z); and mean FD (M). The ADOS score is a standardized assessment tool used to diagnose ASD, with higher scores indicating greater social disability. Although ASD is more prevalent in males than females (Maenner, 2023), we treated sex as a confounder (X) rather than diagnosis-specific variable (Z) because sex-specific differences in functional connectivity have been previously documented (Shanmugan et al., 2022), which could mask ASD-related differences in this cohort. Mean FD is an average of the frame-to-frame displacement calculated from the rigid body motion correction parameters used in quality control and motion correction (Esteban et al., 2019; Power et al., 2014a; Di Martino et al., 2014).

We calculated the average time series for regions of interest defined using Schaefer's 400-node brain parcellation, which associates each node with the resting networks from Yeo-7 (Schaefer et al., 2018). We then regressed the six parameters from motion alignment, the global signal (mean signal across all voxels), white matter (WM), and cerebrospinal fluid (CSF) calculated from fMRIPrep. We used COMBAT (Yu et al., 2018) for site harmonization to account for three protocols. For COMBAT, "site" is a factor with three levels (NYU, KKI-8 channel, KKI-32 channel) and the following covariates: diagnosis, age, sex, handedness, ADOS score, FIQ, stimulant medication status, non-stimulant medication status, and mean FD.

We calculated Fisher z-transformed correlations of every brain region with region 14 ('17networks_LH_DefaultA_pCun_1'), which is a hub of the posterior default mode network (Pham et al., 2022). We defined the indicator of data usability Δ equal to one if a child had more than 5 minutes of data after removing frames with FD $>$ 0.2 mm (Power et al., 2014a). This resulted in $\Delta = 0$ for 98/132 ASD (74.2%) and 119/245 non-ASD children (48.6%).

We compared the naïve estimate with participant removal (retaining $\Delta = 1$), the naïve estimate with no participant removal, and the group difference estimated from MoCo with cross-fitting. We determined the FWER-critical values using simultaneous confidence intervals derived from the multivariate efficient influence functions (Section 7.4.2). For the naïve approaches, the critical values were derived from the sample correlation matrices of residuals of the regression $Y \sim A$. Nuisance regressions involved in our method are estimated using the same super learner library as the simulations. To handle Monte Carlo variability resulting from cross-validation, we generated estimates 50 times and calculated the averaged estimates and z-statistics across runs.

We examined positivity assumptions by making histograms of the inverse probability weights and density ratios that appear in the efficient influence function (see Supplementary Material Section 7.2). Positivity issues arise if inverse weights or density ratios are very large. Assumptions appear to be adequately met, as the values of the estimated density ratios evaluated on the observed data never exceeded 4.

4.5.2 Results

Both MoCo and the naïve approach use data from 377 participants, including 132 with ASD, while the naïve approach with participant removal uses 160 participants, and only 34 with ASD. MoCo reveals four regions that differ in connectivity with the posterior default mode seed region in ASD versus non-ASD at FWER=0.05, including three regions of hyperconnectivity with distant frontal-parietal regions (Figure 4.3). The naïve approach indicates more extensive differences than MoCo, including prominent default mode hypoconnectivity in ASD in long-distance correlations. These are likely spurious differences due to motion, as long-distance correlations tend to be attenuated in high-motion participants (Satterthwaite et al., 2013b). These possible biases are also prominent in the mean connectivity estimates (Supplement Figure 3). The naïve approach also selects some regions of hyperconnectivity in ASD with lateral regions of the frontal lobe. Overall, MoCo results are more similar to the naïve approach with participant removal, although the naïve with participant removal only identifies two regions at FWER=0.05. At FWER=0.20, MoCo selects four additional regions of hypoconnectivity with the far anterior part of the frontal lobe (plus one region of hyperconnectivity with ventral attention), but much less extensive than in the naïve approach.

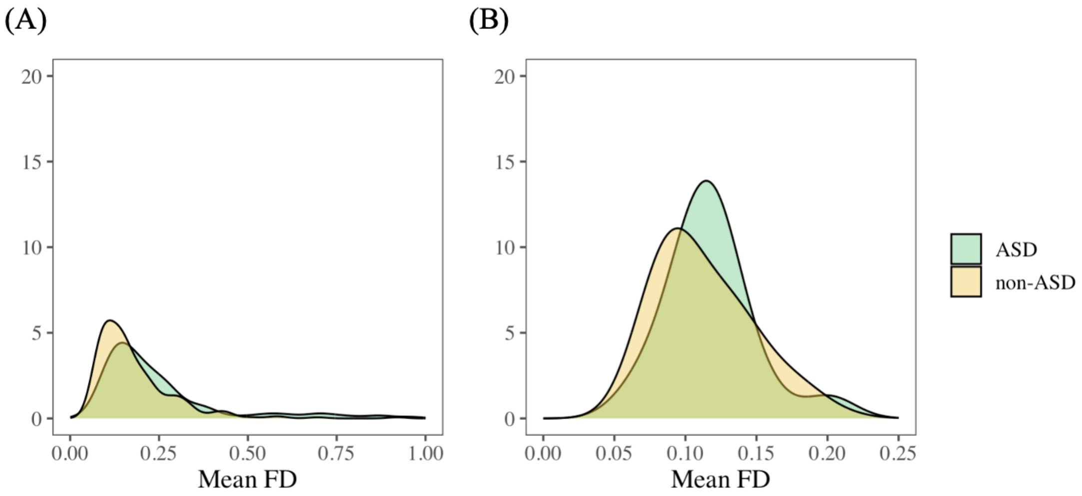


Figure 4.1: Distributions of mean framewise displacement (FD) in the school-age children dataset. Panel A shows the distribution of mean FD over all children. Panel B shows the distribution of mean FD over children who meet the inclusion criteria. The distribution of motion in non-ASD children that pass motion quality control differs from the distribution of motion in children with ASD that pass motion quality control.

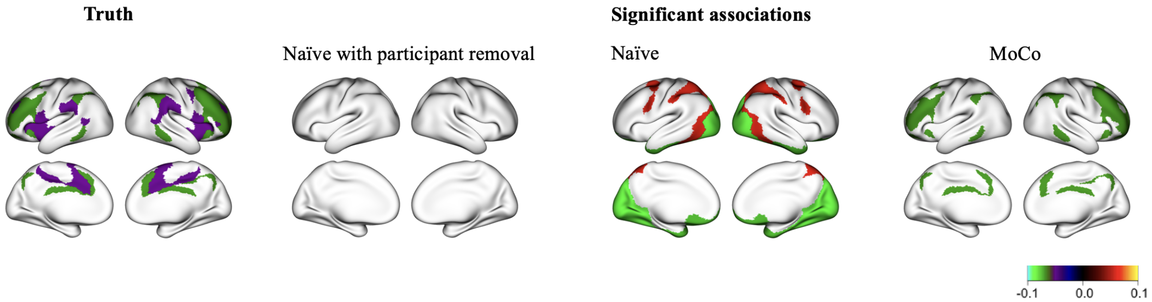


Figure 4.2: Example from one simulated dataset. The true association is marked in dark green and purple, while other regions have zero associations. MoCo identified one of the two true associations correctly. However, the naïve method with participant removal missed both regions; the naïve method with all data caused false positives and failed to detect any of the true associations.

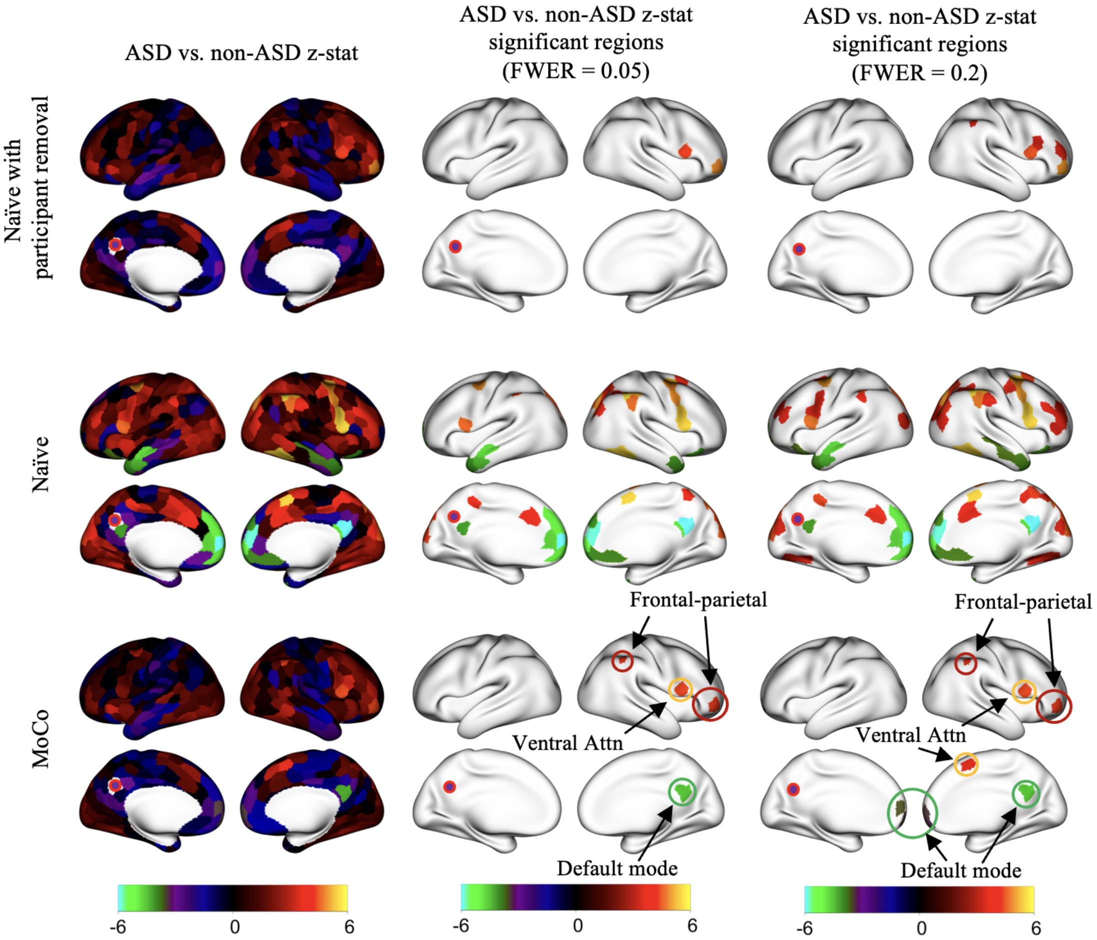


Figure 4.3: Z-statistics for the group difference (ASD − non-ASD) for a seed in the posterior default mode network (fuchsia point) in the ABIDE dataset. Extensive hypoconnectivity between the seed region and anterior parts of the default mode network in the naïve approach are likely due to motion artifacts. MoCo appears to effectively control motion artifacts, as the likely spurious differences are removed, and the results look more similar to the naïve approach with participant removal. At FWER=0.05, MoCo also identifies frontal-parietal hyperconnectivity not detected in the naïve approaches.

	True association		MoCo	naïve with participant removal	naïve
Region 1	0	Bias	**0.0005**	-0.0190	-0.0644
		sd	0.0372	**0.0191**	0.0204
		MSE×10^3	1.3839	**0.7283**	4.5657
		Type I error	**0.0110**	0.1070	0.8670
Region 2	0	Bias	**0.0048**	0.0176	0.0611
		sd	0.0238	0.0234	**0.0221**
		MSE×10^3	**0.5894**	0.8576	4.2200
		Type I error	**0.0100**	0.0730	0.7220
Region 3	0	Bias	**0.0044**	0.0153	0.0553
		sd	0.0183	**0.0179**	0.0182
		MSE×10^3	**0.3554**	0.5542	3.3941
		Type I error	**0.0080**	0.0680	0.8320
Region 4	0	Bias	**-0.0034**	-0.0179	-0.0663
		sd	0.0204	**0.0199**	0.0204
		MSE×10^3	**0.4275**	0.7180	4.8182
		Type I error	**0.0100**	0.1160	0.8870
Region 5	-0.0484	Bias	**0.0065**	0.0213	0.0695
		sd	0.0214	**0.0208**	0.0212
		MSE×10^3	**0.4990**	0.8848	5.2748
		Power	**0.3790**	0.1700	0.1260
Region 6	-0.0682	Bias	**0.0063**	0.0241	0.0796
		sd	0.0203	**0.0178**	0.0214
		MSE×10^3	**0.4523**	0.8979	6.7937
		Power	**0.8690**	0.5280	0.0670

Table 4.2: Simulation results comparing MoCo, the naïve approach with participant removal, and the naïve approach including all participants. MoCo has lower bias in all regions, lower type I error in regions in which the true association is zero, and higher power in the regions where true association is non-zero.

Chapter 5

MoCo: A package for removing motion artifacts in brain phenotype analysis

5.1 Introduction

Participant head motion during scanning sessions is a pervasive issue and has consistently posed a challenge in neuroimaging (Van Dijk et al., 2012; Baum et al., 2018). Head motion causes artifacts in functional connectivity (Maknojia et al., 2019), white matter microstructure (Yendiki et al., 2014), and cortical thickness (Reuter et al., 2015). Careful handling of motion artifacts is especially crucial in research involving children with developmental conditions associated with increased movements, such as Autism Spectrum Disorder (ASD) (Deen and Pelphrey, 2012) and Attention-deficit/hyperactivity disorder (ADHD) (Castellanos and Aoki, 2016; Aoki et al., 2018), etc. Failing to rigorously address motion artifacts may result in spurious differences in brain connectivity measures between the disease groups of interest (Power et al., 2012).

There is extensive literature on methods to mitigate motion artifacts in neuroimaging. Prospectively, techniques such as mock scanner training, providing clear instructions, and head fixation are utilized (Epstein et al., 2007; Van Dijk et al., 2012; Reuter et al., 2015). After image collection, retrospective quality control is conducted. In resting-state fMRI studies, multiple steps are typically applied. During preprocessing, confound regression is commonly applied to remove the effects of motion parameters, global signal, cerebral spinal fluid signal, and white matter from the fMRI time series (Ciric et al., 2017). Despite these efforts, residual relationships between motion and functional connectivity may persist due to potential nonlinear associations (Deen and Pelphrey, 2012). Consequently, motion quality control (QC) procedures are implemented, often involving the exclusion of scans with excessive motion based on various criteria outlined in the literature (Power et al., 2014a; Ciric et al., 2017). In diffusion MRI studies, motion correction is conducted by aligning diffusion volumes to a common reference, and outliers due to movement are corrected using Gaussian Process prediction (Andersson and Sotiropoulos, 2016; Andersson et al., 2016). Scan are recommended for exclusion if they fail inspection based on factors such as registration to the T1-weighted image, image quality, and field-of-view cutoff (Hagler Jr et al., 2019). In structural MRI studies, quality control involves assessing motion artifacts, including ringing, blurring, and adequate gray/white matter contrast, and removing scans with severe artifacts (Reuter et al., 2015; Backhausen et al., 2016).

However, although removing scans may help mitigate motion artifacts across modalities, it can result in significant reductions in sample size and introduce selection bias due to the alteration of the study population. For example, the Adolescent Brain Cognitive Development (ABCD) team performed quality control by manually reviewing structural MRI images and FreeSurfer cortical surface reconstructions (Hagler Jr et al., 2019). In ABCD Release 5.1, 1,332 non-randomly chosen T1 weighted images, 451 were marked as unacceptable due to severe ghosting, blurring, and/or ringing, rendering accurate brain segmentation impossible. In a related rs-fMRI analyses of ABCD data, Cosgrove et al. (2022) reported that motion control resulted in a 41% reduction in participants. The remaining sample exhibited demographic and clinical differences, including fewer females, older age, better neurocognitive skills, and fewer externalizing and neurodevelopmental problems.

To control for potential selection bias, Nebel et al. (2022) treated rs-fMRI scans that were excluded due to motion quality control as missing data. They employed a causal inference-informed approach to estimate the deconfounded difference in functional connectivity between children with ASD and those without ASD. However, the method's overlook of rs-fMRI data from motion QC may compromise the efficiency of the outcome model. Additionally, the absence of integration of covariate balancing into the estimation process might lead to unrealistic assumptions regarding the missingness mechanism. Furthermore, dichotomizing the continuous motion values into inclusion or exclusion indicators may underutilize information. Ran et al. (2024+) developed a motion-controlled estimator to quantify average functional connectivity differences between autistic and non-ASD children. By treating motion values as a mediator between diagnostic groups and functional connectivity, the method standardizes motion relative to low-motion scans across groups, thereby controlling for its impact. However, Ran et al. (2024+) only deals with scans that pass preprocessing quality control. Scans with issues in preprocessing pipelines, such as T1-weighted image segmentation and volume registration, are discarded. In their analysis of ASD data from the Autism Brain Imaging Data Exchange (ABIDE) dataset (Di Martino et al., 2014), 19 children were excluded due to these issues. We believe that the demographic and motion data from these individuals can be utilized to maximize data usage.

In this paper, we introduce an extended version of Ran et al. (2024+), now available as an R package called *MoCo*, to address more generalized scenarios. Our approach leverages demographic information from all participants, regardless of preprocessing quality control status, to address selection biases arising from quality control and missing data. Specifically, for scans with preprocessing issues due to T1-weighted image segmentation and volume registration, we treat the rs-fMRI data as missing but still utilize their demographic and behavioral information. Moreover, MoCo can accommodate multiple imaging modalities, including structural MRI and diffusion MRI, as the same analytical framework applies to them. Notably, the package incorporates an ensemble of machine-learning methods in the estimation process, ensuring efficient and robust estimators of motion-controlled average brain phenotypes and associations. When motion data is unavailable, we provide an enhanced covariate-balanced approach, extending the method proposed by Nebel et al. (2022), to estimate the association.

5.2 Background and theory

5.2.1 Overview of MoCo

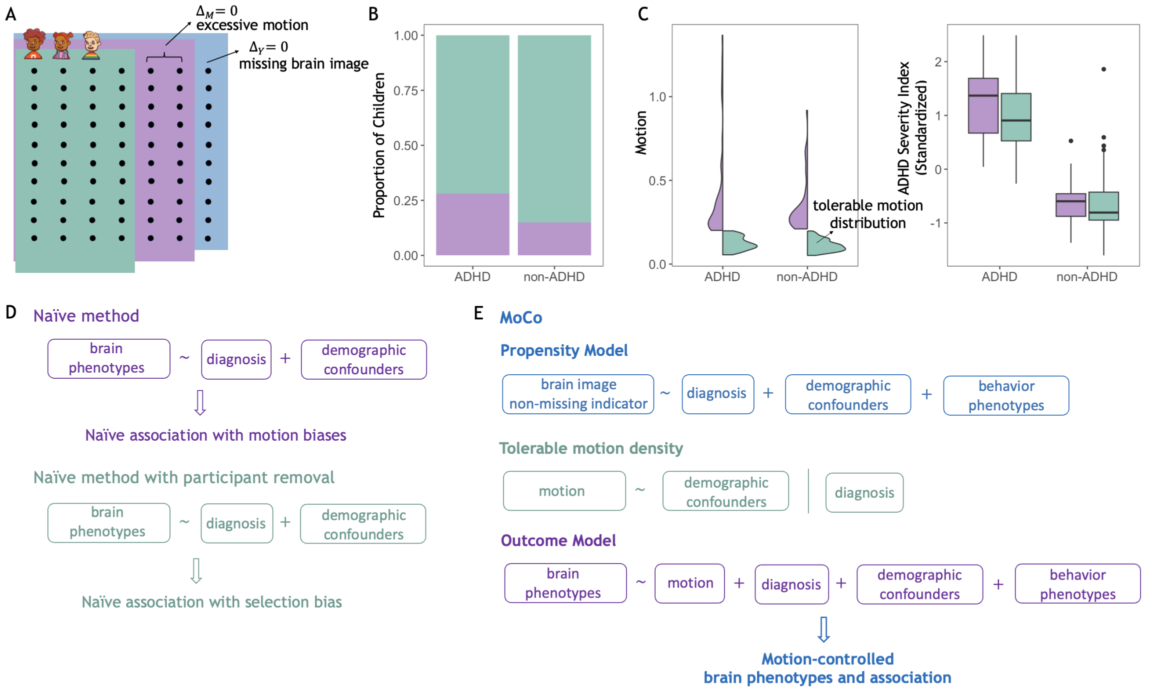


Figure 5.1: Illustration of MoCo using the ADHD200 dataset as an example.

The pipeline of MoCo and the distinction of it from naive methods are depicted in Figure 5.1 using the ADHD200 dataset (Brown et al., 2012) as an example. Panel (A) describes the data structure, where each column represents a child's data, and three sets of scenarios are indicated by color. The blue set includes all children, while the distinction between blue and purple denotes those with missing brain images due to preprocessing quality control. We denote the brain phenotype value for each participant as Y, which represents functional connectivity between a seed region and one other brain region in this example. Children who fail preprocessing quality control have Y as NA, indicated by brain phenotype usability $\Delta_Y = 0$. This occurs when T1-weighted image segmentation and volume registration are problematic, as diagnosed by visual assessment using fMRIPrep (Esteban et al., 2019). The green set, further subset from the purple, represents children with tolerable motion, determined through motion quality control. We use M to denote a numeric motion value, such as mean framewise displacement (FD) in this example. Δ_M is an indicator of whether motion M meets inclusion criteria, with $\Delta_M = 1$ indicating eligibility for inclusion, and 0 represents excessive motion or lack of motion data.

In Panels (B) and (C), we compare the green and purple sets to illustrate the selection bias introduced by motion quality control. Here, we introduce the notation A as a group indicator: in this example, $A = 1$ represents the ADHD diagnostic group, while $A = 0$ represents the non-ADHD control group. In Panel (B), there's a discrepancy in the proportion of excluded children across diagnostic groups, with a higher proportion of ADHD children being excluded. In Panel (C), we observe that children with more excessive motion and more severe ADHD are excluded from the ADHD group compared to the non-ADHD group. Therefore, as shown in panel (D), if we compare compare the naive difference in brain phenotype between the two diagnostic groups using the purple data, the results will be contaminated by motion artifacts. However, if we compare the naive difference using just the green data after motion quality control, there will be selection bias, as scan exclusion due to motion quality control alters the distribution of participant characteristics related to brain phenotype.

Panel (E) illustrates the solution proposed by MoCo. The MoCo framework consists of three primary models: the propensity model, the tolerable motion density, and the outcome model. Details of the MoCo estimator will be covered in the next section. To highlight how MoCo works, it leverages demographic data from all participants to estimate the propensity model, ensuring that the distribution of covariates is accurately assessed and controlled for in subsequent analyses. We denote the vector of demographic confounders as X, which includes variables such as age, sex, and handedness. Behavioral phenotypes are denoted as Z, which include disease severity score, full-scale Intelligence Quotient (FIQ) score, and medication status. All participants' X and Z data are included in the propensity model. In the outcome model, brain phenotypes from all participants who pass preprocessing quality control are used. These brain phenotypes are regressed on motion values and diagnostic groups, while controlling for demographic confounders and behavioral phenotypes. Lastly, to effectively control for the motion effect, we estimate the tolerable motion density using motion values that pass quality control from the control group. This density is then used to standardize motion in the outcome model for both diagnostic groups, ensuring that differences in brain phenotypes between the groups are not attributable to motion.

5.2.2 Defining the target parameter for group comparison in brain phenotype

MoCo is established in a causal context. Let us imagine a scenario where every participant has usable brain phenotype data and reasonable motion levels during scanning. In reality, maintaining stillness and avoiding preprocessing issues, such as T1-weighted image segmentation and volume registration errors, is challenging. However, in this counterfactual scenario, we assume data can be manipulated. This manipulation is not nonsense, as participants are expected to undergo training before the actual scanning. Specifically, we assume all scans are usable with $\Delta_Y = 1$, and each participant's motion is a realization of $M_0 \sim P_{M|\Delta_M=1,A=0,X}$, representing the control group's motion distribution that meets quality control criteria. This ensures data eligibility for analysis.

We denote the counterfactual brain phenotype as $Y(1, M_0)$. Here, 1 stands for $\Delta_Y = 1$, indicating that all participants have usable brain phenotype data, and M_0 represents that they all have an appropriate motion value sampled from the conditional motion distribution. Our proposed motion-controlled target parameter is denoted as

$$\psi_C = E(E_C(Y(1, M_0) \mid A = 1, X)) - E(E_C(Y(1, M_0) \mid A = 0, X)) \tag{5.1}$$

where E_C denotes an expectation with respect to the counterfactual data.

The motion-adjusted association ψ_C estimates the difference in average brain phenotype between diagnosis groups while accounting for the impact of excessive motion. This is accomplished by sampling motion from the conditional distribution that satisfies the inclusion criteria for both groups, such that the association difference is not attributed to motion discrepancies. Furthermore, the parameter accounts for variations in demographic confounders X between diagnostic groups by integrating over the same X distribution in the outer expectation during the comparison. Consequently, the observed difference between the two groups is solely attributed to the diagnosis group A and behavior phenotypes.

The identification of the target parameter using observational data involves three assumptions. The technical statements are provided in the Web Supplement (Section 1). We summarize below:

(A1) *Mean exchangeability (assumption of no missing confounders)*: The demographic confounders X and behavioral phenotypes Z contain all confounding variables of the (M, Δ_Y) and Y relationship.

(A2) *Positivity*: (A2.1) At a population level, there cannot be values of X that are observed exclusively in one diagnostic group. Secondly, it is possible for a control group participant to meet the inclusion criteria across any demographic values X. Thirdly, it is possible to obtain data for both control and diagnostic group participants across all demographic values X and Z. (A2.2) It is possible to observe low motion across any diagnosis, demographic and behavioral values.

(A3) *Causal Consistency*: This is a technical assumption that the observed brain phenotype Y from participants with complete data and a natural motion level m is the same as the brain phenotype that would have been observed in the hypothetical world where participants have tolerable motion in the scanner and have no missing data.

Under (A1)-(A3), the counterfactual motion-controlled target parameter in equation (5.1) is identified by

$$\iiint (\mu_{Y|\Delta_Y=1,A,M,X,Z}(1,m,x,z)p_{Z|A,X}(z \mid 1,x) - \mu_{Y|\Delta_Y=1,A,M,X,Z}(0,m,x,z)p_{Z|A,X}(z \mid 0,x)) \\ p_{M|\Delta=1,A,X}(m \mid 0,x)p_X(x)\,dz\,dm\,dx \tag{5.2}$$

$\mu_{Y|\Delta_Y=1,A,M,X,Z}(a,m,x,z)$ represents the average brain phenotype in the subgroup defined by diagnosis status $A = a$, motion level $M = m$, and covariates $(X,Z) = (x,z)$, among participants with usable brain phenotype values ($\Delta_Y = 1$). Then, this conditional average brain phenotype is standardized by the conditional Z distribution $p_{Z|A,X}(z \mid a,x)$ in each subgroup $(A,X) = (a,x)$. The comparison $\int(\mu_{Y|\Delta_Y=1,A,M,X,Z}(1,m,x,z)p_{Z|A,X}(z \mid 1,x) - \mu_{Y|\Delta_Y=1,A,M,X,Z}(0,m,x,z)p_{Z|A,X}(z \mid$

$0, x))dy$ quantifies a motion- and demographic confounder-conditional group difference in brain phenotype. Next, the target parameter averages the conditional group difference twice: first over the demographic confounder-conditional distribution of motion in the control group ($A = 0$) that passes motion quality control $p_{M|\Delta_M=1,A,X}(m \mid 0, x)$, and then over the marginal distribution of demographic confounders $p_X(x)$ in the population. To understand the motivation for this target parameter, we note that the conditional brain phenotype for *both groups* is averaged over the same conditional distribution of M and the marginal distribution of X. Thus, we are appropriately *controlling* for the impact of motion M and demographic confounders X, while reporting a marginal difference in brain phenotype. Thus, our choice of distributions for this standardization should recover an objective comparison of the groups. Details about the identification can be found in the Supplementary Material Section 1.

5.2.3 Estimation and inference of the motion-controlled group difference

Our target motion-controlled group difference parameter in equation (5.2) requires multiple integrations. To estimate this parameter, we propose to use a sequential regression approach to express the integration as regressions, and then employ regression models to obtain the estimates.

To fit regression models, rather than relying on parametric models like generalized linear regression, which may be susceptible to misspecification, we adopt a more flexible approach known as Super Learner. Super Learner constructs a weighted combination of predictors from a user-specified library of candidate estimators. Examples of these estimators include multivariate adaptive regression splines, LASSO, ridge regression, generalized additive models, generalized linear models, random forest, and XGBoost (Polley et al., 2023). The weights are assigned to minimize the cross-validated risk of the resulting combination. According to large sample theory, the ensemble estimator is expected to yield nuisance parameter estimates that are at least as good as, if not better than, those from any individual candidate regression estimator considered.

To enhance the robustness of the estimator, we incorporate an efficient influence function (EIF)-based correction in addition to evaluating the regression models. Details of the EIF can be found in Supplementary Material Section 2. Evaluating the EIF involves estimating the density of M, for which we propose using the highly adaptive lasso, a flexible semiparametric conditional density estimator (Hejazi et al., 2022b). Furthermore, we integrate cross-fitting into the estimation process to mitigate overfitting and impose less restrictive complexity conditions on the machine learning algorithms embedded in Super Learner. The final estimator exhibits asymptotic linearity under certain assumptions and includes an analytical confidence interval, facilitating subsequent statistical inference.

When multiple Y variables are considered simultaneously, MoCo can make inference adjusting for multiple comparisons. To control the family-wise error rate for multiple tests, MoCo conducts hypothesis testing using simultaneous confidence bands based on EIF. For instance, in seed-based functional connectivity analysis, MoCo can jointly evaluate the functional connectivity of different brain regions with the seed region to identify regions with significant differences between groups. Details of the construction of simultaneous confidence bands can be found in Supplementary Material Section 2.

5.3 Tutorial

In this section, we demonstrate the application of MoCo with a straightforward example analysis. We generate a simulated dataset based on the Autism Brain Imaging Data Exchange (ABIDE) (Di Martino et al., 2014, 2017), which we previously analyzed (cite our preprint). Our seed region of interest is the default mode network (DMN). We are interested in studying the functional connectivity between DMN and the other six regions defined using the Yeo 7 parcellation (Yeo et al., 2011). We use MoCo to compute the motion-controlled mean functional connectivity and associations. Additionally, we identify regions exhibiting significant associations and use ciftiTools for visualization (Pham et al., 2022). The tutorial consists of four steps: (1) reading the dataset; (2) defining variables and specifying parameters for MoCo, and running *moco* function to calculate motion-controlled mean functional connectivity and associations; (3) determining regions with significant associations based on simultaneous confidence bands; (4) visualizing the motion-controlled functional connectivity, associations, and significant regions using ciftiTools. Please note that a rs-fMRI cifti template needs to be downloaded and placed in the `Data` folder for ciftiTools visualization. In addition, plot images are written to a subfolder named `Plots`. To rerun this code, make sure to create these two directories. Additionally, download the cifti template from the supplementary material and place it in the `Data` folder.

5.3.1 Input to the MoCo function

We start by loading the `MoCo` package and the example dataset. The `data` contains simulated covariates and functional connectivity data designed to mimic real-life patterns observed in the ABIDE dataset.

```
# library
library(MoCo)
# load data
data(data)
```

The dataset includes a total of $n = 400$ participants. The left side of Figure 5.2 illustrates the data structure of the input variables. `A, M, Delta_M, Delta_Y` are vectors of length n: Each element of `A` denotes a participant's diagnostic status, with 1 representing ASD and 0 representing non-ASD. `M` represents continuous motion values corresponding to mean FD. Participants are classified as having high motion if

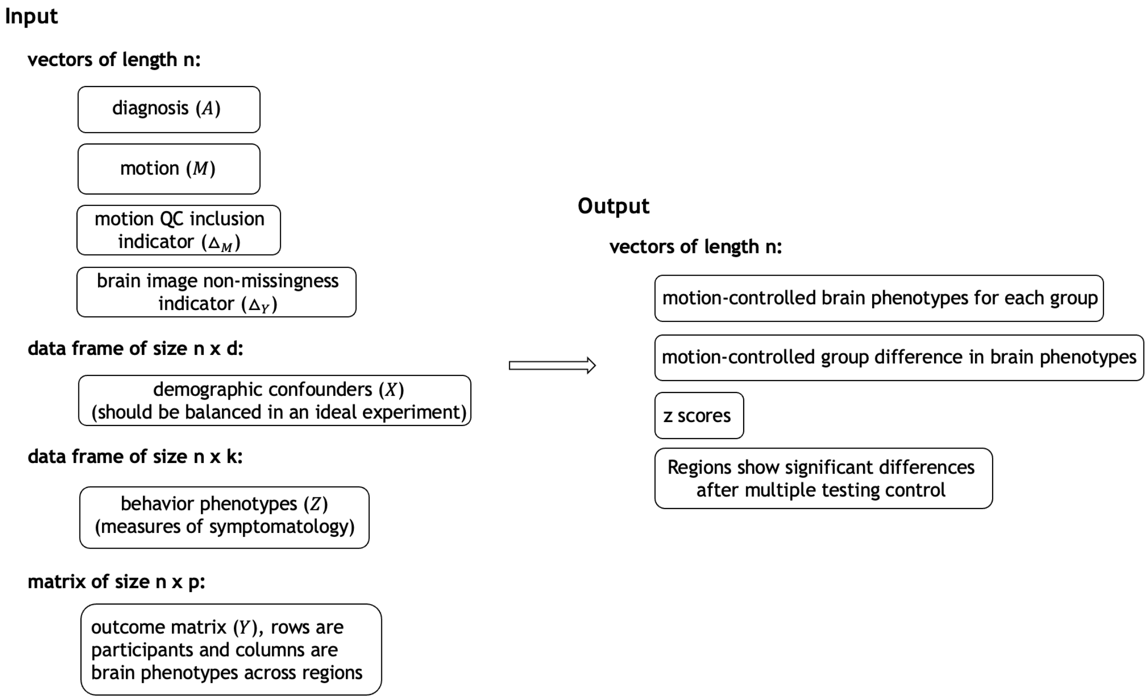


Figure 5.2: Input and output for the MoCo package.

`M` exceeds 0.2 (`Delta_M` = 0), consistent with the threshold used in the analysis of real data. The binary indicator `Delta_Y` is a binary indicator equal to 0 for participants with poor-quality preprocessed images and 1 otherwise. The proportion of participants with `Delta_Y`=0 is 7%.

Demographic confounders `X` and behavioral phenotypes `Z` are represented as data frames, each containing multiple variables. Data frame `X` is of size $n \times 3$, containing three demographic dimensions: sex (`X1`), age (`X2`), and handedness (`X3`). In `X1`, females are coded as 0 and males as 1. `X2` represents age as a continuous numeric value. `X3` indicates handedness, with left-handed individuals coded as 0 and right-handed individuals as 1. Data frame `Z` is of size $n \times 4$, containing four behavior phenotypes. `Z1` represents scores from the Autism Diagnostic Observation Schedule (ADOS), which measures social disability. `Z2` contains the FIQ scores. Medication status is captured in two dimensions: `Z3` indicates stimulant medication use, and `Z4` indicates non-stimulant medication use. For both `Z3` and `Z4`, a value of 0 denotes that the individual is not currently taking the respective medication.

The functional connectivity matrix `Y` has dimensions $n \times 7$. Each row represents the z-transformed functional connectivity derived from rs-fMRI between the seed region and the other regions for a given participant. The 7th column, representing the functional connectivity of the seed region with itself, is filled with NA values to indicate its position. For participants with `Delta_Y` $= 0$, the corresponding rows in `Y` contain all NAs, as their functional connectivity data is not available. The true differences in functional connectivity between each region and region 7 are as follows: for regions 1-4, the association is 0; for region 5, it is -0.0485; and for region 6, it is -0.0682. Region 5 and 6 are set to have significant associations.

5.3.2 Calculating motion-controlled functional connectivity and associations

Now that we have loaded the package and imported the data, we can use MoCo to calculate motion-controlled functional connectivity and associations. We begin by defining the variables and specifying parameters for nuisance parameter calculation.

```
# specify covariates
X = data$X
Z = data$Z
A = data$A
M = data$M
Y = data$Y
Delta_M = data$Delta_M
Delta_Y = data$Delta_Y
```

Note that if `Delta_M` is not available, leave it blank or setting it as `Delta_M = NULL` and specify `thresh`. The motion value `M` will be truncated by `thresh` to produce `Delta_M`. By default, `thresh` is set to `NULL`.

