

**Correction, Validation, and Characterization of Motion  
in Resting-State Functional Magnetic Resonance  
Images of Pediatric Patients**

by

**Jenna Marie Schabdach**

B. S. Electrical Engineering, Drexel University, 2016

M. S. Electrical Engineering, Drexel University, 2016

M. S. Biomedical Informatics, University of Pittsburgh, 2018

Submitted to the Graduate Faculty of

School of Medicine Department of Biomedical Informatics in partial

fulfillment

of the requirements for the degree of

**Doctor of Philosophy**

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH  
SCHOOL OF MEDICINE

This dissertation was presented

by

Jenna Marie Schabdach

It was defended on

March 31, 2020

and approved by

Dr. Douglas Landsittel, Department of Biomedical Informatics

Dr. Ashok Panigrahy, Department of Biomedical Informatics

Dr. Gregory Cooper, Department of Biomedical Informatics

Dr. Rafael Ceschin, Department of Pediatric Radiology, Children's Hospital of Pittsburgh

of UPMC

Dissertation Director: Dr. Douglas Landsittel, Department of Biomedical Informatics

Copyright © by Jenna Marie Schabdach

2020

# **Correction, Validation, and Characterization of Motion in Resting-State Functional Magnetic Resonance Images of Pediatric Patients**

Jenna Marie Schabdach, PhD

University of Pittsburgh, 2020

Approximately 1.35 million children are diagnosed with a congenital heart defect (CHD) annually. As the process of diagnosing and treating CHD has improved, the expected lifespan of CHD patients has increased: at least 12 to 34 million adults worldwide have CHD. A larger population living with CHD means that clinicians are learning more about conditions which CHD patients are at increased risk of getting. Some of these conditions affect the patient's neurodevelopment and neurocognitive status. Diagnosis of the neurocognitive conditions is performed using surveys, but examining the structure and function of the brain could offer more objective diagnoses.

One tool useful for examining a patient's brain is magnetic resonance imaging (MRI). Resting state functional MRI (rs-fMRI) in particular can be used to reveal information about the neuronal networks in the patient's brain. Unfortunately, rs-fMRIs are highly susceptible to motion. Clinical and behavioral techniques can help patients move less during a rs-fMRI scan, though they do not guarantee motion-free images. Various sensors can be used to monitor and correct for motion during the scan, but these sensors are useless when it comes to recovering previously acquired rs-fMRIs corrupted by motion.

We devised a novel approach to volume registration, which is the first step in motion correction. We compare it to traditional volume registration in the context of a complete motion correction pipeline. The registration techniques and pipeline were applied to neurological rs-fMRIs of fetal, neonatal, preadolescent, and adult patients with CHD as well as several phantom images. We identified different motion patterns specific to different demographic groups, and discovered relationships between certain motion patterns and clinical outcomes associated with CHD and different neurodevelopmental conditions.

## Table of Contents

Preface . . . . .	xvii
<b>1.0 Introduction . . . . .</b>	<b>1</b>
1.1 Resting-state Functional Magnetic Resonance Images . . . . .	1
1.2 Motion Effects, Prevention, and Correction . . . . .	2
1.3 Data . . . . .	4
1.4 Experiments . . . . .	5
1.5 Summary . . . . .	6
<b>2.0 rs-fMRIs and Patient Motion . . . . .</b>	<b>8</b>
2.1 Structure of an rs-fMRI . . . . .	8
2.2 Factors Impacting rs-fMRIs . . . . .	9
2.3 Effects of Patient Motion . . . . .	11
2.3.1 The Positional Effects of Motion . . . . .	11
2.3.2 The Spin History Effects of Motion . . . . .	12
2.3.3 The Susceptibility Effects of Motion . . . . .	13
2.4 Measuring Motion . . . . .	14
2.4.1 Measuring Motion: Patient Position . . . . .	14
2.4.2 Measuring Motion: Spurious Signal Correlations . . . . .	16
2.4.3 Acceptable Motion Quantities . . . . .	17
2.5 Motion Prevention . . . . .	17
2.5.1 Pre-Scan: Education . . . . .	18
2.5.2 During Scan: Sedation . . . . .	19
2.5.3 During Scan: Feed and Sleep Protocols . . . . .	21
2.6 Prospective Motion Correction . . . . .	22
2.6.1 Optical Motion Correction . . . . .	22
2.6.2 Non-Visual External Sensors . . . . .	25
2.6.3 Image Signal Motion Monitoring . . . . .	27

2.6.4	General Limitations of Prospective Motion Correction . . . . .	27
2.7	Retrospective Motion Correction . . . . .	28
2.7.1	Volume Registration . . . . .	28
2.7.2	Denoising . . . . .	30
2.7.3	Filtering . . . . .	32
2.7.4	Spin History Distortion Correction . . . . .	34
2.8	Summary . . . . .	34
<b>3.0</b>	<b>METHODS: Motion Correction</b> . . . . .	36
3.1	Directed Acyclic Graph Based Volume Registration . . . . .	36
3.2	Independent Component Analysis . . . . .	39
3.3	Motion Correction Pipeline and Implementation . . . . .	42
3.4	Evaluating Registered and Motion Corrected Sequences Against Gold Standard Usability Thresholds . . . . .	43
<b>4.0</b>	<b>METHODS: Evaluating Motion Patterns</b> . . . . .	44
4.1	Measuring Motion Patterns . . . . .	44
4.1.1	Dice Coefficient . . . . .	44
4.1.2	Mutual Information . . . . .	46
4.1.3	Implementation: Tools and Libraries . . . . .	48
4.2	Patient Classification Using Motion Patterns . . . . .	48
4.2.1	K-means Clustering . . . . .	50
4.2.2	Spectral Clustering . . . . .	51
4.2.3	Agglomerative Clustering . . . . .	54
4.2.4	Visualizing Clustering Results . . . . .	56
<b>5.0</b>	<b>DATA</b> . . . . .	60
5.1	Simulated Sequences . . . . .	60
5.1.1	Background . . . . .	61
5.1.2	SPECTr: Simulated Phantom Emulating Cranial Transformations .	62
5.1.3	Materials . . . . .	62
5.1.4	Simulation Pipeline . . . . .	63
5.1.4.1	Baseline Sequence . . . . .	63

5.1.4.2	Brain Signal . . . . .	65
5.1.4.3	Scanner Noise . . . . .	66
5.1.4.4	Patient Movement . . . . .	68
5.1.5	Implementation . . . . .	69
5.1.6	Simulated Sequences Experiments . . . . .	70
5.2	Clinical Cohorts . . . . .	71
5.2.1	CHD Background . . . . .	71
5.2.2	Study Cohorts . . . . .	76
5.2.2.1	Fetal Subject Population and Images . . . . .	76
5.2.2.2	Neonatal Subject Population and Images . . . . .	81
5.2.2.3	Preadolescent Subject Population and Images . . . . .	84
5.3	Hypothesis . . . . .	0
5.4	Summary . . . . .	1
<b>6.0</b>	<b>RESULTS</b> . . . . .	<b>3</b>
6.1	Volume Registration . . . . .	3
6.1.1	Simulated Data . . . . .	4
6.1.1.1	Volume Registration: Power Thresholds . . . . .	4
6.1.1.2	Volume Registration: Sequence Duration Motion . . . . .	9
6.1.1.3	Volume Registration: Recovered Signal . . . . .	10
6.1.2	Preadolescent Cohort . . . . .	12
6.1.2.1	Volume Registration: Power Thresholds . . . . .	12
6.1.2.2	Volume Registration: Sequence Duration Motion . . . . .	15
6.1.3	Neonatal Cohort . . . . .	19
6.1.3.1	Volume Registration: Power Thresholds . . . . .	19
6.1.3.2	Volume Registration: Sequence Duration Motion . . . . .	23
6.1.4	Fetal Cohort . . . . .	25
6.1.4.1	Volume Registration: Power Thresholds, Fetal Brain . . .	26
6.1.4.2	Volume Registration: Sequence Duration Motion, Fetal Brain	29
6.1.4.3	Volume Registration: Power Thresholds, Placenta . . . . .	32
6.1.4.4	Volume Registration: Sequence Duration Motion, Placenta	39

6.2	Characterizing Motion Across Clinical Groups . . . . .	41
6.2.1	CHD and Control . . . . .	41
6.2.2	Age Groups . . . . .	44
6.3	Summary . . . . .	51
<b>7.0</b>	<b>DISCUSSION . . . . .</b>	<b>52</b>
7.1	Volume Registration . . . . .	52
7.2	Characterizing Motion . . . . .	54
7.3	Expectations vs. Outcomes . . . . .	0
7.4	Relation to Existing Work . . . . .	1
7.4.1	MRI Simulations . . . . .	1
7.4.2	Volume Registration . . . . .	3
7.4.3	Age Group Specific Motion . . . . .	4
7.5	Limitations . . . . .	0
7.5.1	Spin History and Susceptibility Effects of Motion . . . . .	0
7.5.2	Runtime . . . . .	0
7.5.3	Manual Segmentations . . . . .	2
7.6	Future Work . . . . .	3
7.6.1	Theory-based Work . . . . .	3
7.6.2	Application-based Work . . . . .	3
<b>8.0</b>	<b>Conclusions . . . . .</b>	<b>5</b>
<b>Appendix A.</b>	<b>Volume Registration Parameters . . . . .</b>	<b>7</b>
<b>Appendix B.</b>	<b>Supplemental Statistical Tables . . . . .</b>	<b>10</b>
B.1	Simulated Data . . . . .	10
B.2	Preadolescent Cohort . . . . .	14
B.3	Neonatal Cohort . . . . .	17
B.4	Fetal Cohort . . . . .	20
B.4.1	Brain . . . . .	20
B.4.2	Placenta . . . . .	23
<b>Bibliography</b>	<b>. . . . .</b>	<b>24</b>

## List of Tables

1 The number of CHD and healthy control preadolescent subjects scanned on each scanner type at each site. . . . .	86
2 The counts and percentages of image volumes across each type of sequence in the simulated cohort which meet the usability thresholds of $FD < 0.2$ mm and $DVARS < 2.5\%$ . . . . .	7
3 The number of subjects whose sequences of types $S_1$ and $S_2$ had statistically different FD distributions. . . . .	8
4 The number of subjects whose sequences of types $S_1$ and $S_2$ had statistically different DVARS distributions. . . . .	8
5 The average rates and their confidence intervals for the classifications of component voxels as belonging to the DMN ROI for both types of registration. .	11
6 The counts and percentages of image volumes across each type of sequence in the preadolescent cohort which meet the usability thresholds of $FD < 0.2$ mm and $DVARS < 2.5\%$ . . . . .	14
7 The number of preadolescent subjects whose sequences of types $S_1$ and $S_2$ had statistically different FD distributions. . . . .	15
8 The number of preadolescent subjects whose sequences of types $S_1$ and $S_2$ had statistically different DVARS distributions. . . . .	16
9 The number of preadolescent subjects whose sequences of types $S_1$ and $S_2$ had statistically different Dice distributions. . . . .	18
10 The number of preadolescent subjects whose sequences of types $S_1$ and $S_2$ had statistically different MI distributions. . . . .	18
11 The counts and percentages of image volumes across each type of sequence in the neonatal cohort which meet the usability thresholds of $FD < 0.2$ mm and $DVARS < 2.5\%$ . . . . .	21

12	The number of neonatal subjects whose sequences of types $S_1$ and $S_2$ had statistically different FD distributions. . . . .	22
13	The number of neonatal subjects whose sequences of types $S_1$ and $S_2$ had statistically different DVARS distributions. . . . .	22
14	The number of neonatal subjects whose sequences of types $S_1$ and $S_2$ had statistically different Dice distributions. . . . .	24
15	The number of neonatal subjects whose sequences of types $S_1$ and $S_2$ had statistically different MI distributions. . . . .	25
16	The count of and percentage of image volumes across each type of sequence in the fetal brain image data set which meet the usability thresholds of $FD < 0.2$ mm and $DVARS < 2.5\%$ . . . . .	28
17	The number of subjects whose sequences of types $S_1$ and $S_2$ had statistically different FD distributions according to the Kolmogorov-Smirnov test. . . . .	28
18	The number of subjects whose sequences of types $S_1$ and $S_2$ had statistically different DVARS distributions according to the Kolmogorov-Smirnov test. . . . .	29
19	The number of fetal brain images whose sequences of types $S_1$ and $S_2$ had statistically different Dice distributions. . . . .	31
20	The number of fetal brain images whose sequences of types $S_1$ and $S_2$ had statistically different MI distributions. . . . .	32
21	The count of and percentage of image volumes across each type of sequence in the fetal placenta image data set which meet the usability thresholds of $FD < 0.2$ mm and $DVARS < 2.5\%$ . . . . .	34
22	The number of placental images whose sequences of types $S_1$ and $S_2$ had statistically different FD distributions according to the Kolmogorov-Smirnov test. . . . .	35
23	The number of placental images whose sequences of types $S_1$ and $S_2$ had statistically different DVARS distributions according to the Kolmogorov-Smirnov test. . . . .	35
24	Results of t-tests comparing the descriptive statistics of the Dice matrices for the fetal placenta data. . . . .	37

25	Results of t-tests comparing the descriptive statistics of the MI matrices for the fetal placenta data. . . . .	38
26	The number of subjects whose placenta sequences of types $S_1$ and $S_2$ had statistically different Dice distributions. . . . .	40
27	The number of subjects whose placenta sequences of types $S_1$ and $S_2$ had statistically different MI distributions. . . . .	40
28	The composition of age groups in each cluster produced using the MI feature vectors and k-means clustering. . . . .	49
29	The composition of age groups in each cluster produced using the MI feature vectors and spectral clustering. . . . .	50
30	The composition of age groups in each cluster produced using the MI feature vectors and agglomerative clustering. . . . .	50
31	The statistics for the runtimes of the registrations of 17 healthy neonatal subjects demonstrate that runtime of registration is not a trivial concern. . . . .	1
32	Results from the t-tests comparing the counts for the numbers of images meeting the FD, DVARS, and FD and DVARS thresholds for sequence types $S_1$ and $S_2$ . . . . .	10
33	Results of t-tests comparing the descriptive statistics of the correlation ratio matrices for the simulated data. . . . .	11
34	Results of t-tests comparing the descriptive statistics of the Dice matrices for the simulated data. . . . .	12
35	Results of t-tests comparing the descriptive statistics of the MI matrices for the simulated data. . . . .	13
36	Results from the t-tests comparing the counts for the numbers of images meeting the FD, DVARS, and FD and DVARS thresholds for sequence type $S_1$ and sequence type $S_2$ . . . . .	14
37	Results of t-tests comparing the descriptive statistics of the Dice matrices for the preadolescent data. . . . .	15
38	Results of t-tests comparing the descriptive statistics of the MI matrices for the preadolescent data. . . . .	16

39	Results from the t-tests comparing the counts for the numbers of images meeting the FD, DVARS, and FD and DVARS thresholds for sequence type $S_1$ and sequence type $S_2$ .	17
40	Results of t-tests comparing the descriptive statistics of the Dice matrices for the neonatal cohort.	18
41	Results of t-tests comparing the descriptive statistics of the MI matrices for the neonatal data.	19
42	Results from the t-tests comparing the counts for the numbers of images meeting the FD, DVARS, and FD and DVARS thresholds for fetal brain sequence type $S_1$ and sequence type $S_2$ .	20
43	Results of t-tests comparing the descriptive statistics of the Dice matrices for the fetal brain data.	21
44	Results of t-tests comparing the descriptive statistics of the MI matrices for the fetal brain data.	22
45	Results from the t-tests comparing the counts for the numbers of images meeting the FD, DVARS, and FD and DVARS thresholds for fetal placenta sequence type $S_1$ and sequence type $S_2$ .	23

## List of Figures

1	A rs-fMRI can be thought of (a) as a sequence of image volumes or (b) as a single volume where each voxel contains a temporal signal rather than a single numeric value. . . . .	9
2	The patient’s brain activity is effected by several factors before an MRI scanner produces a visually interpretable image sequence. (The rounded rectangles represent the signal at different stages of recording while the rhombuses represent the primary factors.) . . . . .	10
3	The pitch, roll, and yaw rotations describe rotations of an object about three orthogonal axes whose origin is in the object’s center of mass. . . . .	15
4	The traditional approach to volume registration in an rs-fMRI sequence consists of registering all volumes in the sequence to a single reference volume. . . . .	29
5	A rs-fMRI can be viewed as a directed acyclic graph where each volume is a node and the edges connect from each volume $i$ to the following volume $i + 1$ . . . . .	37
6	The DAG-based approach to volume registration in an rs-fMRI sequence consists still registers all volumes in a sequence to a single reference volume, but it also accounts for the temporal relationships between subsequent volumes. . . . .	38
7	An example of $k$ -means clustering performed on the Iris data set. The results of the algorithm are highly dependent on the number of clusters specified. . . . .	49
8	(a) Two different distributions, a 2D Gaussian density and a thin horizontal rectangle are difficult to separate (b) due to their overlap and the penalties built into the cost function of the spectral clustering algorithm. From (Nadler and Galun, 2007). . . . .	53
9	An example of agglomerative clustering on the Iris data set plotted using <b>seaborn</b> ’s <b>clustermap</b> function. . . . .	55
10	Visualization of MNIST data in 3D via TensorFlow Projector as generated using PCA, t-SNE, and UMAP dimensionality reduction techniques. . . . .	59

11	An overview of the SPECTr simulation pipeline. Using atlas data, a simulated phantom containing brain signal, scanner noise, and patient motion is generated.	64
12	An example of the magnitude and phase of an image in k-space. . . . .	67
13	Table of prevalences of congenital heart defects as compiled by (Mozaffarian et al., 2016). . . . .	72
14	Estimated CHD burden in World Health Organization (WHO) regions using incidence rates of approximately 12/1000 and 4/1000 in children and adults, respectively (Webb et al., 2015). . . . .	74
15	Distribution of subject races for all three cohorts. . . . .	77
16	Distribution of subject ethnicities for all three cohorts. . . . .	77
17	Distribution of subject genders for all three cohorts. . . . .	78
18	The distribution of post-conceptual ages at the time of the scan of all fetal subjects. . . . .	79
19	The outer ring represents the group the subject was assigned to based on the fetal examination while the inner ring represents the group the subject was assigned to after birth. . . . .	80
20	The distribution of post-conceptual ages at birth of all neonatal subjects. . .	82
21	The distribution of post-conceptual ages at the time of the scan of all neonatal subjects. . . . .	82
22	The breakdown of subject groups contained in the Control and CHD neonatal cohorts. . . . .	83
23	The distribution of preadolescent subject ages at the time of the scan. . . .	84
24	The distribution of CHD and healthy subjects between the 12 sites enrolled in the preadolescent study. . . . .	85
25	The means and standard deviations of the FD and DVARS metrics for all simulated images before and after each type of registration. . . . .	5
26	The means of the FD and DVARS metrics for all simulated images before and after both registrations. . . . .	6

27	Boxplots of the values of all Dice matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the simulated data. . . . .	9
28	Boxplots of the values of all MI matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the simulated data. . . . .	10
29	The means and standard deviations of the FD and DVARS metrics for all preadolescent images before and after each type of registration. . . . .	13
30	Boxplots of the values of all Dice matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the preadolescent cohort. . . . .	17
31	Boxplots of the values of all MI matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the preadolescent cohort. . . . .	17
32	The means and standard deviations of the FD and DVARS metrics for all neonatal images before and after each type of registration. . . . .	20
33	Boxplots of the values of all Dice matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the neonatal cohort. . . . .	23
34	Boxplots of the values of all MI matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the neonatal cohort. . . . .	24
35	The means and standard deviations of the FD and DVARS metrics for all fetal brain images before and after each type of registration. . . . .	27
36	Boxplots of the values of all Dice matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the fetal-brain images. . . . .	30
37	Boxplots of the values of all MI matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the fetal brain images. . . . .	30

38	The means and standard deviations of the FD and DVARS metrics for all placenta images before and after each type of registration. . . . .	33
39	Boxplots of the values of all Dice matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the placenta images. . . . .	36
40	Boxplots of the values of all MI matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the placenta images. . . . .	37
41	The data for the four metrics used to evaluate motion in the clinical brain images projected using different methods into 2D space labeled by CHD/Control status. . . . .	42
42	The original preadolescent, neonatal, and fetal images clustered by each metric using agglomerative clustering and labeled by disease status. . . . .	43
43	The data for the four metrics used to evaluate motion in the clinical brain images projected using different methods into 2D space and labeled by age group. . . . .	45
44	The preadolescent, neonatal, and fetal images clustered by each metric using agglomerative clustering and labeled by age group. . . . .	46
45	The results of clustering the clinical images using the MI matrices and the age group labels. . . . .	48
46	Iteratively removing clusters containing less than 10% of the data in the data set and repeating the spectral clustering process produced meaningful clusters similar to those produced by the k-means and agglomerative methods. . . .	49
47	The average local motion metrics for all original sequences show that the neonatal subjects move the least, though differences between fetal and preadolescent motion is more clear in (b). . . . .	0
48	The “Pulse Sequence” specifications page in the POSSUM GUI . . . . .	2

## **Preface**

## **DEDICATION**

To the patients being treated in children's hospitals all over the world, their families, and the people working to help them.

## ACKNOWLEDGEMENTS

This project would not have been possible without the assistance and support of many people. First, to my committee members: Dr. Doug Landsittel, Dr. Ashok Panigrahy, Dr. Greg Cooper, and Dr. Rafael Ceschin. Your guidance and suggestions were truly invaluable. Second, to the members of PIRC, especially: Julia Wallace, Vince Lee, and Nancy Beluk, for helping wrangle the data and answering questions about behavioral techniques used during pediatric scans; Dr. Vincent Schmithorst, for the many conversations about the intricacies of MR physics; Billy Reynolds, for the fetal data and for pulling the occasional perfect figure seemingly out of thin air; and Samuel Cho, for helping me develop the data simulation and aptly naming it “SPECTr”. Third, this document would not be the same without the Writing Accountability Groups organized and run by Dr. Moriah Kirdy through the Pitt Writing Center. Moriah’s dedication and organization is (thankfully) contagious.

A successful dissertation is only partly made up of the research and writing. The rest is managing people, places, and paperwork. Many thanks to Toni Porterfield and Tami Robinson for helping me keep my administrative ducks in a row and to Barbara Karnbauer for helping coordinate schedules.

Of course, I would like to thank my friends and family. Ryan, for keeping me grounded, and for rock climbing, burritos, and ice cream. My friends in and around Philadelphia and Pittsburgh, for the laughs, game sessions, and more rock climbing. Mom and Dad, thanks for supporting me the way you always do. Jonathan and David, I’m over the moon you were able to make my defense.

Finally, special thanks to everyone involved with Extra Life, a charity that raises money for kids being helped by Children’s Miracle Network Hospitals. As promised, here is the shout out to those who donated to my page, especially to Leo Chan, Elliot Guiso, Bonnie Young, John Maloney, Justine White, and Mike West.

## 1.0 Introduction

Patient motion is a critical cause of data loss in medical imaging. Many approaches have been developed to prevent and reduce the impact of motion before, during, and after image acquisition. The effectiveness of these techniques varies between patient populations. If the effects of a patient’s movements cannot be removed from an image, that image is considered to have been corrupted by motion and is deemed unusable.

Though patient motion is a problem across the entire medical imaging domain, we focus specifically on resting-state functional magnetic resonance imaging (rs-fMRI). rs-fMRIs measure the blood oxygen level-dependent signal in an organ or organ system. Areas of an organ with more cellular activity require more oxygen than less active areas. When used to examine the brain, the signals recorded by the rs-fMRI are used effectively as approximations of the amount of activity occurring in different areas of the brain. The term “resting-state” means that the patient is not performing any particular task, so any activity that occurs is from underlying networks connecting different areas of the brain.

### 1.1 Resting-state Functional Magnetic Resonance Images

When an area of the brain is active, it uses more oxygen than the surrounding regions. Functional MRIs (fMRI) are sensitive to signals emitted by deoxygenated hemoglobin. The blood oxygen level-dependent (BOLD) signals recorded by the fMRI reveal regions of the brain which are active at the same time. These combinations of regions are called neuronal networks.

Many neuronal networks exist, but most of them are considered to be task-related. In 2001, Raichle et al. suggested the existence of a neuronal network that operates when a person is at rest (Raichle et al., 2001). Their theory was confirmed by Greicius et al. in 2003 (Greicius et al., 2003). Because the patient is not performing a specific task when they are in a resting state, the resting-state networks have the potential to reveal valuable information

about a patient's neurodevelopmental status.

In order to gather enough data to evaluate these networks fully, a series of image volumes must be acquired over several minutes. An fMRI taken of a patient in a resting, task-free state, is called a resting-state fMRI (rs-fMRI). The image volumes within the rs-fMRI sequence have relatively low spatial resolution when compared to structural MRIs, but their temporal resolution is significantly higher: a new image volume is acquired every two to three seconds.