Next, we specify the method for nuisance regressions. Nuisance regressions are regressions for calculating parameters that need to be estimated as an intermediate step in the estimation of motion-controlled functional connectivity. There are two options: basic generalized linear models or Super Learner, an ensemble machine learning algorithm. `MoCo` set the default as SuperLearner, which we highly recommend as it combines multiple prediction algorithms to create a more accurate and robust model. To use Super Learner, users have the option to either specify a recycled set of algorithms for all nuisance regressions using the `SL_library` argument, or set `SL_library` to `NULL` and customize the set of algorithms for each nuisance regression using `SL_library_customize`. By default, a recycled `SL_library` is used, and the default algorithms include multivariate adaptive regression splines, LASSO, ridge regression, generalized additive models, generalized linear models (with and without interactions, and with and without forward stepwise covariate selection), random forest, xgboost, and the mean of the outcome as below. Further details about these algorithms can be found in Polley et al. (2023).

```
# default SL_library in MoCo
SL_library = c("SL.earth","SL.glmnet","SL.gam",
               "SL.glm", "SL.glm.interaction",
               "SL.step","SL.step.interaction",
               "SL.xgboost","SL.ranger","SL.mean")
```

The application of generalized linear models requires the user to specify the `glm_formula` for each nuisance regression. The `glm_formula` requires a comprehensive understanding of each nuisance regression throughout the estimation process. The detailed steps of the estimation, accompanied by an example illustrating the corresponding notation for glm formulas, are available in the Supplementary Material. By default, the formulas are set to `NULL`, as we recommend the use of Super Learner for its robustness and accuracy, which alleviates the need for users to manually define glm formulas.

In addition to nuisance regression, `MoCo` includes two additional parameters:

`HAL_pMX` and `HAL_pMXZ`. These parameters control the use of the highly adaptive lasso, a flexible semiparametric conditional density estimator designed for motion density estimation (Hejazi et al., 2022b). By default, `MoCo` sets these parameters to `TRUE`, indicating the utilization of `haldensify`. Users have the option to customize hyperparameters in haldensify such as `max_degree`, `lambda_seq`, and `num_knots` through the `HAL_options` argument. `MoCo` has initialized default values for these hyperparameters. For further details, please refer to the documentation Hejazi et al. (2022a). If set to `FALSE`, a generalized linear model will be used for motion density estimation. We assume the conditional motion distributions follow a Log-normal distribution. A generalized linear model will be fitted by specifying `pMX` and `pMXZ` in the `glm_formula` argument to estimate the conditional density of the logarithm of motion. The standard deviation is set equal to the sample standard deviation. Subsequently, the estimated conditional density will be transformed into the conditional density of motion.

Examining the overall framework, MoCo employs an efficient influence function-based nonparametric approach to estimate motion-controlled mean functional connectivity. By default, estimators are fitted using cross-fitting, specified with the arguments `cross_fit = TRUE` and `cv_folds = 5`. The parameter `cv_folds` specifies the utilization of K-fold cross-fitting for estimation, with the default value of K being 5. In this setting, the data are divided into five segments, with one segment held out for evaluation while the others are used for fitting. The final result is an average of five fitting runs. Alternatively, users can opt out of cross-fitting by setting `cross_fit = FALSE`.

Finally, for those interested in running the entire cross-fitting process multiple times, the `seed_rgn` argument allows for this specification. By default, it is set to 1, which means that the seed is set to 1 when running the SuperLearner regression. However, as the SuperLearner estimation process is built upon cross-validation, it introduces Monte Carlo variability. To address this, users can input a vector into `seed_rgn`, such as `seed_rgn = 1:10`. In this case, MoCo will execute the entire fitting process 10 times based on each specified `seed_rgn` value and return the average of the results to mitigate potential Monte Carlo errors. It's important to note that this procedure can be time-consuming.

By combining variable definition and parameter specification, the `moco` function is utilized to compute motion-controlled functional connectivity and associations. For illustrative purposes, we choose a simple setting. Here we select a basic `SL_Library`, employ cross-fitting with `cv_folds = 5`, and utilize `glm` for motion density estimation. In the end, we provide a suggested setting for conducting a more comprehensive analysis based on more accurate density estimation using highly adaptive lasso density estimation.

```
# computing motion-controlled functional connectivity and
   associations
result = moco(
  X = data$X,
  Z = data$Z,
  A = data$A,
  M = data$M,
  Y = data$Y,
  Delta_M = data$Delta_M,
  Delta_Y = data$Delta_Y,
  SL_library = c("SL.mean", "SL.glm","SL.glm.interaction"),
  glm_formula = list(pMX = ".",
                     pMXZ = "."),
  HAL_pMX = FALSE,
  HAL_pMXZ = FALSE,
  cross_fit = TRUE,
  cv_folds = 5,
  seed_rgn = 1,
  test = FALSE
```

```
)
```

The `result` object contains a list of 2. The motion-controlled mean functional connectivity is stored in the `est` element. The first row corresponds to the adjusted mean functional connectivity for the non-ASD group ($A = 0$), and the second row corresponds to the ASD group ($A = 1$). The first six columns represent the results for six regions with the seed region, respectively, and the last column is NA, as the seed region is at the 7th position. The motion-controlled association is stored in the `adj_association` vector of length seven, where the first six elements represent the adjusted association for the corresponding region. MoCo achieves satisfactory accuracy when comparing the estimated motion-controlled association with the ground truth.

```
# motion-controlled mean functional connectivity
round(result$est, 4)
# est_A0 -0.2180 -0.1632 -0.1823 0.0535  0.0388 0.0828 NA
# est_A1 -0.2194 -0.1654 -0.1813 0.0513 -0.0084 0.0141 NA

# motion-controlled association
round(result$adj_association, 4)
# -0.0014 -0.0023  0.0010 -0.0022 -0.0472 -0.0687      NA
```

5.3.3 Determining significant associations based on simultaneous confidence band

To identify regions with significant associations, `MoCo` enables conducting multiple hypothesis tests based on simultaneous confidence bands. The test can be performed by setting `test = TRUE` and specifying the desired value for the family-wise error rate (`fwer`). By default, `fwer = 0.05`.

```
result = moco(
  X = data$X,
  Z = data$Z,
```

```
  A = data$A,
  M = data$M,
  Y = data$Y,
  Delta_M = data$Delta_M,
  Delta_Y = data$Delta_Y,
  SL_library = c("SL.mean", "SL.glm","SL.glm.interaction"),
  glm_formula = list(pMX = ".",
                     pMXZ = "."),
  HAL_pMX = FALSE,
  HAL_pMXZ = FALSE,
  cross_fit = TRUE,
  cv_folds = 5,
  seed_rgn = 1,
  test = TRUE,
  fwer = 0.05
)
```

Now, the result will be a list of 4 elements. In addition to the motion-controlled functional connectivity and association, it contains 2 other elements: z-scores and a binary indicator indicating significance for each of the regions.

```
# z-scores
round(result$z_score, 4)
# -0.0586 -0.0853  0.0413 -0.0933 -1.9196 -3.3033      NA

# significant regions
result$significant_regions
# FALSE FALSE FALSE FALSE FALSE  TRUE    NA
```

Recall that according to the ground truth, the fifth and sixth regions have a significant association with the seed region. Upon inspecting the results, when using the simple setting, MoCo successfully captures a significant region without introducing false positives.

Below is the code for a comprehensive setting recommended for running MoCo. This setting uses highly adaptive lasso for conditional density motion estimation, offering more flexible modeling of the conditional motion distribution. It utilizes all default parameters provided by the function, which uses the default Super Learner library for flexible nuisance regression estimation. Users only need to specify the definition of each variable to obtain results.

```
result = moco(
  X = data$X,
  Z = data$Z,
  A = data$A,
  M = data$M,
  Y = data$Y,
  Delta_M = data$Delta_M,
  Delta_Y = data$Delta_Y
)
```

5.3.4 Visualization using ciftiTools

We demonstrate the visualization of the resulting significant regions using the Yeo 7 parcellation with ciftiTools. We provide a demo function for the plot, where users only need to input the path to the workbench (`wb_path`), to the template cifti file in the `Data` folder (`template_path`), and to the `Plots` folder (`plots_path`). The generated plots will be located in the `Plots` folder.

```
# visualization
plot_moco(result,
          wb_path = "/Applications/workbench",
          template_path = "Data/template.dtseries.nii",
          parcellation = "Yeo_7",
          plots_path = "Plots")
```

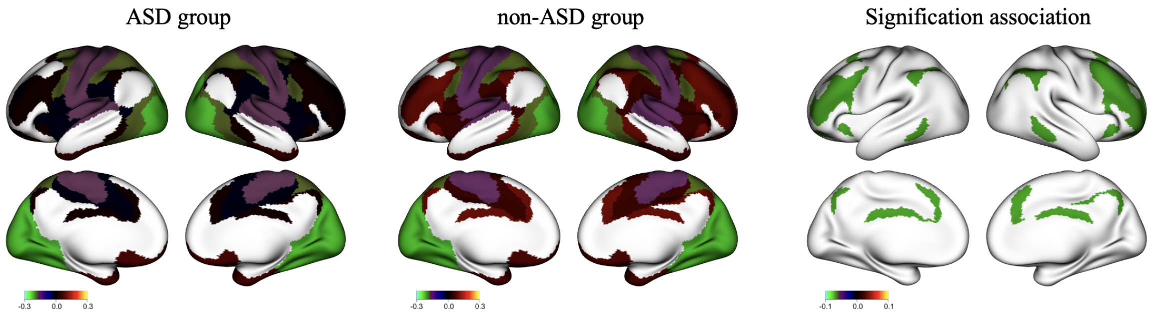


Figure 5.3: Estimated functional connectivity of the ASD group and non-ASD group using MoCo for the example dataset. In this example, MoCo correctly identifies one of the two true associations.

5.3.5 Handling absence of motion data

When motion values are not available, the package offers additional flexibility for estimating the association. By setting `X = NULL, M = NULL`, the method aligns with Nebel et al. (2022). If only `M = NULL` and demographic confounders `X` are provided, the output extends Nebel et al. (2022) by incorporating covariate balancing based on demographic confounders.

5.4 Dataset and analysis

5.4.1 Resting-state fMRI dataset

We illustrate the application of MoCo using resting-state fMRI data from school-age children with Attention Deficit Hyperactivity Disorder (ADHD) (Brown et al., 2012). ADHD is a prevalent mental health condition, impacting approximately 5% to 10% of school-age children (Polanczyk et al., 2014). It is characterized by impulsive, hyperactive, and inattentive behaviors, which may endure into adulthood, posing substantial challenges for affected individuals, their families, and society. Children with ADHD frequently encounter obstacles in learning and in maintaining typical daily routines compared to their peers. In our analysis, we subset to the Kennedy Krieger Institute (KKI), New York University (NYU) and Peking University (Peking) sites with full demographical data, 441 children aged 7 to 18. At the KKI site, rs-fMRI scans were conducted using a Siemens Magnetom TrioTim syngo MR B13 scanner with a repetition time(TR)/echo time (TE) of 2500/30 ms, a flip angle of 75°, and 3.1×3.1×3.5 mm voxels, SENSE phase reduction was set to 1, and the duration of the scan ranged from 5 to 6min. At the NYU site, rs-fMRI scans were conducted using a Siemens Magnetom Avanto syngo MR B17 scanner with a repetition time (TR)/echo time (TE) of 2000/15 ms, a flip angle of 90°, and 3.5×3.5×3 mm voxels, SENSE phase reduction was set to 1, and the scan duration was 6min. For the Peking site, scans were conducted using a Siemens Magnetom TrioTim syngo MR B15 scanner with a TR/TE of 2000/30 ms, a flip angle of 90°, and 3×3×3 mm voxels. SENSE phase reduction was set to 1, and the scan duration was 8min. The first four volumes were removed for both sites. Additionally, all except 2 children from the NYU site underwent the collection of an anatomical T1 scan. Further information on scan parameters and other details can be found on the ADHD-200 webpage.

T1 anatomical and rs-fMRI data were processed with the cifti option for cortical surface registration using fMRIPrep, including anatomical tissue segmentation, surface construction, and surface registration, followed by fMRI motion correction, slice-time correction, boundary-based coregistration, and resampling to the fsaverage template (Esteban et al., 2019). See Supplement Section 3.1 for details. We visually inspected the accuracy of the cortical segmentation and boundary-based registration using the fMRIPrep quality control html files. Two participants from the NYU site do not have rs-fMRI scans, and in total 70 participants failed cortical segmentation or exhibited severe missingness in boundary-based registration, resulting in their rs-fMRI data being marked as missing ($\Delta_Y = 0$). Two example participants who failed this step are included in the Supplement Section 3.2. Most of these cases involved children with partial brain data missing in the alignment of functional and anatomical MRI scans. The remaining cases involved poor-quality T1 images, for example, corrupted by motion, which led to cortical segmentation failure.

Our final study sample includes all 441 children, with 268 non-ADHD children and 173 ADHD children (Table 7.9). In addition to functional connectivity data, we included the following: diagnosis (A); age, sex, and handedness (X); ADHD Index, measure of ADHD Index, the interaction between ADHD Index and measure of ADHD Index, Full-scale Intelligence Quotient (FIQ) score, medication status (Z), and mean framewise displacement (FD) in the analysis (M). The ADHD Index measures ADHD severity. At the KKI and NYU sites, the ADHD Rating Scale IV (ADHD-RS) was used, while the Peking site used the Conners' Parent Rating Scale-Revised, Long Version (CPRS-LV). Although ADHD is more prevalent in males than females (Arnett et al., 2015; Willcutt, 2012), we treated sex as a confounder (X) rather than diagnosis-specific variable (Z) because sex-specific differences in functional connectivity have been previously documented (Shanmugan et al., 2022), which could mask ADHD-related differences in this cohort. Mean FD is an average of the frame-to-frame displacement calculated from the rigid body motion correction parameters. We defined the indicator of data usability Δ_M equal to one if a child mean FD < 0.2 mm (Mummaneni et al., 2023). This resulted in $\Delta_M = 0$ for 50/173 ADHD (28.9%) and 42/268 non-ADHD children (15.7%).

We calculated the average time series for ROIs defined using Schaefer's 400-node brain parcellation, which associates each node with the resting-state networks from Yeo-7 (Schaefer et al., 2018). We then explored multiple preprocessing options (Fox et al., 2009): (1) 9-parameter (9p) nuisance regression: Regressing out the six parameters from motion alignment (2) 36-parameter (36p) nuisance regression: Regressing out the six parameters from motion alignment, their squared values, the lag-1 terms of these 12 regressors, the global signal (mean signal across all voxels), white matter (WM), and cerebrospinal fluid (CSF) as well as their derivatives calculated from fMRIPrep. (3) 36-parameter (36p) spike regression: Applying a preprocessing approach that includes

	ADHD (N=173)	non-ADHD (N=268)	Total (N=441)	p value
$\boldsymbol{\Delta_M}$				< 0.001
Excessive/no motion	50 (28.9%)	42 (15.7%)	92 (20.9%)	
Tolerable motion	123 (71.1%)	226 (84.3%)	349 (79.1%)	
$\boldsymbol{\Delta_Y}$				0.392
Unusable rs-fMRI	25 (14.5%)	47 (17.5%)	72 (16.3%)	
Usable rs-fMRI	148 (85.5%)	221 (82.5%)	369 (83.7%)	
Age				0.106
Mean (SD)	11.678 (2.408)	11.307 (2.316)	11.453 (2.357)	
Range	7.240 - 17.430	7.170 - 17.960	7.170 - 17.960	
Gender				< 0.001
Female	31 (17.9%)	121 (45.1%)	152 (34.5%)	
Male	142 (82.1%)	147 (54.9%)	289 (65.5%)	
Handedness				0.784
Left	5 (2.9%)	9 (3.4%)	14 (3.2%)	
Right	168 (97.1%)	259 (96.6%)	427 (96.8%)	
ADHD index				< 0.001
Mean (SD)	59.876 (13.548)	37.881 (9.946)	46.509 (15.728)	
Range	35.000 - 90.000	18.000 - 81.000	18.000 - 90.000	
FIQ				< 0.001
Mean (SD)	105.815 (12.873)	113.802 (13.698)	110.669 (13.924)	
Range	75.000 - 142.000	80.000 - 153.000	75.000 - 153.000	
Medication				< 0.001
No	112 (64.7%)	261 (97.4%)	373 (84.6%)	
Yes	61 (35.3%)	7 (2.6%)	68 (15.4%)	
Site ID				0.035
KKI	22 (12.7%)	56 (20.9%)	78 (17.7%)	
NYU	56 (32.4%)	87 (32.5%)	143 (32.4%)	
Peking_1	45 (26.0%)	77 (28.7%)	122 (27.7%)	
Peking_2	32 (18.5%)	28 (10.4%)	60 (13.6%)	
Peking_3	18 (10.4%)	20 (7.5%)	38 (8.6%)	

Table 5.1: Demographic characteristics: Continuous variables are described using mean and standard deviation, and diagnostic groups are compared using the Kruskal-Wallis rank-sum test. Binary and categorical variables are reported as frequencies and percentages, and differences between diagnostic groups are assessed using either the Chi-square test or Fisher's exact test.

spike regression following 36p nuisance regression. We used COMBAT (Yu et al., 2018) for site harmonization to account for the seven protocols. For COMBAT, "site" is a factor with seven levels (KKI, NYU, Peking_1 to Peking_5) and the following covariates: diagnosis, age, sex, handedness, ADHD Index, ADHD measure, Interactions of ADHD Index and ADHD measure, FIQ, medication status, and mean FD.

5.4.2 Data analysis and results

We conducted a correlation analysis using a seed region in the default mode network to investigate selection bias and motion impacts on functional connectivity in ADHD. The default mode network (DMN) is a collection of brain regions that tend to co-activate during wakeful rest, including daydreaming or mind wandering. Most studies on ADHD primarily focus on the DMN, and according to Sonuga-Barke and Castellanos (2007), ADHD can be conceptualized as a disorder of the DMN. Hypoconnectivity between anterior and posterior parts of the default mode network was previously found (Castellanos et al., 2008). However, hypoconnectivity in the default mode network also arises from motion artifacts (Power et al., 2012).

We computed Fisher z-transformed correlations of rs-fMRI time series for each brain region with region 14, which is called '17networks_LH_DefaultA_pCun_1', a hub of the posterior default mode network (Pham et al., 2022). We compared the naïve estimate with participant passing motion QC (retaining $\Delta_Y = 1$ and $\Delta_M = 1$), the naïve estimate (retaining $\Delta_Y = 1$), both while adjusting for demographic confounders, and the group difference estimated from our method with cross-fitting. We determined the FWER-critical values using simultaneous confidence intervals derived from the efficient influence functions. For the naïve approaches, the critical values were derived from the sample correlation matrices. Nuisance regressions involved in our method are estimated using the default SuperLearner library in the MoCo package. To handle Monte Carlo variability resulting from cross-validation, we generated estimates 50

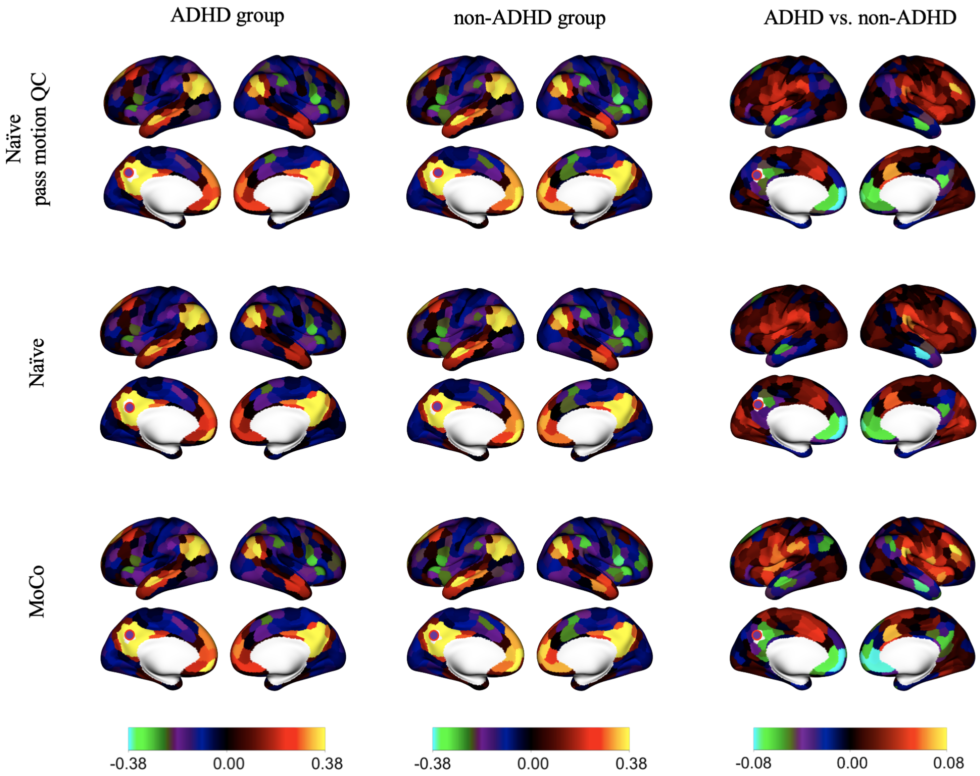


Figure 5.4: Estimated functional connectivity using the naïve approach excluding high-motion participants, the naïve approach, and our proposed method for a seed region in the posterior default mode network (fuchsia point) using 9p nuisance regression.

times and calculated the averaged estimates and z-statistics across runs.

The estimated functional connectivity using the naive approach with participants passing motion QC, the naive approach, and our proposed method are illustrated in Figure 7.12 for the preprocessing pipeline with 9p-parameter regression. The corresponding figure for the 36p-parameter regression and 36p-parameter regression with spike regression is included in the Supplementary Material Section 3.3. The seed region in the posterior default mode network is defined by the fourteenth parcel but is represented by the fuchsia point for clarity. Using Wald tests with simultaneous confidence bands, our method reveals hyperconnectivity in ADHD between the posterior default mode seed region and region in the salience or ventral attention network (see Figure 5.5). The naïve approach including high-motion children seems to result in spurious connectivity differences. These differences include apparent seed region hypoconnectivity in ADHD in long-distance correlations within the default mode network (the connection between the posterior default mode and the medial anterior default mode regions), which is consistent with the view that long-distance correlations are attenuated in high-motion participants (Satterthwaite et al., 2013b). The naïve approach also includes some regions of hyperconnectivity in ADHD in lateral regions of the frontal lobe. The naïve approach with children passing motion QC reduces some of the group differences compared to the naïve approach with all children. However, spurious within default mode connectivity differences still persist. In contrast, MoCo reveals that most default mode regions found to exhibit significant differences in naïve approaches do not show hypoconnectivity after controlling for motion artifacts. Specifically, MoCo only identifies significant hypoconnectivity between the seed region and restricted anterior default mode regions, as well as hyperconnectivity between the seed region and the salience and ventral attention networks. In the supplementary materials, we include z-statistics for the group differences when using 36 parameters, and 36 parameters and spike regression in preprocessing. MoCo demonstrates consistent results regardless of preprocessing (Figure 5.6), whereas the naive approaches are sensitive to the preprocessing pipeline with fewer significant regions identified with the stringent preprocessing.

Figure 5.5: Z-statistics for the group difference (ADHD − non-ADHD) for a seed in the posterior default mode network (fuchsia point) in the ADHD200 dataset. The "naive approach" includes participants that pass the preprocessing pipeline ($\Delta_Y = 1$). The naive approach that pass motion QC are participants that also pass motion QC ($\Delta_Y = 1$ and $\Delta_M = 1$). MoCo uses imaging data from all participants that pass preprocessing (including those with excessive motion) and uses demographic data from all participants (including those that failed preprocessing). Both naive approaches appear to generate spurious findings, suggesting extensive anterior-posterior DMN hypoconnectivity. Based on the results from MoCo, most of these regions do not differ in hypoconnectivity when standardizing motion between groups. Network labels from (Schaefer et al., 2018).

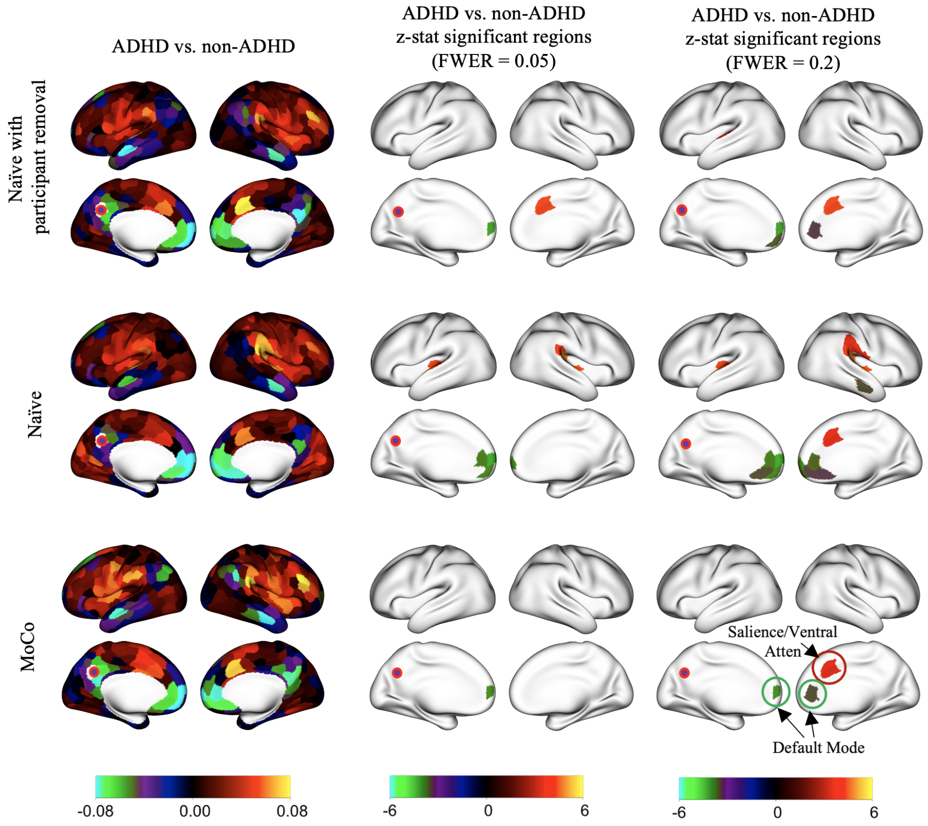
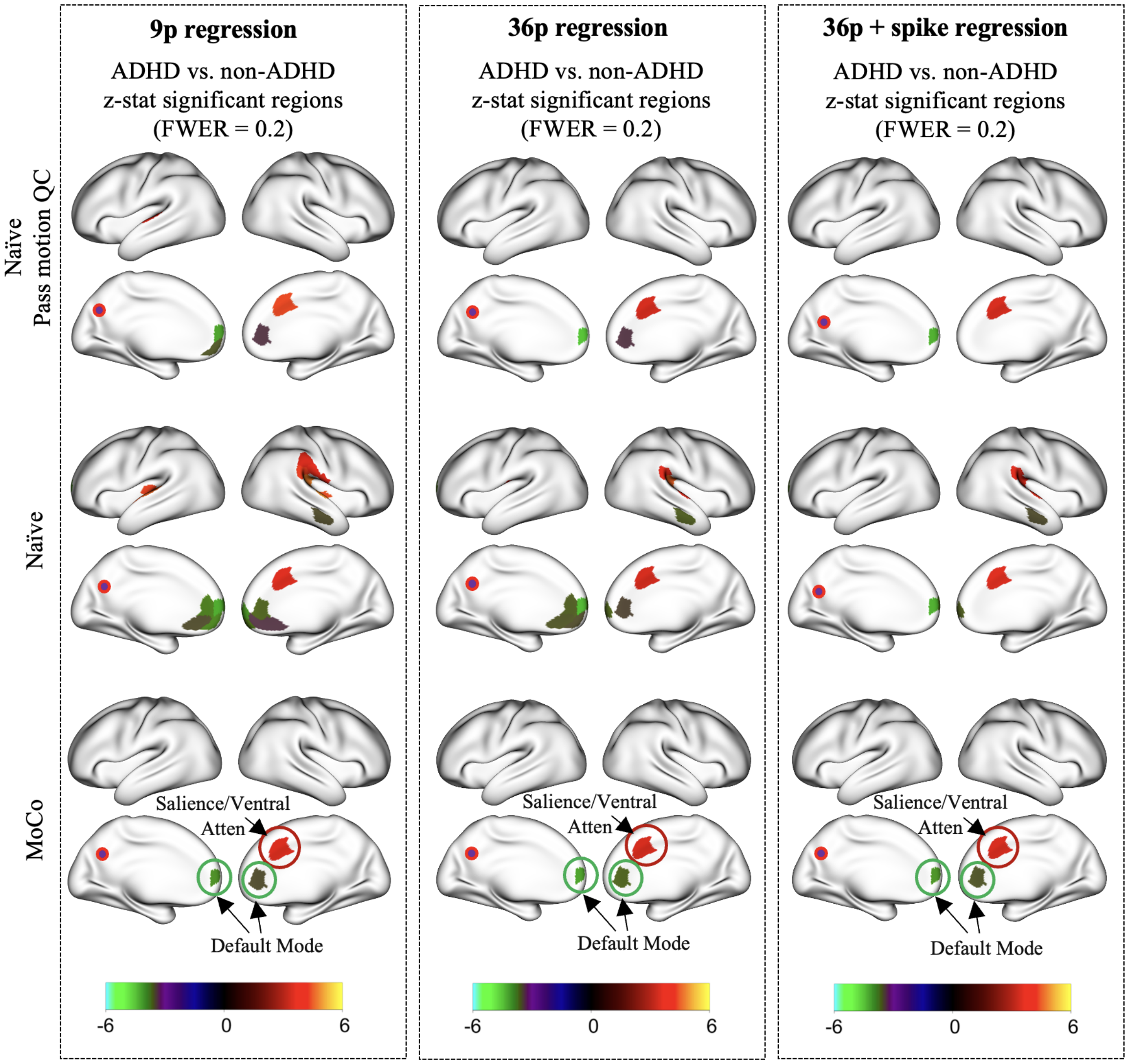


Figure 5.6: MoCo consistently produces stable results regardless of preprocessing pipelines, while the naive approaches are sensitive to the preprocessing pipeline, identifying fewer significant regions under more stringent preprocessing conditions.



Chapter 6

Conclusion and Future Work

6.1 Conclusion

In this dissertation, we developed several models and algorithms for brain connectivity analysis to address the challenges of analyzing large-scale and complex functional connectivity data in neuroimaging, as well as mitigating potential artifacts introduced by motion during scanning. The key conclusions from this work are summarized below.

Unveiling Hidden Sources of Dynamic Functional Connectome through a Novel Regularized Blind Source Separation Approach We propose a general latent source separation method for brain dynamic connectome. There has been a strong interest in advancing the understanding of the dynamic reorganization of the human brain. The proposed dyna-LOCUS method aligns with the broader framework of analyzing dFC using clustering methods such as k-means or decomposition methods such as ICA and PCA to unveil basis patterns underlying the observed dynamic connectome (Miller et al., 2016). However, the distinguishing features of dyna-LOCUS set it apart from the existing methods, enabling dyna-LOCUS to offer valuable contributions to the field. Specifically, compared to the clustering methods such as k-means or fuzzy k-means, dyna-LOCUS is tailored toward a distinct objective and produces different types of results. k-means methods cluster the observed dFC data into a number of clusters, where the cluster centroids represent specific “states” of the overall whole-brain connectivity patterns observed over time. In comparison, dyna-LOCUS decomposes the observed dFCs to reveal the latent connectivity sources/traits underlying the overall whole-brain connectivity, where each trait represents a subset of brain connections that tend to occur together during neural processing. Furthermore, dyna-LOCUS uncovers the temporal expression profiles of the latent connectivity traits, Therefore, results from dyna-LOCUS provide new information to reveal the underlying connectivity traits that constitute the whole-brain connectivity “states” and further reveal the key connectivity traits that drive each whole-brain connectivity “state”. Compared to decomposition methods such as ICA or PCA, dyna-LOCUS shares the similar goal of decomposing dFC matrices to reveal underlying latent sources. However, dyna-LOCUS incorporates several innovative methodology strategies that lead to advancements over the existing ICA/PCA methods. Specifically, dyna-LOCUS incorporates innovations such as a low-rank structure to improve estimation efficiency and accuracy, an angle-based sparsity regularization for reliably recovering source signal maps by reducing spurious edges, and a temporal smoothness regularization to improve the estimation of temporal expression profiles of connectivity traits. A related work that models dFC using sparse low-rank matrices is (Cai et al., 2017). In this work, (Cai et al., 2017) decomposes dFC as a linear combination of a set of dynamic sparse connectivity patterns (dSCPs), modeled via rank-one matrices with sparsity imposed on the rank-one vector the dsCPs. There are several distinctions between dyna-LOCUS and dSCPs. In comparison to dSCPs that uses rank-one matrices, dyna-LOCUS uses a more flexible low-rank factorization where the rank parameters are source-specific to accommodate different types of connectivity traits and can be selected using our proposed adaptive selection method. To achieve sparsity, dyna-

LOCUS utilizes a novel angle-based element-wise sparsity regularization, specifically designed to achieve sparsity in the connectivity sources. Furthermore, (Cai et al., 2017) focuses on identifying whole-brain dFC states which are combinations of the underlying connectivity sources and investigating group differences in these dFC states. Hence, (Cai et al., 2017) selects the number of sources based on classification accuracy for subject groups. In comparison, our paper focuses on reliably uncovering latent connectivity sources and revealing their temporal dynamics, interactions, and synchronization with our proposed temporal metrics. We select the number of sources based on the reproducibility of the extracted sources.

Another related method is the recently proposed LOCUS method for decomposing static functional connectivity matrices (Wang and Guo, 2023). Both dyna-LOCUS and LOCUS are designed to investigate brain functional connectivity(FC) and share similarities such as modeling connectivity patterns using a low-rank structure and sparsity regularization. However, the two methods have major differences in terms of the type of functional connectivity data the methods are applied to, the results generated from the methods, and the analytical approaches. LOCUS is designed for decomposing stationary functional connectivity measures that are derived based on the whole fMRI time series in a scanning session, ignoring the temporal changes in brain connectivity. In comparison, dyna-LOCUS uncovers latent sources underlying the series of dynamic FC measures obtained using fMRI BOLD signals within short time windows that slide across the session. In addition to recovering the spatial compositions of the underlying dynamic connectivity sources/traits, dyna-LOCUS also produces temporal trait loading series and several novel temporal dynamic metrics to measure the energy and variation of each dynamic trait and to reveal the interaction and synchronization between traits. Furthermore, dyna-LOCUS offers an efficient and reliable approach to identifying whole-brain dynamic connectivity states and elucidating the contribution of each connectivity trait to the dynamic FC states. These temporal dynamic insights into functional connectivity are not attainable using the LOCUS method. There are also technical differences between the two methods. dyna-LOCUS models the temporal expression profile of each connectivity trait and incorporates temporal smoothness regularization. This leads to variations in optimization strategies, particularly for updating the loading matrix in the decomposition model. The LOCUS method updates the trait loading matrix using a simple regression method by regressing subjects' static FC against the estimated static sources. In comparison, dyna-LOCUS updates its temporal loading matrix using a new and more sophisticated strategy that takes into account the temporal smoothness regularization (Appendix B). The aforementioned distinguishing features underscore the unique capabilities of dyna-LOCUS in probing dynamic connectivity, thereby offering novel contributions beyond the LOCUS method.