The BOLD signals in rs-fMRI image sequences are analyzed using a process called functional connectivity analysis. Functional connectivity analysis identifies patterns and networks of brain activity. Several functional connectivity analysis studies have led to the discoveries of links between specific disruptions in the networks and neurodevelopmental diseases such as autism and attention deficit hyperactivity disorder (Assaf et al., 2010) (Zang et al., 2007). With further refinements of both acquisition techniques and characterization of these functional networks, clinicians may be able to use rs-fMRI to evaluate the neurodevelopmental status of CHD patients and to identify patients who may benefit from specific therapies or neuroprotective interventions.

Though it can gather high-quality data on relatively short timescales, the rs-fMRI suffers from two major limitations: rs-fMR images have low physical resolution and are highly susceptible to motion. The first limitation can be addressed by obtaining an MR image with high physical resolution and registering the rs-fMRI to this structural image, but the second limitation is a significant problem.

## 1.2 Motion Effects, Prevention, and Correction

There are three effects of motion on an rs-fMRI scan: the positional effects, the spin history effects, and the susceptibility effects. The impact of motion on the position of the patient means that a given voxel will not record the electromagnetic signal from the same location in the brain for the duration of the scan. When the patient moves, the molecules which were in the area activated by the MRI scanner also move. Molecules that were not

activated are now in the area where the signal is being recorded. This shift results in an artificial decrease in the recorded signal. At the next activation, any molecules which moved out of the previous area of activation may undergo further activation, which can result in an artificial increase in signal. These spin history effects impact the image sequence for several frames but dissipate. The recorded signal is also impacted by the change in the susceptibility properties of the tissue being recorded by the voxel. The difference in the susceptibility of different tissues is most prominent at the tissue interfaces. When the patient moves, the tissue interfaces move and contribute or detract from the signal at new locations.

In general, the best way to prevent the effects of motion is to prevent motion itself. Various clinical, behavioral, and technical protocols have been developed in an attempt to prevent patient motion from impacting the rs-fMRI image as it is acquired. Sedation can be used to immobilize a patient during a scan but requires additional personnel to perform safely and involves an increased time commitment from the patient. Sedation is also not recommended for use in young children and fetal patients. Behavioral and educational techniques can be employed to prepare a patient for stressors he may experience during a medical imaging scan. However, these approaches do not prevent the patient from moving out of boredom, discomfort, or distress. Several groups have developed techniques to compensate for motion as an image is acquired, but these techniques often require additional scanner-compatible equipment and can only be utilized during the scan. Sedation or intra-scan motion monitoring approaches are difficult to integrate with MR scanners due to the constraints of MR safety requirements.

Additional processing is needed to remove motion from an image after the scan is acquired. Many methods have been developed to mitigate the effects of motion after the rs-fMRI is acquired. While different post-acquisition motion correction pipelines utilize different processing techniques, they begin with global volume registration. Global volume registration is the process used to align all volumes in an rs-fMRI sequence into the same physical space. Traditionally, all volumes in the sequence are registered directly to one volume. This approach can be effective in images where the subject remains relatively still throughout the scan but is not as successful in images containing high quantities of patient movement.

We have developed an alternative volume registration framework that takes into account the spatiotemporal relationships between sequential volumes in the rs-fMRI sequence and uses these relationships during the registration process. Herein, we evaluate it further in the context of a complete motion correction pipeline across healthy and disease populations at various stages of life. In addition to reducing the effects of patient motion on image quality, we are also interested in the patient motion itself. We believe there are relationships between different motion patterns, patient age, and clinical outcomes, and we have explored these relationships throughout our experiments.

### 1.3 Data

The disease population we used for our study is a population with a variety of congenital heart defects. Congenital heart defects (CHDs) have many presentations, and all presentations cause problems in a patient’s heart structure and the structure of the surrounding vessels. It has been found that the development of cardiac problems *in utero* is often linked to delays in patient neurodevelopment. Research in the area of CHD and neurodevelopment has often focused on younger populations. The treatments for CHDs have evolved over the past fifty years with the result that many CHD patients live to adulthood. Every year, approximately 1.35 million children are born with a congenital heart defect, and it is estimated that about 12 to 34 million adults are living with CHD (van der Linde et al., 2011). Researchers have recognized the burden of neurocognitive disorders on the aging CHD population and are now starting to investigate the relationships between CHD and neurocognitive outcomes.

The process for objectively identifying neurocognitive disorders is still under development. Psychologists have developed and validated surveys to estimate a patient’s neurocognitive status. These surveys vary with the child’s age. Initially, a parent fills out the survey on behalf of his infant or toddler child. When the child has reached certain developmental milestones, the parent and child might both fill out different portions of a different survey. After a certain age, the child can fill out his own survey. Psychologists may meet with the

patient and his parents to determine a diagnosis. These survey-based methods are highly subjective, and objective methods based on rs-fMRIs are being explored.

Eventually, clinicians will be able to develop a lifespan approach to managing CHD and neurocognitive disorders. As a community, we are still in the data-gathering stage of this research. We cannot afford to lose rs-fMRI scans of healthy or CHD patients in any stage of life because of motion. For these reasons, cohorts of healthy and CHD patient images are an ideal data set for our motion correction work.

We were able to access data from three clinical cohorts. The first cohort is a set of healthy and CHD neonatal subjects scanned at our primary study site. The second cohort is a set of healthy and CHD preadolescent subjects enrolled in a multicenter study. Some subjects were scanned at our primary study site, while others were scanned at one of the 11 other participating sites. The third cohort is a set of healthy and CHD fetal subjects scanned at our primary site.

Many rs-fMRI studies struggle to obtain a sufficient number of low motion scans to come to statistically significant conclusions. We develop a tool to address this challenge by simulating rs-fMRIs using an existing average brain and functional region of interest templates. The simulated sequences contain brain signals, scanner noise, and motion. We then use this tool to simulate additional images with which to test the registration frameworks and to determine the effects of volume registration on brain signals.

## 1.4 Experiments

Both the traditional and novel registration techniques are applied to our clinical and simulated images. We compare the original and registered images with respect to the amount of motion removed from each sequence. We use two metrics to measure the local effects of motion and two metrics to measure the global similarity of the sequences.

The metrics used to measure local motion are the change in patient position (framewise displacement, FD) and the change in overall signal (the derivative of the variance of the root mean square of the signal, DVARS) between one timepoint and the next. These metrics

are only calculated between every pair of chronological image volumes in a sequence. They are then compared to a pair of thresholds commonly recognized in the field of motion in rs-fMRIs to determine the usability of a sequence.

The three similarity metrics are the correlation ratio, the mutual information, and the Dice coefficient. The similarity metrics are used to compare every volume in a sequence to every other volume in the same sequence, which produces a two-dimensional matrix for each metric. Changes in the patterns of the matrices for each metric are used to determine the global impact of registration on the sequence.

While correcting motion within an rs-fMRI is important both for clinical use and research applications, we are also interested in the motion itself. Through discussion with radiologists and researchers who work with rs-fMRIs, we have developed the following hypothesis: **Neonatal patients on average exhibit less motion than preadolescent patients, who exhibit less motion than fetal patients.** In addition, we apply unsupervised machine learning techniques to the age group level cohorts to identify patients with common patterns of motion in each cohort.

## 1.5 Summary

The remainder of this document is laid out as follows. We elaborate on resting-state functional magnetic resonance images (rs-fMRIs) and their use for investigating functional brain networks in Chapter 2. In Chapter 3, we perform a breadth-wise review of the effects of motion, methods to prevent motion, and methods to mitigate the effects of motion. Chapter 4 transitions into methods for analyzing MRIs, machine learning techniques, and our approach to statistical analysis.

Chapter 5 contains information about the simulation process developed to generate simulated images with known brain signals, background noise, and motion. It discusses congenital heart disease, its relationship with neurological conditions, and methods for evaluating neurological conditions. It also contains information about the scans and demographics information for each clinical cohort.

Chapter 6 contains the results of our experiments and statistical analyses for the registration experiments and the machine learning experiments. Chapter 7 contains a discussion of these results, and Chapter 8 contains a metadiscussion of this study as a whole.

## 2.0 rs-fMRIs and Patient Motion

This chapter discusses rs-fMRIs and how they are affected by patient motion. Specific topics include the structure of rs-fMRIs, sources of motion, quantifying motion, and a review of current methods for preventing and managing motion in rs-fMRIs.

### 2.1 Structure of an rs-fMRI

rs-fMRIs are discrete representations of continuous data. A new image volume of the patient’s brain is acquired every two to three seconds. The image volume is composed of a three-dimensional version of a pixel called a voxel (volume element). Just as the “distance” between each image volume encompasses a certain amount of time, each voxel encompasses a small volume of physical space. The transformations between the continuous physical and temporal dimensions and the discrete physical and temporal dimensions are the spatial and temporal resolutions. The concept of a 4D rs-fMRI image is illustrated in two different ways in Figure 1. The first representation is an ordered list of 3D image volumes where each voxel contains a single numeric value. The second representation is a single 3D image volume, where the value of each voxel is a temporal signal.

An rs-fMRI is considered to have a relatively low spatial resolution but high temporal resolution. The physical size of a single voxel seems small at about  $4\text{ mm}^3$ , but this resolution is not granular enough to capture details about activity within small structures of the brain. The activity information recorded during an rs-fMRI must be combined with the detailed anatomic information from a structural MRI to know precisely which areas of the brain are active at each point in time. A structural MRI volume takes much longer to acquire than an rs-fMRI volume, which can be obtained every two to three seconds. Unfortunately, the patient’s position and neural activity can change faster than the image volume can be acquired. As a result, a temporal resolution of two to three seconds is not fast enough to actively compensate for sources of noise which confound the BOLD signal.

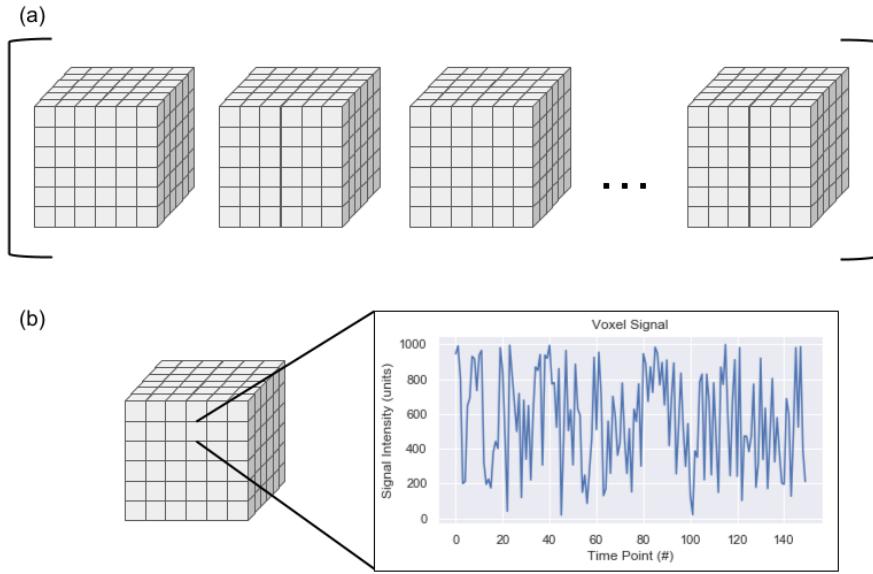


Figure 1: A rs-fMRI can be thought of (a) as a sequence of image volumes or (b) as a single volume where each voxel contains a temporal signal rather than a single numeric value.

## 2.2 Factors Impacting rs-fMRIs

The BOLD signal present inside a patient’s brain is not recorded with complete accuracy by an MRI scanner. Even if the same patient exhibited precisely the same BOLD signal during two different scans, the recorded image sequences would vary slightly. Several factors impact the image sequence viewed by a radiologist, and we give an overview of them in Figure 2.

The first factor in Figure 2 is the patient’s physiology. An rs-fMRI is not sensitive enough to detect brain activity on a neuronal level. Instead, it measures the changes in the amounts of deoxygenated hemoglobin in the brain. The deoxygenated hemoglobin quantities are highly correlated with brain activity because active areas of the brain use more oxygen than inactive areas, but the BOLD signal is still only an approximation of brain activity.