Our dyna-LOCUS analysis of rs-fMRI data from the PNC study has led to exciting insights into the latent sources that underlie the brain's dynamic functional connectome. Among the 30 dyna-LOCUS extracted latent connectivity traits, an impressive 18 of them demonstrate substantial or almost perfect reproducibility with the resampling of study participants. These traits represent consistent subsystems in neural processing and brain organizations, reflecting the cohesive interactions among specific neural circuits. The interplay among these subsystems results in the dynamic reconfiguration of the brain's functional connectivity patterns over time. The sparse source signal maps generated by dyna-LOCUS allow us to reliably identify the key neural connections and brain nodes that drive each of these subsystems. In addition to unveiling the composition of the connectivity traits, dyna-LOCUS offers important insights into the temporal expression characteristics of each trait, providing a deeper understanding of how these traits manifest over time. For instance, Trait 6 (CB), known for its cerebellum-driven nature, emerges as a stable trait with a modest expression level across time. This stability suggests a consistent and enduring involvement of the cerebellum-driven connectivity pattern in the brain's dynamic organization. In contrast, Trait 27 (SM-DMN-Aud), encompassing long-distance connections between the sensorimotor and DMN networks, exhibits both the highest energy and the greatest temporal variation. This trait displays periods of heightened expression, followed by periods where it becomes less prominent or even disappears. This dynamic behavior suggests the presence of transient interactions between the sensorimotor and DMN networks, which may play crucial roles during specific cognitive processes or behavioral states. The identification of such distinct dynamic features within connectivity traits sheds light on the complexity and diversity of brain dynamics. Understanding the temporal expression patterns of connectivity traits enriches our knowledge of how different brain networks interact and adapt over time, unveiling the intricate mechanisms that underlie various cognitive functions and behaviors.

The dyna-LOCUS analysis of the PNC study has uncovered previously unknown latent connectivity sources within the dynamic connectome. Particularly noteworthy is the discovery of Trait 14 (EC-AUD-DMN-FRP), which is mainly driven by the executive control network and its connections with other cognitive networks. This executive connectivity subsystem exhibits the highest expression level among the connectivity traits, and it also demonstrates the best stability among the highly expressed traits. These findings highlight the consistent and prevalent presence of this executive connectivity subsystem throughout different time points in the dynamic functional connectome. Additionally, the executive connectivity subsystem demonstrates significant interaction with other connectivity traits. Notably, Trait 14 is involved in two out of the top six synchronized pairs of connectivity traits (Figure 2.8), indicating its active cooperation with other connectivity subsystems within the connectome. Moreover, our analysis has uncovered seven distinct whole-brain dFC states (Figure 2.10). Among these states, dFC state 1 emerges as the most frequently observed, with Trait 14 standing out as the most influential connectivity subsystem, displaying the highest expression level among all connectivity traits in this dFC state. These findings collectively emphasize the central role of Trait 14, the executive connectivity subsystem, in the dynamic functional connectome. Upon conducting a deeper investigation into the development of Trait 14, we made an intriguing discovery regarding distinct developmental trajectories between genders. In females, early development of this executive connectivity subsystem is observed, whereas in males, its development occurs later, mostly from adolescence to early adulthood. As young adults, males display a more pronounced presence of the executive connectivity subsystem compared to females. This gender-specific variation adds another dimension to our understanding of the executive connectivity subsystem.

Investigating latent neurocircuitry traits underlying longitudinal brain functional connectome We develop a longitudinal-LOCUS method to study changes in brain connectomes over time. This method decomposes longitudinal FC measures using blind source separation with low-rank structures and angle-based sparsity regularization. We present an efficient iterative node-rotation algorithm to solve the non-convex optimization problem for learning longitudinal-LOCUS. Simulations demonstrate superior accuracy in recovering latent sources and mixing coefficients compared to existing methods. We applied it to the Adolescent Brain Cognitive Development (ABCD) data to investigate developmental changes in neural circuits and their differences between genders. In this work, we employ an L1 norm-based sparsity penalty and our framework is adaptable to alternative norms like L2, MCP, and SCAD. Longitudinal-LOCUS can be extended to support various connectivity measures including structural connectivity from DTI and functional connectivity using mutual information. The angle-based sparsity regularization in the method extends its applicability to tensor-decomposition methods with low-rank structures, providing a robust alternative to existing penalties. To facilitate wider use, we intend to release an R package on CRAN, making the method more accessible for neuroimaging research.

Nonparametric Motion Control in Functional Connectivity Studies in

Children with Autism Spectrum Disorder and MoCo

We introduce MoCo, a method for controlling motion in fMRI studies to estimate the difference in functional connectivity between ASD and non-ASD children when all children have a tolerable level of motion. Our theoretical framework enables the use of flexible machine-learning techniques for parameter estimation. The method allows the estimation of simultaneous confidence intervals for controlling FWER across hundreds of brain connections. It shows improved statistical power in identifying the true association and a lower type I error rate in the absence of an association. Our approach avoids selection bias caused by motion quality control exclusion criteria.

Our findings differ greatly from the naïve approach, which suggested hypoconnectivity across many DMN regions when including all participants. The naïve approach with participant removal suggests these differences were due to motion artifacts, but it is difficult to disentangle this from power loss and selection biases, as only 34 ASD children passed motion quality control. MoCo contributes to the ASD literature by flexibly modeling motion artifacts while including all the phenotypic variability in the study sample, providing stronger evidence that the hypoconnectivity differences were due to motion artifacts. MoCo recovered more regions than the naïve approach with motion removal (four versus two at FWER=0.05), although the overall picture suggests relatively minor differences between autistic and non-ASD children in this study sample.

An important decision in the modeling process is to designate variables as possible confounders X or variables related to diagnosis group Z. In the data analysis, we consider age, sex, and handedness as possible confounders X, all of which can potentially affect the associations between ASD and functional connectivity due to an imbalanced study design. We treat FIQ as a diagnosis-specific variable, which on average was lower in the ASD group. However, FIQ is highly variable in autism, and whether or not it should be considered as a possible confounder or as a diagnosis-specific variable is debatable. In our dataset, the child with the highest FIQ was also diagnosed with autism (Supplement Table 5). Neural diversity in autism is associated with strengths like unique perspectives, problem-solving skills, intense focus, attention to detail, and other traits that extend beyond a single measure of intelligence. A limitation in our data analysis is that we only considered a limited set of behavioral and diagnosis-specific covariates, which was driven by the covariates available in the two ABIDE study sites. Additional research into the associations between functional connectivity and neural diversity may help elucidate the neurological underpinnings in ASD.

6.2 Future Work

Based on the findings of this dissertation, there are several potential avenues for future research.

Although dyna-LOCUS and longitudinal-LOCUS are reliable methods for decomposing dFC/longitudinal-FC measures, a potential limitation is that both decompose the FC matrices as a linear combination of latent connectivity traits, following the common assumption in blind source separation. However, this assumption may not always capture the complexity of real-world systems. A valuable direction for future work is to extend these methods to accommodate nonlinear mixing cases, thereby broadening their applicability across diverse data scenarios and enhancing their capacity to capture complex relationships. Another potential extension is to incorporate the spatial information of the nodes into modeling the latent coordinates in the low-rank factorization. This can further increase the accuracy and reliability in recovering the connectivity traits by accounting for the spatial dependence between the nodes.

For dyna-LOCUS, we are currently utilizing the PNC data to examine the results of dynamic functional connectivity matrix decomposition. As we are embracing many sources of data nowadays, there is significant potential for a cross-dataset large-scale study to compare the dynamic connectivity traits derived from datasets such as PNC, Human Connectome Project (HCP), and ABCD. Such a study would require a consistent preprocessing pipeline for all datasets and harmonization across sites. Our group has completed the preprocessing and is working on dynamic connectivity calculation. A thorough comparison will be part of future work.

For longitudinal functional connectivity analysis, instead of modeling the original functional connectivity for consecutive visits, we could directly decompose the difference in functional connectivity maps to identify the underlying neural circuitry reflected in the difference map. Another option is to use a dyna-LOCUS-like model for this analysis. Since the longitudinal imaging data also involves multiple time points, this approach allows the study of temporal profiles while assuming that the connectivity traits remain unchanged over time.

For MoCo, although we extend the method in Chapter 5 to incorporate behavioral and demographic information from participants with missing fMRI data and other neuroimaging modalities, several future directions remain possible. First, since the method uses machine learning, particularly the Super Learner package, it is important to investigate its robustness with regard to the seed when running SuperLearner, which uses cross-validation for model weight tuning. One potential improvement is to develop a probability measure for each brain region's significance in the context of multiple testing. Second, we currently fit functional connectivity for a seed-based analysis rather than simultaneously analyzing the functional connectivity matrix. A matrix-variate approach could be designed to exploit low-rank or sparse structures, potentially improving efficiency. Third, instead of applying the method to correlations, we could directly apply the logic to time series collected from fMRI before calculating functional connectivity. This would require modeling the time series data and addressing how to handle vector outcomes.

Chapter 7

Appendix

7.1 Supplementary Material of Chapter 2

7.1.1 Latent dynamic connectivity traits extracted from the PNC study using dyna-LOCUS

dyna-LOCUS uncovers 30 dynamic latent connectivity traits by decomposing dynamic functional connectivity (dFC) measures derived from resting state fMRI (rs-fMRI) data from the Philadelphia Neurodevelopmental Cohort (PNC) project. Figure A.1 (Part I - Part V) presents the source signal maps for all 30 connectivity traits ranked by their reliability index. In the figure, the top 0.5% brain connections with the highest magnitude of source signal intensity in each of the connectivity traits are mapped onto the brain. Node contribution indices that help identify key brain nodes and networks that drive each connectivity trait are also shown in boxplots arranged by networks.

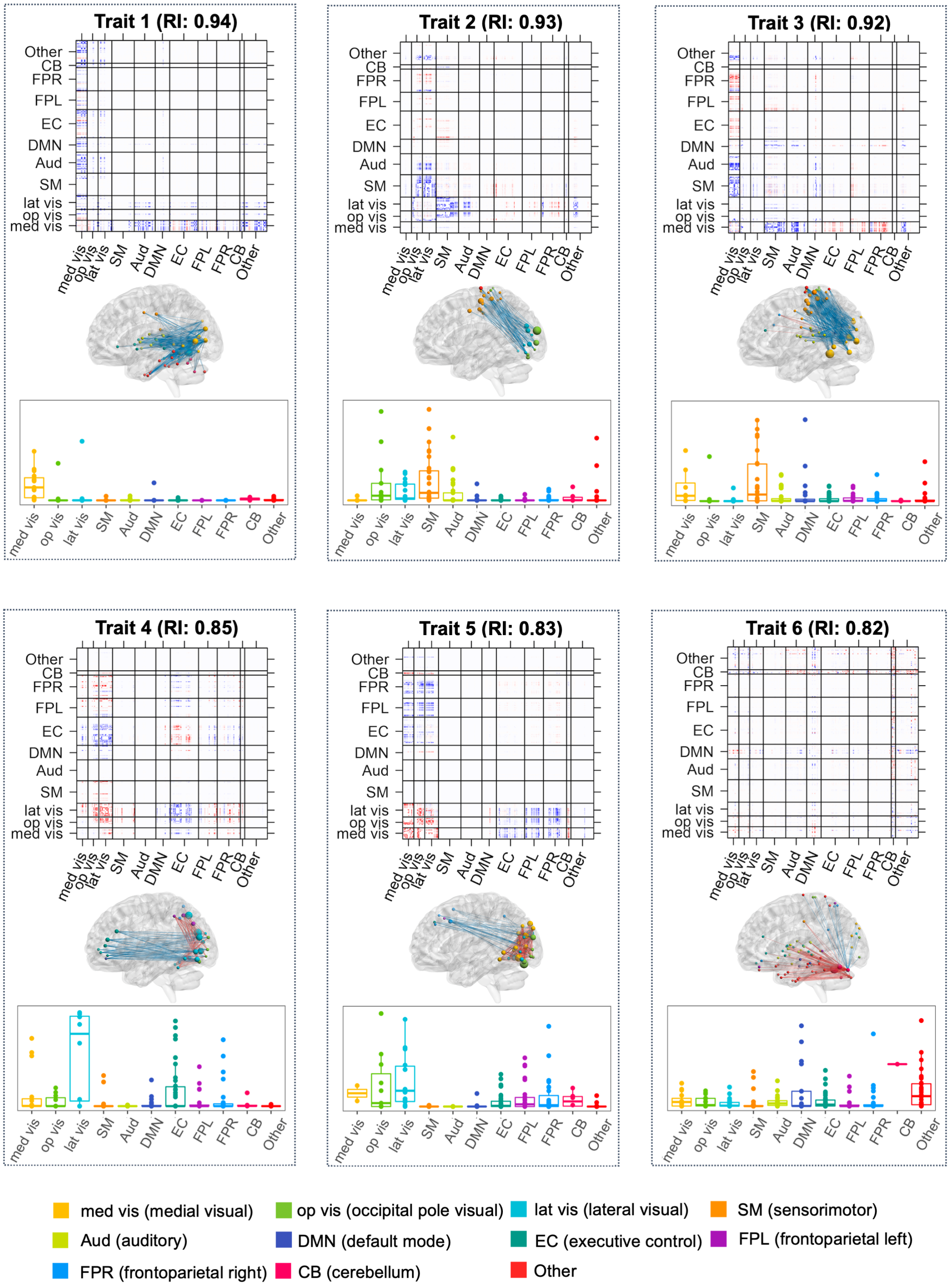


Figure 7.1: Part I: Thirty connectivity traits extracted from the PNC study ordered by their reliability index. The top 0.5% brain connections and top contributing nodes are displayed in the brain maps. Node contribution indices that help identify key brain nodes and networks driving each connectivity trait are shown in the boxplot arranged by networks.

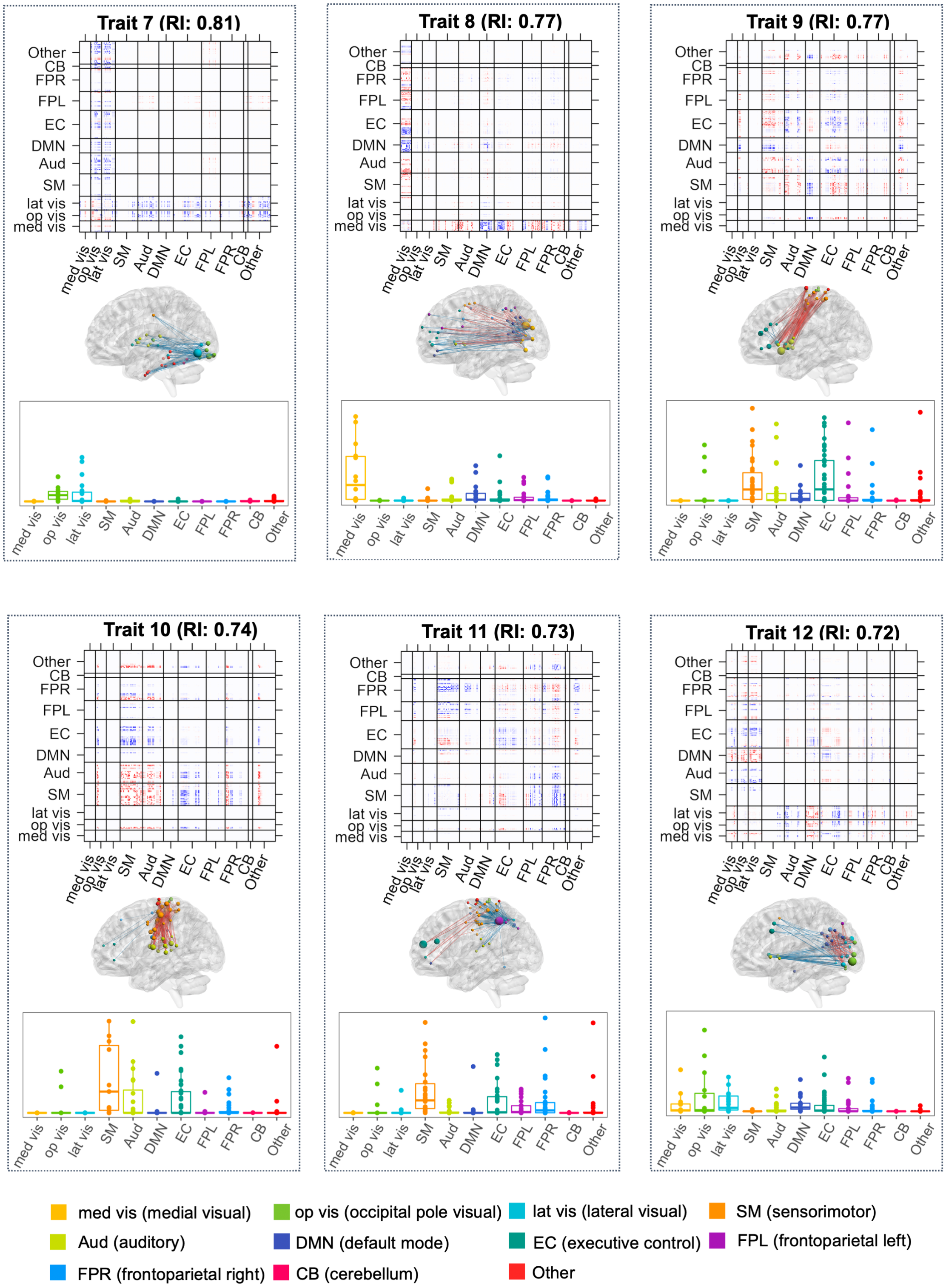


Figure 7.1: Part II: Thirty connectivity traits extracted from the PNC study ordered by their reliability index. The top 0.5% brain connections and top contributing nodes are displayed in the brain maps. Node contribution indices that help identify key brain nodes and networks driving each connectivity trait are shown in the boxplot arranged by networks.

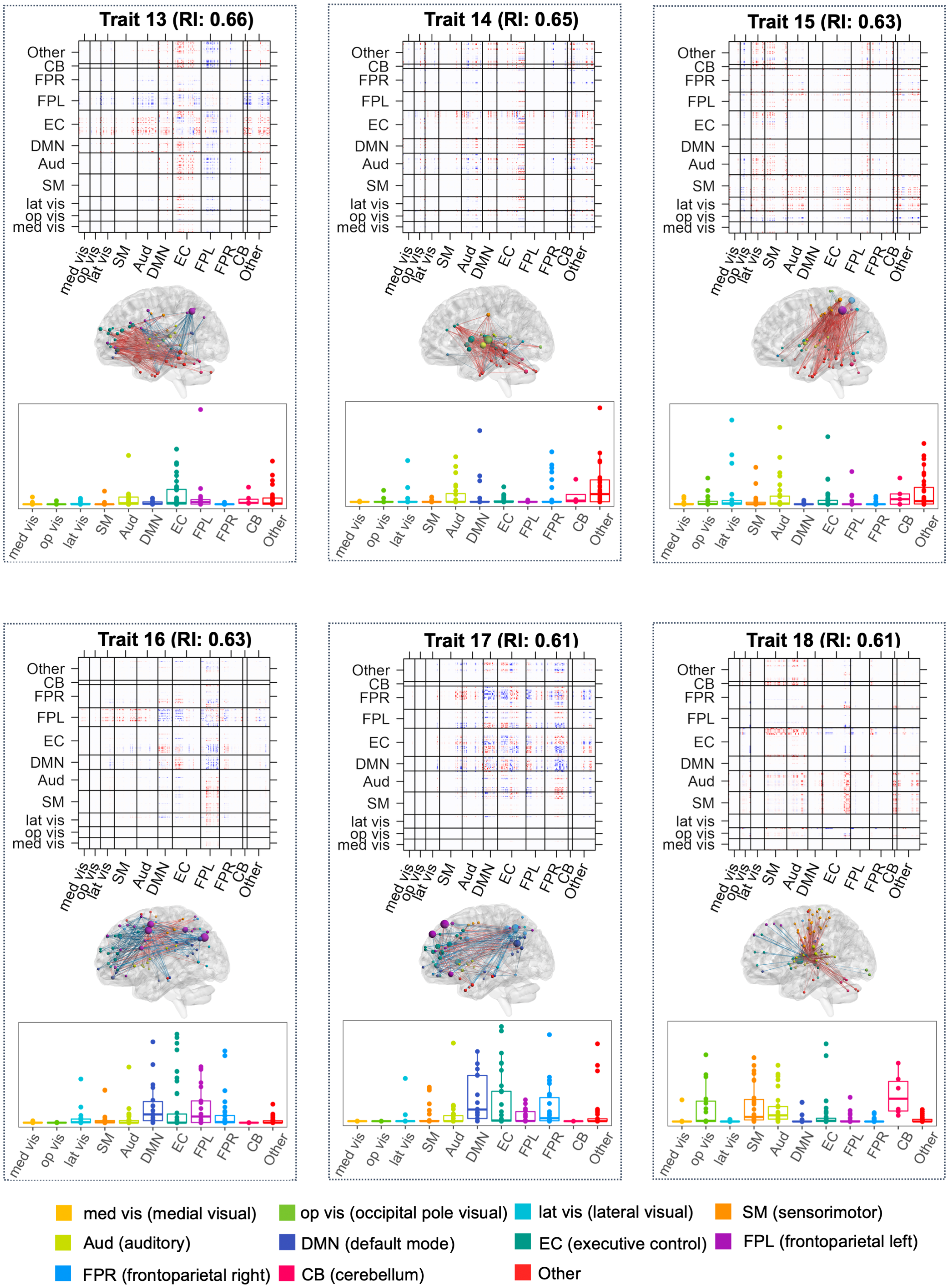


Figure 7.1: Part III: Thirty connectivity traits extracted from the PNC study ordered by their reliability index. The top 0.5% brain connections and top contributing nodes are displayed in the brain maps. Node contribution indices that help identify key brain nodes and networks driving each connectivity trait are shown in the boxplot arranged by networks.

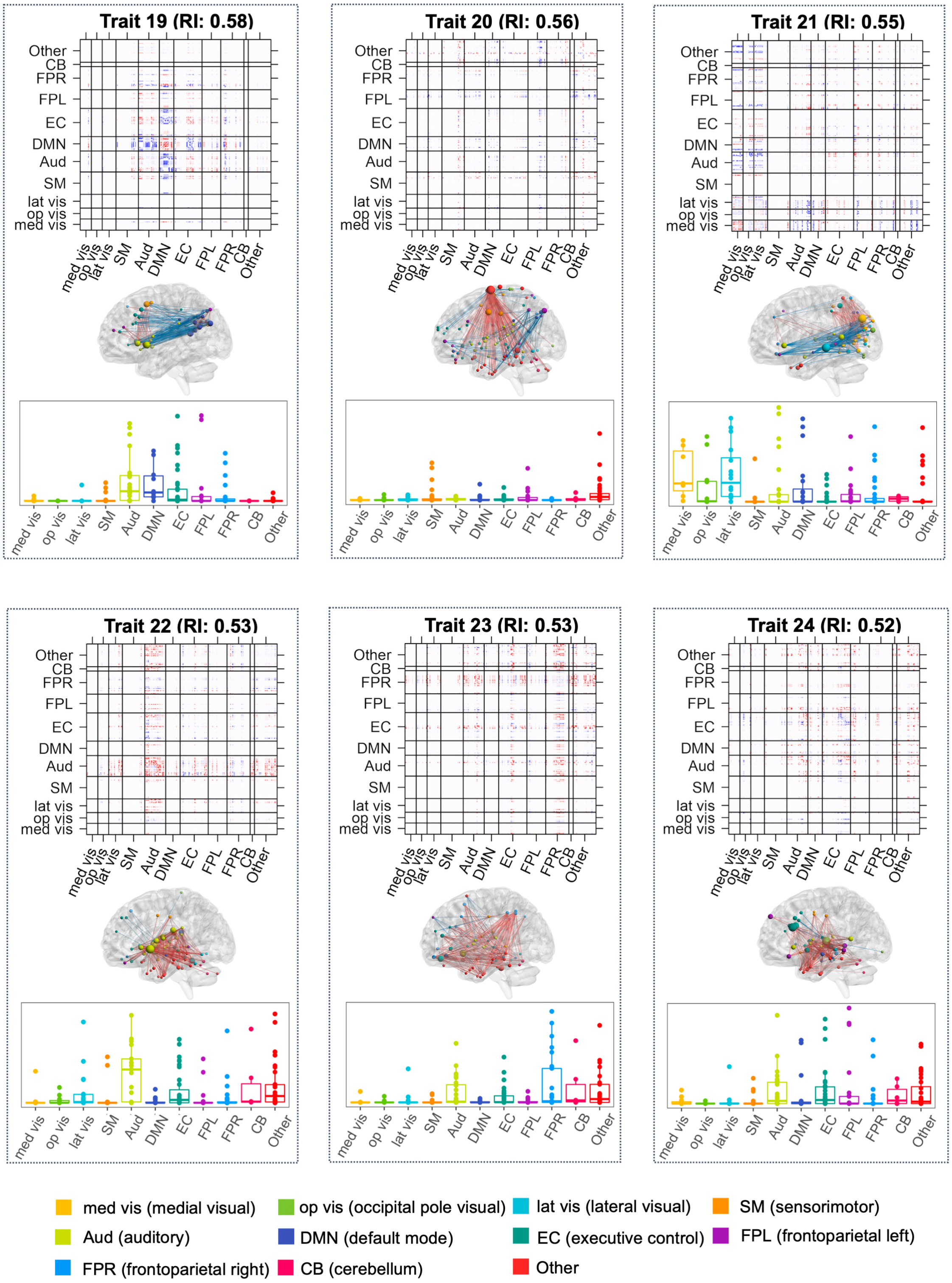
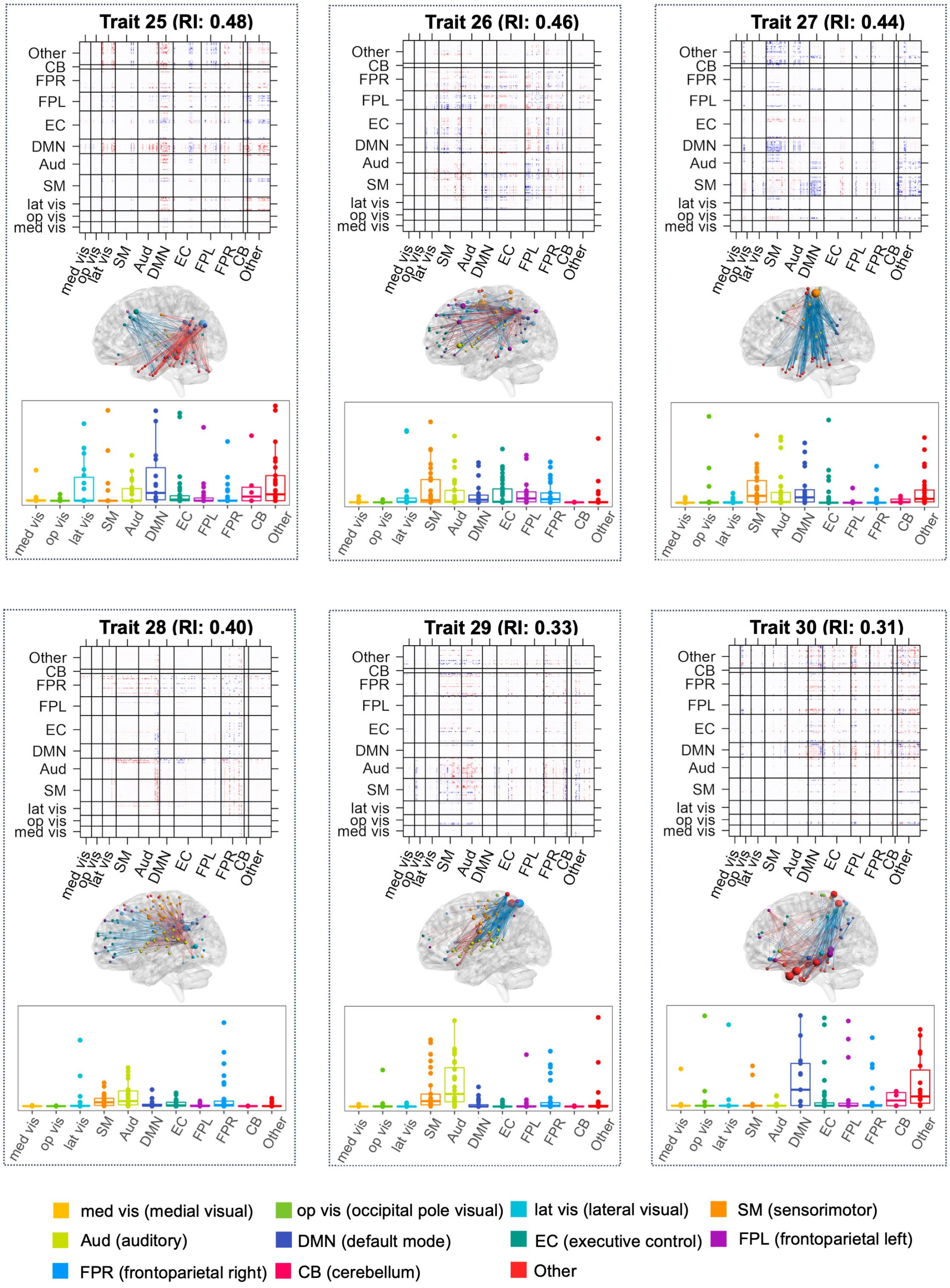


Figure 7.1: Part IV: Thirty connectivity traits extracted from the PNC study ordered by their reliability index. The top 0.5% brain connections and top contributing nodes are displayed in the brain maps. Node contribution indices that help identify key brain nodes and networks driving each connectivity trait are shown in the boxplot arranged by networks.



Trait 25 (RI: 0.48)

Trait 26 (RI: 0.46)

Trait 27 (RI: 0.44)

Trait 28 (RI: 0.40)

Trait 29 (RI: 0.33)

Trait 30 (RI: 0.31)

Figure 7.1: Part V: Thirty connectivity traits extracted from the PNC study ordered by their reliability index. The top 0.5% brain connections and top contributing nodes are displayed in the brain maps. Node contribution indices that help identify key brain nodes and networks driving each connectivity trait are shown in the boxplot arranged by networks.

7.1.2 Latent dynamic connectivity traits extracted from the PNC study using existing methods

In this section, we present the 30 dynamic latent connectivity traits uncovered by connICA (Amico et al., 2017) (Figure 7.2) and dictionary learning (DL) (Figure 7.3) by decomposing dynamic functional connectivity (dFC) measures derived from resting-state fMRI (rs-fMRI) data from the Philadelphia Neurodevelopmental Cohort (PNC) project. For each method, the traits are matched one-to-one with dyna-LOCUS traits presented in Supplementary Materials Section 1 and ordered accordingly.

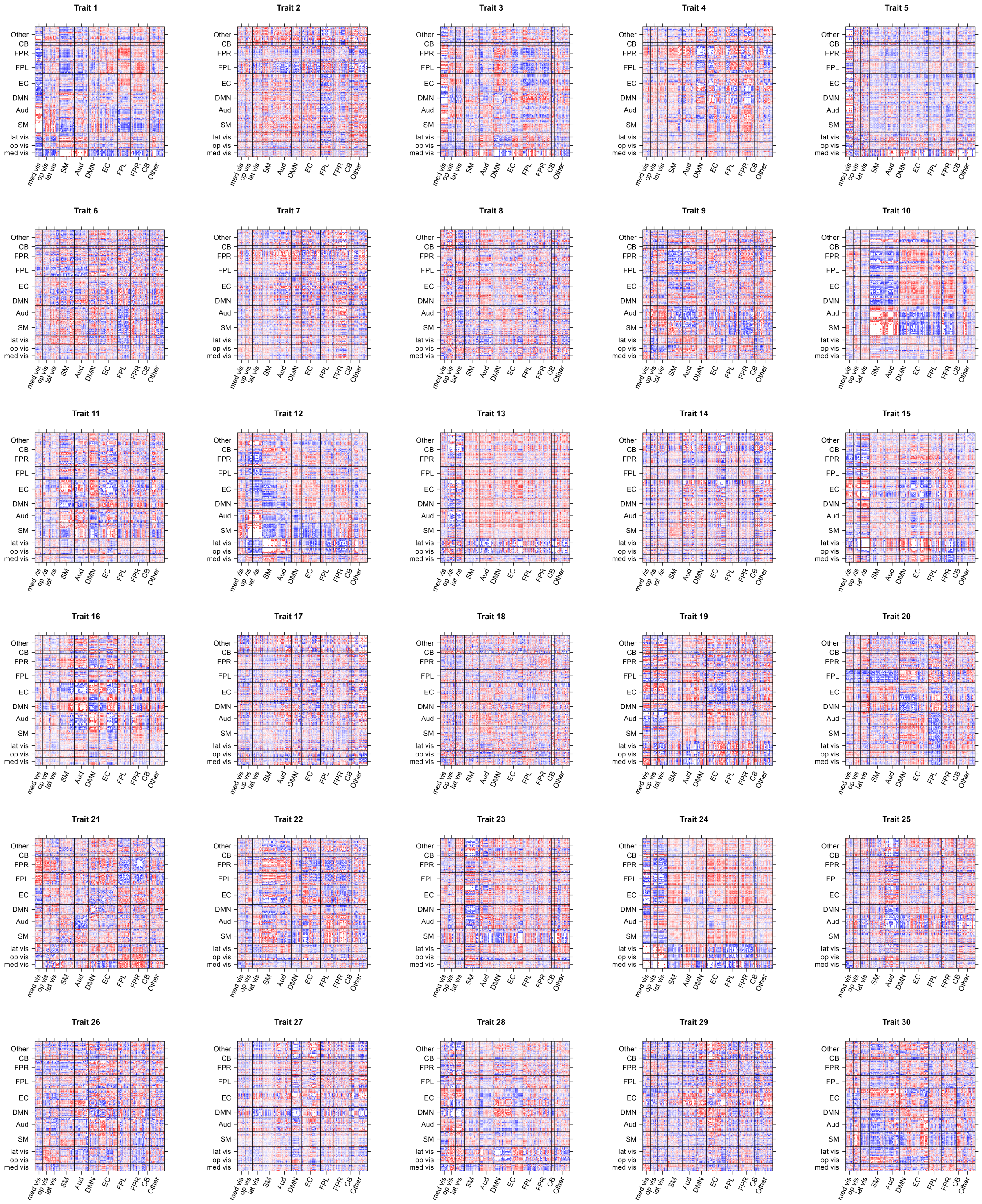


Figure 7.2: 30 dynamic latent connectivity traits uncovered by connICA using dFC measures derived from rs-fMRI data from the PNC project. These traits are matched one-to-one with the traits extracted by dyna-LOCUS and ordered accordingly.

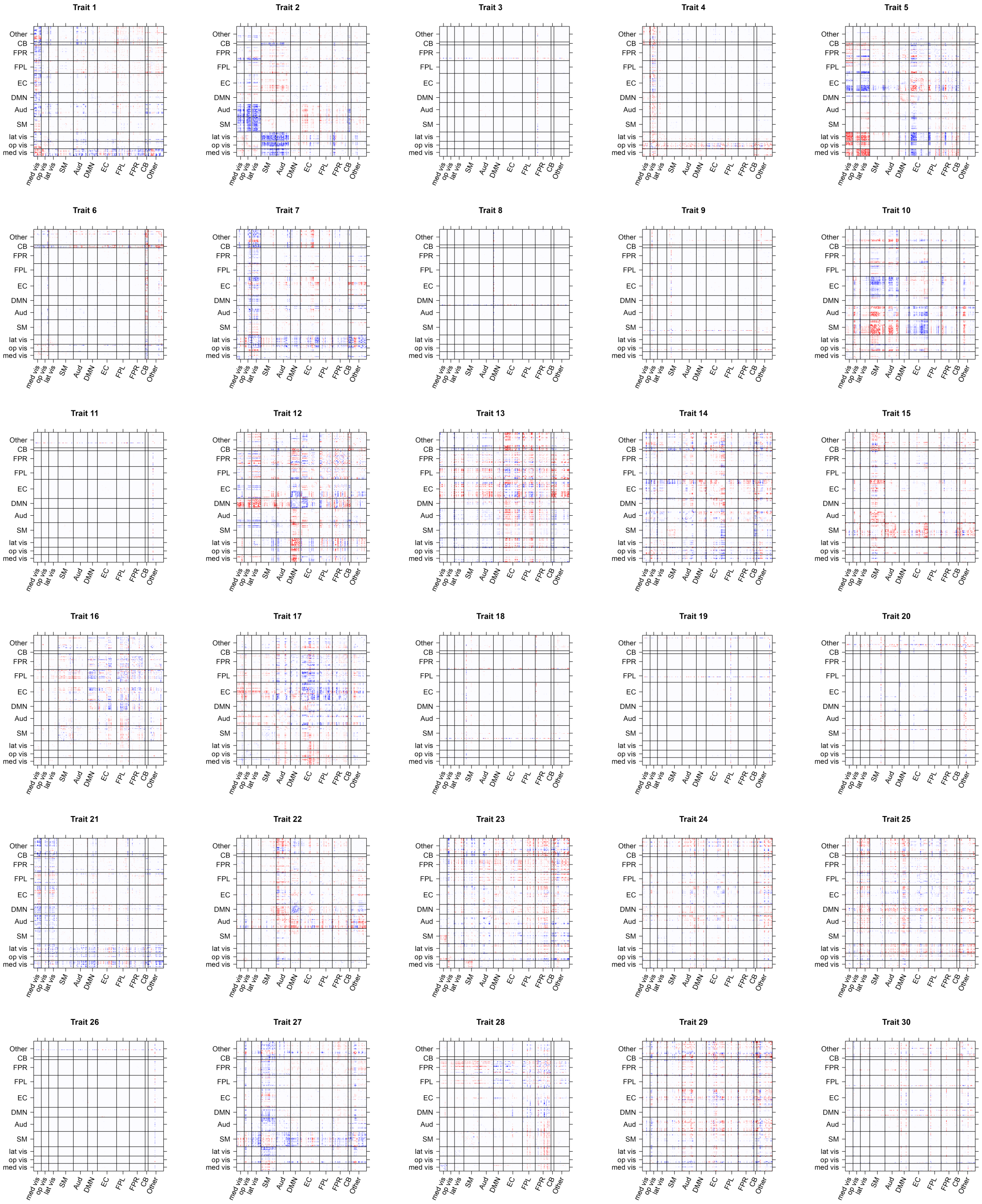

Trait 1
Trait 2
Trait 3
Trait 4
Trait 5
Trait 6
Trait 7
Trait 8
Trait 9
Trait 10
Trait 11
Trait 12
Trait 13
Trait 14
Trait 15
Trait 16
Trait 17
Trait 18
Trait 19
Trait 20
Trait 21
Trait 22
Trait 23
Trait 24
Trait 25
Trait 26
Trait 27
Trait 28
Trait 29
Trait 30

Figure 7.3: 30 dynamic latent connectivity traits uncovered by dictionary learning (DL) using dFC measures derived from rs-fMRI data from the PNC project. These traits are matched one-to-one with the traits extracted by dyna-LOCUS and ordered accordingly.

7.1.3 Gender and age differences in the connectivity traits

In this section, we first validate dyna-LOCUS's finding of the age and gender effect for the executive function related connectivity traits extracted from the PNC study. Among the 30 extracted traits, Trait 14 (EC-Aud-DMN-FPR) consists of connections within the executive control (EC) network and connections between the EC network and several other networks. The regression analysis of the logarithm of energy of individual's trait loadings on Trait 14 shows a significant interaction effect between age and gender. To validate this finding from dyna-LOCUS, Specifically, we generate 100 replication data samples, each consisting of a subset of 376 participants (75% of the full sample size) randomly sampled from the PNC data. We then perform the same regression analysis and test for the age and gender interaction effect in each of the replication samples. Figure 7.4 shows the p-values across the 100 replication samples. When there isn't age by gender interaction effect, we expect the p-value would approximately follow a uniform distribution between 0 and 1 (Moore et al., 2018), meaning they should be evenly distributed within this range. The violin plot reveals that the p-values are predominantly smaller than expected under the null hypothesis. In approximately 40% of the replication samples, the results are significant with p-values less than 0.05. And the results in approximately 60% of the replication samples are significant at the significance level of 0.1. This indicates the presence of an interaction effect between age and gender for Trait 14, validating the finding reported in the paper.

Furthermore, we present gender and age effects for all 30 connectivity traits extracted from the PNC study by dyna-LOCUS and the existing connICA and dictionary learning (DL) methods. We match the traits from connICA and DL to those from dyna-LOCUS. We then employ regression models to model the logarithm of energy of an individual's trait loadings in terms of age, gender, and their interaction. The p-values for these effects from each method are presented in Table S7.1. Due

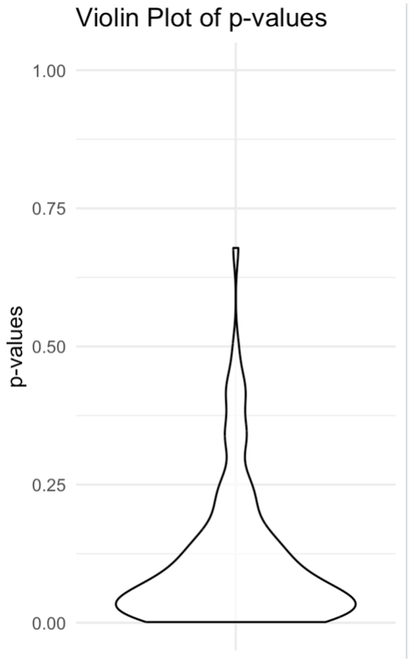


Figure 7.4: Validation of the age-by-gender interaction effect the executive function related connectivity trait (Trait 14 EC-Aud-DMN-FPR) using the data resampling method. The violin plot displays the p-values of this effect in 100 replication data samples from the PNC study.

to the simultaneous hypothesis testing across all 30 traits, we implement multiple comparison corrections using the false discovery rate (FDR) method. Effects that are significant after controlling for the FDR at the 0.05 level are marked with asterisks in the table.

Trait	dyna-LOCUS			connICA			DL		
	gender	age	interaction	gender	age	interaction	gender	age	interaction
1	0.466	0.240	0.889	0.055	0.609	0.801	0.378	0.446	0.642
2	0.921	0.682	0.369	0.517	0.023	0.498	0.658	0.162	0.280
3	0.938	0.159	0.296	0.489	0.203	0.577	0.089	0.010*	0.356
4	0.846	0.400	0.928	0.794	0.887	0.712	0.909	0.052	0.053
5	0.819	0.008*	0.526	0.437	0.051	0.463	0.831	0.060	0.218
6	0.333	0.063	0.175	0.414	0.208	0.617	0.326	0.293	0.169
7	0.794	0.887	0.712	0.003*	0.230	0.026	0.322	0.156	0.295
8	0.437	0.051	0.463	0.576	0.005*	0.586	0.031	0.000*	0.042
9	0.697	0.226	0.267	0.456	0.382	0.359	0.892	0.004*	0.933
10	0.055	0.609	0.801	0.938	0.159	0.296	0.503	0.066	0.706
11	0.104	0.233	0.276	0.823	0.574	0.517	0.780	0.003*	0.058
12	0.280	0.488	0.384	0.191	0.797	0.479	0.044	0.721	0.309
13	0.003*	0.230	0.026	0.090	0.002*	0.324	0.212	0.085	0.862
14	0.493	0.000*	0.030	0.466	0.240	0.889	0.146	0.000*	0.018
15	0.000*	0.531	0.598	0.545	0.023	0.210	0.132	0.434	0.886
16	0.130	0.677	0.458	0.036	0.001*	0.781	0.004	0.279	0.672
17	0.191	0.797	0.479	0.000*	0.531	0.598	0.834	0.000*	0.085
18	0.517	0.023	0.498	0.130	0.677	0.458	0.264	0.013*	0.252
19	0.036	0.001*	0.781	0.333	0.063	0.175	0.041	0.121	0.243
20	0.024	0.308	0.004	0.024	0.308	0.004	0.008	0.014*	0.056
21	0.456	0.382	0.359	0.394	0.171	0.261	0.052	0.079	0.102
22	0.545	0.023	0.210	0.697	0.226	0.267	0.075	0.360	0.976
23	0.090	0.002*	0.324	0.104	0.233	0.276	0.474	0.318	0.398
24	0.489	0.203	0.577	0.921	0.682	0.369	0.975	0.131	0.253
25	0.166	0.074	0.407	0.819	0.008*	0.526	0.054	0.027	0.732
26	0.576	0.005*	0.586	0.493	0.000*	0.030	0.121	0.001*	0.025
27	0.823	0.574	0.517	0.280	0.488	0.384	0.423	0.471	0.564
28	0.257	0.008*	0.007	0.257	0.008*	0.007	0.646	0.381	0.925
29	0.394	0.171	0.261	0.166	0.074	0.407	0.552	0.343	0.452
30	0.414	0.208	0.617	0.846	0.400	0.928	0.793	0.019	0.061

Table 7.1: Age and Gender effects for all 30 connectivity traits extracted from the PNC study by dyna-LOCUS and the existing connICA and dictionary learning (DL) methods. Effects that are significant after controlling for the false discovery rate (FDR) at the 0.05 level are marked with asterisks in the table.

7.1.4 Additional simulation studies with varying levels of sparsity in the source signals

In this section, we evaluate the performance of dyna-LOCUS across varying sparsity levels of source signals. In addition to the simulation scenario presented in the main manuscript where the sources have a high sparsity level, we consider two additional settings where the sources exhibit decreasing levels of sparsity as the number of connections increases. Table 7.2 and Figure 7.5 depict simulation results for a medium sparsity level, while Table 7.3 and Figure 7.6 present results for a low sparsity level. Overall, dyna-LOCUS exhibits better accuracy in recovering the underlying source signals and their respective temporal loadings compared to the connICA and sparse dictionary learning (DL) methods. Results from these two additional simulation settings are consistent with the findings reported in the main manuscript for the high sparsity level setting.

Term	N	Var.	**dyna-LOCUS**	connICA	DL
Latent Source Corr. (SD)	20	Low	0.936 (0.006)	0.801 (0.023)	0.908 (0.011)
		Mid	0.911 (0.011)	0.762 (0.014)	0.890 (0.007)
		High	0.829 (0.011)	0.675 (0.015)	0.783 (0.020)
	50	Low	0.939 (0.002)	0.788 (0.001)	0.929 (0.004)
		Mid	0.925 (0.006)	0.771 (0.005)	0.916 (0.002)
		High	0.901 (0.008)	0.734 (0.010)	0.887 (0.003)
Loading Matrix Corr. (SD)	20	Low	0.987 (0.005)	0.799 (0.014)	0.917 (0.017)
		Mid	0.959 (0.005)	0.782 (0.010)	0.930 (0.007)
		High	0.889 (0.006)	0.730 (0.012)	0.857 (0.024)
	50	Low	0.988 (0.003)	0.792 (0.002)	0.940 (0.003)
		Mid	0.960 (0.005)	0.772 (0.007)	0.922 (0.002)
		High	0.911 (0.006)	0.731 (0.011)	0.881 (0.003)

Table 7.2: Simulation results for comparing dyna-LOCUS with connICA and DL for source signals with a medium sparsity level, based on 100 simulation runs conducted under three variance (Var.) settings. Values presented are mean and standard deviation of correlations between the true and estimated latent sources and loading/mixing matrices.

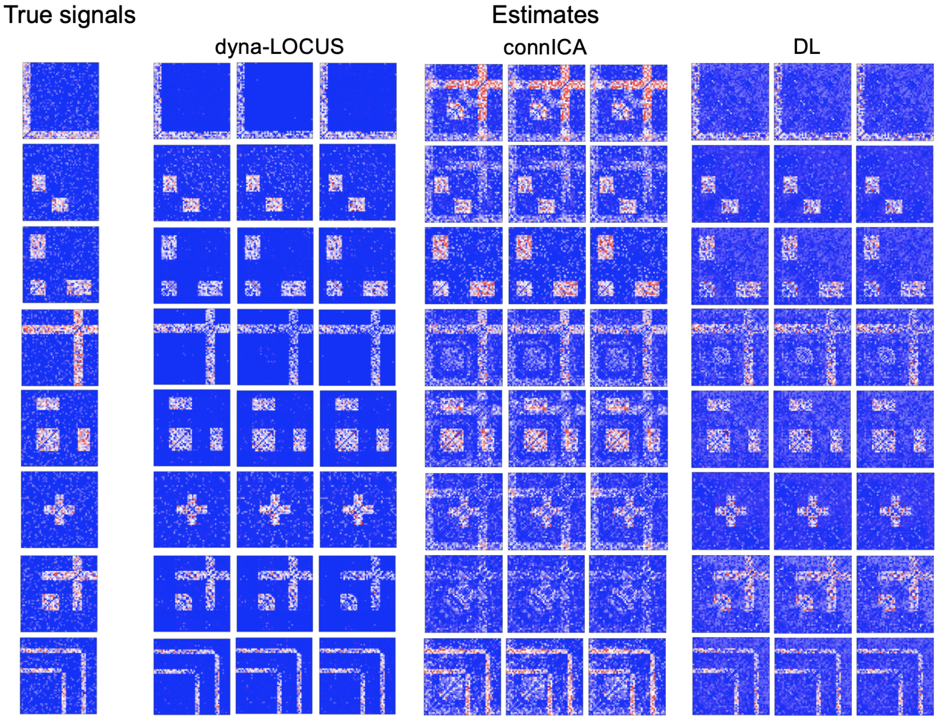


Figure 7.5: Results for the simulation study where the source signals exhibit a medium sparsity level. Figures depict the true source signals and the estimated signals by dyna-LOCUS, connICA, and DL in three randomly selected simulation runs.

Term	N	Var.	**dyna-LOCUS**	connICA	DL
Latent Source Corr. (SD)	20	Low	0.956 (0.002)	0.806 (0.017)	0.842 (0.029)
		Mid	0.911 (0.005)	0.765 (0.020)	0.893 (0.012)
		High	0.809 (0.012)	0.688 (0.020)	0.786 (0.024)
	50	Low	0.958 (0.001)	0.787 (0.005)	0.903 (0.006)
		Mid	0.937 (0.004)	0.772 (0.006)	0.913 (0.007)
		High	0.885 (0.012)	0.742 (0.003)	0.883 (0.014)
Loading Matrix Corr. (SD)	20	Low	0.996 (0.002)	0.802 (0.011)	0.834 (0.029)
		Mid	0.966 (0.004)	0.783 (0.013)	0.931 (0.013)
		High	0.893 (0.011)	0.741 (0.016)	0.857 (0.032)
	50	Low	0.997 (0.000)	0.791 (0.007)	0.922 (0.004)
		Mid	0.969 (0.004)	0.775 (0.008)	0.916 (0.007)
		High	0.914 (0.010)	0.742 (0.003)	0.883 (0.013)

Table 7.3: Simulation results for comparing dyna-LOCUS with connICA and DL for source signals with a low sparsity level, based on 100 simulation runs conducted under three variance (Var.) settings. Values presented are mean and standard deviation of correlations between the true and estimated latent sources and loading/mixing matrices.

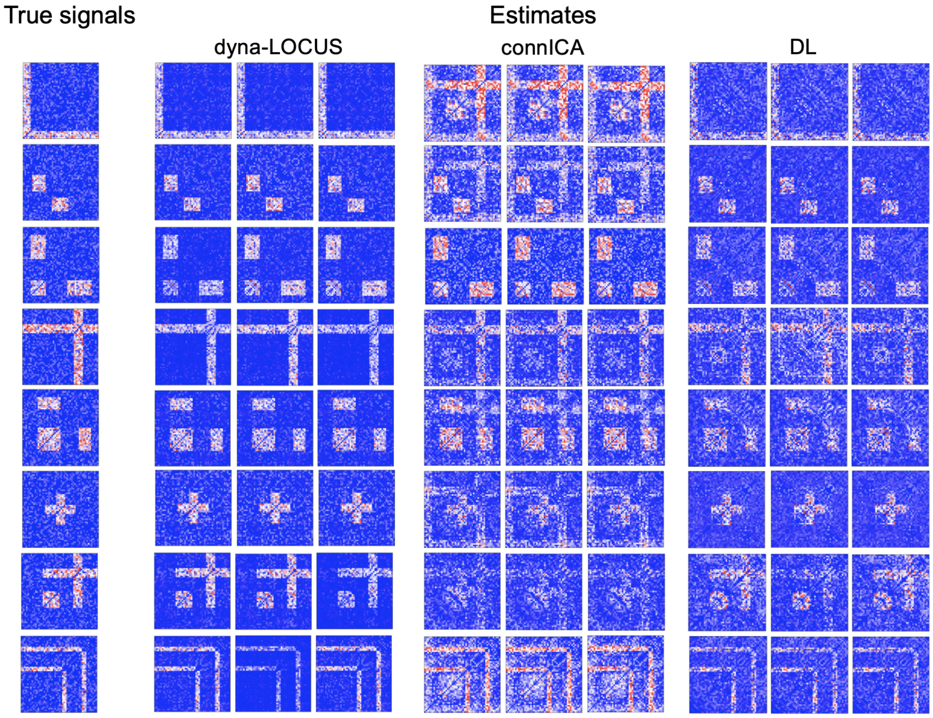


Figure 7.6: Results for the simulation study where the source signals exhibit a low sparsity level. Figures depict the true source signals and the estimated signals by dyna-LOCUS, connICA and DL in three randomly selected simulation runs.

7.1.5 Additional simulation study using source signals derived from the PNC study

In this section, we simulate dFC data using true source signals derived from connectivity traits extracted from the PNC study. We compare the performance of dyna-LOCUS with two other source separation methods: connICA (Amico et al., 2017) which is a recently developed connectivity ICA method, and the dictionary learning (DL) method (Mairal et al., 2009) which is a popular sparse decomposition method with l_1 sparsity penalization.

We specify $V = 264, q = 3$, and consider two sample sizes $N = 20, 50$ and $T = 26$ windows. We generate three latent connectivity source signals based on the connectivity traits derived directly from the PNC study (Figure 7.7). The mixing coefficients are also sampled from the estimates from the PNC study. Furthermore, we add zero mean

Gaussian noises to the mixture of signals where the variance is specified based on signal-to-noise ratio observed from PNC data. Specifically, we consider three variance settings with $\sigma^2 = 2^2, 3^2$, and 4^2, corresponding to low, medium, and high variance levels, respectively. In summary, we have 2×3 simulation settings with different combinations of sample sizes and variance levels. For each setting, we generate 100 simulation runs.

Following previous work (Beckmann et al., 2005; Wang and Guo, 2019, 2023), we evaluate the performance of each method based on the correlations between the truth and the model-based estimates. We further examine the standard deviation of the correlations across 100 simulation runs to evaluate the stability of the results.

Term	N	Var.	**dyna-LOCUS**	connICA	DL
Latent Source Corr. (SD)	20	Low	0.981 (0.000)	0.844 (0.001)	0.918 (0.001)
		Mid	0.925 (0.001)	0.722 (0.001)	0.824 (0.002)
		High	0.829 (0.002)	0.610 (0.002)	0.755 (0.003)
	50	Low	0.994 (0.000)	0.927 (0.000)	0.969 (0.000)
		Mid	0.983 (0.000)	0.854 (0.001)	0.955 (0.001)
		High	0.949 (0.001)	0.774 (0.001)	0.924 (0.001)
Loading Matrix Corr. (SD)	20	Low	0.996 (0.000)	0.996 (0.000)	0.996 (0.000)
		Mid	0.988 (0.001)	0.987 (0.000)	0.990 (0.000)
		High	0.970 (0.001)	0.969 (0.001)	0.979 (0.001)
	50	Low	0.997 (0.000)	0.996 (0.000)	0.997 (0.000)
		Mid	0.991 (0.000)	0.991 (0.000)	0.993 (0.000)
		High	0.981 (0.000)	0.980 (0.000)	0.986 (0.000)

Table 7.4: Simulation results for comparing dyna-LOCUS with connICA and DL based on 100 simulation runs conducted under three variance (Var.) levels. Values presented are mean and standard deviation of correlations between the true and estimated latent sources and loading/mixing matrices.

Results in Table 7.4 show that while the three methods show comparable accuracy for estimating the mixing coefficients, dyna-LOCUS consistently yields more accurate results in uncovering the latent connectivity sources. Figure 7.7 illustrates the true source signals alongside the estimated signals generated by dyna-LOCUS, connICA, and DL. In comparison with the two existing methods, dyna-LOCUS generates more accurate results with fewer false positive findings. Specifically, connICA, being a decomposition method without sparsity constraints and the low-rank structure, tends to yield noisy and inaccurate estimates. As a sparse decomposition method, DL produces sparse estimates for the source signals. However, it doesn't model the source signals using the low-rank structure and disregards the interdependence among brain connections. Instead, it treats connections as independent parameters, leading to a large number of parameters for DL to learn. As a result, DL yields less accurate results compared to the proposed dyna-LOCUS. For instance, for the second source signal, dyna-LOCUS successfully recovers the connections between the Visual Networks and EC, FPL, and FPR networks, while DL produces very weak or no signals for these connections and generates false positive findings between lat vis and med vis networks. Similarly, dyna-LOCUS successfully recovers the connections between SM and DMN, EC and FPR for the first source signals and the connections between the visual networks and FPL and FPR in the third source signals, while DL fails to achieve the same level of recovery.

7.1.6 Latent dynamic connectivity traits extracted from the PNC study using LOCUS

In this section, we present 30 static latent connectivity traits uncovered by LOCUS when decomposing static functional connectivity measures derived from resting-state fMRI (rs-fMRI) data from the PNC project (Figure 7.8). The traits are matched one-to-one with dyna-LOCUS traits in Supplementary Materials Section 1 and ordered accordingly.

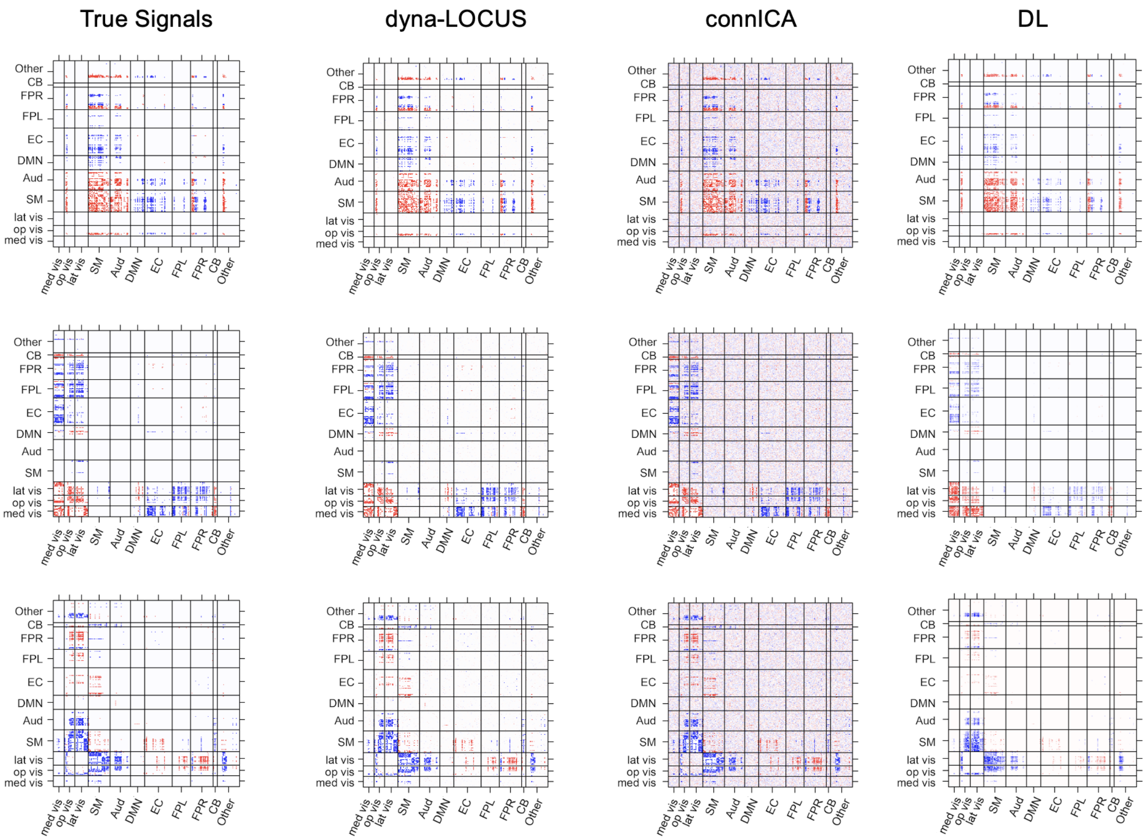


Figure 7.7: Results from the simulation study with source signals derived from the PNC study. The figures illustrate the true source signals and the estimated signals by dyna-LOCUS, connICA, and dictionary learning(DL) in three randomly selected simulation runs conducted under the low level variance setting.

7.1.7 An alternative estimation algorithm

The proposed node-rotation algorithm presented in the Algorithm 1 of Appendix B of the paper is developed based on the block multiconvexity of the objective function. It has the appealing theoretical property that updating each block of the parameters can be performed via convex optimization. Though being a highly efficient algorithm with analytic solutions, the node-rotation algorithm does involve rotating across the nodes to update the latent coordinates of each node $\boldsymbol{x}_\ell(v)(v = 1, \ldots, V)$. Here, we present an alternative estimation method to further accelerate computation in learning dyna-LOCUS, especially for studies with large sample sizes and brain atlases involving a large number of nodes. Instead of iteratively updating each node, the alternative

Figure 7.8: 30 static latent connectivity traits uncovered by LOCUS using static FC measures derived from rs-fMRI data from the PNC project. These traits are matched one-to-one with the dyna-LOCUS results, as detailed in Supplementary Materials Section 1.

algorithm updates the latent coordinates of all the nodes $\boldsymbol{X}_\ell$ simultaneously using an eigenvalue decomposition. Specifically, the objective function for dyna-LOCUS is:

$$\min_{\widetilde{\boldsymbol{A}},\{\boldsymbol{X}_\ell,\boldsymbol{D}_\ell\}} \sum_{l=1}^{q} ||\widetilde{\boldsymbol{Y}}'\widetilde{\boldsymbol{a}}_\ell - \mathcal{L}(\boldsymbol{X}_\ell \boldsymbol{D}_\ell \boldsymbol{X}_\ell')||_F^2 + \phi \sum_{\ell=1}^{q} \sum_{u<v} |\boldsymbol{x}_\ell(u)' \boldsymbol{D}_\ell \boldsymbol{x}_\ell(v)| + \lambda ||\boldsymbol{W}\widetilde{\boldsymbol{A}}||_F^2. \tag{7.1}$$

We initialize the algorithm with $\hat{\widetilde{\boldsymbol{A}}}^{(0)}, \{\hat{\boldsymbol{X}}_\ell^{(0)}, \hat{\boldsymbol{D}}_\ell^{(0)}\}$ derived from estimates based on existing methods such as connICA. The algorithm then iteratively updates the parameters through the following steps: *Step 1: Updating* $\boldsymbol{X}_\ell, \boldsymbol{D}_\ell$*.* At the kth iteration, denote $\boldsymbol{b}_\ell = \mathcal{L}(\boldsymbol{X}_\ell \boldsymbol{D}_\ell \boldsymbol{X}_\ell')$. We first derive a sparse solution for $\boldsymbol{b}_\ell$ by solving the following objective function:

$$\min_{\boldsymbol{b}_\ell} ||\widetilde{\boldsymbol{Y}}'\widetilde{\boldsymbol{a}}_\ell - \boldsymbol{b}_\ell||_2^2 + \phi||\boldsymbol{b}_\ell||_1 \tag{7.2}$$

An analytical solution $\hat{\boldsymbol{b}}_\ell$ can be obtained (Fan and Li, 2001). We map it back to the connectivity matrix form using $\mathcal{L}^{-1}$, and then conduct eigenvalue decomposition on the matrix $\mathcal{L}^{-1}(\hat{\boldsymbol{b}}_\ell)$ to obtain $\{\hat{\boldsymbol{X}}_\ell^{(k)}, \hat{\boldsymbol{D}}_\ell^{(k)}\}$. *Step 2: Updating* $\widetilde{\boldsymbol{A}}$*.* We update the mixing matrix $\hat{\widetilde{\boldsymbol{A}}}^{(k)}$ based on the estimates of $\{\hat{\boldsymbol{X}}_\ell^{(k)}, \hat{\boldsymbol{D}}_\ell^{(k)}\}$, following the same procedure as the node-rotation algorithm.

We compare the results of the node-rotation algorithm and the alternative estimation algorithm using 50 datasets obtained by bootstrapping the PNC data, with each dataset containing 514 subjects. Figure 7.9(A) presents the correlations of the connectivity traits and their temporal loadings obtained using the two algorithms across the 50 datasets. The results produced by the alternative algorithm are consistent with those from the node-rotation algorithm. Figure 7.9(B) shows the computation time across the 50 datasets. On average, the alternative algorithm reduces the computational time by 19.2%.

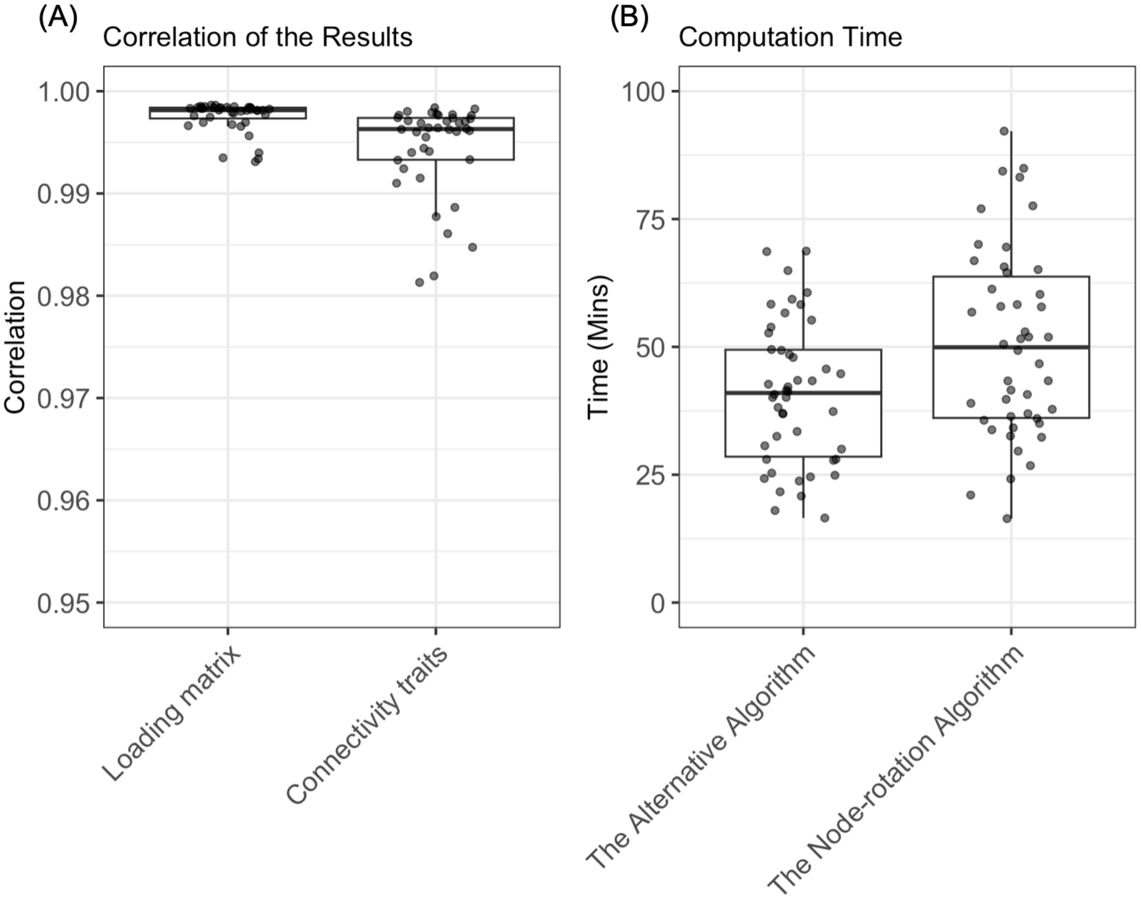


Figure 7.9: Comparison of the results and computation time between the node-rotation algorithm and the alternative algorithm based on 50 bootstrap datasets from the PNC study.

7.2 Supplementary Material of Chapter 3

7.2.1 The derivation of the final optimization function

The original objective function for the Longitudinal LOCUS model with the sparsity regularization is,

$$\min \sum_{g=1}^{G} \sum_{t=1}^{T} ||\widetilde{\boldsymbol{Y}}_{gt} - \widetilde{\boldsymbol{A}}_{gt}\boldsymbol{S}_{gt}||_F^2 + \phi \sum_{g=1}^{G} \sum_{t=1}^{T} |\boldsymbol{S}_{gt}| \tag{7.3}$$

where $\boldsymbol{S}_{gt} = \begin{bmatrix} \boldsymbol{s}'_{gt1} \\ \dots \\ \boldsymbol{s}'_{gtq} \end{bmatrix} \in \mathcal{R}_{q\times p}$ and $\boldsymbol{s}_{gt\ell} = \mathcal{L}(\boldsymbol{X}_\ell \boldsymbol{D}_{gt\ell} \boldsymbol{X}'_\ell)$.

Given the orthogonality on $\widetilde{\boldsymbol{A}}_{gt}$, for group g visit t, we have

$$\begin{aligned} ||\widetilde{\boldsymbol{Y}}_{gt} - \widetilde{\boldsymbol{A}}_{gt}\boldsymbol{S}_{gt}||_F^2 &= \left\| \widetilde{\boldsymbol{A}}_{gt} \left(\widetilde{\boldsymbol{A}}'_{gt} \widetilde{\boldsymbol{Y}}_{gt} - \boldsymbol{S}_{gt} \right) \right\|_F^2 \\ &= \left\| \left[\widetilde{\boldsymbol{Y}}'_{gt} \widetilde{\boldsymbol{a}}_{gt1}, \dots, \widetilde{\boldsymbol{Y}}'_{gt} \widetilde{\boldsymbol{a}}_{gtq} \right]' - [\boldsymbol{s}_{gt1}, \dots, \boldsymbol{s}_{gtq}]' \right\|_F^2 = \sum_{\ell=1}^{q} \left\| \widetilde{\boldsymbol{Y}}'_{gt} \widetilde{\boldsymbol{a}}_{gt\ell} - \boldsymbol{s}_{gt\ell} \right\|_2^2 \\ &= \sum_{\ell=1}^{q} ||\widetilde{\boldsymbol{Y}}'_{gt} \widetilde{\boldsymbol{a}}_{gt\ell} - \mathcal{L}\left(\boldsymbol{X}_\ell \boldsymbol{D}_{gt\ell} \boldsymbol{X}'_\ell \right)||_2^2 \end{aligned}$$

Therefore, the final objective function is

$$\min_{\widetilde{\boldsymbol{A}}_{gt}, \{\boldsymbol{X}_\ell, \boldsymbol{D}_{gt\ell}\}} \sum_{g=1}^{G} \sum_{t=1}^{T} \sum_{l=1}^{q} ||\widetilde{\boldsymbol{Y}}'_{gt} \widetilde{\boldsymbol{a}}_{gt\ell} - \mathcal{L}(\boldsymbol{X}_\ell \boldsymbol{D}_{gt\ell} \boldsymbol{X}'_\ell)||_2^2 + \phi \sum_{g=1}^{G} \sum_{t=1}^{T} \sum_{\ell=1}^{q} \sum_{u<v} |\boldsymbol{x}_\ell(u)' \boldsymbol{D}_{gt\ell} \boldsymbol{x}_\ell(v)| \tag{7.4}$$

where $\widetilde{\boldsymbol{a}}_{gt\ell}$ is the ℓth column of $\widetilde{\boldsymbol{A}}_{gt}$.