The second factor in Figure 2 represents the changes to the signal that occur due to patient motion. During every medical imaging scan, the patient will naturally perform

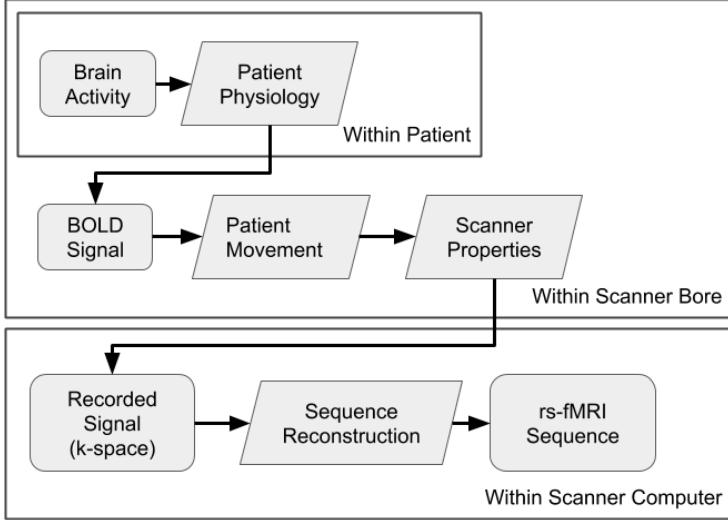


Figure 2: The patient’s brain activity is effected by several factors before an MRI scanner produces a visually interpretable image sequence. (The rounded rectangles represent the signal at different stages of recording while the rhombuses represent the primary factors.)

small, automatic movements due to regular bodily functions. Minuscule movements caused by cardiac activity may disrupt scans with high spatial resolution or with high sensitivity to the movement of blood molecules. More significant movements caused by respiration result in motion artifacts in images of the thoracic and abdominal cavities.

Other motions occur on a larger and more conscious scale. It is important to note that different populations may exhibit more of certain macro-motions than others. The patient may fidget or shift his gaze when he becomes bored in the scanner. If the patient falls asleep during a scan, there may be slight movement as the body relaxes and re-tenses if the patient wakes. Certain MRI protocols are known to produce loud sounds: during one of these protocols, the patient may become surprised and react by jumping. Additionally, claustrophobic patients or patients who feel secure around specific people that are not allowed in the scanner room may become agitated.

Both the small-scale, automatic motions and the large-scale, reflexive motions corrupt the BOLD signal. These effects of patient motion will be discussed in-depth in the next

section.

The third factor in Figure 2 is the unique properties of the scanner. The acquired image sequence will vary slightly even in machines made by the same company because each scanner has a unique primary magnetic field,  $B_0$ . The  $B_0$  inhomogeneities can be measured using a primary field map. These field maps can be used to correct signals displaced by the  $B_0$  field, though they cannot be used to recover signals corrupted by the field.

The final factor is also related to the properties of the MRI scanner. The image sequence produced by a scanner is dependent on the scanner vendor. Different vendors use different proprietary algorithms to convert the signal recorded by the scanner in k-space into a visually interpretable image sequence. Between the  $B_0$  inhomogeneities and the scanner vendor differences, the differences in the scanners used to acquire the sequences must be resolved before images from different scanners can be compared.

## 2.3 Effects of Patient Motion

Due to their low spatial and high temporal resolutions, rs-fMRIs are highly susceptible to all types of motion outlined in the previous section. In general, motion affects an acquired image sequence in three ways. The first and most obvious effect is the position of the patient changes throughout the sequence. The second effect is due to the way changes in the patient's position affect the signal recorded by the scanner. The third effect is related to the magnetic fields in patient tissue and changes in their orientation within  $B_0$ . These three effects will hereafter be referred to as the positional effect, the spin history effect, and the susceptibility effect of motion.

### 2.3.1 The Positional Effects of Motion

The technique used for analyzing neuronal network activity rs-fMRIs, called functional connectivity analysis, assumes that the contents of one voxel at every time point during the sequence all contain signals from a single point in the brain. This assumption is vital in the

process of inferring networks of neuronal activity.

While rs-fMRIs have a spatial resolution on the order of millimeters, neuronal activity occurs on the spatial resolution of microns. As a result, each voxel in an rs-fMRI volume contains information from a number of neurons. The smallest movement of the patient can alter the voxel to which a cluster of neurons contributes. These seemingly insignificant changes can alter the position of the patient enough to cause the voxels to record signals from different brain regions or even tissue types. This change in voxel location within the brain violates the assumption that single voxels record from the same location within the brain for the duration of the sequence.

### 2.3.2 The Spin History Effects of Motion

In addition to changing the recorded position of the patient, motion impacts the established spin gradients, which introduces artifacts into the image sequence.

During an ideal MRI scan, the patient is sitting in the scanner and all molecules become aligned with the primary magnetic field  $B_0$  in a relaxed state. Then, a radiofrequency (RF) pulse is applied to the field. The purpose of the pulse is to excite the molecules in a particular volume of physical space change their orientation to match the induced perpendicular field. When the pulse ends, the molecules precess back to their orientation in  $B_0$ . As they do, their small magnetic fields induce electric currents on the RF coil. The currents are recorded by the scanner as signals in frequency space. The volume of the space intended to be excited is known, and the signal produced by the induced electric current is used in conjunction to reconstruct the image in voxel space.

However, when the patient moves, the volume of space which was thought to be excited is not actually excited: some other volume of space, which may or may not overlap with the intended volume of space, is excited instead. Because the MRI scanner has no way to know this assumption is not correct, it does not know that not all of the molecules in its intended area are relaxed and correctly aligned to the  $B_0$  field at the end of the RF pulse. The scanner proceeds with the next RF pulse, which excites a new set of “relaxed molecules”, some of which are still excited from the previous pulse. As a result, the signals produced in the

second RF pulse are different than they should be. Signals that are artificially decreased result in dark shadows within motion affected volumes of the sequence, while signals that are artificially increased result in bright spots.

The previous few paragraphs in this section describe how motion disrupts the magnetic spin gradients present in the patient during an rs-fMRI scan. The spin gradients need time to recover to the correct magnetic field orientation, and up to eight to ten seconds may pass before the recovery is complete (Power et al., 2014). While the spin gradients are reorienting, the recorded BOLD signal will vary more than expected between temporally neighboring volumes. These variations are more difficult to quantify than the positional effects of motion.

The nature of the  $B_0$  field in an MRI scanner can be recorded using a field map. Acquiring a field map while the patient is in the scanner records information about how the patient interacts with the  $B_0$  field. These field maps of the patient in the scanner can be used to perform distortion correction on the acquired sequence. A study of the effects of distortion correction on rs-fMRIs of 40 healthy subjects suggested that distortion correction using field maps can increase the functional connectivity in rs-fMRIs (Togo et al., 2017).

### 2.3.3 The Susceptibility Effects of Motion

The susceptibility of a material describes how the material will behave when placed in a magnetic field. Most materials are either paramagnetic or diamagnetic. Paramagnetic materials are attracted to and align with magnetic fields. Diamagnetic have the opposite interaction: they are repelled from and become anti-aligned with magnetic fields. Additionally, paramagnetic materials contribute to the magnetic field, while diamagnetic materials detract from it.

Both types of materials cause distortions in the magnetic field. These distortions are prominent in stronger magnetic fields and at the interface of two different material types. In MRI scanners, the differences in susceptibility between soft tissue and bone or between soft tissue and air can produce artifacts in the acquired sequence. These artifacts are amplified when the patient moves. When the patient moves, the interfaces between tissues and air

distort the  $B_0$  field. The distortions in the  $B_0$  field change the electromagnetic signal recorded by the scanner and lead to spurious correlations during the analysis of the rs-fMRI sequence.

Susceptibility artifacts can be reduced using susceptibility maps. Susceptibility maps are short acquisitions that use multiple echo periods to detect small changes in the location of susceptibility interfaces in an MRI sequence. They are not part of most rs-fMRI acquisition protocols.

## 2.4 Measuring Motion

Even though we described three effects of motion on rs-fMRIs in the previous section, these effects impact the sequence in two areas: the position of the patient and spurious signal correlations throughout the sequence.

### 2.4.1 Measuring Motion: Patient Position

The effect of motion on patient position is measured as the difference in the positions of subsequent image volumes. The difference in position is determined using metrics calculated by performing rigid volume registration on the two volumes. In rigid volume registration, one volume is chosen as the reference volume and the other is the moving volume. The reference volume remains stationary while the moving volume is translated and rotated in three-dimensional space on top of it. The registration is complete when the position of the patient in the moving volume matches the position in the reference volume.

The moving volume can undergo linear or nonlinear transformations. Linear transformations include translation, rotation, and affine transformations along all three spatial dimensions as well as a scaling transformation. These transformations move the image volume as a whole: all voxels in the moving image remain in the same location relative to their neighbors. On the other hand, nonlinear transformations can warp the contents of the moving volume so that it better matches the contents of the reference volume. Nonlinear transformations are more complex than linear transformation. They involve additional image processing steps

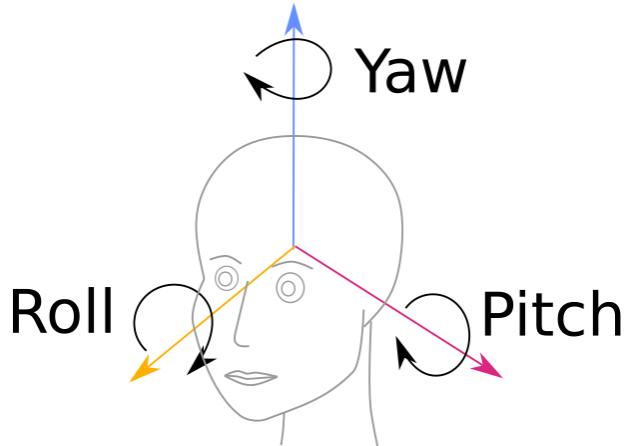


Figure 3: The pitch, roll, and yaw rotations describe rotations of an object about three orthogonal axes whose origin is in the object’s center of mass.

such as smoothing and voxel interpolation.

Even in cases when nonlinear transformations are used, the registration process begins with the translation and rotation transformations. The three translation and three rotation parameters used to achieve the best alignment are used to calculate the positional change between the image volumes. The positional change between temporally neighboring volumes is called the framewise displacement (FD).

Several researchers have proposed slightly different methods for calculating the FD. Power et al., Jenkinson et al., and Dosenbach et al. each propose a slightly different method for calculating the FD (Power et al., 2012) (Jenkinson et al., 2002) (Dosenbach et al., 2017). All three FD calculations produce correlated metrics: the FD metric proposed by Power et al. produces measurements approximately twice as large as the metric proposed by Jenkinson et al., and Dosenbach et al. reported a high correlation between their FD and Powers FD (Yan et al., 2013b) (Dosenbach et al., 2017).

In the remainder of this document, the abbreviation FD refers to Power et al.’s version

of the FD metric, which is calculated as:

$$FD(J_i) = |\Delta d_{ix}| + |\Delta d_{iy}| + |\Delta d_{iz}| + |\Delta \alpha_i| + |\Delta \beta_i| + |\Delta \gamma_i| \quad (2.1)$$

where  $J_i$  is the image volume  $i$ ,  $|\Delta d_{i*}|$  are the magnitude of change in position along the  $x$ ,  $y$ , and  $z$  axes between volumes  $i$  and  $i - 1$ , and  $|\Delta \alpha_i|$ ,  $|\Delta \beta_i|$ , and  $|\Delta \gamma_i|$  are the magnitude of change in pitch, roll, and yaw between volumes  $i$  and  $i - 1$ , respectively. A diagram outlining how pitch, roll, and yaw relate to the human head can be seen in Figure 3.

#### 2.4.2 Measuring Motion: Spurious Signal Correlations

Both the spin history and susceptibility effects of motion contribute to alterations in the signal recorded and reconstructed by the MRI scanner. Without  $B_0$  field maps and susceptibility maps, it is difficult to separate the impact of each of these factors on the recorded signal. We assume at this point that both the spin history and the susceptibility effects contribute equally to changes in the recorded signal intensity.

One popular metric to measure changes in the recorded signal due to patient motion was developed by Smyser et al. in 2010. Their metric is called DVARS, which measures the temporal derivative of the root mean squared variance over the voxels between two volumes (Smyser et al., 2010). Power et al. explain the steps to calculate DVARS in a separate study (Power et al., 2012). The DVARS value is calculated in two steps. The first step uses backward differences to approximate the derivative of the BOLD signal change between volumes  $J_i$  and  $J_{i-1}$  at every point  $\vec{x}$  contained in both image volumes:

$$\frac{\partial}{\partial t} J_i(\vec{x}) \approx J_i(\vec{x}) - J_{i-1}(\vec{x}). \quad (2.2)$$

The second step calculates the root mean square of the approximated derivatives for all  $N$  points  $\vec{x}$ :

$$DVARS(J_i) = \sqrt{\frac{1}{N} \sum_{\vec{x} \in J_i, J_{i-1}} \left( \frac{\partial}{\partial t} J_i(\vec{x}) \right)^2}. \quad (2.3)$$

DVARS measures the change in BOLD signal intensity, which is highly related to motion-induced spin gradient changes.

#### 2.4.3 Acceptable Motion Quantities

Even though the effects of motion on the patient position and the recorded signal can be measured, we still need gold standard criteria to determine whether an image containing motion can be used. Patients move slightly due to breathing and cardiac function, and the BOLD signal naturally fluctuates over time. Some motion is expected; however, we need to know how much motion can be present in the image before it is considered to be corrupted by it. Power et al. established thresholds for FD and DVARS to determine the usability of a pair of images:

- FD less than or equal to 0.2 mm from the previous volume, and
- DVARS less than or equal to 25 units on a normalized scale of [0, 1000] signal units (Power et al., 2014)

Image volumes that meet these criteria are considered to be low-motion.

The minimum duration of low-motion data is highly debated. van Dijk et al. established that approximately five minutes of low-motion data is sufficient for use in functional connectivity analysis (van Dijk et al., 2012). However, a recent study by Laumann et al. suggests that at least 10 minutes of low-motion data is essential for obtaining high-quality results (Laumann et al., 2015). From a practical standpoint, it is difficult to obtain even five minutes of low motion data from certain patient populations, so radiology technicians and neuroimaging study designers are often content with the five minute time standard.

## 2.5 Motion Prevention

During an MRI scan, thick foam pads of various sizes and shapes are used to isolate and immobilize the area of interest on the patient. While these pads impede most motion, especially in compliant patients, additional techniques and protocols are often used to prevent

patients from moving during the image acquisition process. Not all of these techniques are suitable for all patient populations, and some techniques have been designed specifically for certain populations.

### **2.5.1 Pre-Scan: Education**

Educational material can be used to help the patient understand what to expect during an MRI scan as well as to teach the patient different behavioral coping strategies. The education materials can be used either before or upon arrival at the imaging facility. Most of the formal literature focuses on informative, distraction, and behavioral techniques to use during pediatric MRI scans, though many of the following approaches could be adapted for use with adults.

In a review of the available literature, Alexander found several commonly used techniques to educate pediatric patients before and comfort or distract pediatric patients during radiology procedures (Alexander, 2012). Tools such as educational coloring books and short videos can expose patients to the types of equipment they can expect to see using a familiar, engaging medium. Pediatric patients can learn coping strategies to employ during the scan, such as breathing techniques, imagery, and positive statements. Alexander notes that allowing a pediatric patient to choose a behavioral coping strategy gives the patient a sense of control and may encourage the patient to cooperate during the MRI acquisition.

Mock scanners and MRI simulators can also help the patient feel more comfortable during the scan. Barnea-Goraly et al. showed that both a commercial MRI simulator and a low-tech mock scanner desensitized pediatric patients between four and ten years of age to the MRI scanner with the results that 92.3% of the acquired images could be used in high-resolution anatomical studies (Barnea-Goraly et al., 2014).

Several groups have investigated the role of auditory and visual distraction during an MRI acquisition. Headphones with music and stories or MR compatible video goggles can distract patients from the tedium of the scan (Alexander, 2012) (Barnea-Goraly et al., 2014) (Harned and Strain, 2001). Khan et al. found that a relatively simple moving light show can help distract younger patients (Khan et al., 2007). Garcia-Palacios et al. performed a case

study comparing the efficacy of music and immersive virtual reality tools as distractions during a mock scan (Garcia-Palacios et al., 2007). They suggest that immersive virtual reality may help decrease patient anxiety during a scan more effectively than music alone. As virtual reality technology improves, it may join headphones and MR compatible video goggles as an available distraction method.

Another valuable source of distraction for pediatric patients could be the patient's parent or parents. Having a parent involved with the scanning process may calm the patient and encourage him to cooperate; however, parental distress can further upset an anxious patient and complicate the scanning process (Alexander, 2012).

These techniques for educating the patient and helping the patient cope with the anxiety that can accompany an MRI scan all depend on the ability of the patient to understand instructions and communicate with the scan team. Due to the gap in communication abilities, these techniques are not useful for young patients such as neonates, infants, toddlers, and possibly elementary school-aged children. Other patient populations, such as those with developmental delays and neurobehavioral disorders, may also have difficulty adhering to these protocols. Even in patients with developed and intact communication skills, the techniques outlined here do not actively prevent the patient from moving during the scan: they only help the patient feel more comfortable with the MRI environment.

### **2.5.2 During Scan: Sedation**

Sedation can be used to help a patient tolerate an MRI scan. Murphy and Brunberg retrospectively analyzed seven weeks of data from the MR department and found that 14.2% of their adult patients required some form of sedation (Murphy and Brunberg, 1997). In a study about claustrophobia and MR acquisitions, Dewey et al. report that out of 55,734 patients who underwent MRI scans, a total of 1004 patients experienced claustrophobia, and 610 of these patients required intravenous sedation before their scans (Dewey et al., 2007). Even though sedation allowed the patients mentioned in this paragraph to undergo an MRI scan, the authors of both studies note that sedation can result in adverse events and advise the reader to avoid patient sedation if possible.

Sedation can be used with pediatric patients, though the risks are more significant than with adult patients. Studies have shown that sedation for pediatric imaging can lead to hypoxemia and inappropriate sedation levels during image acquisition (Malviya et al., 2000). Pediatric patients can also expect “motor imbalance and gastrointestinal effects,” as well as agitation and restlessness for several hours after waking from sedation.

A report from the American Academy of Pediatrics and the American Academy of Pediatric Dentistry outlines the minimum set of criteria needed for a pediatric patient to be sedated for a procedure (Coté and Wilson, 2016):

- The patient must be a suitable candidate for sedation based on their medical history and medical needs.
- The patient’s health status must be evaluated and verified by the sedation team before the procedure.
- Informed consent must be obtained before the procedure.
- Instructions for what to expect and how to transport the patient home safely must be provided to the patient’s responsible adult.
- At least one responsible adult must be with the patient at the medical facility. Furthermore, the report recommends that two adults are present for patients who travel to and from the facility using car seats. This practice ensures that one adult can monitor the patient after the procedure while the other adult drives.
- The patient’s food and drink intake before the procedure should be taken into account to minimize the risk of pulmonary aspiration.
- The clinician administering the sedation must have immediate access to emergency facilities, personnel, and equipment and should monitor the patient for adverse events, including respiratory events, seizures, vomiting, and allergic reactions.
- There must be a clear protocol outlined for immediately accessing these emergency services.
- Emergency equipment and drugs appropriate for the patient’s size and age must be immediately available in case the patient needs to be resuscitated.
- The information about the procedure must be correctly documented.

- The facility should have a dedicated recovery area, and the status of the patient should be recorded when he is discharged. The patient should not be discharged if his levels of consciousness and oxygen saturation do not meet recognized guidelines.
- The patient may be held at the facility for prolonged monitoring after the procedure.

This report clearly states that the levels of monitoring suggested above should serve as minimum levels of involvement: clinicians should increase patient monitoring as needed for complex cases. Rutman has a similar and detailed perspective on patient monitoring during and after sedation, adding that two independent medical personnel should be present during the scan, and one of them should be present until the patient is discharged (Rutman, 2009). Rutman also notes that all sedation and monitoring equipment must be MR compatible, which is a simple but important safety constraint. This constraint makes sedation less advisable if the appropriate equipment is not available.

Sedation in neonatal and infant populations is not recommended. The U. S. Food and Drug Administration (FDA) issued a warning in late 2016 about repeated use of sedation or general anesthesia for patients under three years of age or pregnant women during their third trimester (United States Food and Drug Administration, 2016). The warning states that while a single, relatively short exposure to sedative and anesthetic drugs is unlikely to impact the patient, the effects of prolonged exposure to these drugs are still being studied. Studies of sedative and anesthetic drugs in multiple animal models have shown that these drugs can lead to loss of nerve cells in the brain when the animals undergo prolonged, repeated exposure to them during this period of brain development. More data is needed to determine if this effect translates to humans.

### **2.5.3 During Scan: Feed and Sleep Protocols**

Neither sedation nor educational and behavioral techniques are appropriate to use with neonatal patients, but rs-fMRIs in neonates and infants are invaluable in studying early brain development and neurological diseases (Smyser and Neil, 2015). A set of protocols has been developed specifically for scanning neonates without sedation. These protocols are referred to as “feed and sleep” or “feed and bundle” protocols.

Windram et al. describe a protocol in which the infant is deprived of food for four hours before the scan (Windram et al., 2011). At the scanning facility, the patient is fed by his mother, swaddled, and placed in a vacuum-bag immobilizer for the duration of the scan.

Rather than deprive the patient of food before the scan, Gale et al.’s protocol recommends timing the scan so that the patient is fed after arrival on-site and less than 45 minutes before the scan (Gale et al., 2013). The patient’s ears are protected from the noise of the MR scanner by a layer of dental putty followed by headphones and held in place by a hat. The patient is then swaddled and placed in the scanner once he is asleep. Additional foam padding is used to cushion the patient’s head and provides extra noise protection.

Mathur et al. describe a protocol similar to the previous two: the patient’s feeding schedule is adjusted so that he feeds 30-45 minutes before the scan time, and he is swaddled, given ear protection, and placed in a vacuum-bag immobilizer (Mathur et al., 2008).

When performed correctly, these protocols are generally successful, and the neonatal patient will sleep for the duration of the MRI scan. However, the patient may shift slightly while asleep or may wake up and move mid-scan.

## 2.6 Prospective Motion Correction

Since motion cannot be completely eliminated from rs-fMRI scans, different approaches have developed for correcting for the effects of motion after the scan. These approaches can be divided into two groups: those that monitor the patient’s motion during the scan and those that work solely on the acquired sequences.

### 2.6.1 Optical Motion Correction

Several groups have developed methods for actively accounting for changes in the patient’s position during an MRI scan. Optical-based methods record the patient’s position using a combination of markers placed on the patient, and one or more MR compatible optical cameras placed the scanner bore. The changes in the patient position from one timepoint

to the next are used to update the MR parameters in real-time. Real-time updates of the MR parameters result in decreased spatial and spin-history effects of motion in the acquired sequences.

The first report of successful prospective motion correction using optical cameras and markers was by Zaitsev et al. in 2006 (Zaitsev et al., 2006). Their dual-camera system was located outside of the MRI scanner and focused on the patient inside the system. Four reflective markers were attached to a modified mouthpiece initially designed for patient immobilization. Changes in the translation and rotation of the patient were recorded and processed during the exam. The processed changes were sent in real-time to the MRI scanner, which used them to update the gradient orientations, RF frequencies, and RF phases at every time point during the acquisition process.

Aksoy et al. simplify this approach by using a single in-bore optical camera and replacing the 3D markers with a small 2D chessboard grid (Aksoy et al., 2008). Intrinsic camera properties, as well as information about the camera's placement within the MRI scanner, were recorded prior to the scan as part of a calibration process. During the scan, patient movements recorded using the optical camera were used to calculate the relationship between the patient's position at the current time point in the physical space and the patient's position at the initial time point in the MR space. The transformation needed to translate between these two positions was calculated on a laptop and passed to the MRI scanner to correct for motion in real-time. The camera used to record the position of the chessboard is mounted on the head coil. If the patient moves his head significantly, the camera will only be able to record the position of part of the chessboard marker. This limitation makes it difficult for the computer vision processing to identify the independent features on the standard chessboard.

Forman et al. modified the chessboard marker to improve its use for high-motion patients (Forman et al., 2011). To differentiate between the different blocks in the chessboard, they added a unique, machine-readable symbol to each black block in the chessboard. The symbols were chosen to be unique even in the event of rotation so that the identification of each block would be robust to rotation movements. The chessboard marker was embedded with MR-detectable agar so that the position of the marker could be detected in the MRI scan as well as by the in-bore camera. At each point during the scan, the image recorded by the in-bore

camera was sent to a computer independent from the MRI controller. The independent computer detected the blocks of the chessboard and identified their spatial locations using the symbols contained within them. Their positions were checked by confirming the locations of the symbols with respect to each other. The confirmed locations of the corners of the black boxes were used to estimate the position of the patient, which was then sent to the MRI controller so that the magnetic gradients and RF hardware could be updated for the time point. The authors note that the latency of the system is a significant limitation to their system, but overall they experienced an increase in the accuracy of the estimates of the patient's position.