7.3 Supplementary Material of Chapter 4

7.3.1 Identifying motion-controlled association of ASD and brain connectivity

Theorem 5. *Under (A1)-(A3), the counterfactual* $\theta_{C,a}$ *is identified by* θ_a*, where* $\theta_a = \iiint \mu_{Y|A,Z,M,X}(a,z,m,x) p_{Z|A,X}(z \mid a,x) p_{M|\Delta=1,A,X}(m \mid 0,x) p_X(x) dz dm dx.$

Proof Let $\theta_{\mathrm{C},a} = E\{E_{\mathrm{C}}[Y(M_0) \mid A = a, X]\}$, where $M_0 \sim P_{M|\Delta=1,A,X}(m \mid 0,x)$. We have:

$$
\begin{aligned}
&\theta_{\mathrm{C},a} = E\{E_{\mathrm{C}}[Y(M_0) \mid A = a, X]\}\\
&= \int E_{\mathrm{C}}(Y(M_0) \mid A = a, X = x) p_X(x) dx\\
&\overset{\text{tower rule}}{=} \iint E_{\mathrm{C}}(Y(m) \mid M_0 = m, A = a, X = x) p_{M|\Delta=1,A,X}(m \mid 0,x) p_X(x) dm dx\\
&\overset{\text{defn}}{=} \iint E_{\mathrm{C}}(Y(m) \mid A = a, X = x) p_{M|\Delta=1,A,X}(m \mid 0,x) p_X(x) dm dx\\
&\overset{\text{tower rule}}{=} \iiint E_{\mathrm{C}}(Y(m) \mid A = a, X = x, Z = z) p_{Z|A,X}(z \mid a,x) p_{M|\Delta=1,A,X}(m \mid 0,x) p_X(x) dz dm dx\\
&\overset{\text{assumption (A2)}}{=} \iiint E_{\mathrm{C}}(Y(m) \mid M = m, A = a, X = x, Z = z) p_{Z|A,X}(z \mid a,x) p_{M|\Delta=1,A,X}(m \mid 0,x) p_X(x) dz dm dx\\
&\overset{\text{assumption (A3)}}{=} \iiint E(Y \mid M = m, A = a, X = x, Z = z) p_{Z|A,X}(z \mid a,x) p_{M|\Delta=1,A,X}(m \mid 0,x) p_X(x) dz dm dx\ .
\end{aligned}
$$

The fourth equality results from the fact that by the construction of M_0, we have that $Y(m) M_0 \mid A, X$ for all m. The sixth equality results from the assumption that $Y(m) M \mid A, X, Z$.

7.3.2 Efficient influence function Theorem 3.1

Overview of efficiency theory

The efficient influence function can be used to characterize the nonparametric efficiency bound, i.e., the smallest asymptotic variance of any regular, *asymptotically linear* estimator of θ_a (Bickel et al., 1993). An estimator $\theta_{n,a}$ of θ_a based on $O_1, \ldots, O_n \overset{\text{i.i.d}}{\sim} P$

is said to be asymptotically linear if there exists a function $o \mapsto \tilde{D}_{P,a}(o)$ such that $E[\tilde{D}_{P,a}(O)] = 0$, $E[\tilde{D}^2_{P,a}(O)] < \infty$, and $\theta_{n,a} = \theta_a + n^{-1}\sum_{i=1}^n \tilde{D}_{P,a}(O_i) + o_P(n^{-1/2})$. We refer to $\tilde{D}_{P,a}$ as the *influence function* of $\theta_{n,a}$. Under this representation, the asymptotic study of $\theta_{n,a}$ reduces to the study of the sample mean $n^{-1}\sum_{i=1}^n \tilde{D}_{P,a}(O_i)$ whose large sample behavior can be described using standard results such as the weak law of large numbers and the central limit theorem. The latter implies $n^{1/2}(\theta_{n,a} - \theta_a)$ converges to a mean-zero normal random variable with variance equal to $E[\tilde{D}^2_{P,a}(O)]$. Due to the fact that the asymptotic variance of an asymptotically linear estimator is characterized by the variance of the influence function, the influence function that has the smallest variance amongst all influence functions of regular estimators is called the *efficient influence function.* An estimator with an influence function equal to the efficient influence function is, by definition, asymptotically efficient.

Proof of Theorem 3.1

Proof Let $\mathcal{P}$ be the model for the true distribution. Let $o = (a, m, \delta, x, z, y)$ denote values of the observed vector of variable $O = (A, M, \Delta, X, Z, Y)$, and $O \sim P \in \mathcal{P}$. We use $\Psi_a : \mathcal{P} \to \mathbb{R}$ to denote a parameter as a functional that maps the distribution P in the model $\mathcal{P}$ to the real number θ_a.

$$\begin{aligned}\Psi_a(P) &= \iiint \mu_{Y|A,M,X,Z}(a, m, x, z) p_{Z|A,X}(z \mid a, x) p_{M|\Delta=1,A,X}(m \mid 0, x) p_X(x) dz dm dx \\ &= \int y dP_{Y|A,M,X,Z}(y \mid a, m, x, z) dP_{Z|A,X}(z \mid a, x) dP_{M|\Delta=1,A,X}(m \mid 0, x) dP_X(x) \ .\end{aligned}$$

We consider a collection of submodels through P at $\epsilon = 0$ in the direction S, $\{P_\epsilon \in \mathcal{P},\ dP_\epsilon = (1 + \epsilon S)dP\}$ where S is an element of the Hilbert space $L^2_0(P)$, the space of all functions of O such that $\int S(o)dP(o) = 0$, $\int S(o)^2 dP(o) < \infty$ equipped with inner product $< f, g >= \int f(o)g(o)dP(o)$. We consider the derivative of the parameter mapping along the path P_ϵ. We can view this derivative as a bounded functional on $L_0^2(P)$, which, by the Reisz representation theorem, will have an inner-product form $< s, D_P >$ for a unique element $D_P \in L_0^2(P)$. D_P is referred to as the *canonical gradient* of Ψ_a. This gradient will also be the efficient influence function of regular asymptotically linear estimators of Ψ_a. Thus, to derive the efficient influence function, we may study the following derivative of Ψ_a and write that derivative in an inner product form (Levy, 2019). Below, we use the notation $\int f(o)dP_0(o)$ interchangably with $\int f(o)P_0(do)$ to denote the Lebesgue integral of a P_0-measurable function f with respect to probability measure P_0.

$$\frac{\partial}{\partial \epsilon}\Psi_a\left(P_\epsilon\right)\bigg|_{\epsilon=0} = \frac{\partial}{\partial \epsilon}\int y dP_{\epsilon,Y|A,M,X,Z}(y \mid a,m,x,z) dP_{\epsilon,Z|A,X}(z \mid a,x) dP_{\epsilon,M|\Delta=1,A,X}(m \mid 0,x) dP_{\epsilon,X}(x)\bigg|_{\epsilon=0}$$

$$= \int y S_{Y|A,M,X,Z}(y \mid a,m,x,z) dP_{Y|A,M,X,Z}(y \mid a,m,x,z) dP_{Z|A,X}(z \mid a,x) dP_{M|\Delta=1,A,X}(m \mid 0,x) dP_X \tag{2.1}$$

$$+ \int y S_{Z|A,X}(z \mid a,x) dP_{Y|A,M,X,Z}(y \mid a,m,x,z) dP_{Z|A,X}(z \mid a,x) dP_{M|\Delta=1,A,X}(m \mid 0,x) dP_X(x) \tag{2.2}$$

$$+ \int y S_{M|\Delta=1,A,X}(m \mid 0,x) dP_{Y|A,M,X,Z}(y \mid a,m,x,z) dP_{Z|A,X}(z \mid a,x) dP_{M|\Delta=1,A,X}(m \mid 0,x) dP_X(x) \tag{2.3}$$

$$+ \int y S_X(x) dP_{Y|A,M,X,Z}(y \mid a,m,x,z) dP_{Z|A,X}(z \mid a,x) dP_{M|\Delta=1,A,X}(m \mid 0,x) dP_X(x) \tag{2.4}$$

where

$$dP_{\epsilon,Y|A,M,X,Z}(y \mid a,m,x,z) = \frac{\int_\delta (1+\epsilon S(o))P(a,m,d\delta,x,z,y)}{\int_{\delta,y}(1+\epsilon S(o))P(a,m,d\delta,x,z,dy)},$$

$$S_{Y|A,M,X,Z}(y \mid a,m,x,z) = \frac{\partial \log dP_{\epsilon,Y|A,M,X,Z}(y \mid a,m,x,z)}{\partial \epsilon}\bigg|_{\epsilon=0} = E(S(O) \mid y,a,m,x,z) - E(S(O) \mid a,m,x,z),$$

$$dP_{\epsilon,Z|A,X}(y \mid a,x) = \frac{\int_{\delta,m,y}(1+\epsilon S(o))P(a,dm,d\delta,x,z,dy)}{\int_{\delta,m,y,z}(1+\epsilon S(o))P(a,dm,d\delta,x,dz,dy)},$$

$$S_{Z|A,X}(z \mid a,x) = \frac{\partial \log dP_{\epsilon,Z|A,X}(y \mid a,x)}{\partial \epsilon}\bigg|_{\epsilon=0} = E(S(O) \mid z,a,x) - E(S(O) \mid a,x),$$

$$dP_{\epsilon,M|\Delta=1,A,X}(m \mid 0,x) = \frac{\int_{z,y}\, _{0,1}(a,\delta)(1+\epsilon S(o))P(a,m,\delta,x,dz,dy)}{\int_{m,z,y}\, _{0,1}(a,\delta)(1+\epsilon S(o))P(a,dm,\delta,x,dz,dy)},$$

$$S_{M|\Delta=1,A,X}(m \mid 0,x) = \frac{\partial \log P_{\epsilon,M|\Delta=1,A,X}(m \mid 0,x)}{\partial \epsilon}\bigg|_{\epsilon=0} = E(S(O)_{0,1}(A,\Delta) \mid m,a,\delta,x) - E(S(O)_{0,1}(A,\Delta) \mid a,\delta,x),$$

$$dP_{\epsilon,X}(x) = \int_{\delta,a,m,y,z}(1+\epsilon S(o))P(da,dm,d\delta,x,dz,dy),$$

$$S_X(x) = E(S(O) \mid x)\ .$$

Evaluating the derivative for the term (2.1), we have:

$$
\begin{aligned}
&(2.1) \int yS_{Y|A,M,X,Z}(y \mid a,m,x,z)P_{Y|A,M,X,Z}(y \mid a,m,x,z)dP_{Z|A,X}(z \mid a,x)dP_{M|\Delta=1,A,X}(m \mid 0,x)dP_X(x) \\
&= \int \frac{{}_{a}(a')}{\pi_{a'}(x)} yS_{Y|A,M,X,Z}\left(y \mid a',m,x,z\right) dP_{Y|A,M,X,Z}\left(y \mid a',m,x,z\right) dP_{Z|A,X}\left(z \mid a',x\right) dP_{M|\Delta=1,A,X}(m \mid 0,x) \\
&\quad dP_{A,X}\left(a',x\right) \\
&= \int \frac{{}_{a}(a')}{\pi_{a'}(x)} \frac{p_{M|\Delta=1,A,X}(m \mid 0,x)}{p_{M|A,X,Z}\left(m \mid a',x,z\right)} yS_{Y|A,M,X,Z}\left(y \mid a',m,x,z\right) dP_{Y|A,M,X,Z}\left(y \mid a',m,x,z\right) \\
&\quad dP_{M|A,X,Z}\left(m \mid a',x,z\right) dP_{Z|A,X}\left(z \mid a',x\right) dP_{A,X}\left(a',x\right) \\
&= \int \frac{{}_{a}(a')}{\pi_{a'}(x)} \frac{p_{M|\Delta=1,A,X}(m \mid 0,x)}{p_{M|A,X,Z}\left(m \mid a',x,z\right)} yS_{Y|A,M,X,Z}\left(y \mid a',m,x,z\right) dP(o) \\
&\overset{*}{=} \int \frac{{}_{a}(a')}{\pi_{a'}(x)} \frac{p_{M|\Delta=1,A,X}(m \mid 0,x)}{p_{M|A,X,Z}\left(m \mid a',x,z\right)} \left(y - \mu_{Y|A,M,X,Z}(a,m,x,z)\right) S_{Y|A,M,X,Z}\left(y \mid a',m,x,z\right) dP(o) \\
&\overset{**}{=} \int \frac{{}_{a}(a')}{\pi_{a}(x)} \frac{p_{M|\Delta=1,A,X}(m \mid 0,x)}{p_{M|A,X,Z}\left(m \mid a,x,z\right)} \left(y - \mu_{Y|A,M,X,Z}(a,m,x,z)\right) S(o)dP(o) \ .
\end{aligned}
$$

The reason for (*) is:

$$
\int \frac{{}_{a}(a')}{\pi_{a'}(x)} \frac{p_{M|\Delta=1,A,X}(m \mid 0,x)}{p_{M|A,X,Z}\left(m \mid a',x,z\right)} \mu_{Y|A,M,X,Z}(a,m,x,z) S_{Y|A,M,X,Z}\left(y \mid a',m,x,z\right) dP(o) = 0 \ .
$$

The reason for (**) is:

$$
\begin{aligned}
&S_{Y,A,M,X,Z}\left(y,a',m,x,z\right) = S_{Y|A,M,X,Z}\left(y \mid a',m,x,z\right) + S_{A,M,X,Z}\left(a',m,x,z\right) \ , \\
&\int \frac{{}_{a}(a')}{\pi_{a}(x)} \frac{p_{M|\Delta=1,A,X}(m \mid 0,x)}{p_{M|A,X,Z}\left(m \mid a,x,z\right)} \left(y - \mu_{Y|A,M,X,Z}(a,m,x,z)\right) S_{A,M,X,Z}\left(a',m,x,z\right) dP(o) = 0 \ , \\
&S(o) = S_{Y,A,M,X,Z}\left(y,a',m,x,z\right) + S_{\Delta|Y,A,M,X,Z}\left(\delta \mid y,a',m,x,z\right) \ .
\end{aligned}
$$

The same logic can be applied to evaluate term (2.2)-(2.4). We have,

$$
\begin{aligned}
&(2.2) \int y S_{Z|A,X}(z \mid a,x) dP_{Y|A,M,X,Z}(y \mid a,m,x,z) dP_{Z|A,X}(z \mid a,x) dP_{M|\Delta=1,A,X}(m \mid 0,x) dP_X(x) \\
&= \int \eta_{\mu|A,X,Z}(a,x,z) S_{Z|A,X}(z \mid a,x) dP_{Z|A,X}(z \mid a,x) dP_X(x) \\
&= \int \frac{{}_{a}(a')}{\pi_{a'}(x)} \eta_{\mu|A,X,Z}\left(a',x,z\right) S_{Z|A,X}\left(z \mid a',x\right) dP_{Z|A,X}\left(z \mid a',x\right) dP\left(a',x\right) \\
&= \int \frac{{}_{a}(a')}{\pi_{a'}(x)} \left(\eta_{\mu|A,X,Z}\left(a',x,z\right) - \xi_{a,\eta|X}(x)\right) S_{Z|A,X}\left(z \mid a',x\right) dP(o) \\
&= \int \frac{{}_{a}(a')}{\pi_{a}(x)} \left(\eta_{\mu|A,X,Z}(a,x,z) - \xi_{a,\eta|X}(x)\right) S(o) dP(o) \ .
\end{aligned}
$$

$$
\begin{aligned}
&(2.3) \int y S_{M|\Delta=1,A,X}(m \mid 0,x) dP_{Y|A,M,X,Z}(y \mid a,m,x,z) dP_{Z|A,X}(z \mid a,x) dP_{M|\Delta=1,A,X}(m \mid 0,x) dP_X(x) \\
&= \int \eta_{\mu|A,M,X}(a,m,x) S_{M|\Delta=1,A,X}(m \mid 0,x) dP_{M|\Delta=1,A,X}(m \mid 0,x) dP_X(x) \\
&= \int \frac{{}_{0,1}(a',\Delta)}{\bar{\pi}_0(x)} \eta_{\mu|A,M,X}(a,m,x) S_{M|\Delta=1,A,X}(m \mid 0,x) dP_{M,\Delta=1,A,X}\left(m, \delta=1, a'=0, x\right) \\
&= \int \frac{{}_{0,1}(a',\Delta)}{\bar{\pi}_0(x)} \left(\eta_{\mu|A,M,X}(a,m,x) - \xi_{a,\eta|X}(x)\right) S_{M|\Delta=1,A,X}(m \mid 0,x) dP(o) \\
&= \int \frac{{}_{0,1}(a',\Delta)}{\bar{\pi}_0(x)} \left(\eta_{\mu|A,M,X}(a,m,x) - \xi_{a,\eta|X}(x)\right) S(o) dP(o) \ .
\end{aligned}
$$

$$
\begin{aligned}
&(2.4) \int y S_X(x) dP_{Y|A,M,X,Z}(y \mid a,m,x,z) dP_{Z|A,X}(z \mid a,x) dP_{M|\Delta=1,A,X}(m \mid 0,x) dP_X(x) \\
&= \int \xi_{a,\eta|X}(x) S_X(x) dP_X(x) \\
&= \int \left(\xi_{a,\eta|X}(x) - \theta_a\right) S(o) dP(o) \ .
\end{aligned}
$$

Putting the results together, we have:

$$
\begin{aligned}
\frac{\partial}{\partial \epsilon} \Psi_a\left(P_\epsilon\right)\bigg|_{\epsilon=0} = &\int \frac{{}_{a}(a')}{\pi_a(x)} \frac{p_{M|\Delta=1,A,X}(m \mid 0,x)}{p_{M|A,X,Z}\left(m \mid a,x,z\right)} \left(y - \mu_{Y|A,M,X,Z}(a,m,x,z)\right) S(o) dP(o) \\
&+ \int \frac{{}_{a}(a')}{\pi_a(x)} \left(\eta_{\mu|A,X,Z}(a,x,z) - \xi_{a,\eta|X}(x)\right) S(o) dP(o) \\
&+ \int \frac{{}_{0,1}(a',\Delta)}{\bar{\pi}_0(x)} \left(\eta_{\mu|A,M,X}(a,m,x) - \xi_{a,\eta|X}(x)\right) S(o) dP(o) \\
&+ \int \left(\xi_{a,\eta|X}(x) - \theta_a\right) S(o) dP(o) \ .
\end{aligned}
$$

Thus, we have expressed the derivative of Ψ_a along a path P_ϵ as an inner product between S and the gradient:

$$\begin{aligned}
D_{P,a}(O_i) = {} & \frac{{}_{a}(A_i)}{\pi_a(X_i)} \frac{p_{M|\Delta=1,A,X}(M_i \mid 0, X_i)}{p_{M|A,X,Z}(M_i \mid A_i, X_i, Z_i)} \left\{Y_i - \mu_{Y|A,M,X,Z}(a, M_i, X_i, Z_i)\right\} \\
& + \frac{{}_{a}(A_i)}{\pi_a(X_i)} \left\{\eta_{\mu|A,Z,X}(a, X_i, Z_i) - \xi_{a,\eta|X}(X_i)\right\} \\
& \quad + \frac{{}_{0,1}(A_i, \Delta_i)}{\bar{\pi}_0(X_i)} \left\{\eta_{\mu|A,M,X}(a, M_i, X_i) - \xi_{a,\eta|X}(X_i)\right\} \\
& \quad\quad + \xi_{a,\eta|X}(X_i) - \theta_a \ .
\end{aligned}$$

As the tangent space of our model is $L_0^2(P_0)$, there is only a single gradient for Ψ_a. Thus, this gradient is by definition the efficient gradient and the efficient influence function for regular asymptotically Normal estimators of $\Psi_a(P_0)$.

7.3.3 Consistency of the one-step estimator

Proof of Theorem 3.2

Part I For fixed a, we define $D^*_{P,a} = D_{P,a} + \theta_a$. $D^*_{P,a}$ corresponds to the first three lines and the first term of the fourth line in equation (3.5) of the main manuscript. We rewrite $D^*_{P,a}$ as

$$\begin{aligned}
D^*_{P,a}(O) = {} & \frac{{}_{a}(A)}{\pi_a(X)} \frac{p_{M|\Delta=1,A,X}(M \mid 0, X)}{p_{M|A,X,Z}(M \mid A, X, Z)} \left\{Y - \mu_{Y|A,M,X,Z}(a, M, X, Z)\right\} \\
& + \frac{{}_{a}(A)}{\pi_a(X)} \eta_{\mu|A,Z,X}(a, X, Z) \\
& \quad + \frac{{}_{0,1}(A, \Delta)}{\bar{\pi}_0(X)} \left\{\eta_{\mu|A,M,X}(a, M, X) - \xi_{a,\eta|X}(X)\right\} \\
& \quad\quad + \xi_{a,\eta|X}(X) - \frac{{}_{a}(A)}{\pi_a(X)} \xi_{a,\eta|X}(X) \ .
\end{aligned}$$

Consider a probability distribution $P' \in \mathcal{P}$. Next we show that, if any one of the conditions in assumption (ii) holds, then $E[D^*_{P',a}(O)] - \theta_a = 0$.

$$\begin{aligned}
E\left[D^*_{P',a}(O)\right] - \theta_a = {} & E\left[\frac{{}_a(A)}{\pi'_a(X)}\frac{p'_{M|\Delta=1,A,X}(M \mid 0,X)}{p'_{M|A,X,Z}(M \mid A,X,Z)}\left\{Y - \mu'_{Y|A,M,X,Z}(a,M,X,Z)\right\}\right] \\
& + E\left[\frac{{}_a(A)}{\pi'_a(X)}\eta'_{\mu|A,Z,X}(a,X,Z)\right] \\
& + E\left[\frac{{}_{0,1}(A,\Delta)}{\bar{\pi}'_0(X)}\left\{\eta'_{\mu|A,M,X}(a,M,X) - \xi'_{a,\eta|X}(X)\right\}\right] \\
& + E\left[\xi'_{a,\eta|X}(X) - \frac{{}_a(A)}{\pi'_a(X)}\xi'_{a,\eta|X}(X)\right] \\
& - \theta_a \ .
\end{aligned}$$

We derive the precise expression for each term:

$$\begin{aligned}
& E\left[\frac{{}_a(A)}{\pi'_a(X)}\frac{p'_{M|\Delta=1,A,X}(M \mid 0,X)}{p'_{M|A,X,Z}(M \mid A,X,Z)}\left\{Y - \mu'_{Y|A,M,X,Z}(a,M,X,Z)\right\}\right] \\
& = \int \frac{{}_a(a^*)}{\pi'_a(X)}\frac{p'_{M|\Delta=1,A,X}(m \mid 0,x)}{p'_{M|A,X,Z}(m \mid a^*,x,z)}\left\{y - \mu'_{Y|A,M,X,Z}(a,m,x,z)\right\} p_{A,M,X,Z,Y}(a^*,m,x,z,y)da^*dmdzdxdy \\
& = \int \frac{{}_a(a^*)}{\pi'_a(X)}\frac{p'_{M|\Delta=1,A,X}(m \mid 0,x)}{p'_{M|A,X,Z}(m \mid a,x,z)}\left\{\int y p_{Y|A,M,X,Z}(y \mid a,m,x,z)dy - \mu'_{Y|A,M,X,Z}(a,m,x,z)\right\} \\
& \quad p_{A,M,X,Z}(a,m,x,z)da^*dmdzdx \\
& = \int \frac{\int {}_a(a^*)\pi_a(x)da^*}{\pi'_a(X)}\frac{p'_{M|\Delta=1,A,X}(m \mid 0,x)}{p'_{M|A,X,Z}(m \mid a,x,z)}\left\{\mu_{Y|A,M,X,Z}(a,m,x,z) - \mu'_{Y|A,M,X,Z}(a,m,x,z)\right\} \\
& \quad p_{M|A,X,Z}(m \mid a,x,z)p_{Z|A,X}(z \mid a,x)p_X(x)dmdzdx \\
& = \int \frac{\pi_a(x)}{\pi'_a(x)}\frac{p'_{M|\Delta=1,A,X}(m \mid 0,x)}{p'_{M|A,X,Z}(m \mid a,x,z)}\left\{\mu_{Y|A,M,X,Z}(a,m,x,z) - \mu'_{Y|A,M,X,Z}(a,m,x,z)\right\} \\
& \quad p_{M|A,X,Z}(m \mid a,x,z)p_{Z|A,X}(z \mid a,x)p_X(x)dmdzdx \ . \qquad (3.1)
\end{aligned}$$

$$\begin{aligned}
& E\left[\frac{{}_a(A)}{\pi'_a(X)}\eta'_{\mu|A,Z,X}(a,X,Z)\right] \\
& = \int \frac{{}_a(a^*)}{\pi'_a(X)}\eta'_{\mu|A,Z,X}(a,x,z)p_{A,X,Z}(a^*,x,z)da^*dzdx \\
& = \int \frac{{}_a(a^*)}{\pi'_a(X)}\eta'_{\mu|A,Z,X}(a,x,z)p_{Z|A,X}(a,x)\pi_a(x)p_X(x)da^*dzdx \\
& = \int \frac{\int {}_a(a^*)\pi_a(x)da^*}{\pi'_a(X)}\eta'_{\mu|A,Z,X}(a,x,z)p_{Z|A,X}(a,x)p_X(x)dzdx \\
& = \int \frac{\pi_a(x)}{\pi'_a(x)}\eta'_{\mu|A,Z,X}(a,x,z)p_{Z|A,X}(a,x)p_X(x)dzdx \ . \qquad (3.2)
\end{aligned}$$

$$\begin{aligned}
&E\left[\frac{{}_{0,1}(A,\Delta)}{\bar{\pi}'_0(X)}\left\{\eta'_{\mu|A,M,X}(a,M,X)-\xi'_{a,\eta|X}(X)\right\}\right]\\
&=\int\frac{{}_{0,1}(a^*,\Delta)}{\bar{\pi}'_0(X)}\eta'_{\mu|A,M,X}(a,m,x)p_{A,\Delta,M,X}(a^*,\Delta,m,x)da^*d\Delta dmdx\\
&\qquad-\int\frac{{}_{0,1}(a^*,\Delta)}{\bar{\pi}'_0(X)}\xi'_{a,\eta|X}(x)p_{A,\Delta,X}(a^*,\Delta,x)da^*d\Delta dx\\
&=\int\frac{{}_{0,1}(a^*,\Delta)}{\bar{\pi}'_0(X)}\eta'_{\mu|A,M,X}(a,m,x)p_{M|\Delta=1,A,X}(m\mid 0,x)\bar{\pi}_0(x)p_X(x)da^*d\Delta dmdx\\
&\qquad-\int\frac{{}_{0,1}(a^*,\Delta)}{\bar{\pi}'_0(X)}\xi'_{a,\eta|X}(x)\bar{\pi}_0(x)p_X(x)da^*d\Delta dx\\
&=\int\frac{\int_{0,1}(a^*,\Delta)\bar{\pi}_0(x)da^*d\Delta}{\bar{\pi}'_0(X)}\eta'_{\mu|A,M,X}(a,m,x)p_{M|\Delta=1,A,X}(m\mid 0,x)p_X(x)dmdx\\
&\qquad-\int\frac{\int_{0,1}(a^*,\Delta)\bar{\pi}_0(x)da^*d\Delta}{\bar{\pi}'_0(X)}\xi'_{a,\eta|X}(x)p_X(x)dx\\
&=\int\frac{\bar{\pi}_0(x)}{\bar{\pi}'_0(x)}\eta'_{\mu|A,M,X}(a,m,x)p_{M|\Delta=1,A,X}(m\mid 0,x)p_X(x)dmdx\\
&\qquad-\int\frac{\bar{\pi}_0(x)}{\bar{\pi}'_0(x)}\eta'_{\mu|A,M,X}(a,m,x)p'_{M|\Delta=1,A,X}(m\mid 0,x)p_X(x)dmdx\\
&=\int\frac{\bar{\pi}_0(x)}{\bar{\pi}'_0(x)}\eta'_{\mu|A,M,X}(a,m,x)\left\{p_{M|\Delta=1,A,X}(m\mid 0,x)-p'_{M|\Delta=1,A,X}(m\mid 0,x)\right\}p_X(x)dmdx\;. \qquad (3.3)
\end{aligned}$$

$$\begin{aligned}
&E\left[\xi'_{a,\eta|X}(X)-\frac{{}_{a}(A)}{\pi'_a(X)}\xi'_{a,\eta|X}(X)\right]\\
&=\int\xi'_{a,\eta|X}(x)p_X(x)dx-\int\frac{{}_{a}(a^*)}{\pi'_a(X)}\xi'_{a,\eta|X}(x)p_{A,X}(a^*,x)da^*dx\\
&=\int\xi'_{a,\eta|X}(x)p_X(x)dx-\int\frac{\pi_a(x)}{\pi'_a(x)}\xi'_{a,\eta|X}(x)p_X(x)dx\\
&=\int\xi'_{a,\eta|X}(x)\left\{1-\frac{\pi_a(x)}{\pi'_a(x)}\right\}p_X(x)dx\;. \qquad (3.4)
\end{aligned}$$

$$-\theta_a=-\int\xi_{a,\eta|X}(x)p_X(x)dx\;. \qquad (3.5)$$

Therefore,

$$E[D^*_{P',a}(O)] - \theta_a = (3.1) + (3.2) + (3.3) + (3.4) + (3.5)\ .$$

First, note that

$$\begin{aligned}(3.4) + (3.5) &= \int \xi'_{a,\eta|X}(x)\left\{1 - \frac{\pi_a(x)}{\pi'_a(x)}\right\} p_X(x)dx - \int \xi_{a,\eta|X}(x)p_X(x)dx \\ &= \int \left\{\xi'_{a,\eta|X}(x) - \xi_{a,\eta|X}(x)\right\}\left\{1 - \frac{\pi_a(x)}{\pi'_a(x)}\right\} p_X(x)dx \\ &\quad - \int \xi_{a,\eta|X}(x)\frac{\pi_a(x)}{\pi'_a(x)} p_X(x)dx\ .\end{aligned} \tag{3.6}$$

Using the definition of $\xi_{a,\eta|X}(x)$ in the main manuscript, we obtain

$$\begin{aligned}&\int \xi_{a,\eta|X}(x)\frac{\pi_a(x)}{\pi'_a(x)} p_X(x)dx = \\ &\quad\int \frac{\pi_a(x)}{\pi'_a(x)}\frac{p_{M|\Delta=1,A,X}(m \mid 0,x)}{p_{M|A,X,Z}(m \mid a,x,z)}\left\{\mu_{Y|A,M,X,Z}(a,m,x,z) - \mu'_{Y|A,M,X,Z}(a,m,x,z)\right\} p_{M|A,X,Z}(m \mid a,x,z) \\ &\qquad p_{Z|A,X}(z \mid a,x)p_X(x)dmdzdx \\ &\quad+ \int \frac{\pi_a(x)}{\pi'_a(x)} p_{M|\Delta=1,A,X}(m \mid 0,x)\mu'_{Y|A,M,X,Z}(a,m,x,z)p_{Z|A,X}(z \mid a,x)p_X(x)dmdzdx\ .\end{aligned} \tag{3.7}$$

Thus

$$\begin{aligned}(3.1) + (3.4) + (3.5) =&(3.6) - (3.7) \\ &+ \int \frac{\pi_a(x)}{\pi'_a(x)}\left\{\frac{p'_{M|\Delta=1,A,X}(m \mid 0,x)}{p'_{M|A,X,Z}(m \mid a,x,z)} - \frac{p_{M|\Delta=1,A,X}(m \mid 0,x)}{p_{M|A,X,Z}(m \mid a,x,z)}\right\} \times \\ &\left\{\mu_{Y|A,M,X,Z}(a,m,x,z) - \mu'_{Y|A,M,X,Z}(a,m,x,z)\right\} \\ &\quad p_{M|A,X,Z}(m \mid a,x,z)p_{Z|A,X}(z \mid a,x)p_X(x)dmdzdx\ .\end{aligned} \tag{3.8}$$

We also have

$$
\begin{aligned}
(3.2)-(3.7) = \int \frac{\pi_a(x)}{\pi'_a(x)} &\left\{p'_{M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x)\right\} \mu'_{Y|A,M,X,Z}(a,m,x,z) \\
&p_{Z|A,X}(z \mid a,x) p_X(x) dmdzdx \ ,
\end{aligned}
$$

which yields

$$
\begin{aligned}
&(3.2)+(3.3)-(3.7) \\
&= \int \frac{\bar{\pi}_0(x)}{\bar{\pi}'_0(x)} \eta'_{\mu|A,M,X}(a,m,x) \Big(p_{M|\Delta=1,A,X}(m \mid 0,x) - p'_{M|\Delta=1,A,X}(m \mid 0,x)\Big) p_X(x) dmdx \\
&\quad + \int \frac{\pi_a(x)}{\pi'_a(x)} \left\{p'_{M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x)\right\} \mu_{Y|A,M,X,Z}(a,m,x,z) p_{Z|A,X}(z \mid a,x) p_X(x) dmdzdx \\
&\quad + \int \frac{\pi_a(x)}{\pi'_a(x)} \left\{p'_{M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x)\right\} \left\{\mu'_{Y|A,M,X,Z}(a,m,x,z) - \mu_{Y|A,M,X,Z}(a,m,x,z)\right\} \\
&\quad p_{Z|A,X}(z \mid a,x) p_X(x) dmdzdx \\
&= - \int \frac{\bar{\pi}_0(x)}{\bar{\pi}'_0(x)} \eta'_{\mu|A,M,X}(a,m,x) \left\{p'_{M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x)\right\} p_X(x) dmdx \\
&\quad + \int \frac{\pi_a(x)}{\pi'_a(x)} \eta_{\mu|A,M,X}(a,m,x) \left\{p'_{M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x)\right\} p_X(x) dmdx \\
&\quad + \int \frac{\pi_a(x)}{\pi'_a(x)} \left\{p'_{M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x)\right\} \{\mu'_{Y|A,M,X,Z}(a,m,x,z) - \mu_{Y|A,M,X,Z}(a,m,x,z)\} \\
&\quad p_{Z|A,X}(z \mid a,x) p_X(x) dmdzdx \\
&= \int \left\{\frac{\pi_a(x)}{\pi'_a(x)} \eta_{\mu|A,M,X}(a,m,x) - \frac{\bar{\pi}_0(x)}{\bar{\pi}'_0(x)} \eta'_{\mu|A,M,X}(a,m,x)\right\} \\
&\qquad \left\{p'_{M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x)\right\} p_X(x) dmdx \qquad (3.9) \\
&\quad + \int \frac{\pi_a(x)}{\pi'_a(x)} \left\{p'_{M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x)\right\} \left\{\mu'_{Y|A,M,X,Z}(a,m,x,z) - \mu_{Y|A,M,X,Z}(a,m,x,z)\right\} \\
&\quad p_{Z|A,X}(z \mid a,x) p_X(x) dmdzdx \ . \qquad (3.10)
\end{aligned}
$$

Putting everything together yields

$$
(3.1)+(3.2)+(3.3)+(3.4)+(3.5) = (3.6)+(3.8)+(3.9)+(3.10) \ .
$$

Thus,

$$
\begin{aligned}
& E\left[D^*_{P',a}(O)\right] - \theta_a \\
&= \int \left\{\xi'_{a,\eta|X}(x) - \xi_{a,\eta|X}(x)\right\} \left(1 - \frac{\pi_a(x)}{\pi'_a(x)}\right) p_X(x)dx \\
&\quad + \int \frac{\pi_a(x)}{\pi'_a(x)} \left\{ \frac{p'_{M|\Delta=1,A,X}(m \mid 0,x)}{p'_{M|A,X,Z}(m \mid a,x,z)} - \frac{p_{M|\Delta=1,A,X}(m \mid 0,x)}{p_{M|A,X,Z}(m \mid a,x,z)} \right\} \times \\
&\quad \left\{\mu_{Y|A,M,X,Z}(a,m,x,z) - \mu'_{Y|A,M,X,Z}(a,m,x,z)\right\} p_{M|A,X,Z}(m \mid a,x,z) p_{Z|A,X}(z \mid a,x) p_X(x) dmdzdx \\
&\quad + \int \left\{ \frac{\pi_a(x)}{\pi'_a(x)} \eta_{\mu|A,M,X}(a,m,x) - \frac{\bar{\pi}_0(x)}{\bar{\pi}'_0(x)} \eta'_{\mu|A,M,X}(a,m,x) \right\} \\
&\qquad \left\{p'_{M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x)\right\} p_X(x)dmdx \\
&\quad + \int \frac{\pi_a(x)}{\pi'_a(x)} \left\{p'_{M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x)\right\} \left\{\mu'_{Y|A,M,X,Z}(a,m,x,z) - \mu_{Y|A,M,X,Z}(a,m,x,z)\right\} \\
&\quad p_{Z|A,X}(z \mid a,x) p_X(x) dmdzdx \ .
\end{aligned}
$$

Part II Let P_n denote the empirical measure of $O_1, \ldots, O_n$. Let $P_n^{'}$ be any estimator of P_0 compatible with the nuisance models used in the estimation of $D_{P,a}$, we have:

$$
\begin{aligned}
\Psi_a(P_n^{'}) - \Psi_a(P) &= (P_n^{'} - P)D_{P_n^{'},a} + R_2(P, P_n^{'}) \\
&= -PD_{P_n^{'},a} + R_2(P, P_n^{'}) \\
&= -P_n D_{P_n^{'},a} + (P_n - P)D_{P,a} + (P_n - P)(D_{P_n^{'},a} - D_{P,a}) + R_2(P, P_n^{'}) \ ,
\end{aligned}
$$

where $R_2(P, P_n^{'})$ is the second-order remainder term. As the one-step estimator is defined as $\Psi_a(P_n^{'}) + P_n D_{P_n^{'},a}$, we have:

$$
\theta^+_{n,a} - \theta_a = (P_n - P)D_{P,a} + (P_n - P)(D_{P_n^{'},a} - D_{P,a}) + R_2(P, P_n^{'}) \ .
$$

Next, we need to show $(P_n - P)(D_{P_n^{'},a} - D_{P,a}) = o_P(1)$ and $R_2(P, P_n^{'}) = o_P(1)$. The former will hold under conditions (B3) and (B4) of the theorem.