Several companies have developed commercial products for prospective motion correction in neurological images. KintetiCor's system uses a high-resolution camera and a physical marker to detect motion (KinetiCor Biometric Intelligence, 2019). The camera's resolution allows it to detect respiratory and cardiac motion through changes in skin displacement on the patient's forehead. The physical marker consists of a pair of rectangles containing several concentric circles that are connected via a bridge across the nose. Any patient movement is reflected in the movement of the markers, which is also tracked through the camera. Both the camera system and the marker are MR compatible. Another company, TracInnovations, uses a stereo camera system to track all patient motion (TracInnovations, 2019). At the start of the scan, the stereo camera obtains a point cloud (a set of single point coordinates in space) specifying the patient's position at that time. The points in the point cloud are averaged together to create a primary marker. Small facial motions, cardiac motion, and respiratory motion are monitored using the point cloud. Larger head motions are monitored using both the point cloud and the primary marker. These two systems both allow prospective motion correction to be turned on or off. If the prospective motion correction is off, the system will still acquire the motion parameters so that the motion can be corrected retrospectively.

The methods and technologies discussed above have a few limitations due to the optical camera setups. For precise real-time motion correction, the camera or cameras must be carefully placed so that the position of the marker on the patient can be recorded. They must have a clear line of sight, which means they will be in the same room as the MRI scanner, if not within the scanner bore. The cameras and markers must be MR compatible,

and the positions of the cameras and markers in physical space relative to the visual markers on the patient must be known. These positions are vital for the calculations used to measure the motions. Even if the motion measurements are accurate, the changes in position that are recorded and used to adapt the scan parameters will only be valid for rigid body motion of the body part to which the markers are attached: any distortion of soft tissue will not be accurately accounted for during the motion correction unless the camera system was explicitly built for and trained to do so.

Systems using markers attached to patient suffer from these limitations, but also from limitations due to the markers themselves. If the marker is not attached correctly to the patient, the marker can slip and move independently from the patient. The whole purpose of the markers is that they provide a known set of visual features that can be used to determine how a patient moved because they moved with the subject.

### 2.6.2 Non-Visual External Sensors

Cameras are not the only type of external sensor that can be used to measure motion during an rs-fMRI scan.

There is a class of sensors that can take advantage of the electrophysical properties of an MRI scanner. These sensors include wired nuclear magnetic resonance field probes, wireless inductivity coupled markers, and off-resonance markers. The fact that these sensors directly interact with the magnetic field of the MR scanner means that protocols using these sensors must be modified to account for them. As a result of the protocol modification, the scan time might need to be extended.

As mentioned earlier in this chapter, respiration and cardiac activity are sources of patient motion. The most straightforward way to reduce the effect of respiratory motion is to instruct the patient to hold their breath at specific intervals during the scan. One alternative to breath-holding uses the periodic nature of respiration. Since respiration is relatively periodic, it can be monitored and accounted for within a scan protocol via gating. A comparative study of breath-holding and respiratory gating in MRIs of the coronary arteries suggests that images acquired using respiratory gating had 76% better quality than

the images acquired using breath-holding (Hofman et al., 1995).

In cases where only respiration is taken into account, gating prevents an image from being acquired unless the patient is in the specified state. In the case of respiration, the specified state is either complete inhalation or exhalation. The state of a patient’s respiration can be tracked using respiration bellows. After acquiring the MRI sequence, volumes in the sequence can be grouped depending on when they were recorded in the breathing cycle. By only using volumes recorded during the same stage of the breathing cycle, the effects of respiratory motion can be mitigated. This approach to gating will increase the amount of scanner time needed.

When both cardiac and respiratory motions are considered, the patient’s respiration is monitored through a bellows or pressure belt, and their cardiac activity is monitored on a delay during the scan through a pulse oximeter on the patient’s finger (Hu et al., 1995). The cardiac and respiratory data are synchronized with the acquired image after the scan. The synchronized signals are then used to remove changes in the image associated with physiological motion, which “substantially reduces image-to-image fluctuations” (?).

For macro-scale motions, electromagnetically sensitive trackers can be placed on the patient and used to monitor changes in position during the scan. Afacan et al. used a tracker produced by Robin Medical Inc. (Baltimore, MD) to measure the position and orientation of subjects relative to the center of the scanner bore (Afacan et al., 2016). The tracker consisted of two sensors that were attached to the patient’s forehead. The sensors recorded their  $x$ ,  $y$ , and  $z$  coordinates as well as two vectors indicating their orientations at every magnetic activation. These measurements were processed and displayed for a technician in real-time during each scan.

Ultimately, using external sensors to monitor the patient for inherent physiological motion can improve the quality of the obtained MR images. However, the addition of extra sensors complicates the process and set up of rs-fMRI scans.

### **2.6.3 Image Signal Motion Monitoring**

Dosenbach et al. have developed a tool to evaluate motion in rs-fMRI sequences as they are acquired (Dosenbach et al., 2017). It registers each volume to the initial volume of the rs-fMRI sequence immediately after the new volume is recorded. The parameters produced by this registration are used to calculate the framewise displacement between pairs of volumes, which is then compared to a set of displacement thresholds associated with the scan quality. The number of volumes that meet each threshold is used to determine how many more volumes are needed to obtain five minutes of low motion volumes. This method for assessing the quality of a scan in real-time is useful for ensuring images are acquired with a sufficient number of low-motion volumes. It can also aid the technologists in determining whether to prematurely terminate a scan, which may be desirable if the amount of time needed to obtain enough low motion volumes is greater than the amount of time remaining for the patient in the scanner.

### **2.6.4 General Limitations of Prospective Motion Correction**

All types of prospective motion correction introduce a delay into the scanning process. The delay is due to the additional processing of some metrics to determine the patient's position, the transmission of these metrics to the MR scanner, and the adjustments the scanner makes to its next set of measurements. These alterations to the image acquisition during prospective motion correction actively change the image as it is acquired. Maclarens et al. note that while prospective motion correction reduces inhomogeneities in the  $B_0$  field, the  $B_0$  field will still change when the patient moves and may change while the motion correction is occurring (Maclarens et al., 2013).

In order to view a scan not impacted by prospective motion correction, the patient often must undergo a second scan. It may be wise to build the second image acquisition into the same scan period as the prospectively motion-corrected scan: unsuccessful prospective motion correction has the potential to drastically corrupt the acquired scan (Zaitsev et al., 2017).

Finally, though prospective motion correction has great power for managing motion

during a scan, it cannot be used to recover motion-corrupted data in existing data sets.

## 2.7 Retrospective Motion Correction

Many groups have put significant effort into developing techniques for motion correction after the scan is acquired. Here, we discuss several commonly used techniques: volume registration, denoising, and filtering.

### 2.7.1 Volume Registration

The rs-fMR image is stored in computer memory as a set of 3D matrices. The values in corresponding cells of each matrix are considered to be aligned in this digital space (voxel space). The voxel space is defined by the imaging protocol and relates to the physical space through the spatial resolution of the image. Even though the spatial and voxel spaces for the image align, the contents of the image volumes may be misaligned due to patient movement. Because we cannot assume that an image is entirely motion-free, we cannot directly compare the contents of each image volume in the rs-fMRI sequence. However, we can use image registration to align the contents of the image volumes to reduce the impact of motion on patient position.

Image registration is the process of morphing the contents of one image so that they overlap optimally with another image. The morphing operations include translation, rotation, scaling, skewing, and nonlinear adjustments. The linear and affine operations in this list should be used to perform rigid body registrations for organs such as the brain. Nonlinear operations can be used to fine-tune the alignment of more pliable organs such as the liver. All morphing operations are applied to one image repeatedly until its contents optimally match those of the static reference image as determined by a chosen similarity metric.

One of the earliest examples of image registration was described by Friston et al. in 1995 (Friston et al., 1995). They performed image registration on positron emission tomography (PET) scans and MRI scans of a human brain. During the registration process, one scan

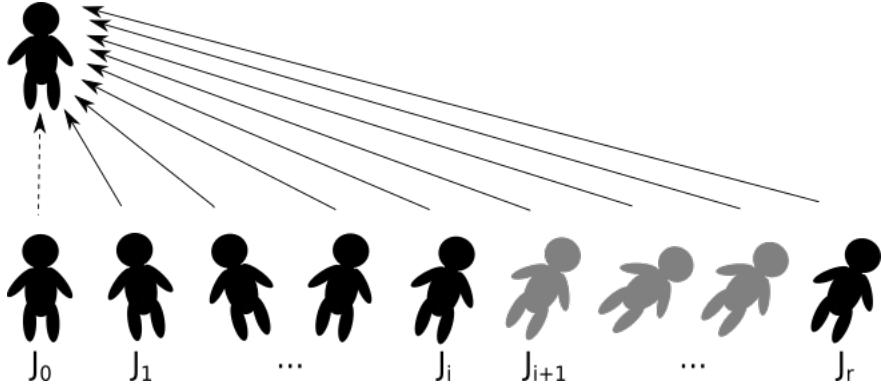


Figure 4: The traditional approach to volume registration in an rs-fMRI sequence consists of registering all volumes in the sequence to a single reference volume.

was designated as the “reference” image, which remained stationary, and the other scan was designated as the “object” image, which was transformed to match the reference image. Constraining the alignment process to transform a single image into the coordinates of the other image rather than transforming both images into an independent coordinate frame simplifies the registration process.

When performing image registration on a sequence of image volumes, one volume must be chosen as the reference volume for the entire sequence. All other volumes in the sequence are registered to this volume. An example of this process can be seen in Figure 4. In subsequent work, Friston et al. used the first volume in the rs-fMRI sequence as the universal reference image (Friston et al., 1996). Common choices for the reference volume include the volume with the least positional difference to all other volumes in the sequence, a volume produced by averaging all volumes in the sequence, or the first volume in the sequence (Friston et al., 1996) (Liao et al., 2005). In our implementation, we chose to use the first volume in the sequence as the reference volume.

One drawback to this traditional approach to volume registration is that it only minimizes the differences between all the image volumes in the sequence and the reference volume. The keyword here is “minimizes”: minimizing differences between image volumes does not mean that there are no differences between the image volumes. Image registration is an

optimization problem, and its goal is to find the overlap between a pair of volumes with as few differences as possible either within a defined number of iterations or until the optimization cost does not change above a certain tolerance for a certain amount of time. These practical constraints on optimization problems mean that there may still be differences between other pairs of image volumes in the sequence that do not include the reference volume.

Variations on Friston et al.’s framework have been developed over the last two decades. Liao et al. suggested that an rs-fMRI sequence could be viewed as a hidden Markov model, and reflected this idea in their suggested registration framework (Liao et al., 2016). They still use the first volume in the image sequence as the reference volume. Their framework uses the transformation of the previous volume to the reference volume to initialize the transformation for the current volume and the reference volume.

It has been demonstrated that image registration across the entire image sequence reduces the effects of motion on the image sequence, though they do note that motion also affects the image due to changes in the spin history of the image. These effects are not correctable by global volume registration alone and will be discussed later in this chapter.

### 2.7.2 Denoising

Denoising techniques can be applied to an rs-fMRI after global volume registration is completed. They consist of regressions of various confounding variables.

It is relatively common for researchers to use rigid realignment parameters calculated during volume registration as signals to remove via regression. The realignment parameters are the translations along and the rotations about the  $x$ ,  $y$ , and  $z$  axes (Power et al., 2012). These six parameters and their first-order derivatives are suggested as regression parameters by several researchers (Power et al., 2012) (Satterthwaite et al., 2012) (van Dijk et al., 2012).

While the realignment parameters and their derivative for the current image volumes help reduce the effects of motion on an image sequence, these parameters are not sufficient on their own. As the effects of motion on rs-fMRIs have been studied, it has been established that motion at one point in the sequence can affect the next several volumes. Researchers have decided to address this effect by incorporating the rigid realignment parameters from

surrounding timepoints into the regression parameters for any given current image volume (Power et al., 2014) (Satterthwaite et al., 2013) (Yan et al., 2013b).

Patriat et al. performed a robust comparison of different regression parameters on their MotSim motion data set (Patriat et al., 2017). They included rigid realignment parameters, but also used parameters obtained by performing principal component analysis (PCA) on the image sequences. PCA generates a set of linear, uncorrelated components that reflect the main features of a patient’s motion. The list of parameter combinations included

- 12mot: The six rigid realignment parameters and their first derivatives,
- 12for: The first 12 principal components of the whole brain before realignment,
- 12back: The first 12 principal components of the whole brain after realignment,
- 12both: The first 12 principal components of the whole brain both before and after realignment,
- 24mot: the six rigid realignment parameters of the current volume, the six rigid realignment parameters of the previous volume, and the square of these rigid realignment parameters,
- 24both: the first 24 principal components of the whole brain before and after realignment.

They found that the features extracted from the image sequence using PCA explained more variance in the image sequence (measured using  $R^2$ ) than the rigid realignment parameters. They showed that increasing the number of regressors increased the amount of variance explained, but with diminishing returns. While their work is promising, their experiment was performed on a simulated data set using healthy subject data and required an accurate estimate of the subject’s head motion.

In addition to the positional effects of motion, the spin history and susceptibility effects of motion must be considered. Signals associated with these effects are called nuisance signals. One common nuisance signal is the global signal of the image sequence. Global signal regression (GSR) corrects for variance between temporal signals within a voxel and for the mean BOLD signal across all voxels (Power et al., 2014) (Satterthwaite et al., 2013) (Yan et al., 2013a) (Yan et al., 2013b). GSR has been shown to reduce spuriously increased long-distance correlations in functional connectivity studies, but may inadvertently weaken

shorter-distance connections (Jo et al., 2013) (Power et al., 2014) (Satterthwaite et al., 2012). Other nuisance signals that can be removed via regression are signals white matter and cerebral spinal fluid (CSF) (Power et al., 2014) (Satterthwaite et al., 2013). The impact of removing these signals from the image sequence is limited: in some cases, removing white matter and CSF signals do not reduce the effects of motion on the BOLD signal (Yan et al., 2013b) (Jo et al., 2010).

Another set of signals that may be removed via regression are components identified using techniques such as principal component analysis (PCA) or independent component analysis (ICA) (Pruim et al., 2015) (Salimi-Khorshidi et al., 2014) (Behzadi et al., 2007). Though both PCA and ICA decompose a set of data into a list of signals, the properties of the lists are different. PCA produces a list of orthogonal signals that best represent the features of a data set ordered from most representative to least representative. ICA treats the problem of decomposing the image sequence into different source signals as a blind source separation problem. ICA separates components by maximizing their statistical independence but requires additional help to identify the correct number of components. The set of components produced by ICA are assumed to be independent, non-Gaussian, and less complex than the original signal. For more details about ICA, refer to Section 3.2. Regression of each of these sets of signals has been shown to reduce the effects of motion in the sequence, but neither removes them entirely (Power et al., 2015) (Parkes et al., 2017).

### 2.7.3 Filtering

Not all patients move in the same ways or at the same rate. In some cases, a patient will perform a large and sudden movement and then settle into a position close to their original position. The volumes containing the motion can be removed from the rs-fMRI sequence to create smaller, motion-free sequences. Several studies have found that subsequences of data of the same patient can be concatenated to produce a single, longer sequence without impacting the functional connectivity analysis. The process of detecting and removing motion-corrupted volumes has been formalized into several filtering techniques. Three popular filtering techniques are scrubbing, spike regression, and despiking.

**Scrubbing** begins by examining the image sequence for “high-motion” image volumes (Power et al., 2012). Here, “high-motion” is defined as image volumes that have 0.5 mm FD and 0.5% DVARS difference from their preceding volume. The volumes containing at least this quantity of motion, the preceding volume, and two subsequent volumes, are removed from the sequence, resulting in a set of discontinuous subsequences. The discontinuous subsequences can be combined without negatively impacting the functional connectivity analysis of the BOLD signal (Fair et al., 2007) (van Dijk et al., 2010). As long as the reconnected subsequences compose a sequence at least 125 frames (approximately 5 minutes) in length, the scrubbed sequence can be used for analysis (Power et al., 2012). Many techniques for temporally filtering high-motion rs-fMRIs have been developed, though all of them result in the loss of image volumes (Barnes et al., 2011) (Fransson et al., 2007) (Jones et al., 2010) (Kennedy and Courchesne, 2008) (Smyser et al., 2010) (Smyser et al., 2011).

**Spike regression** also identifies volumes containing large quantities of motion, though it treats them differently than scrubbing does. (Satterthwaite et al., 2013). It identifies spikes in the image sequence and replaces frames affected by the spikes with interpolated volumes.

A spike is defined as a frame in the sequence where the FD or the FD and the DVARS between it and its previous frame surpass a given set of thresholds. The value of the thresholds greatly impacts the cleaned image sequence. A low threshold will identify more spikes and produce a cleaner image sequence but remove more data, while a high threshold will retain more data but contain more motion artifacts. The thresholds used by Satterthwaite et al. on their adolescent data set are 0.25 mm FD and 1.4% signal units of DVARS.

After identifying the spikes in an image sequence, they are modeled as signals to remove through regression. The simplest model of a spike is as a motion event at a single time point. However, it has been established that patient motion may affect several frames in the image sequence. Satterthwaite et al. examined six combinations of frames around the frame in which the spike was based during their analysis of spike regression. They found that removing spikes identified using just the FD threshold and modeled as only effecting a single frame resulted in the fewest different neural connections in the highest and lowest motion images as well as the lowest correlation between functional connectivity and patient

motion (Satterthwaite et al., 2013).

**Despiking** is different from these techniques. It treats the image sequence as a single image volume where each voxel contains an intensity signal. Each intensity signal is examined for sudden changes or spikes. The value of each detected spike is replaced with an interpolated value calculated using the preceding and subsequent points in the voxel’s signal (Jo et al., 2013) (Patel et al., 2014). Despiking does not remove volumes, but could accidentally remove valuable signals if they appear as outliers in the image voxel signals.

#### 2.7.4 Spin History Distortion Correction

Many post-acquisition methods have been developed specifically to correct for distortions due to the impact of motion on the magnetic field. The usability of these dynamic distortion correction methods has been studied in a few specific cases, but their generalizability has yet to be confirmed in a broader range of fMRI studies (Zaitsev et al., 2017).

### 2.8 Summary

Resting-state fMRIs are four-dimensional images that record the BOLD signal in active areas of the brain. The BOLD signal can be used to evaluate the functional connectivity of different underlying networks in a patient’s brain. Since rs-fMRIs are highly sensitive to motion, clinicians and psychologists have devised techniques to inform patients about what they can expect during an MRI scan as well as different coping mechanisms to help them remain calm during the scan. These techniques do not prevent the patient from moving, but approaches that do are not always appropriate to use during an rs-fMRI scan. Techniques and algorithms to prospectively and retrospectively remove motion from rs-fMRIs have also been developed, though they are not always successful in removing the effects of motion. Ultimately, the amount of motion present in the rs-fMRI sequence dictates whether or not the sequence can be used in clinical or research applications.

In the next chapter, we will discuss rs-fMRIs in the context of our chosen population of

CHD patients.

## 3.0 METHODS: Motion Correction

In the previous chapter, we discussed several techniques used to retrospectively correct motion. Motion correction pipelines may use denoising and filtering, but all pipelines begin with volume registration. In this chapter, we discuss a different approach to volume registration, how it compares to traditional volume registration, and how volume registration fits into a motion correction pipeline.

### 3.1 Directed Acyclic Graph Based Volume Registration

As discussed previously, the major drawback to Friston et al.’s approach to volume registration is that it only minimized the positional differences between the reference volume and the rest of the sequence. This drawback demonstrates an inability for the traditional approach to account for relationships in the patient’s position throughout the scan. Intuitively, we know that the patient’s position at any volume in the scan is more similar to his position in the immediately previous or subsequent volume than to another randomly chosen volume in the image.

In our proposed framework, we wish to account for these spatiotemporal relationships between temporally neighboring volumes in the sequence. To accomplish this goal, we start by viewing the rs-fMRI sequence as a directed acyclic graph (DAG). A DAG consists of a set of nodes and edges. Each edge has a direction associated with it and connects a pair of nodes. Since a DAG contains no cycles, there is no possible path back to a node once it has been traversed.

In the case of an rs-fMRI, each volume can be considered a node. The relationship between each pair of temporally neighboring volumes is represented as a directed edge connecting the node for the first volume to the node for the next volume. The acyclic nature of the DAG means that once a patient was in a specific position, he never returns to that precise position with the exact same neurons firing. The position of the subject and his brain

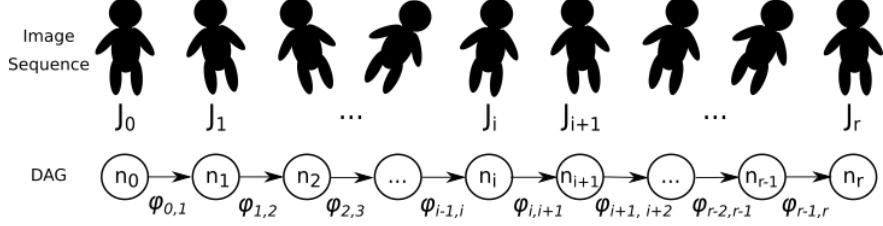


Figure 5: A rs-fMRI can be viewed as a directed acyclic graph where each volume is a node and the edges connect from each volume  $i$  to the following volume  $i + 1$ .

activity, as measured by the BOLD signal, may be similar in subsequent image volumes, but it will never be precisely the same. The perspectives of an rs-fMRI sequence as a set of images and as a DAG can be seen in Figure 5.

The cost of transitioning from one node to the next in our DAG has a parallel representation to the combination of the positional transformation needed to align volume  $i$  to volume  $i + 1$  and the signal change between the volumes. This representation can be written as

$$J_{i+1} = \phi_{i,i+1} J_i + \delta s_{i,i+1} + \epsilon \quad (3.1)$$

where  $J_i$  and  $J_{i+1}$  are volumes  $i$  and  $i + 1$ ,  $\phi_{i,i+1}$  is a matrix of transformation parameters that must be applied to  $J_i$  to achieve the patient's position in  $J_{i+1}$ ,  $\delta s_{i,i+1}$  is the natural change in BOLD signal, and  $\epsilon$  is the change in BOLD signal due to motion. Currently, there is no way to estimate the natural change in the BOLD signal and the change in the BOLD signal due to motion without incorporating additional information about the MRI scanner and the patient that is not included in an rs-fMRI. We simplify our representation of the relationship between two volumes to:

$$J_{i+1} = \phi_{i,i+1} J_i + \epsilon^* \quad (3.2)$$

where  $\epsilon^*$  is the change in the BOLD signal that cannot be accounted for after aligning the patient's position in the two volumes. Here, we use the notation  $\epsilon^*$  to represent the generic

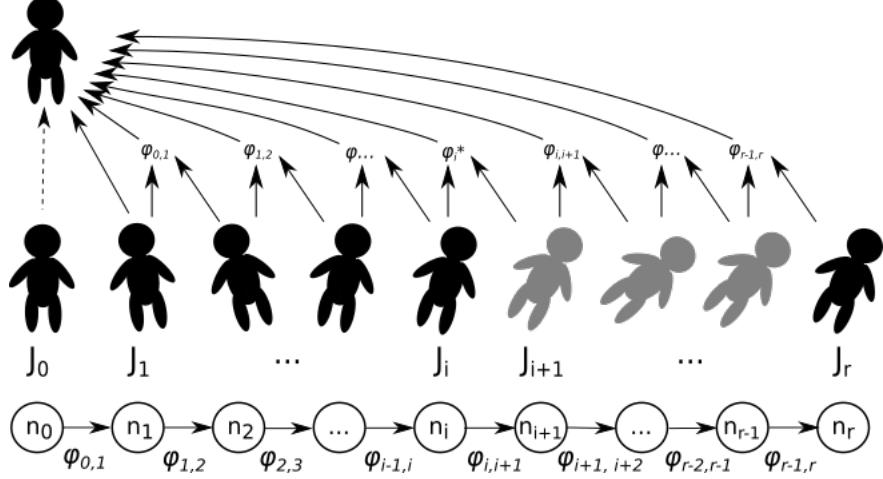


Figure 6: The DAG-based approach to volume registration in an rs-fMRI sequence consists still registers all volumes in a sequence to a single reference volume, but it also accounts for the temporal relationships between subsequent volumes.

error change in a BOLD signal across any pair of volumes.

After aligning two volumes  $i$  and  $i + 1$ , we will then align volumes  $i + 1$  and  $i + 2$ :

$$\begin{aligned}
 J_{i+2} &= \phi_{i+1,i+2} J_{i+1} + \epsilon^* \\
 &= \phi_{i+1,i+2} (\phi_{i,i+1} J_i + \epsilon^*) + \epsilon^* \\
 &= \phi_{i+1,i+2} \phi_{i,i+1} J_i + \epsilon^{*'}
 \end{aligned} \tag{3.3}$$

Traditional volume registration assumes that

$$\phi_{i,i+2} = \phi_{i+1,i+2} \phi_{i,i+1} \tag{3.4}$$

and calculates  $\phi_{i,i+2}$  directly. We argue that this assumption is not true in all cases. Rather than directly calculate  $\phi_{0,i}$  and use it to align volume  $i$  to the reference volume as the traditional method does, we calculate each component  $\phi$  that is a factor of  $\phi_{0,i}$ . Each component  $\phi_{i,i+1}$  is combined with the preceding  $\phi_{0,i}$ s to recursively align volume  $i + 1$  to the reference volume without making the large and often inaccurate transformations required by directly calculating  $\phi_{0,i+1}$ . This process is outlined in Figure 6.