As for the second order reminder term, we have:

$$
R_2(P, P_n^{'}) = \Psi_a(P_n^{'}) + PD_{P_n^{'},a} - \Psi_a(P) = E\left[D^*_{P_n^{'},a}(O)\right] - \theta_a \ .
$$

Referring back to Part I, we have

$$
\begin{aligned}
& R_2(P, P_n^{'}) \\
= & \int \left\{\xi_{n,a,\eta|X}(x) - \xi_{a,\eta|X}(x)\right\} \left(1 - \frac{\pi_a(x)}{\pi_{n,a}(x)}\right) p_X(x)dx \\
& + \int \frac{\pi_a(x)}{\pi_{n,a}(x)} \left\{ \frac{p_{n,M|\Delta=1,A,X}(m \mid 0, x)}{p_{n,M|A,X,Z}(m \mid a, x, z)} - \frac{p_{M|\Delta=1,A,X}(m \mid 0, x)}{p_{M|A,X,Z}(m \mid a, x, z)} \right\} \times \\
& \left\{\mu_{Y|A,M,X,Z}(a, m, x, z) - \mu_{n,Y|A,M,X,Z}(a, m, x, z)\right\} p_{M|A,X,Z}(m \mid a, x, z) p_{Z|A,X}(z \mid a, x) p_X(x) dmdzdx \\
& + \int \left\{ \frac{\pi_a(x)}{\pi_{n,a}(x)} \eta_{\mu|A,M,X}(a, m, x) - \frac{\bar{\pi}_0(x)}{\bar{\pi}_{n,0}(x)} \eta_{n,\mu|A,M,X}(a, m, x) \right\} \\
& \qquad \left\{p_{n,M|\Delta=1,A,X}(m \mid 0, x) - p_{M|\Delta=1,A,X}(m \mid 0, x)\right\} p_X(x) dmdx \\
& + \int \frac{\pi_a(x)}{\pi_{n,a}(x)} \left\{p_{n,M|\Delta=1,A,X}(m \mid 0, x) - p_{M|\Delta=1,A,X}(m \mid 0, x)\right\} \left\{\mu_{n,Y|A,M,X,Z}(a, m, x, z) - \mu_{Y|A,M,X,Z}(a, m, x, z)\right\} \\
& p_{Z|A,X}(z \mid a, x) p_X(x) dmdzdx \ .
\end{aligned}
$$

Using Cauchy-Schwartz inequality and assumptions (B1) and (B2) of the theorem, we can show $R_2(P, P_n^{'}) = o_P(1)$. To be more specific, for the first line of the term,

$$
\begin{aligned}
& \int \left\{\xi_{n,a,\eta|X}(x) - \xi_{a,\eta|X}(x)\right\} \left(1 - \frac{\pi_a(x)}{\pi_{n,a}(x)}\right) p_X(x)\, dx \\
\leq & \int \left| \left\{\xi_{n,a,\eta|X}(x) - \xi_{a,\eta|X}(x)\right\} \left(\frac{\pi_{n,a}(x) - \pi_a(x)}{\pi_{n,a}(x)}\right) p_X(x) \right| dx \\
\leq & \left\{ \sup_x \frac{1}{\pi_{n,a}(x)} \right\} \int \left| \left\{\xi_{n,a,\eta|X}(x) - \xi_{a,\eta|X}(x)\right\} \left\{\pi_{n,a}(x) - \pi_a(x)\right\} \right| p_X(x)\, dx \ .
\end{aligned}
$$

Then applying assumption (B1) to the supremum and Cauchy-Schwarz to the integration, we have

$$
\begin{aligned}
& \leq \frac{1}{\epsilon_1} \|\xi_{n,\eta|a,X} - \xi_{a,\eta|X}\| \|\pi_{n,a} - \pi_a\| \\
& = o_P(1) \ ,
\end{aligned}
$$

where the last line follows from assumption (B2). The same reasoning can be extended to the other terms in $R_2(P, P_n^{'})$.

For the second line of the term,

$$
\begin{aligned}
&\int \frac{\pi_a(x)}{\pi_{n,a}(x)} \left\{ \frac{p_{n,M\mid\Delta=1,A,X}(m \mid 0,x)}{p_{n,M\mid A,X,Z}(m \mid a,x,z)} - \frac{p_{M\mid\Delta=1,A,X}(m \mid 0,x)}{p_{M\mid A,X,Z}(m \mid a,x,z)} \right\} \times \\
&\left\{\mu_{Y\mid A,M,X,Z}(a,m,x,z) - \mu_{n,Y\mid A,M,X,Z}(a,m,x,z)\right\} p_{M\mid A,X,Z}(m \mid a,x,z) p_{Z\mid A,X}(z \mid a,x) p_X(x) dmdzdx \\
&= \int \frac{\pi_a(x)}{\pi_{n,a}(x)} \left\{ \frac{p_{n,M\mid\Delta=1,A,X}(m \mid 0,x) - p_{M\mid\Delta=1,A,X}(m \mid 0,x)}{p_{n,M\mid A,X,Z}(m \mid a,x,z)} \right\} \times \\
&\left\{\mu_{Y\mid A,M,X,Z}(a,m,x,z) - \mu_{n,Y\mid A,M,X,Z}(a,m,x,z)\right\} p_{M\mid A,X,Z}(m \mid a,x,z) p_{Z\mid A,X}(z \mid a,x) p_X(x) dmdzdx \\
&+ \int \frac{\pi_a(x)}{\pi_{n,a}(x)} \left\{ \frac{p_{M\mid A,X,Z}(m \mid a,x,z) - p_{n,M\mid A,X,Z}(m \mid a,x,z)}{p_{n,M\mid A,X,Z}(m \mid a,x,z)} \right\} \times \\
&\left\{\mu_{Y\mid A,M,X,Z}(a,m,x,z) - \mu_{n,Y\mid A,M,X,Z}(a,m,x,z)\right\} p_{M\mid\Delta=1,A,X}(m \mid 0,x) p_{Z\mid A,X}(z \mid a,x) p_X(x) dmdzdx \\
&\leq \left\{ \sup_{m,a,x,z} \frac{1}{\pi_{n,a}(x) p_{n,M\mid A,X,Z}(m \mid a,x,z)} \right\} \int \Bigg| \left\{ p_{n,M\mid\Delta=1,A,X}(m \mid 0,x) - p_{M\mid\Delta=1,A,X}(m \mid 0,x) \right\} \times \\
&\qquad \left\{\mu_{Y\mid A,M,X,Z}(a,m,x,z) - \mu_{n,Y\mid A,M,X,Z}(a,m,x,z)\right\} \Bigg| p_{M\mid A,X,Z}(m \mid a,x,z) p_{Z\mid A,X}(z \mid a,x) p_X(x)\, dx \\
&+ \left\{ \sup_{m,a,x,z} \frac{1}{\pi_{n,a}(x) p_{n,M\mid A,X,Z}(m \mid a,x,z)} \right\} \int \Bigg| \left\{ p_{M\mid A,X,Z}(m \mid a,x,z) - p_{n,M\mid A,X,Z}(m \mid a,x,z) \right\} \times \\
&\qquad \left\{\mu_{Y\mid A,M,X,Z}(a,m,x,z) - \mu_{n,Y\mid A,M,X,Z}(a,m,x,z)\right\} \Bigg| p_{M\mid\Delta=1,A,X}(m \mid 0,x) p_{Z\mid A,X}(z \mid a,x) p_X(x)\, dx \\
&\leq \frac{1}{\epsilon_1 \epsilon_3} \| p_{n,M\mid\Delta=1,A,X}(m \mid 0,x) - p_{M\mid\Delta=1,A,X}(m \mid 0,x) \| \| \mu_{Y\mid A,M,X,Z}(a,m,x,z) - \mu_{n,Y\mid A,M,X,Z}(a,m,x,z) \| \\
&+ \frac{1}{\epsilon_1 \epsilon_3} \| p_{M\mid A,X,Z}(m \mid a,x,z) - p_{n,M\mid A,X,Z}(m \mid a,x,z) \| \| \mu_{Y\mid A,M,X,Z}(a,m,x,z) - \mu_{n,Y\mid A,M,X,Z}(a,m,x,z) \| \\
&= o_p(1) \ .
\end{aligned}
$$

For the third line of the term,

$$
\begin{aligned}
&\int \left\{ \frac{\pi_a(x)}{\pi_{n,a}(x)} \eta_{\mu|A,M,X}(a,m,x) - \frac{\bar{\pi}_0(x)}{\bar{\pi}_{n,0}(x)} \eta_{n,\mu|A,M,X}(a,m,x) \right\} \\
&\qquad \left\{ p_{n,M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x) \right\} p_X(x) dm dx \\
&= \int \{ \frac{\pi_a(x) - \pi_{n,a}(x)}{\pi_{n,a}(x)} \eta_{\mu|A,M,X}(a,m,x) + \frac{\bar{\pi}_{n,0}(x) - \bar{\pi}_0(x)}{\bar{\pi}_{n,0}(x)} \eta_{\mu|A,M,X}(a,m,x) \\
&\qquad + \frac{\bar{\pi}_0(x)}{\bar{\pi}_{n,0}(x)} \left\{ \eta_{\mu|A,M,X}(a,m,x) - \eta_{n,\mu|A,M,X}(a,m,x) \right\} \} \times \\
&\qquad \left\{ p_{n,M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x) \right\} p_X(x) dm dx \\
&\le \left\{ \sup_x \frac{1}{\pi_{n,a}(x)} \right\} \int \left| \{ \pi_a(x) - \pi_{n,a}(x) \} \left\{ p_{n,M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x) \right\} \right| \\
&\qquad \eta_{\mu|A,M,X}(a,m,x) p_X(x) dm dx \\
&+ \left\{ \sup_x \frac{1}{\bar{\pi}_{n,0}(x)} \right\} \int \left| \{ \bar{\pi}_{n,0}(x) - \bar{\pi}_0(x) \} \left\{ p_{n,M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x) \right\} \right| \\
&\qquad \eta_{\mu|A,M,X}(a,m,x) p_X(x) dm dx \\
&+ \left\{ \sup_x \frac{1}{\bar{\pi}_{n,0}(x)} \right\} \int \left| \left\{ \eta_{\mu|A,M,X}(a,m,x) - \eta_{n,\mu|A,M,X}(a,m,x) \right\} \left\{ p_{n,M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x) \right\} \right| \\
&\qquad \bar{\pi}_0(x) p_X(x) dm dx \\
&\le \frac{1}{\epsilon_1} \| \pi_a(x) - \pi_{n,a}(x) \| \| p_{n,M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x) \| \\
&+ \frac{1}{\epsilon_2} \| \bar{\pi}_{n,0}(x) - \bar{\pi}_0(x) \| \| p_{n,M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x) \| \\
&+ \frac{1}{\epsilon_2} \| \eta_{\mu|A,M,X}(a,m,x) - \eta_{n,\mu|A,M,X}(a,m,x) \| \| p_{n,M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x) \| \\
&= o_p(1)\ .
\end{aligned}
$$

For the last line of the term,

$$
\begin{aligned}
&\int \frac{\pi_a(x)}{\pi_{n,a}(x)} \left\{ p_{n,M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x) \right\} \left\{ \mu_{n,Y|A,M,X,Z}(a,m,x,z) - \mu_{Y|A,M,X,Z}(a,m,x,z) \right\} \\
&\qquad p_{Z|A,X}(z \mid a,x) p_X(x)\, dm dz dx \\
&\le \left\{ \sup_x \frac{1}{\pi_{n,a}(x)} \right\} \int \Big| \Big\{ p_{n,M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x) \Big\} \times \\
&\qquad\qquad \Big\{ \mu_{n,Y|A,M,X,Z}(a,m,x,z) - \mu_{Y|A,M,X,Z}(a,m,x,z) \Big\} \Big| p_{Z|A,X}(z \mid a,x) \pi_a(x) p_X(x)\, dm dx \\
&\le \frac{1}{\epsilon_1} \| p_{n,M|\Delta=1,A,X}(m \mid 0,x) - p_{M|\Delta=1,A,X}(m \mid 0,x) \| \| \mu_{n,Y|A,M,X,Z}(a,m,x,z) - \mu_{Y|A,M,X,Z}(a,m,x,z) \| \\
&= o_p(1)\ .
\end{aligned}
$$

Thus $R_2(P, P_n^{'}) = o_p(1)$, thereby concluding the proof.

7.3.4 Theorem 3.3

Assumptions of theorem

Assumption (B1) guarantees that estimated propensities and motion densities are appropriately bounded so that the one-step estimator is never ill-defined. Assumption (C1) stipulates convergence rate conditions on nuisance estimates in terms of $L^2(P)$ norms. This assumption would be satisfied, for example, if each nuisance estimate achieved a rate of at least $n^{-1/4}$ with respect to $L^2(P)$ norm. However, it is also possible for slower convergence rates attained by some nuisance estimators to be compensated for by faster convergence rates attained by others. We note that the $n^{-1/4}$ rate is slower than the standard parametric rate, which potentially allows for the use of flexible regression techniques. On the other hand, to achieve this rate if X and/or Z are high-dimensional may require additional smoothness assumptions on the underlying nuisance parameters. For example, the highly adaptive LASSO estimator achieves a sufficiently fast rate of convergence *if* the underlying nuisance parameters have a *bounded variation norm* (Benkeser and van der Laan, 2016), an assumption that becomes more restrictive in higher dimensions. Assumptions (B3) and (C2) are necessary to ensure the negligibility of an empirical process term (Van Der Vaart et al., 1996). Assumption (C2) can be eliminated by utilizing cross-fitting techniques (van der Laan et al., 2011; Chernozhukov et al., 2018).

Proof of Theorem 3.3

The proof of Theorem 3.3 closely parallels the proof of Theorem 3.2.

$$\theta^{+}_{n,a} - \theta_a = (P_n - P)D_{P,a} + (P_n - P)(D_{P'_n,a} - D_{P,a}) + R_2(P, P^{'}_n) \ .$$

For asymptotic linearity, we need to show $(P_n - P)(D_{P'_n,a} - D_{P,a}) = o_P(1/\sqrt{n})$ and $R_2(P, P^{'}_n) = o_P(1/\sqrt{n})$. The former holds under conditions (B3) and (C2) of the theorem. The structure of the second-order term is presented in the proof of Theorem 3.2, Part I, and the convergence of the second-order term is established using logic similar to the proof of Theorem 3.2, Part II.

7.3.5 Simulation details

Sex X_1 was generated from a Bernoulli distribution with a probability of 0.75 of female sex, age X_2 was generated by truncating a Gamma distribution with a shape of 25 and a rate of 2.5 to values between 8 and 13, and right-handedness X_3 was simulated from a Bernoulli distribution with success probability of 0.92. Given covariates $X = x$, diagnosis A was drawn from a Bernoulli distribution with a success probability of $\text{expit}(-0.11 + 0.71x_1 - 0.08x_2 - 0.19x_3)$, where these coefficients were derived from a logistic regression fitted to the real data. Diagnosis-specific covariates Z included four components: ADOS Z_1 was assigned a value of 0 for simulated non-ASD participants, while for simulated ASD participants, a value was drawn from a Poisson distribution with mean 11.86; FIQ Z_2 was sampled from a $\text{Normal}(114.6, 11.6^2)$ distribution for simulated non-ASD participants and a $\text{Normal}(104.2, 17.4^2)$ distribution for simulated ASD participants, where these means and standard deviations were calculated based on the real data; stimulant and non-stimulant medication, Z_3 and Z_4, respectively, were assigned a value of 0 for simulated non-ASD participants and had a value drawn from a Bernoulli distribution with success probabilities of 0.2 and 0.17, respectively for $A = 1$ participants. Given $A = a, X = x, Z = z$, the natural logarithm of mean framewise displacement M was generated from a $\text{Normal}(-1.26 + 0.095a + 0.104x_1 - 0.0535x_2 - 0.12x_3 + 0.00675z_1 - 0.000255z_2 + 0.324z_3 + 0.064z_4, 0.56^2)$ distribution. We defined a tolerable motion level $\Delta = 1$ as the indicator that $M \leq 0.2$.

The conditional means of the six functional connectivity outcomes were

$$\mu_{Y_1|A,M,X,Z}(a,m,x,z) = -0.22 + 0 \times a - 0.98m - 0.06x_1 + 0.012x_2 + 0.03x_3,$$

$$\mu_{Y_2|A,M,X,Z}(a,m,x,z) = -0.20 + 0 \times a + 0.92m + 0.06x_1 - 0.009x_2 - 0.03x_3,$$

$$\mu_{Y_3|A,M,X,Z}(a,m,x,z) = -0.37 + 0 \times a + 0.86m + 0.04x_1 + 0.002x_2 + 0.04x_3,$$

$$\mu_{Y_4|A,M,X,Z}(a,m,x,z) = 0.17 + 0 \times a - 1.02m - 0.06x_1 + 0.002x_3 + 0.04x_3,$$

$$\begin{aligned}\mu_{Y_5|A,M,X,Z}&(a,m,x,z)\\ &= -0.20 - 0.03a + 1.50m - 0.61m^2 + 0.02(x_1 - z_4) - 0.002x_2 + 0.03(x_3 - z_3)\\ &\qquad - 0.0005z_1 + 0.0003z_2,\end{aligned}$$

$$\begin{aligned}\mu_{Y_6|A,M,X,Z}&(a,m,x,z)\\ &= -0.16 - 0.05a + 1.67m - 0.64m^2 + 0.03(x_1 - z_4) - 0.001x_2 + 0.02(x_3 - z_3)\\ &\qquad - 0.0005z_1 + 0.0003z_2.\end{aligned}$$

7.3.6 Additional simulations to confirm theoretical properties of estimators

We evaluate the statistical properties of our estimators established by our theorems through Monte Carlo simulation. These simulations were conducted purely to confirm theoretical properties of the estimators and are not tied to the real data analysis context in any way. In this simulation, we generated covariate X from Bernoulli(1/2). Given $X = x$, we generated a binary variable A according to a Bernoulli distribution with $\pi_1(x) = \text{expit}(x - 1/4)$. Given $A = a$, we drew Z from Bernoulli(expit($5a/4 - 1/2$)). Given $A = a, X = x, Z = z$, we drew M from a normal distribution $N(1 + a + x/2 - z/4, 1)$. Finally, given $A = a, X = x, Z = z, M = m$, we drew Y is drawn from a normal distribution $N(-1 + x/2 - z/3 - a/4 + m/5, 1)$. We defined a tolerable M level $\Delta = 1$ as the indicator that $M \leq 2$. The true values of these parameters were -0.717 and -1.068, respectively, with variance bound $Var\{D_{a,P}(O)\}$ equal to 3.453 and

7.151, respectively.

For each sample size $n \in \{200, 500, 1000, 2000, 4000\}$, we generated 1000 datasets based on the above data generating process, and used the resultant data to compute our proposed estimators along with their corresponding confidence interval, either with cross-fitting or without the use of cross-fitting. We evaluated the estimators based on their bias (scaled by $n^{1/2}$), their standard error (scaled by $n^{1/2}$), the ratio of the scaled standard error to the square root of the efficient variance, and the coverage of 95% Wald-style confidence interval.

We first evaluated estimators under the conditions of the theorem where all nuisance parameters are consistently estimated at appropriate rates. To achieve this, we used the logistic regression model for $\pi_a(x)$ and main term linear model $\mu_{Y|A,M,X,Z}(a, m, x, z)$ and fully saturated (all possible interactions) regression models for the remaining nuisance parameters. In this scenario, the one-step estimators are expected to be consistent and asymptotically linear.

When all the nuisance parameters are estimated consistently at appropriate rates, the bias of the estimators decreases to 0 as the sample size increases, and the coverage of the 95% confidence interval increases to 0.95 (Table 7.5). Even at a sample size of 200, the bias remains approximately 2% of the true value, and the worst-case coverage of 95% confidence intervals is 86%. Similar results are achieved without the use of cross-fitting (Table 7.6).

We also studied the impact of inconsistent estimation of different combinations of nuisance parameters. We examined five situations in which only specific combinations of the nuisance parameters were correctly specified, as defined in column 1 of Table 7.7. Based on its multiple robustness properties, our one-step estimators are expected to maintain consistency across these five scenarios. Nuisance regressions that were incorrectly specified were modeled using only the intercept. Conditional densities that were incorrectly specified were modeled using a Gaussian distribution in which the

n	$\theta_{n,0}^{\text{cf}}$				$\theta_{n,1}^{\text{cf}}$			
	$n^{1/2}$ bias	$n^{1/2}$ sd	sd ratio	cover	$n^{1/2}$ bias	$n^{1/2}$ sd	sd ratio	cover
200	-0.208	2.148	1.122	0.910	-0.143	2.391	1.382	0.860
500	-0.169	1.926	0.996	0.942	-0.165	2.319	1.172	0.912
1000	-0.149	1.918	0.987	0.945	-0.119	2.325	1.100	0.929
2000	-0.058	1.980	1.035	0.941	-0.072	2.421	1.093	0.928
4000	0.011	1.887	1.000	0.946	0.050	2.185	0.961	0.960

Table 7.5: Confirming theoretical properties of estimators: All nuisance parameters are consistently estimated at appropriate rates with the use of MoCo (with cross-fitting).

n	$\theta_{n,0}^{+}$				$\theta_{n,1}^{+}$			
	$n^{1/2}$ bias	$n^{1/2}$ sd	sd ratio	cover	$n^{1/2}$ bias	$n^{1/2}$ sd	sd ratio	cover
200	-0.136	2.035	1.053	0.946	-0.168	2.427	1.249	0.882
500	-0.149	2.007	1.035	0.929	-0.238	2.338	1.122	0.920
1000	-0.102	1.940	1.011	0.943	-0.169	2.331	1.080	0.935
2000	-0.052	1.995	1.049	0.933	-0.144	2.403	1.071	0.934
4000	0.037	1.867	0.989	0.956	0.066	2.251	0.982	0.951

Table 7.6: Confirming theoretical properties of estimators: All nuisance parameters are consistently estimated at appropriate rates without the use of cross-fitting.

mean was equal to the sample mean and the standard deviation equal to the sample standard deviation.

Our results indicate that as sample size increases, the bias and standard error decrease across all settings considered, which supports the multiple robustness theory (Table 7.7 and Table 7.8).

Setting	n	$\text{bias}_{\theta^{\text{cf}}_{n,0}}$	$\text{sd}_{\theta^{\text{cf}}_{n,0}}$	$\text{bias}_{\theta^{\text{cf}}_{n,1}}$	$\text{sd}_{\theta^{\text{cf}}_{n,1}}$
(B2.1)	200	0.0500	0.1289	0.0681	0.1455
	500	0.0341	0.1016	0.0558	0.0923
	1000	0.0103	0.0710	0.0470	0.0710
	2000	-0.0007	0.0475	0.0380	0.0504
	4000	-0.0026	0.0170	0.0331	0.0368
(B2.2)	200	-0.0575	0.1164	-0.0510	0.1621
	500	-0.0130	0.0865	-0.0206	0.1125
	1000	-0.0087	0.0594	-0.0252	0.0800
	2000	-0.0044	0.0405	-0.0223	0.0555
	4000	-0.0008	0.0264	-0.0147	0.0400
(B2.3)	200	-0.0412	0.1199	-0.0561	0.1275
	500	-0.0021	0.0892	-0.0073	0.1006
	1000	-0.0019	0.0612	-0.0043	0.0743
	2000	-0.0011	0.0429	-0.0027	0.0520
	4000	0.0008	0.0272	0.0019	0.0375
(B2.4)	200	-0.0585	0.1189	-0.0488	0.1487
	500	-0.0099	0.0915	-0.0087	0.1035
	1000	-0.0092	0.0608	-0.0170	0.0756
	2000	-0.0039	0.0429	-0.0161	0.0532
	4000	-0.0004	0.0272	-0.0105	0.0383
(B2.5)	200	-0.0599	0.1217	-0.0505	0.1556
	500	-0.0118	0.0885	-0.0143	0.1107
	1000	-0.0107	0.0582	-0.0248	0.0794
	2000	-0.0054	0.0397	-0.0230	0.0565
	4000	-0.0008	0.0264	-0.0148	0.0399

Table 7.7: Bias and standard deviation(sd) of MoCo (with cross-fitting). The settings column indicates which nuisance parameters are consistently estimated based on assumption (B2) in Theorem 3.2 as outlined in the main manuscript.

Setting	n	$\text{bias}_{\theta^+_{n,0}}$	$\text{sd}_{\theta^+_{n,0}}$	$\text{bias}_{\theta^+_{n,1}}$	$\text{sd}_{\theta^+_{n,1}}$
(B2.1)	200	0.0195	0.1164	0.0382	0.1646
	500	0.0106	0.1045	0.0371	0.1020
	1000	0.0002	0.0695	0.0363	0.0735
	2000	-0.0059	0.0478	0.0330	0.0518
	4000	-0.0058	0.0188	0.0308	0.0367
(B2.2)	200	-0.0277	0.1275	-0.0064	0.1982
	500	-0.0125	0.0872	-0.0296	0.1214
	1000	-0.0066	0.0596	-0.0212	0.0851
	2000	-0.0048	0.0409	-0.0194	0.0603
	4000	0.0016	0.0255	-0.0138	0.0413
(B2.3)	200	-0.0234	0.1248	-0.0094	0.1707
	500	-0.0058	0.0883	-0.0112	0.1041
	1000	-0.0035	0.0608	-0.0013	0.0770
	2000	-0.0025	0.0425	-0.0003	0.0552
	4000	0.0021	0.0259	0.0034	0.0389
(B2.4)	200	-0.0275	0.1289	-0.0014	0.1824
	500	-0.0101	0.0895	-0.0221	0.1119
	1000	-0.0062	0.0614	-0.0142	0.0803
	2000	-0.0042	0.0428	-0.0135	0.0574
	4000	0.0018	0.0260	-0.0096	0.0399
(B2.5)	200	-0.0277	0.1275	-0.0064	0.1982
	500	-0.0125	0.0872	-0.0296	0.1214
	1000	-0.0066	0.0596	-0.0212	0.0851
	2000	-0.0048	0.0409	-0.0194	0.0603
	4000	0.0016	0.0255	-0.0138	0.0413

Table 7.8: Bias and standard deviation(sd) of the one-step estimator without the use of cross-fitting. The settings column indicates which nuisance parameters are consistently estimated based on assumption (B2) in Theorem 3.2 as outlined in the main manuscript.

7.3.7 Additional details of data analysis

We subset to 377 scans corresponding to 8 to 13-year-old children from the Kennedy Krieger Institute (KKI) and New York University (NYU) from ABIDE I and ABIDE II . The ADOS score is a standardized assessment tool used to diagnose ASD, with higher scores indicating greater social disability.

Although ASD is more prevalent in males than females, we treated sex as a confounder (X) rather than diagnosis-specific variable (Z) because sex-specific differences in functional connectivity have been previously documented (Shanmugan et al., 2022), which could mask ASD-related differences in this cohort. Mean FD is an average of the frame-to-frame displacement calculated from the rigid body motion correction parameters used in quality control and motion correction (Esteban et al., 2019; Power et al., 2014a).

Resting-state fMRI scans were acquired using one of three protocols: 1) a 3T Philips Achieva scanner, 8-channel head coil, repetition time (TR)/echo time (TE)=2500/30 ms, flip angle 75°, 3×3×3 mm voxels, SENSE phase reduction=3, 2 dummy scans, scan duration varied from 5m10s (one subject) to 6m45s (one subject) with most scans at either 5m20s (n=46) or 6m30s (n=99) (KKI); 2) the same but with a 32-channel head coil with most scans 6m30s (n=61) (KKI); 3) a 3T Siemens Allegra scanner, 8-channel head coil, TR/TE=2000/15 ms, flip angle=90°, 3x3x4 mm voxels, scan duration=6:00 (NYU), and we removed the first two volumes. All children also had an anatomical T1 scan collected.

T1 anatomical and rs-fMRI data were processed with the cifti option for cortical surface registration using fMRIPrep, including anatomical tissue segmentation, surface construction, and surface registration, followed by fMRI motion correction, slice-time correction, boundary-based coregistration, and resampling to the fsaverage template (Esteban et al., 2019). The detailed output from fMRIPrep is included in Section 7.4.3. We visually inspected the accuracy of the cortical segmentation using the fMRIPrep quality control html files. We excluded 19 participants due to issues with the cortical segmentation. Issues with fMRIprep included image homogeneity issues, outliers in brain morphology, and motion during the T1 scan. An example participant that failed this step is included in Figure 7.14.

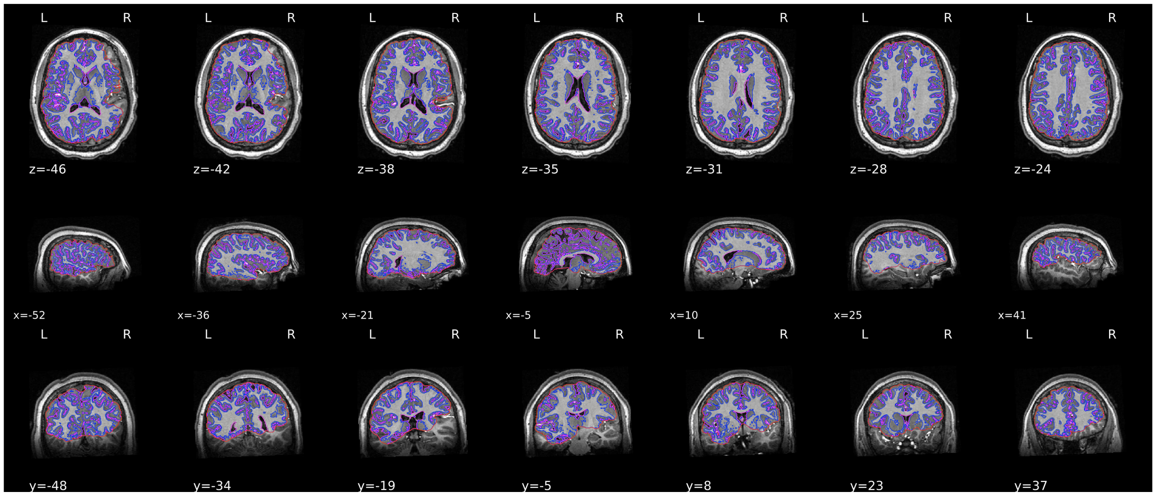


Figure 7.10: An example participant whose cortical segmentation failed in fMRIPrep. The template T1-weighted image is shown with contours outlining the detected brain mask and brain tissue segmentations. It is apparent from the middle and lower rows that large parts of the brain, including most of the temporal lobe and parts of the occipital lobe, were incorrectly excluded from the segmentation.

Our final study sample was 377 children, with 245 non-autistic children and 132 ASD children. Table 7.9 displays characteristics of the analysis cohort by ASD diagnosis status.

We calculated the average time series for regions of interest defined using Schaefer's 400-node brain parcellation, which associates each node with the resting networks from Yeo-7 (Schaefer et al., 2018).

fMRIprep data preprocessing for ABIDE

Results included in this manuscript come from preprocessing performed using *fMRIPrep* 21.0.2 (Esteban et al. (2018b); Esteban et al. (2018a); RRID:SCR_016216), which is based on *Nipype* 1.6.1 (Gorgolewski et al. (2011); Gorgolewski et al. (2018);

	ASD (N=132)	non-ASD (N=245)	Total (N=377)	p value
Δ				< 0.001
Unusable	98 (74.2%)	119 (48.6%)	217 (57.6%)	
Usable	34 (25.8%)	126 (51.4%)	160 (42.4%)	
Age				0.395
Mean (SD)	10.281 (1.590)	10.413 (1.356)	10.367 (1.441)	
Range	8.014 - 13.950	8.010 - 13.720	8.010 - 13.950	
Gender				0.009
Female	24 (18.2%)	75 (30.6%)	99 (26.3%)	
Male	108 (81.8%)	170 (69.4%)	278 (73.7%)	
Handedness				0.454
Left	12 (9.1%)	17 (6.9%)	29 (7.7%)	
Right	120 (90.9%)	228 (93.1%)	348 (92.3%)	
ADOS				< 0.001
Mean (SD)	11.856 (3.831)	0.000 (0.000)	4.151 (6.098)	
Range	4.000 - 22.000	0.000 - 0.000	0.000 - 22.000	
FIQ				< 0.001
Mean (SD)	104.182 (17.368)	114.649 (11.569)	110.984 (14.729)	
Range	67.000 - 148.000	80.000 - 144.000	67.000 - 148.000	
Currently on Stimulants				< 0.001
No	105 (79.5%)	245 (100.0%)	350 (92.8%)	
Yes	27 (20.5%)	0 (0.0%)	27 (7.2%)	
Currently on Nonstimulants				< 0.001
No	109 (82.6%)	244 (99.6%)	353 (93.6%)	
Yes	23 (17.4%)	1 (0.4%)	24 (6.4%)	
Site ID				< 0.001
ABIDEI-KKI	19 (14.4%)	32 (13.1%)	51 (13.5%)	
ABIDEI-NYU	39 (29.5%)	41 (16.7%)	80 (21.2%)	
ABIDEII-KKI	51 (38.6%)	152 (62.0%)	203 (53.8%)	
ABIDEII-NYU	23 (17.4%)	20 (8.2%)	43 (11.4%)	

Table 7.9: Demographic characteristics: Continuous variables are described using mean and standard deviation, and diagnostic groups are compared using the Kruskal-Wallis rank-sum test. Binary and categorical variables are reported as frequencies and percentages, and differences between diagnostic groups are assessed using either the Chi-square test or Fisher's exact test.

RRID:SCR_002502). The text below is automatically produced by fMRIprep.

Anatomical data preprocessing A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset. The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection` (Tustison et al., 2010), distributed with ANTs 2.3.3 (Avants et al., 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using `fast` (FSL 6.0.5.1:57b01774, RRID:SCR_002823, Zhang et al., 2001). Brain surfaces were reconstructed using `recon-all` (FreeSurfer 6.0.1, RRID:SCR_001847, Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al., 2017). Volume-based spatial normalization to two standard spaces (MNI152 NLin6 Asym, MNI152 NLin 2009c Asym) was performed through nonlinear registration with `antsRegistration` (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following templates were selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* [Fonov et al. (2009), RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym], *FSL's MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model* [Evans et al. (2012), RRID:SCR_002823; TemplateFlow ID: MNI152NLin6Asym].

Functional data preprocessing For each of the 1 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using `mcflirt` (FSL 6.0.5.1:57b01774, Jenkinson et al., 2002). BOLD runs were slice-time corrected to 1.22s (0.5 of slice acquisition range 0s-2.45s) using `3dTshift` from AFNI (Cox and Hyde, 1997, RRID:SCR_005927). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head-motion. These resampled

BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer) which implements boundary-based registration (Greve and Fischl, 2009). Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD) and three region-wise global signals. FD was computed following Power (absolute sum of relative motions, Power et al. (2014b)) calculated using the implementation in *Nipype* (following the definitions by Power et al., 2014b). The three global signals were extracted within the CSF, the WM, and the whole-brain masks. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. The BOLD time-series were resampled onto the following surfaces (FreeSurfer reconstruction nomenclature): *fsaverage*. *Grayordinates* files (Glasser et al., 2013b) containing 91k samples were also generated using the highest-resolution `fsaverage` as intermediate standardized surface space. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

Many internal operations of *fMRIPrep* use *Nilearn* 0.8.1 (Abraham et al., 2014, RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in *fMRIPrep*'s documentation.

Copyright Waiver The above boilerplate text was automatically generated by fMRIPrep with the express intention that users should copy and paste this text into their manuscripts *unchanged*. It is released under the CC0 license.

Positivity assumptions in the ABIDE data

For the dataset, the positivity assumption (A1.1) holds as no demographic characteristics X can perfectly predict ASD. To assess the positivity assumptions (A1.2), we examine the values of four ratios present in the estimation process and the efficient influence function:

$$\frac{p_{n,M|\Delta=1,A,X}(M_i \mid 0, X_i)}{p_{n,M|\Delta=1,A,X,Z}(M_i \mid A_i, X_i, Z_i)} \text{ for } \Delta_i = 1 \tag{7.5}$$

$$\frac{p_{n,M|A,X}(M_i \mid A_i, X_i)}{p_{n,M|A,X,Z}(M_i \mid A_i, X_i, Z_i)} \tag{7.6}$$

$$\frac{p_{n,M|\Delta=1,A,X}(M_i \mid 0, X_i)}{p_{n,M|A,X,Z}(M_i \mid 0, X_i, Z_i)} \tag{7.7}$$

$$\frac{p_{n,M|\Delta=1,A,X}(M_i \mid 0, X_i)}{p_{n,M|A,X,Z}(M_i \mid 1, X_i, Z_i)}. \tag{7.8}$$

There are no very large values for these ratios, which indicates the satisfaction of the positivity assumption. Figure 7.11 displays the histogram depicting the distribution of ratio values.

7.3.8 Estimated functional connectivity

The estimated functional connectivity using the naïve approach excluding high-motion participants, the naïve approach including all participants, and MoCo are illustrated in Figure 7.12. The seed region in the posterior default mode network is defined by the fourteenth parcel but is represented by the fuchsia point for clarity.

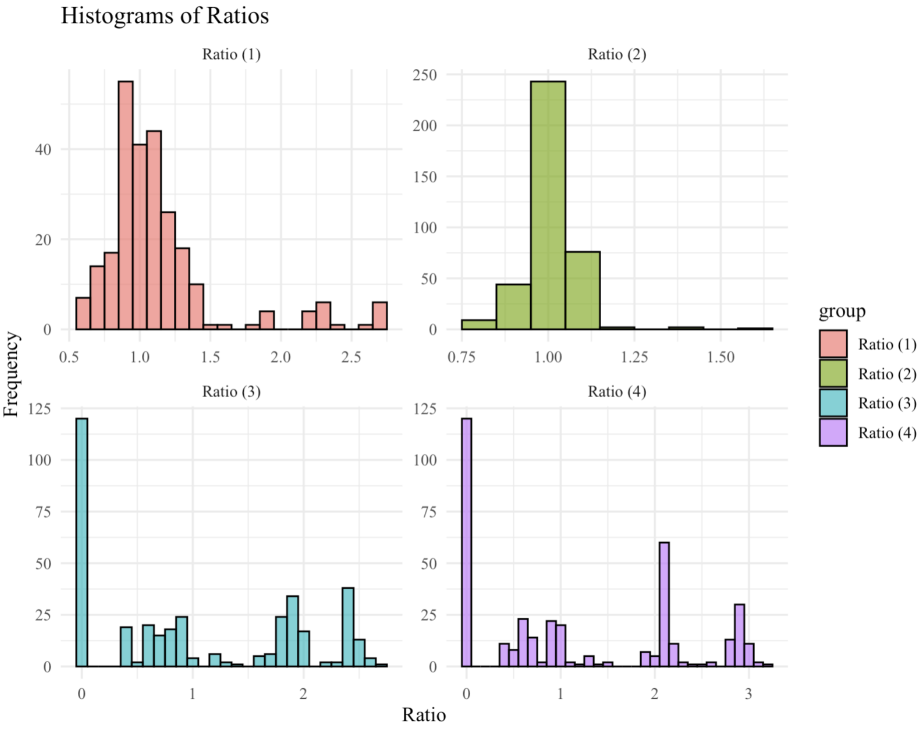


Figure 7.11: Histogram depicting the distribution of ratio values. Positivity assumptions appear reasonable since the ratios are not too large. Here, the ratios result in reasonable weights in the pseudo-regressions.

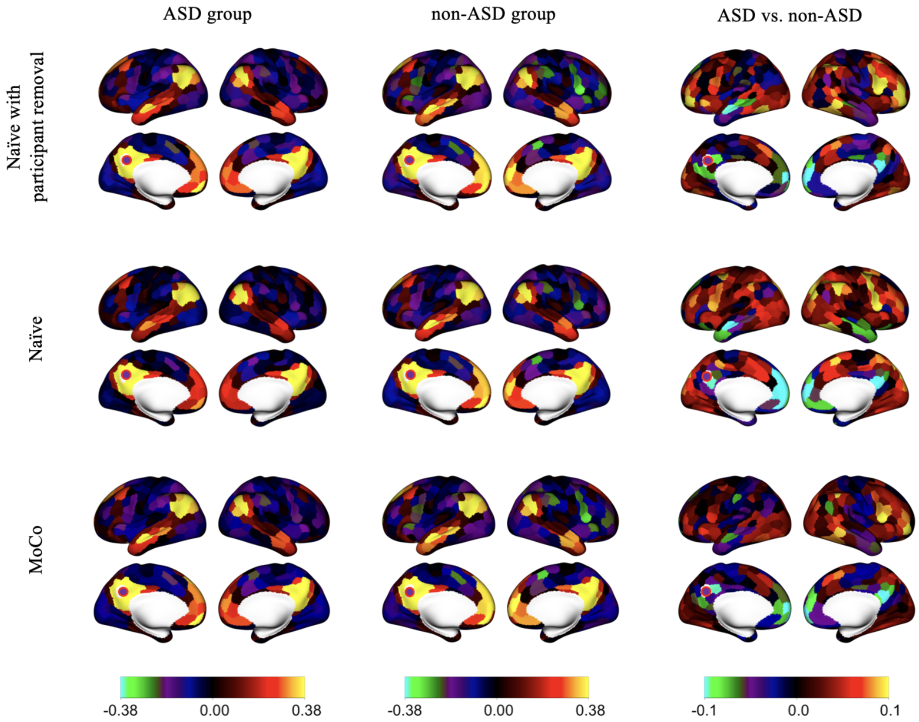


Figure 7.12: Estimated functional connectivity using the naïve approach excluding high-motion participants, the naïve approach, and MoCo for a seed region in the posterior default mode network (fuchsia point).

7.4 Supplementary Material of Chapter 5

7.4.1 Assumptions and proof for identification

The identification of the target parameter using observational data involves three assumptions. Below, we present the mathematical formulation in accordance with the sequence in the main paper:

(A1) *Mean exchangeability*: for all m such that $P\{p_{M\mid\Delta_M=1,0,X}(m \mid X) > 0\} > 0$, we have $E_{\mathrm{C}}\{Y(1,m) \mid A=a,X,Z\} = E_{\mathrm{C}}\{Y(1,m) \mid \Delta_Y=1, M=m, A=a, X, Z\}$, a.e.-$P$;

(A2) *Positivity*: (A2.1) for all possible x, we also have $P(A=a \mid X=x) > 0$ for $a=0,1$, and $P(\Delta_M=1 \mid A=0, X=x) > 0$. For all possible (x,z), we have $P(\Delta_Y=1 \mid A=a, X=x, Z=z) > 0$ for $a=0,1$. (A2.2) for every (x,z,m) such that $p_X(x)p_{Z\mid A,X}(z \mid a,x)p_{M\mid\Delta_M=1,A,X}(m \mid 0,x) > 0$, we also have that $p_{M\mid\Delta_Y=1,A,X,Z}(m \mid a,x,z) > 0$ for $a=0,1$.

(A3) *Causal Consistency*: for any participant with brain phenotype data ($\Delta_Y=1$) and motion value $M=m$, the observed brain phenotype measurement Y is equal to the counterfactual brain phenotype measurement $Y(1,m)$.

Theorem 6. *We denote $E\{E_C[Y(1,M_0) \mid A=a,X]\}$ as $\theta_{C,a}$. Under (A1)-(A3), the counterfactual $\theta_{C,a}$ is identified by θ_a, where*
$\theta_a = \iiint \mu_{Y\mid\Delta_Y=1,A,M,X,Z}(a,m,x,z)p_{Z\mid A,X}(z \mid a,x)p_{M\mid\Delta=1,A,X}(m \mid 0,x)p_X(x)dzdmdx.$

Proof Let $\theta_{\mathrm{C},a} = E\{E_{\mathrm{C}}[Y(1,M_0) \mid A=a,X]\}$, where $M_0 \sim P_{M\mid\Delta=1,A,X}(m \mid 0,x)$. We have:

$$
\begin{aligned}
&\theta_{\mathrm{C},a} = E\{E_{\mathrm{C}}[Y(1, M_0) \mid A = a, X]\} \\
&= \int E_{\mathrm{C}}(Y(1, M_0) \mid A = a, X = x)p_X(x)dx \\
&\overset{\text{tower rule}}{=} \iint E_{\mathrm{C}}(Y(1, m) \mid M_0 = m, A = a, X = x)p_{M\mid\Delta_M=1,A,X}(m \mid 0, x)p_X(x)dmdx \\
&\overset{\text{defn}}{=} \iint E_{\mathrm{C}}(Y(1, m) \mid A = a, X = x)p_{M\mid\Delta_M=1,A,X}(m \mid 0, x)p_X(x)dmdx \\
&\overset{\text{tower rule}}{=} \iiint E_{\mathrm{C}}(Y(1, m) \mid A = a, X = x, Z = z)p_{Z\mid A,X}(z \mid a, x)p_{M\mid\Delta_M=1,A,X}(m \mid 0, x)p_X(x)dzdmdx \\
&\overset{\text{assumption (A1)}}{=} \iiint E_{\mathrm{C}}(Y(1, m) \mid \Delta_Y = 1, M = m, A = a, X = x, Z = z)p_{Z\mid A,X}(z \mid a, x)p_{M\mid\Delta_M=1,A,X}(m \mid 0, x) \\
&\qquad p_X(x)dzdmdx \\
&\overset{\text{assumption (A3)}}{=} \iiint E(Y \mid \Delta_Y = 1, M = m, A = a, X = x, Z = z)p_{Z\mid A,X}(z \mid a, x)p_{M\mid\Delta_M=1,A,X}(m \mid 0, x) \\
&\qquad p_X(x)dzdmdx\ .
\end{aligned}
$$

The fourth equality results from the fact that by the construction of M_0, we have that $Y(1, m)M_0 \mid A, X$ for all m. The sixth equality results from the assumption that $Y(1, m)(\Delta_Y = 1, M) \mid A, X, Z$.

Under Theorem 1.1, the motion-controlled target parameter $\psi_C = \theta_{C,1} - \theta_{C,0}$ can be identified by $\theta_1 - \theta_0$, that is,

$$
\begin{aligned}
\iiint \big(&\mu_{Y\mid\Delta_Y=1,A,M,X,Z}(1, m, x, z)p_{Z\mid A,X}(z \mid 1, x) - \mu_{Y\mid\Delta_Y=1,A,M,X,Z}(1, m, x, z)p_{Z\mid A,X}(z \mid 1, x)\big) \\
&p_{M\mid\Delta_M=1,A,X}(m \mid 0, x)p_X(x)\, dz\, dm\, dx.
\end{aligned}
\tag{7.9}
$$

7.4.2 Details about the estimation and inference of the motion-controlled group difference

EIF and the one-step estimator

We develop a nonparametric efficient one-step estimator of θ_a based on the efficient influence function (EIF). The EIF-based estimator enjoys desirable properties such as asymptotic linearity with closed-form confidence intervals, and multiple robustness, which allows for some components of the data distribution to be inconsitently estimated while preserving the consistency of the estimator.

We define $\pi_{\Delta_Y=1|A,X,Z}(a,z,x) = P(\Delta_Y = 1 \mid A = a, Z = z, X = x)$ as the probability of a participant in the group a with covariate value (z,x) having nonmissing brain pheonotype data. We introduce the shorthand $\pi_a(x) = P(A = a \mid X = x)$, and $\bar{\pi}_0(x) = \pi_0(x)P(\Delta_M = 1 \mid A = 0, X = x)$ as the probability that $A = 0$ and $\Delta_M = 1$ conditional on $X = x$. We denote the indicator function ${}_a(A_i)$ equal to 1 if $A_i = a$ and zero otherwise; ${}_{a,1}(A_i, \Delta_{Y,i})$ equal to 1 if $A_i = a$ and $\Delta_{Y,i} = 1$ and equals zero otherwise. We also define for $a = 0, 1$

$$r_a(m,x,z) = \frac{p_{M|\Delta_M=1,A,X}(m \mid 0,x)}{p_{M|\Delta_Y=1,A,X,Z}(m \mid a,x,z)} \ , \tag{7.10}$$

$$\eta_{\mu|A,X,Z}(a,x,z) = \int \mu_{Y|\Delta_Y=1,A,M,X,Z}(a,m,x,z)p_{M|\Delta_M=1,A,X}(m \mid 0,x)dm \tag{7.11}$$

$$\eta_{\mu|A,M,X}(a,m,x) = \int \mu_{Y|\Delta_Y=1,A,M,X,Z}(a,m,x,z)p_{Z|A,X}(z \mid a,x)dz \ , \tag{7.12}$$

The efficient influence function of θ_a, for $a = 0, 1$ evaluated on a typical observation O_i is

$$\begin{aligned} D_{P,a}(O_i) =& \frac{{}_a(A_i, \Delta_{Y,i})}{\pi_{\Delta_Y=1|A,X,Z}(a, X_i, Z_i)\pi_a(X_i)} r_a(M_i, X_i, Z_i)\{Y_i - \mu_{Y|\Delta_Y=1,A,M,X,Z}(A_i, M_i, X_i, Z_i)\} \\ &+ \frac{{}_a(A_i)}{\pi_a(X_i)}\{\eta_{\mu|A,Z,X}(a, X_i, Z_i) - \xi_{a,\eta|X}(X_i)\} \\ &\quad + \frac{{}_{0,1}(A_i, \Delta_{M,i})}{\bar{\pi}_0(X_i)}\left\{\eta_{\mu|A,M,X}(a, M_i, X_i) - \xi_{a,\eta|X}(X_i)\right\} \\ &\quad\quad + \xi_{a,\eta|X}(X_i) - \theta_a. \end{aligned} \tag{7.13}$$

The one-step estimator of θ_a is defined as $\theta^+_{n,a} = \theta_{n,a} + \frac{1}{n}\sum_{i=1}^n D_{n,a}(O_i)$. Consequently, the motion-controlled average group difference is given by $\theta^+_{n,1} - \theta^+_{n,0}$.

Under similar assumptions as Ran (2024+), the one-step estimator is asymptotically linear: $\theta^{+}_{n,a} - \theta_a = \frac{1}{n}\sum_{i=1}^{n} D_{P,a}(O_i) + o_P(n^{-1/2})$, and

$$n^{1/2}(\theta^{+}_{n,a} - \theta_a) \Rightarrow N(0, E[D_{P,a}(O)^2]) \ .$$

Cross-fit one-step estimator

The cross-fitting process involves randomly dividing the data set into K parts. $K-1$ parts of the data are used to estimate the nuisance parameters appearing in the efficient influence function. Consider the example of $\xi_{a,\eta|X}$. We denote by $\xi_{n,k,a,\eta|X}$ the estimate of $\xi_{a,\eta|X}$ obtained when the k-th part of the data is withheld from the nuisance estimation stage. Similarly, we denote by $D_{a,n,k}$ the efficient influence function evaluated at the nuisance parameters estimated without using the k-th part of the data. Denote by $\mathcal{I}_k$ the indices of observations in the k-th part of the data and denote the number of observations in this set by n_k. The cross-fit estimate of θ_a is $\theta^{\text{cf}}_{n,a} = \frac{1}{K}\sum_{k=1}^{K}\left[\frac{1}{n_k}\sum_{i\in\mathcal{I}_k}\xi_{n,k,a,\eta|X}(X_i) + \frac{1}{n_k}\sum_{i\in\mathcal{I}_k} D_{a,n,k}(O_i)\right]$.

Simultaneous inference for associations

When making inferences with multiple Y values simultaneously, we control the family-wise error rate for multiple tests by conducting hypothesis testing using simultaneous confidence bands (Ruppert et al., 2003). We illustrate the procedure with testing associations between ADHD and brain connectivity across hundreds of brain regions. Let $j = 1, \ldots, J$ index the region, and $\theta_{a,j}$ denote the motion-controlled average functional connectivity in diagnosis group a between a seed region and region j. Similarly, $D_{P,a,j}$ represents the efficient influence function for diagnosis group a and region j, and by $\tau^2_{n,j}$ denotes the region-specific estimate of the asymptotic variance.

Leveraging the asymptotic linearity of the one-step estimator,

$$\begin{pmatrix} \theta^+_{n,1,1} - \theta^+_{n,0,1} \\ \vdots \\ \theta^+_{n,1,J} - \theta^+_{n,0,J} \end{pmatrix} - \begin{pmatrix} \theta_{1,1} - \theta_{0,1} \\ \vdots \\ \theta_{1,J} - \theta_{0,J} \end{pmatrix} \Rightarrow N\left\{ \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \mathrm{Cov}\begin{pmatrix} D_{P,1,1}(O) - D_{P,0,1}(O) \\ \vdots \\ D_{P,1,J}(O) - D_{P,0,J}(O) \end{pmatrix} \right\}, \tag{7.14}$$

where $\theta^+_{n,a,j}$ is the estimator at $A = a$ at location $j \in \{1, ..., J\}$. An approximate $1-\alpha$ simultaneous confidence interval is $(\theta^+_{n,1,1} - \theta^+_{n,0,1}, \ldots, \theta^+_{n,1,J} - \theta^+_{n,0,J})^\top \pm z_{\max,1-\alpha}(\tau_{n,1}, \ldots, \tau_{n,J})^\top$, where $z_{\max,1-\alpha}$ is the $1-\alpha$ quantile of the random variable $\max_{1\leq j\leq J}\{n^{1/2}|(\theta^+_{n,1,j} - \theta^+_{n,0,j}) - (\theta_{1,j} - \theta_{0,j})|/\tau_{n,j}\}$, which depends on the covariance matrix in (7.14).

To approximate $z_{\max,1-\alpha}$, Monte-Carlo integration is performed by taking 10^5 independent draws of a J-dimensional mean-zero multivariate normal random variable with covariance matrix equal to an empirical estimate of the correlation matrix derived from the covariance matrix on the right-hand side of (7.14). This correlation matrix can be estimated via the empirical correlation of the vector $(D_{n,1,1}(O) - D_{n,0,1}(O), \ldots, D_{n,1,J}(O) - D_{n,0,J}(O))^\top$. For each of the 10^5 random draws, the maximal absolute value of the components of the vector is calculated. The critical value $z_{\max,1-\alpha}$ is approximated by calculating the empirical $(1-\alpha)$-quantile of these maximum values. Wald hypothesis tests controlling family-wise error rate at level α are conducted by rejecting the null hypothesis of no association between diagnosis group and functional connectivity in the j-th region whenever $n^{1/2}|\theta^+_{n,1,j} - \theta^+_{n,0,j}|/\tau_{n,j}$ is larger than the estimated value of $z_{\max,1-\alpha}$.

7.4.3 Additional details of data analysis

fMRIprep data preprocessing for ADHD200

Results included in this manuscript come from preprocessing performed using *fMRIPrep* 21.0.2 (Esteban et al. (2018b); Esteban et al. (2018a); RRID:SCR_016216), which is based on *Nipype* 1.6.1 (Gorgolewski et al. (2011); Gorgolewski et al. (2018); RRID:SCR_002502). The text below is automatically produced by fMRIprep.

Anatomical data preprocessing A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset.The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection` (Tustison et al., 2010), distributed with ANTs 2.3.3 (Avants et al., 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using `fast` (FSL 6.0.5.1:57b01774, RRID:SCR_002823, Zhang et al., 2001). Brain surfaces were reconstructed using `recon-all` (FreeSurfer 6.0.1, RRID:SCR_001847, Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al., 2017). Volume-based spatial normalization to two standard spaces (MNI152NLin2009cAsym, MNI152NLin6Asym) was performed through nonlinear registration with `antsRegistration` (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following templates were selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* [Fonov et al. (2009), RRID:SCR_008796; TemplateFlow

ID: MNI152NLin2009cAsym], *FSL's MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model* [Evans et al. (2012), RRID:SCR_002823; TemplateFlow ID: MNI152NLin6Asym].

Functional data preprocessing For each of the 1 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using `mcflirt` (FSL 6.0.5.1:57b01774, Jenkinson et al., 2002). BOLD runs were slice-time corrected to 1.22s (0.5 of slice acquisition range 0s-2.45s) using `3dTshift` from AFNI (Cox and Hyde, 1997, RRID:SCR_005927). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head-motion. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer) which implements boundary-based registration (Greve and Fischl, 2009). Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al. (2014b)) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al. (2002)). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al., 2014b). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, the aCompCor masks are subtracted a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's *aseg* segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013a). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space.* First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*.

The BOLD time-series were resampled onto the following surfaces (FreeSurfer reconstruction nomenclature): *fsaverage*. *Grayordinates* files (Glasser et al., 2013b) containing 91k samples were also generated using the highest-resolution `fsaverage` as intermediate standardized surface space. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

Many internal operations of *fMRIPrep* use *Nilearn* 0.8.1 (Abraham et al., 2014, RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in *fMRIPrep*'s documentation.

Copyright Waiver

The above boilerplate text was automatically generated by fMRIPrep with the express intention that users should copy and paste this text into their manuscripts *unchanged*. It is released under the CC0 license.

Examples of failed preprocessing in fMRIPrep

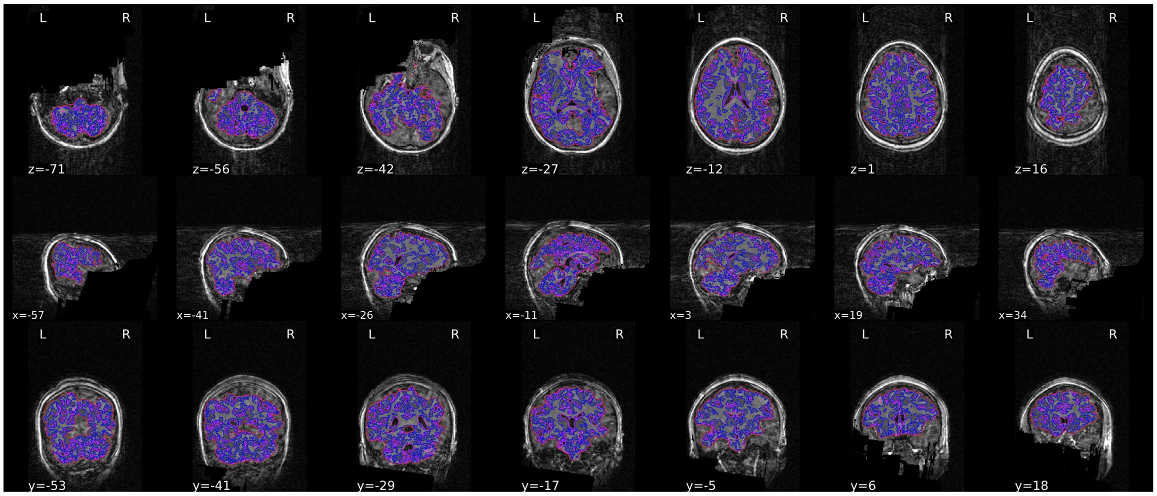


Figure 7.13: An example participant whose cortical segmentation failed in fMRIPrep. The template T1-weighted image with contours delineating the detected brain mask and brain tissue segmentations is shown. The top, middle and bottom rows indicate that large part of brain, including large part of frontal lobe, some part of temporal lobe and occipital lobe, were incorrectly excluded from the segementation.

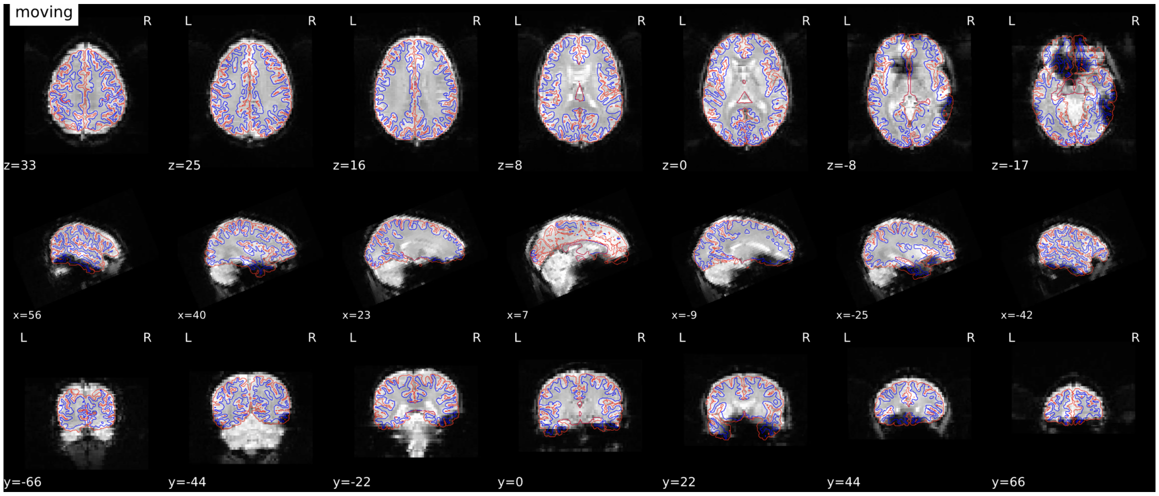


Figure 7.14: An example participant whose fMRI alignment issue in fMRIPrep. The alignment of functional and anatomical MRI data is shown and Freesurfer boundary-based registration is used to generate transformation from EPI-space to T1w-space. It is apparent from the top, middle and bottom rows that some parts of the brain tissue, including parts of the frontal, parietal and temporal lobe, were missing from functional data.

MoCo result using 36p and 36p spike regression

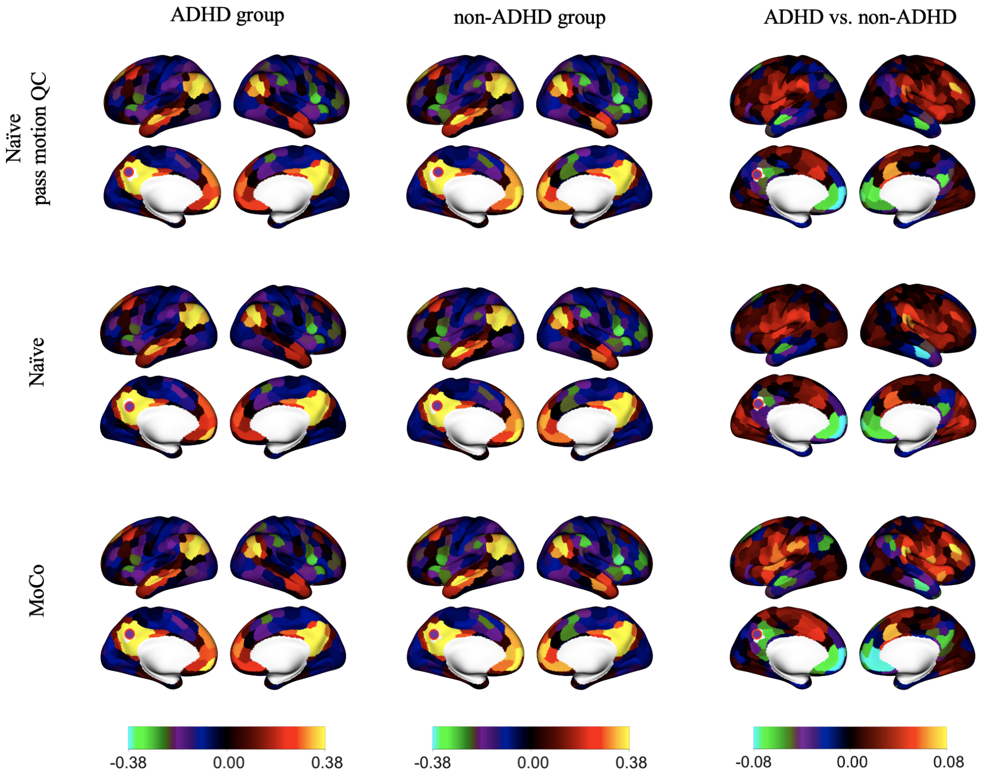


Figure 7.15: Estimated functional connectivity using the naïve approach excluding high-motion participants, the naïve approach, and MoCo for a seed region in the posterior default mode network (fuchsia point). Data are preprocessed using 36-parameter regression.

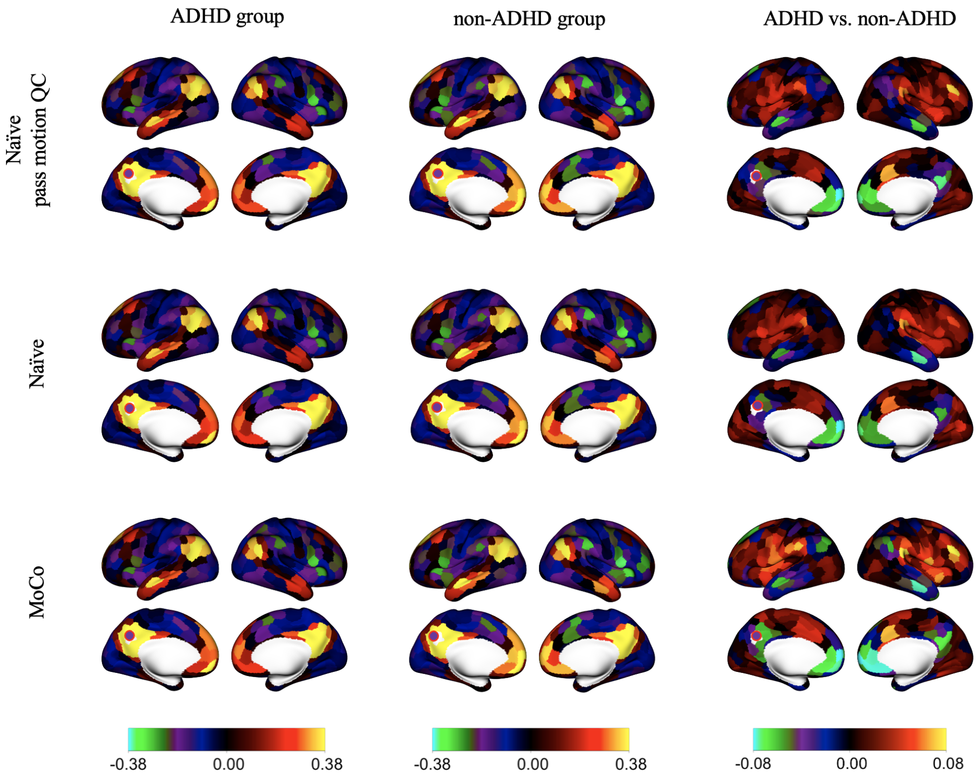


Figure 7.16: Estimated functional connectivity using the naïve approach excluding high-motion participants, the naïve approach, and MoCo for a seed region in the posterior default mode network (fuchsia point). Data are preprocessed using 36-parameter regression with spike regression.

Figure 7.17: Z-statistics for the group difference (ADHD − non-ADHD) for a seed in the posterior default mode network (fuchsia point) in the ADHD200 dataset. The "naive approach" includes participants that pass the preprocessing pipeline ($\Delta_Y = 1$). The naive approach that pass motion QC are participants that also pass motion QC ($\Delta_Y = 1$ and $\Delta_M = 1$). MoCo uses imaging data from all participants that pass preprocessing (including those with excessive motion) and uses demographic data from all participants (including those that failed preprocessing). Both naive approaches appear to generate spurious findings, suggesting extensive anterior-posterior DMN hypoconnectivity. Based on the results from MoCo, most of these regions do not differ in hypoconnectivity when standardizing motion between groups. Network labels from (Schaefer et al., 2018). Data are preprocessed using 36-parameter regression.

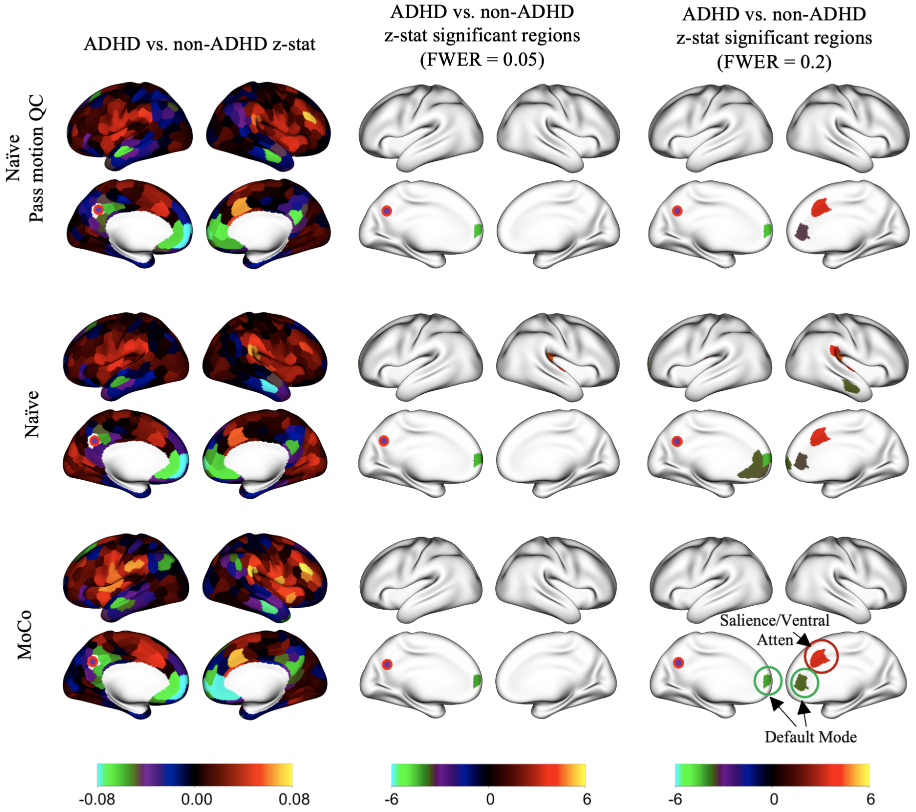
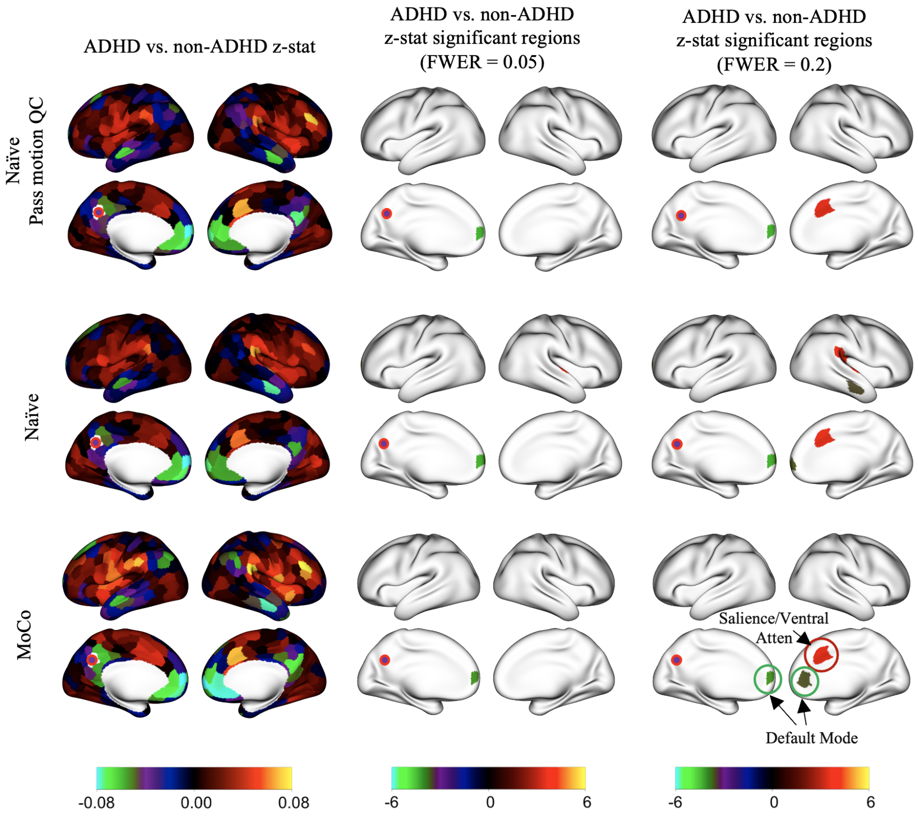


Figure 7.18: Z-statistics for the group difference (ADHD − non-ADHD) for a seed in the posterior default mode network (fuchsia point) in the ADHD200 dataset. The "naive approach" includes participants that pass the preprocessing pipeline ($\Delta_Y = 1$). The naive approach that pass motion QC are participants that also pass motion QC ($\Delta_Y = 1$ and $\Delta_M = 1$). MoCo uses imaging data from all participants that pass preprocessing (including those with excessive motion) and uses demographic data from all participants (including those that failed preprocessing). Both naive approaches appear to generate spurious findings, suggesting extensive anterior-posterior DMN hypoconnectivity. Based on the results from MoCo, most of these regions do not differ in hypoconnectivity when standardizing motion between groups. Network labels from (Schaefer et al., 2018). Data are preprocessed using 36-parameter regression with spike regression.



Bibliography

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., and Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 8.

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., and Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex*, 24(3):663–676.

Amico, E., Marinazzo, D., Di Perri, C., Heine, L., Annen, J., Martial, C., Dzemidzic, M., Kirsch, M., Bonhomme, V., Laureys, S., and Joaquín, G. (2017). Mapping the functional connectome traits of levels of consciousness. *Neuroimage*, 148:201–211.

Andersson, J. L., Graham, M. S., Zsoldos, E., and Sotiropoulos, S. N. (2016). Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion mr images. *Neuroimage*, 141:556–572.

Andersson, J. L. and Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion mr imaging. *Neuroimage*, 125:1063–1078.

Aoki, Y., Cortese, S., and Castellanos, F. X. (2018). Research review: Diffusion tensor imaging studies of attention-deficit/hyperactivity disorder: meta-analyses and reflections on head motion. *Journal of Child Psychology and Psychiatry*, 59(3):193–202.

Arnett, A. B., Pennington, B. F., Willcutt, E. G., DeFries, J. C., and Olson, R. K. (2015). Sex differences in adhd symptom severity. *Journal of Child Psychology and Psychiatry*, 56(6):632–639.

Avants, B., Epstein, C., Grossman, M., and Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41.

Backhausen, L. L., Herting, M. M., Buse, J., Roessner, V., Smolka, M. N., and Vetter, N. C. (2016). Quality control of structural mri images applied using freesurfer—a hands-on workflow to rate motion artifacts. *Frontiers in neuroscience*, 10:558.

Baker, A. P., Brookes, M. J., Rezek, I. A., Smith, S. M., Behrens, T., Smith, P. J. P., and Woolrich, M. (2014). Fast transient networks in spontaneous human brain activity. *Elife*, 3:e01867.

Bastos, A. M. and Schoffelen, J.-M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience*, 9:175.

Baum, G. L., Roalf, D. R., Cook, P. A., Ciric, R., Rosen, A. F., Xia, C., Elliott, M. A., Ruparel, K., Verma, R., Tunç, B., et al. (2018). The impact of in-scanner head motion on structural connectivity derived from diffusion mri. *Neuroimage*, 173:275–286.

Beckmann, C. F., DeLuca, M., Devlin, J. T., and Smith, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):1001–1013.

Beckmann, C. F. and Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE transactions on medical imaging*, 23(2):137–152.

Behzadi, Y., Restom, K., Liau, J., and Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fmri. *NeuroImage*, 37(1):90–101.

Benkeser, D. and van der Laan, M. (2016). The highly adaptive lasso estimator. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 689–696. IEEE.

Best, J. R. and Miller, P. H. (2010). A developmental perspective on executive function. *Child development*, 81(6):1641–1660.

Betzel, R. F., Satterthwaite, T. D., Gold, J. I., and Bassett, D. S. (2017). Positive affect, surprise, and fatigue are correlates of network flexibility. *Scientific reports*, 7(1):1–10.

Beuriat, P.-A., Cohen-Zimerman, S., Smith, G. N., Krueger, F., Gordon, B., and Grafman, J. (2020). A new insight on the role of the cerebellum for executive functions and emotion processing in adults. *Frontiers in Neurology*, 11:593490.

Bickel, P. J., Klaassen, C. A., Ritov, Y., and Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models*, volume 4. Springer.

Birn, R. M. (2012). The role of physiological noise in resting-state functional connectivity. *Neuroimage*, 62(2):864–870.

Birn, R. M., Cornejo, M. D., Molloy, E. K., Patriat, R., Meier, T. B., Kirk, G. R., Nair, V. A., Meyerand, M. E., and Prabhakaran, V. (2014). The influence of physiological noise correction on test–retest reliability of resting-state functional connectivity. *Brain connectivity*, 4(7):511–522.

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

Brown, M. R., Sidhu, G. S., Greiner, R., Asgarian, N., Bastani, M., Silverstone, P. H., Greenshaw, A. J., and Dursun, S. M. (2012). Adhd-200 global competition: diagnosing adhd using personal characteristic data can outperform resting state fmri measurements. *Frontiers in systems neuroscience*, 6:69.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186–198.

Cai, B., Zille, P., Stephen, J. M., Wilson, T. W., Calhoun, V. D., and Wang, Y. P. (2017). Estimation of dynamic sparse connectivity patterns from resting state fmri. *IEEE transactions on medical imaging*, 37(5):1224–1234.

Calhoun, V., Adali, T., Pearlson, G., and Pekar, J. (2001). A method for making group inferences from functional mri data using independent component analysis. *Human brain mapping*, 14(3):140–151.

Castellanos, F. X. and Aoki, Y. (2016). Intrinsic functional connectivity in attention-deficit/hyperactivity disorder: a science in development. *Biological psychiatry: cognitive neuroscience and neuroimaging*, 1(3):253–261.

Castellanos, F. X., Margulies, D. S., Kelly, C., Uddin, L. Q., Ghaffari, M., Kirsch, A., Shaw, D., Shehzad, Z., Di Martino, A., Biswal, B., et al. (2008). Cingulate-precuneus interactions: a new locus of dysfunction in adult attention-deficit/hyperactivity disorder. *Biological psychiatry*, 63(3):332–337.

Chai, L. R., Khambhati, A. N., Ciric, R., Moore, T. M., Gur, R. C., Gur, R. E., Satterthwaite, T. D., and Bassett, D. S. (2017). Evolution of brain network dynamics in neurodevelopment. *Network Neuroscience*, 1(1):14–30.

Chang, C. and Glover, G. H. (2010). Time–frequency dynamics of resting-state brain connectivity measured with fmri. *Neuroimage*, 50(1):81–98.

Chen, A. A., Beer, J. C., Tustison, N. J., Cook, P. A., Shinohara, R. T., Shou, H., and Initiative, A. D. N. (2022). Mitigating site effects in covariance for machine learning in neuroimaging data. *Human brain mapping*, 43(4):1179–1195.

Chen, P.-Y., Chiou, J.-M., Yang, Y.-F., Chen, Y.-T., Hsieh, H.-L., Chang, Y.-L., and Tseng, W.-Y. I. (2016). Heterogeneous aging effects on functional connectivity in different cortical regions: A resting-state functional mri study using functional data analysis. *PloS one*, 11(9):e0162028.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Chung, M. K. (2018). Statistical challenges of big brain network data. *Statistics & probability letters*, 136:78–82.

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., et al. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage*, 154:174–187.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Colaço, D. (2024). When remediating one artifact results in another: control, confounders, and correction. *History and Philosophy of the Life Sciences*, 46(1):5.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20:273–297.

Cosgrove, K. T., McDermott, T. J., White, E. J., et al. (2022). Limits to the generalizability of resting-state functional magnetic resonance imaging studies of youth: An examination of abcd study® baseline data. *Brain imaging and behavior*, 16(4):1919–1925.

Cox, R. W. and Hyde, J. S. (1997). Software tools for analysis and visualization of fmri data. *NMR in Biomedicine*, 10(4-5):171–178.

Cribben, I., Haraldsdottir, R., Atlas, L. Y., Wager, T. D., and Lindquist, M. A. (2012). Dynamic connectivity regression: determining state-related changes in brain connectivity. *Neuroimage*, 61(4):907–920.

Cribben, I., Wager, T., and Lindquist, M. (2013). Detecting functional connectivity change points for single-subject fmri data. *Frontiers in computational neuroscience*, 7:143.

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*, 9(2):179–194.

Damaraju, E., Allen, E. A., Belger, A., Ford, J. M., McEwen, S., Mathalon, D., Mueller, B., Pearlson, G., Potkin, S., Preda, A., et al. (2014). Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *NeuroImage: Clinical*, 5:298–308.

De Bie, H. M., Boersma, M., Wattjes, M. P., Adriaanse, S., Vermeulen, R. J., et al. (2010). Preparing children with a mock scanner training protocol results in high quality structural and functional mri scans. *European journal of pediatrics*, 169:1079–1085.

De Lacy, N., Doherty, D., King, B., Rachakonda, S., and Calhoun, V. (2017). Disruption to control network function correlates with altered dynamic connectivity in the wider autism spectrum. *NeuroImage: Clinical*, 15:513–524.

DeCarli, C., Kawas, C., Morrison, J. H., Reuter-Lorenz, P. A., Sperling, R. A., and Wright, C. B. (2012). Session ii: mechanisms of age-related cognitive change and targets for intervention: neural circuits, networks, and plasticity. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 67(7):747–753.

Deco, G., Jirsa, V. K., and McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience*, 12(1):43–56.

Deen, B. and Pelphrey, K. (2012). Perspective: brain scans need a rethink. *Nature*, 491(7422):S20–S20.

Demirtaş, M., Tornador, C., Falcón, C., López-Solà, M., Hernández-Ribas, R., Pujol, J., Menchón, J. M., Ritter, P., Cardoner, N., Soriano-Mas, C., et al. (2016). Dynamic functional connectivity reveals altered variability in functional connectivity among patients with major depressive disorder. *Human brain mapping*, 37(8):2918–2930.

Di Martino, A., O'connor, D., Chen, B., Alaerts, K., Anderson, J. S., Assaf, et al. (2017). Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii. *Scientific data*, 4(1):1–15.

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, et al. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667.

Díaz, I., Hejazi, N. S., Rudolph, K. E., and van Der Laan, M. J. (2021). Nonparametric efficient causal mediation with intermediate confounders. *Biometrika*, 108(3):627–641.

Eavani, H., Satterthwaite, T. D., Gur, R. E., Gur, R. C., and Davatzikos, C. (2013). Unsupervised learning of functional network dynamics in resting state fmri. In *International conference on information processing in medical imaging*, pages 426–437. Springer.

Ekman, M., Derrfuss, J., Tittgemeyer, M., and Fiebach, C. J. (2012). Predicting errors from reconfiguration patterns in human brain networks. *Proceedings of the National Academy of Sciences*, 109(41):16714–16719.

Epstein, J. N., Casey, B., Tonev, S. T., Davidson, M., Reiss, A. L., Garrett, A., Hinshaw, S. P., Greenhill, L. L., Vitolo, A., Kotler, L. A., et al. (2007). Assessment and prevention of head motion during imaging of patients with attention deficit hyperactivity disorder. *Psychiatry Research: Neuroimaging*, 155(1):75–82.

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, James D. andGoncalves, M., DuPre, E., Sitek, K. R., Gomez, D. E. P., Lurie, D. J., Ye, Z., Poldrack, R. A., and Gorgolewski, K. J. (2018a). fmriprep. *Software*.

Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Erramuzpe Aliaga, A., Kent, J., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S., Wright, J., Durnez, J., Poldrack, R., and Gorgolewski, K. J. (2018b). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A., et al. (2019). fmriprep: a robust preprocessing pipeline for functional mri. *Nature methods*, 16(1):111–116.

Evans, A., Janke, A., Collins, D., and Baillet, S. (2012). Brain templates and atlases. *NeuroImage*, 62(2):911–922.

Fair, D. A., Cohen, A. L., Dosenbach, N. U., et al. (2008). The maturing architecture of the brain's default network. *Proceedings of the National Academy of Sciences*, 105(10):4028–4032.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360.

Feczko, E., Conan, G., Marek, S., Tervo-Clemmens, B., Cordova, M., Doyle, O., Earl, E., Perrone, A., Sturgeon, D., Klein, R., et al. (2021). Adolescent brain cognitive development (abcd) community mri collection and utilities. *BioRxiv*, pages 2021–07.

Fonov, V., Evans, A., McKinstry, R., Almli, C., and Collins, D. (2009). Unbiased non-linear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47, Supplement 1:S102.

Fox, M. D. and Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature reviews neuroscience*, 8(9):700–711.

Fox, M. D., Zhang, D., Snyder, A. Z., and Raichle, M. E. (2009). The global signal and observed anticorrelated resting state brain networks. *Journal of neurophysiology*, 101(6):3270–3283.

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain connectivity*, 1(1):13–36.

Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. J. (1993). Functional connectivity: the principal-component analysis of large (PET) data sets. *Journal of cerebral blood flow and metabolism*, 13:5.

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S., and Turner, R. (1996). Movement-related effects in fmri time-series. *Magnetic resonance in medicine*, 35(3):346–355.

Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R., Heeringa, S., Jernigan, T., Potter, A., Thompson, W., and Zahs, D. (2018). Recruiting the abcd sample: Design considerations and procedures. *Developmental cognitive neuroscience*, 32:16–22.

Geng, H., Xu, P., Sommer, I. E., Luo, Y.-J., Aleman, A., and Ćurčić-Blake, B. (2020). Abnormal dynamic resting-state brain network organization in auditory verbal hallucination. *Brain Structure and Function*, 225:2315–2330.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., et al. (2013a). The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., and Jenkinson, M. (2013b). The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80:105–124.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., and Ghosh, S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, 5:13.

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., Jarecka, D., Johnson, H., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault, B., Salo, T., Jordan, K., Goncalves, M., Waskom, M., Clark, D., Wong, J., Loney, F., Modat, M., Dewey, B. E., Madison, C., Visconti di Oleggio Castello, M., Clark, M. G., Dayan, M., Clark, D., Keshavan, A., Pinsard, B., Gramfort, A., Berleant, S., Nielson, D. M., Bougacha, S., Varoquaux, G., Cipollini, B., Markello, R., Rokem, A., Moloney, B., Halchenko, Y. O., Wassermann, D., Hanke, M., Horea, C., Kaczmarzyk, J., de Hollander, G., DuPre, E., Gillman, A., Mordom, D., Buchanan, C., Tungaraza, R., Pauli, W. M., Iqbal, S., Sikka, S., Mancini, M., Schwartz, Y., Malone, I. B., Dubois, M., Frohlich, C., Welch, D., Forbes, J., Kent, J., Watanabe, A., Cumba, C., Huntenburg, J. M., Kastman, E., Nichols, B. N., Eshaghi, A., Ginsburg, D., Schaefer, A., Acland, B., Giavasis, S., Kleesiek, J., Erickson, D., Küttner, R., Haselgrove, C., Correa, C., Ghayoor, A., Liem, F., Millman, J., Haehn, D., Lai, J., Zhou, D., Blair, R., Glatard, T., Renfro, M., Liu, S., Kahn, A. E., Pérez-García, F., Triplett, W., Lampe, L., Stadler, J., Kong, X.-Z., Hallquist, M., Chetverikov, A., Salvatore, J., Park, A., Poldrack, R., Craddock, R. C., Inati, S., Hinds, O., Cooper, G., Perkins, L. N., Marina, A., Mattfeld, A., Noel, M., Snoek, L., Matsubara, K., Cheung, B., Rothmei, S., Urchs, S., Durnez, J., Mertz, F., Geisler, D., Floren, A., Gerhard, S., Sharp, P., Molina-Romero, M., Weinstein, A., Broderick, W., Saase, V., Andberg, S. K., Harms, R., Schlamp, K., Arias, J., Papadopoulos Orfanos, D., Tarbert, C., Tambini, A., De La Vega, A., Nickson, T., Brett, M., Falkiewicz, M., Podranski, K., Linkersdörfer, J., Flandin, G., Ort, E., Shachnev, D., McNamee, D., Davison, A., Varada, J., Schwabacher, I., Pellman, J., Perez-Guevara, M., Khanuja, R., Pannetier, N., McDermottroe, C., and Ghosh, S. (2018). Nipype. *Software*.

Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the national academy of sciences*, 100(1):253–258.

Greve, D. N. and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1):63–72.

Hagler Jr, D. J., Hatton, S., Cornejo, M. D., Makowski, C., Fair, D. A., Dick, A. S., Sutherland, M. T., Casey, B., Barch, D. M., Harms, M. P., et al. (2019). Image processing and analysis methods for the adolescent brain cognitive development study. *Neuroimage*, 202:116091.

Hejazi, N. S., Benkeser, D. C., and van der Laan, M. J. (2022a). haldensify: Highly adaptive lasso conditional density estimation. `https://github.com/nhejazi/haldensify`. R package version 0.2.5.

Hejazi, N. S., van der Laan, M. J., and Benkeser, D. (2022b). haldensify: Highly adaptive lasso conditional density estimation inr. *Journal of Open Source Software*, 7(77):4522.

Hlinka, J., Paluš, M., Vejmelka, M., Mantini, D., and Corbetta, M. (2011). Functional connectivity in resting-state fmri: is linear correlation sufficient? *Neuroimage*, 54(3):2218–2225.

Hoff, G. A.-J., Van den Heuvel, M., Benders, M. J., Kersbergen, K. J., and De Vries, L. S. (2013). On development of functional brain connectivity in the young brain. *Frontiers in human neuroscience*, 7:650.

Huang, S., Li, J., Sun, L., Ye, J., Fleisher, A., Wu, T., Chen, K., Reiman, E., Initiative, A. D. N., et al. (2010). Learning brain connectivity of alzheimer's disease by sparse inverse covariance estimation. *NeuroImage*, 50(3):935–949.

Hull, J. V., Dokovna, L. B., Jacokes, Z. J., et al. (2017). Resting-state functional connectivity in autism spectrum disorders: a review. *Frontiers in psychiatry*, 7:205.

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Della Penna, S., Duyn, J. H., Glover, G. H., Gonzalez-Castillo, J., et al. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, 80:360–378.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent component analysis*, volume 46. John Wiley & Sons.

Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430.

Iannilli, E., Broy, F., Kunz, S., and Hummel, T. (2017). Age-related changes of gustatory function depend on alteration of neuronal circuits. *Journal of neuroscience research*, 95(10):1927–1936.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2):825–841.

Kang, J., Wang, L., Yan, C., Wang, J., Liang, X., and He, Y. (2011). Characterizing dynamic functional connectivity in the resting brain using variable parameter regression and kalman filtering approaches. *Neuroimage*, 56(3):1222–1234.

Karahanoğlu, F. I. and Van De Ville, D. (2015). Transient brain activity disentangles fmri resting-state dynamics in terms of spatially and temporally overlapping networks. *Nature communications*, 6(1):7751.

Keeratimahat, K. and Nichols, T. E. (2022). Discussion on "distributional independent component analysis for diverse neuroimaging modalities" by ben wu, subhadip pal, jian kang, and ying guo. *Biometrics*, 78(3):1113–1117.

Kemmer, P. B., Wang, Y., Bowman, F. D., Mayberg, H., and Guo, Y. (2018). Evaluating the strength of structural connectivity underlying brain functional networks. *Brain Connectivity*, 8(10):579–594.

Kiviniemi, V., Vire, T., Remes, J., Elseoud, A. A., Starck, T., Tervonen, O., and

Nikkinen, J. (2011). A sliding time-window ica reveals spatial variability of the default mode network in time. *Brain connectivity*, 1(4):339–347.

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., and Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, 13(2):e1005350.

Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, 1(1):76–85.

Lang, E. W., Tomé, A. M., Keck, I. R., Górriz-Sáez, J., and Puntonet, C. G. (2012). Brain connectivity analysis: a short survey. *Computational intelligence and neuroscience*, 2012(1):412512.

Lee, H., Lee, D. S., Kang, H., Kim, B.-N., and Chung, M. K. (2011). Sparse brain network recovery under compressed sensing. *IEEE Transactions on Medical Imaging*, 30(5):1154–1165.

Leonardi, N., Richiardi, J., Gschwind, M., Simioni, S., Annoni, J.-M., Schluep, M., Vuilleumier, P., and Van De Ville, D. (2013). Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest. *NeuroImage*, 83:937–950.

Leonardi, N., Shirer, W. R., Greicius, M. D., and Van De Ville, D. (2014). Disentangling dynamic networks: Separated and joint expressions of functional connectivity patterns in time. *Human brain mapping*, 35(12):5984–5995.

Levy, J. (2019). Tutorial: Deriving the efficient influence curve for large models. *arXiv preprint arXiv:1903.01706*.

Lindquist, M. A., Xu, Y., Nebel, M. B., and Caffo, B. S. (2014). Evaluating dynamic bivariate correlations in resting-state fmri: a comparison study and a new approach. *NeuroImage*, 101:531–546.

Lv, H., Wang, Z., Tong, E., Williams, L. M., Zaharchuk, G., Zeineh, M., Goldstein-Piekarski, A. N., Ball, T. M., Liao, C., and Wintermark, M. (2018). Resting-state functional mri: everything that nonexperts have always wanted to know. *American Journal of Neuroradiology*, 39(8):1390–1399.

Maenner, M. J. (2023). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2020. *MMWR. Surveillance Summaries*, 72.

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2009). Online dictionary learning for sparse coding. In *Proceedings of the 26th annual international conference on machine learning*, pages 689–696.

Maknojia, S., Churchill, N. W., Schweizer, T. A., and Graham, S. (2019). Resting state fmri: Going through the motions. *Frontiers in neuroscience*, 13:462471.

Marek, S., Tervo-Clemmens, B., Calabro, F. J., et al. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, 603(7902):654–660.

Marusak, H. A., Calhoun, V. D., Brown, S., Crespo, L. M., Sala-Hamrick, K., Gotlib, I. H., and Thomason, M. E. (2017). Dynamic functional connectivity of neurocognitive networks in children. *Human brain mapping*, 38(1):97–108.

Mayberg, H. S. (2003). Modulating dysfunctional limbic-cortical circuits in depression: towards development of brain-based algorithms for diagnosis and optimised treatment. *British medical bulletin*, 65(1):193–207.

Miller, R. L., Yaesoubi, M., Turner, J. A., Mathalon, D., Preda, A., Pearlson, G., Adali, T., and Calhoun, V. D. (2016). Higher dimensional meta-state analysis reveals reduced resting fmri connectivity dynamism in schizophrenia patients. *PloS one*, 11(3):e0149849.

Minka, T. (2000). Automatic choice of dimensionality for pca. *Advances in neural information processing systems*, 13.

Monti, R. P., Lorenz, R., Braga, R. M., Anagnostopoulos, C., Leech, R., and Montana, G. (2017). Real-time estimation of dynamic functional connectivity networks. *Human brain mapping*, 38(1):202–220.

Moore, D. S., McCabe, G. P., and Craig, B. A. (2018). *Introduction to the Practice of Statistics*. W. H. Freeman.

Mummaneni, A., Kardan, O., Stier, A. J., Chamberlain, T. A., Chao, A. F., Berman, M. G., and Rosenberg, M. D. (2023). Functional brain connectivity predicts sleep duration in youth and adults. *Human Brain Mapping*, 44(18):6293–6307.

Muñoz, I. D. and Van Der Laan, M. (2012). Population intervention causal effects based on stochastic interventions. *Biometrics*, 68(2):541–549.

Nebel, M. B., Lidstone, D. E., Wang, L., et al. (2022). Accounting for motion in resting-state fmri: What part of the spectrum are we characterizing in autism spectrum disorder? *NeuroImage*, 257:119296.

Nielsen, A. N., Greene, D. J., Gratton, C., Dosenbach, N. U., Petersen, S. E., and Schlaggar, B. L. (2019). Evaluating the prediction of brain maturity from functional connectivity after motion artifact denoising. *Cerebral Cortex*, 29(6):2455–2469.

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological methods*, 19(4):459.

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., Jack Jr, C., Jagust, W. J., Shaw, L. M., Toga, A. W., et al. (2010). Alzheimer's disease neuroimaging initiative (adni) clinical characterization. *Neurology*, 74(3):201–209.

Pham, D. D., Muschelli, J., and Mejia, A. F. (2022). ciftitools: A package for reading, writing, visualizing, and manipulating cifti files in r. *NeuroImage*, 250:118877.

Polanczyk, G. V., Willcutt, E. G., Salum, G. A., Kieling, C., and Rohde, L. A. (2014). Adhd prevalence estimates across three decades: an updated systematic review and meta-regression analysis. *International journal of epidemiology*, 43(2):434–442.

Polley, E., LeDell, E., Kennedy, C., and van der Laan, M. (2023). *SuperLearner: Super Learner Prediction.* R package version 2.0-28.1.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. *Neuroimage*, 59(3):2142–2154.

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., et al. (2011). Functional network organization of the human brain. *Neuron*, 72(4):665–678.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2014a). Methods to detect, characterize, and remove motion artifact in resting state fmri. *Neuroimage*, 84:320–341.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2014b). Methods to detect, characterize, and remove motion artifact in resting state fmri. *NeuroImage*, 84(Supplement C):320–341.

Rashid, B., Damaraju, E., Pearlson, G. D., and Calhoun, V. D. (2014). Dynamic connectivity states estimated from resting fmri identify differences among schizophrenia, bipolar disorder, and healthy control subjects. *Frontiers in human neuroscience*, 8:897.

Ren, P., Lo, R. Y., Chapman, B. P., Mapstone, M., Porsteinsson, A., Lin, F., Initiative, A. D. N., et al. (2016). Longitudinal alteration of intrinsic brain activity in the striatum in mild cognitive impairment. *Journal of Alzheimer's disease*, 54(1):69–78.

Reuter, M., Tisdall, M. D., Qureshi, A., Buckner, R. L., van der Kouwe, A. J., and Fischl, B. (2015). Head motion during mri acquisition reduces gray matter volume and thickness estimates. *Neuroimage*, 107:107–115.

Richardson, T. S. and Robins, J. M. (2013). Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 128(30):2013.

Rubinov, M. and Sporns, O. (2011). Weight-conserving characterization of complex functional brain networks. *Neuroimage*, 56(4):2068–2079.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression*. Number 12. Cambridge university press.

Sakoğlu, Ü., Pearlson, G. D., Kiehl, K. A., Wang, Y. M., Michael, A. M., and Calhoun, V. D. (2010). A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 23(5):351–366.

Satterthwaite, T. D., Connolly, J. J., Ruparel, K., Calkins, M. E., Jackson, C., Elliott, M. A., Roalf, D. R., Hopson, R., Prabhakaran, K., Behr, M., et al. (2016). The philadelphia neurodevelopmental cohort: A publicly available resource for the study of normal and abnormal brain development in youth. *Neuroimage*, 124:1115–1119.

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., and Wolf, D. H. (2013a). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, 64(1):240–256.

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., et al. (2013b). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage*, 64:240–256.

Satterthwaite, T. D., Elliott, M. A., Ruparel, K., Loughead, J., Prabhakaran, K., Calkins, M. E., Hopson, R., Jackson, C., Keefe, J., Riley, M., et al. (2014). Neuroimaging of the philadelphia neurodevelopmental cohort. *Neuroimage*, 86:544–553.

Satterthwaite, T. D., Wolf, D. H., Roalf, D. R., Ruparel, K., Erus, G., Vandekar, S., Gennatas, E. D., Elliott, M. A., Smith, A., Hakonarson, H., et al. (2015). Linked sex differences in cognition and functional connectivity in youth. *Cerebral cortex*, 25(9):2383–2394.

Satterthwaite, T. D., Wolf, D. H., Ruparel, K., Erus, G., Elliott, M. A., Eickhoff, S. B., Gennatas, E. D., Jackson, C., Prabhakaran, K., Smith, A., et al. (2013c). Heterogeneous impact of motion on fundamental patterns of developmental changes in functional connectivity during youth. *Neuroimage*, 83:45–57.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., and Yeo, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral cortex*, 28(9):3095–3114.

Schmahmann, J. D. (2004). Disorders of the cerebellum: ataxia, dysmetria of thought, and the cerebellar cognitive affective syndrome. *The Journal of neuropsychiatry and clinical neurosciences*, 16(3):367–378.

Seeley, W. W., Crawford, R. K., Zhou, J., Miller, B. L., and Greicius, M. D. (2009). Neurodegenerative diseases target large-scale human brain networks. *Neuron*, 62(1):42–52.

Shanmugan, S., Seidlitz, J., Cui, Z., Adebimpe, A., Bassett, D. S., et al. (2022). Sex differences in the functional topography of association networks in youth. *Proceedings of the National Academy of Sciences*, 119(33):e2110416119.

Shi, R. and Guo, Y. (2016). Investigating differences in brain functional networks using hierarchical covariate-adjusted independent component analysis. *The annals of applied statistics*, 10(4):1930.

Shine, J. M., Koyejo, O., Bell, P. T., Gorgolewski, K. J., Gilat, M., and Poldrack, R. A. (2015). Estimation of dynamic functional connectivity using multiplication of temporal derivatives. *NeuroImage*, 122:399–407.

Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., and Greicius, M. D. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral cortex*, 22(1):158–165.

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., and Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the national academy of sciences*, 106(31):13040–13045.

Sobel, M. E. and Lindquist, M. A. (2014). Causal inference for fmri time series data with systematic errors of measurement in a balanced on/off study of social evaluative threat. *Journal of the American Statistical Association*, 109(507):967–976.

Solo, V., Poline, J.-B., Lindquist, M. A., Simpson, S. L., Bowman, F. D., Chung, M. K., and Cassidy, B. (2018). Connectivity in fmri: blind spots and breakthroughs. *IEEE transactions on medical imaging*, 37(7):1537–1550.

Sonuga-Barke, E. J. and Castellanos, F. X. (2007). Spontaneous attentional fluctuations in impaired states and pathological conditions: a neurobiological hypothesis. *Neuroscience & Biobehavioral Reviews*, 31(7):977–986.

Sorg, C., Riedl, V., Mühlau, M., Calhoun, V. D., Eichele, T., Läer, L., Drzezga, A., Förstl, H., Kurz, A., Zimmer, C., and Wohlschläger, A. M. (2007). Selective changes of resting-state networks in individuals at risk for alzheimer's disease. *Proceedings of the National Academy of Sciences*, 104(47):18760–18765.

Srinivasan, R., Winter, W. R., Ding, J., and Nunez, P. L. (2007). Eeg and meg coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics. *Journal of neuroscience methods*, 166(1):41–52.

Thompson, W. H., Brantefors, P., and Fransson, P. (2017a). From static to temporal network theory: Applications to functional brain connectivity. *Network Neuroscience*, 1(2):69–99.

Thompson, W. H., Richter, C. G., Plavén-Sigray, P., and Fransson, P. (2017b). A simulation and comparison of dynamic functional connectivity methods. *bioRxiv*, page 212241.

Thorndike, R. (1953). Who belongs in the family? *Psychometrika*, 18(4):267–276.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C. (2010). N4itk: Improved n3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289.

Uddin, L. Q., Clare Kelly, A., Biswal, B. B., Xavier Castellanos, F., and Milham, M. P. (2009). Functional connectivity of default mode network components: correlation, anticorrelation, and causality. *Human brain mapping*, 30(2):625–637.

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

Van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007). Super learner. *Statistical applications in genetics and molecular biology*, 6(1).

Van der Laan, M. J., Rose, S., et al. (2011). *Targeted learning: causal inference for observational and experimental data*, volume 4. Springer.

van der Laan, M. J., Rose, S., Zheng, W., and van der Laan, M. J. (2011). Cross-validated targeted minimum-loss-based estimation. *Targeted learning: causal inference for observational and experimental data*, pages 459–474.

Van Der Vaart, A. W., Wellner, J. A., van der Vaart, A. W., and Wellner, J. A. (1996). *Weak convergence*. Springer.

Van Dijk, K. R., Sabuncu, M. R., and Buckner, R. L. (2012). The influence of head motion on intrinsic functional connectivity mri. *Neuroimage*, 59(1):431–438.

Volkow, N. D., Koob, G. F., Croyle, R. T., Bianchi, D. W., Gordon, J. A., Koroshetz, W. J., Pérez-Stable, E. J., Riley, W. T., Bloch, M. H., Conway, K., et al. (2018).

The conception of the abcd study: From substance use to a broad nih collaboration. *Developmental cognitive neuroscience*, 32:4–7.

Wang, Y. and Guo, Y. (2019). A hierarchical independent component analysis model for longitudinal neuroimaging studies. *NeuroImage*, 189:380–400.

Wang, Y. and Guo, Y. (2023). Locus: A regularized blind source separation method with low-rank structure for investigating brain connectivity. *The Annals of Applied Statistics*, 17(2):1307–1332.

Wang, Y., Kang, J., Kemmer, P. B., and Guo, Y. (2016). An efficient and reliable statistical method for estimating functional connectivity in large scale brain networks using partial correlation. *Frontiers in neuroscience*, 10:123.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack Jr, C. R., Jagust, W., Morris, J. C., et al. (2017). The alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement. *Alzheimer's & Dementia*, 13(5):561–571.

Willcutt, E. G. (2012). The prevalence of dsm-iv attention-deficit/hyperactivity disorder: a meta-analytic review. *Neurotherapeutics*, 9(3):490–499.

Williams, L. M. (2016). Precision psychiatry: a neural circuit taxonomy for depression and anxiety. *The Lancet Psychiatry*, 3(5):472–480.

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52.

Wu, B., Pal, S., Kang, J., and Guo, Y. (2022). Rejoinder to discussions of "distributional independent component analysis for diverse neuroimaging modalities". *Biometrics*, 78(3):1122–1126.

Wu, G.-R., Stramaglia, S., Chen, H., Liao, W., and Marinazzo, D. (2013). Mapping the voxel-wise effective connectome in resting state fmri. *PloS one*, 8(9):e73670.

Yaesoubi, M., Allen, E. A., Miller, R. L., and Calhoun, V. D. (2015). Dynamic coherence analysis of resting fmri data to jointly capture state-based phase, frequency, and time-domain information. *Neuroimage*, 120:133–142.

Yan, C.-G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R. C., Di Martino, A., Li, Q., Zuo, X.-N., Castellanos, F. X., and Milham, M. P. (2013). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *Neuroimage*, 76:183–201.

Yang, Z., Craddock, R. C., Margulies, D. S., Yan, C.-G., and Milham, M. P. (2014). Common intrinsic connectivity states among posteromedial cortex subdivisions: Insights from analysis of temporal dynamics. *Neuroimage*, 93:124–137.

Yendiki, A., Koldewyn, K., Kakunoori, S., Kanwisher, N., and Fischl, B. (2014). Spurious group differences due to head motion in a diffusion mri study. *Neuroimage*, 88:79–90.

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*.

Yerys, B. E., Gordon, E. M., Abrams, D. N., Satterthwaite, T. D., Weinblatt, R., et al. (2015). Default mode network segregation and social deficits in autism spectrum disorder: Evidence from non-medicated children. *NeuroImage: Clinical*, 9:223–232.

Yu, M., Linn, K. A., Cook, P. A., et al. (2018). Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fmri data. *Human brain mapping*, 39(11):4213–4227.

Zhang, L., Fu, Z., Zhang, W., Huang, G., Liang, Z., Li, L., Biswal, B. B., Calhoun, V. D., and Zhang, Z. (2021). Accessing dynamic functional connectivity using l0-regularized sparse-smooth inverse covariance estimation from fmri. *Neurocomputing*, 443:147–161.

Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57.

Zhou, H., Li, L., and Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502):540–552.

Zivich, P. N. and Breskin, A. (2021). Machine learning for causal inference: on the use of cross-fit estimators. *Epidemiology (Cambridge, Mass.)*, 32(3):393.

Zuo, X.-N., Kelly, C., Adelstein, J. S., Klein, D. F., Castellanos, F. X., and Milham, M. P. (2010). Reliable intrinsic connectivity networks: test–retest evaluation using ica and dual regression approach. *Neuroimage*, 49(3):2163–2177.