### 3.2 Independent Component Analysis

The purpose of image registration is purely to ensure the position of the patient throughout the entire rs-fMRI is consistent. After registration, the image still contains BOLD source signals and noise signals caused by factors other than brain activity. The challenge of separating these combined signals is called blind source separation (BSS).

We chose to focus on an independent component analysis (ICA) approach for solving the BSS problem. The specific technique we use has been described by Beckmann and Smith as probabilistic ICA. This section aims to provide an overview of the probabilistic ICA technique. For further details, please refer to the technical reports by the FMRIB group (Beckmann and Smith, 2004) (Woolrich et al., 2004) (Beckmann et al., ) (Smith et al., 2004).

Probabilistic ICA is a linear regression model that performs mixing in the original data space and assumes the true BOLD signal has been obfuscated by Gaussian noise. These constraints mean that BSS can be solved in three steps:

1. Estimate a joint subspace consisting of source and noise signals and a noise subspace orthogonal to the joint subspace,
2. Estimate the independent sources in the joint subspace, and
3. Assess the statistical significance of the independent sources.

Probabilistic ICA treats the voxel intensity values in every frame of the image sequence as a matrix of  $V$  voxels across  $n$  time points. For each voxel  $v_i \in V$ , the observed signal in that voxel can be modeled as

$$\vec{x}_i = A\vec{s}_i + \mu + \vec{\eta}_i \quad (3.5)$$

This equation allows three different types of signals to contribute to the observed voxel values  $\vec{x}_i$  for a given voxel across all  $n$  timepoints in the sequence. The first type of signal is a vector of non-Gaussian source signals  $\vec{s}_i$  across all  $n$  timepoints. The source signals are modulated by mixing matrix  $A$  whose shape is the number of time points  $n$  by the number of source signals  $q$ . The second type of signal is an offset denoted by  $\mu$ . The offset constrains the

observed signals to be centered around the mean of all observed signals. The third type of signal  $\vec{\eta}_i$  is a vector of noise throughout the duration of the sequence.

To summarize, probabilistic ICA explicitly assumes that the observed signal in a given voxel can be divided into non-Gaussian source signals, isotropic Gaussian noise signals, and some offset. This assumption makes it easier to separate a source signal from a noise signal: noise signals have Gaussian distributions while source signals do not. **The goal of probabilistic ICA is to identify the source signals,  $\vec{s}$ .**

With this combination of signals in mind, we can write the covariance matrix of the observed data  $x$  as

$$R_x = \langle x_i x_i^T \rangle = AA^T + \sigma^2 I \quad (3.6)$$

where  $A$  is the mixing matrix,  $\sigma^2$  is the standard deviation of the noise, and  $I$  is  $n \times n$  identity matrix. The covariance matrix of the observed data  $R_x$  can be calculated, but  $A$  and  $\sigma^2$  are both unknown. The noisy observed data is transformed with respect to the noise sources using a process called whitening. The whitening with respect to noise enforces the assumption of noise following an isotropic Gaussian distribution with a mean of zero and a standard deviation of  $\sigma^2$ .

The mixing matrix  $A$  can be estimated using maximum likelihood estimation. Beckmann and Smith use singular value decomposition of the observed data  $X = U(N\Lambda)^{\frac{1}{2}}V$  to model the estimator of  $A$ :

$$\hat{A}_{ML} = U_q(\Lambda_q - \sigma^2 I_q)^{\frac{1}{2}}Q^T \quad (3.7)$$

where  $U_q$  contains the eigenvectors associated with the  $q$  largest eigenvalues,  $\Lambda_q$  contains the  $q$  largest eigenvalues, and  $Q$  is a  $qxq$  orthogonal rotation matrix in the whitened observation space such that  $QQ^T = I$ . The eigenvectors and eigenvalues can be calculated from  $X$ , but  $\sigma$  and  $Q$  remain unknown. As noted earlier, the matrix  $Q$  is an orthogonal rotation matrix which, when applied to the whitened data  $\tilde{x}$ , has the same effect of applying an unmixing matrix to the observed data:

$$W\vec{x} = Q\tilde{x} = \hat{s} \quad (3.8)$$

Both matrix-vector multiplications serve to estimate individual source signals  $\hat{s}$ . The estimated source signals are identified by projecting the whitened data  $\tilde{x}$  onto each row  $r$  of the unmixing matrix  $Q$  a total of  $q$  times:

$$\hat{s}_r = Q_{r,:}\tilde{x} \quad (3.9)$$

where the  $Q_{r,:}$  represents row  $r$  of matrix  $Q$ . (*Note: A key assumption in this step is that the rows of the unmixing matrix are mutually orthogonal so that they cover the entire space of signal sources. Additional steps described by Beckmann and Smith can be taken to incorporate prior information about the voxels into this step (Beckmann and Smith, 2004).*)

At this point, the standard deviation of the noise  $\sigma^2$  and the source signals are unknown. We can solve the following system of equations jointly to resolve these two unknown quantities:

$$\hat{s}_{ML} = (\hat{A}^T \hat{A})^{-1} \hat{A}^T x = \hat{W}x = Q\tilde{x} \quad (3.10)$$

$$\hat{\sigma}_{ML}^2 = \frac{1}{n - q} \sum_{l=q+1}^p \lambda_l. \quad (3.11)$$

Solving these equations is an iterative process. First, the mixing matrix and source signals are estimated. These estimations are used to calculate the corresponding estimator of the standard deviation of the noise. Then, the residual noise  $\hat{\eta}_i$  at each voxel  $v_i$  is calculated:

$$\hat{\eta}_i = (I - \hat{W}^T \hat{W})x_i. \quad (3.12)$$

Recalling from Equation 3.5 how probabilistic ICA views a signal, Equation 3.12 becomes:

$$\hat{\eta}_i = (I - \hat{W}^T \hat{W})A + (I - \hat{W}^T \hat{W})\eta \quad (3.13)$$

When the correct number of sources has been identified, the estimated mixing matrix fully spans the source signal space. Then, the residual noise is only related to the true noise:

$$\hat{\eta}_i = 0 + (I - \hat{W}^T \hat{W})\eta \quad (3.14)$$

Upon reaching this stage in the probabilistic ICA technique, the source signals have been approximated. The source signals are called spatial independent component maps. Normalizing the values in these maps by the variance of the noise produces  $Z$ -statistic maps.  $Z$ -statistic maps can be analyzed to identify voxels with statistically significant activations. These activations are attributed to the BOLD signal.

One of the major limitations of ICA is that it is highly data-driven. It assumes the dataset contains a sufficiently large number of images, each with a sufficiently large number of voxels. Even assuming an ideal data set, the true value of the mixing matrix is dependent on the observed data (Beckmann and Smith, 2004). Fluctuations in the data can lead to deviations of the residual noise in certain voxels from the true noise. These deviations can produce *type-I* and *type-II* errors when examining the  $Z$ -statistic maps to identify statistically significantly activated voxels.

Additionally, the developers of probabilistic ICA note that not all noise follows the isotropic Gaussian assumption. Noise based on the patient's physiology is likely to be structured in a way that is non-Gaussian. The non-Gaussian noise signals can still be separated from the BOLD source signals, but only if these noise signals are not highly correlated with the source signals.

### 3.3 Motion Correction Pipeline and Implementation

Both the traditional and novel volume registration techniques were applied independently to each image from the subject cohorts described in Chapter 5. After registration, three versions of each image existed: the original BOLD sequence, the sequence modified using traditional volume registration, and the sequence modified using the novel registration method.

The registration algorithms applied to rigid tissue types used affine registration with two degrees of granularity. When applied to soft tissue types (i.e., placenta), three nonlinear transformations with increasing granularities were performed after the affine registrations. The exact parameters used for each volume registration can be seen in Appendix A A. The

registration frameworks were implemented in Python using the nipype (Neuroimaging in Python Pipelines and Interfaces) library (Gorgolewski et al., 2011). Volume registration used the ANTs (Advanced Normalization Tools) tools as a backend (Avants et al., 2014).

After performing volume registration to ensure the patient is in the same physical space throughout the image sequence, the image sequence may still contain artifacts due to motion. Our registered sequences underwent motion correction via a well-established motion correction pipeline. We chose to use the independent component analysis (ICA) pipeline outlined by Beckmann and Smith (Beckmann and Smith, 2004). The motion-corrected sequences produced by FMRIB’s MELODIC tool were saved alongside the original and registered sequences.

### 3.4 Evaluating Registered and Motion Corrected Sequences Against Gold Standard Usability Thresholds

The main goal of motion correction is to reduce the effects of motion on the image so that it is usable. The gold standards for rs-fMRI usability as established by Power et al. are that the FD and DVARS metrics must change less than 0.2 mm and 2.5% normalized voxel units between at least 50% of the neighboring volumes. The FD and DVARS metrics between each pair of subsequent image volumes were calculated for the original, registered, and motion-corrected sequences. The metrics for each sequence were then compared to the gold standard image usability thresholds. This comparison answers the critical question of how each registration framework impacts an established motion correction pipeline.

Additionally, a smaller comparison of the registered sequences was conducted. This comparison evaluates the immediate impact of the registration algorithm on the image sequence. It is highly unlikely that an entire image sequence would meet the Power et al. usability thresholds after only the initial step of a motion correction pipeline. Still, it is valuable to examine the impact of a volume registration algorithm at each stage of the pipeline.

**Implementation.** We calculated the FD and DVARS metrics defined by Power et al. using the FSLMotionOutliers tool (Power et al., 2012).

## 4.0 METHODS: Evaluating Motion Patterns

In the previous chapter, we described the methods we use to mitigate the positional effects of motion in rs-fMRI sequences. We briefly discuss how the registered sequences were evaluated with respect to gold standard usability criteria. Here, we expand on our analysis of the motion extracted from the sequences.

### 4.1 Measuring Motion Patterns

While the Power et al. usability thresholds for the FD and DVARS metrics quantify the volume-to-volume motion well, they do not quantify the overall motion contained in the image sequence. The FD and DVARS metrics, as well as other imaging metrics, can be used to compare every volume in an image sequence to every other volume in the image sequence to quantify global motion more effectively. As the FD and DVARS metrics have been discussed previously, we focus in this section on two other image metrics: the Dice coefficient and the mutual information. These five metrics were applied to each whole sequence to measure patient motion and image signal consistency throughout the entire scan.

#### 4.1.1 Dice Coefficient

The Dice coefficient was proposed by Lee R. Dice in 1945 (Dice, 1945). Dice examined several existing metrics for measuring association, and finding them lacking, proposed his own “coincidence index”. His coincidence index measures the association between a set of samples  $a$  where condition  $A$  is true and a set of samples  $b$  where condition  $B$  is true:

$$Index = \frac{2h}{a + b} \tag{4.1}$$

In this equation,  $h$  represents the number of samples where both conditions  $A$  and  $B$  are true. His index can take on any value between 1.0 and 0.0 such that a value of 1.0 means that

conditions  $A$  and  $B$  are true for all samples. Similarly, a value of 0.0 means that conditions  $A$  and  $B$  are never both true for any sample. While this index is a count of samples that meet both conditions and not an actual probability, Dice suggests that the chi-squared test can be used to determine if the combinations of conditions in the samples from a set of data are meaningful or due to random chance.

Many medical imaging researchers have adapted the Dice coefficient to measure the overlap between pairs of images. Zijdenbos et al. trained an artificial neural network to semiautomatically segment brain MRIs and compared the generated segmentations to manual segmentations using the Dice coefficient (Zijdenbos et al., 1994). Zou et al. used the Dice similarity coefficient in their analysis of the reproducibility of manually segmented MRIs and the accuracy of automatic segmentations of the same images for prostate and brain tumor datasets (Zou et al., 2004). Liao et al. used it to measure the accuracy of a volume registration framework for aligning manual segmentations of multiple organs in fetal images (Liao et al., 2016). Bharatha et al. performed a study on pre- and intra-operative images of the prostate. They segmented the images and used the segmentations to generate deformable finite element models of the organs. They compared the registered segmentations and finite element models using the Dice coefficient (Bharatha et al., 2001).

It should be noted that the Dice coefficient, as used in these contexts, is a measure of similarity of items from two categories where each item belongs to one of two binary classes. The two classes of interest in the case of rs-fMRIs are “brain” or “not brain”. Medical images do not naturally have binary values. All studies mentioned in the previous paragraph required a domain expert to manually segment each image that was analyzed using the Dice coefficient. The manually segmented images are considered the gold standard to which automatic segmentations or registered images can be compared.

The images in our dataset have been manually curated to remove the skull and other anatomical features outside of it, but the images still contain a continuous range of voxel intensity values. The Dice coefficient cannot be directly applied to these image sequences, even though each voxel only belongs to one of two classes. The volumes first must undergo thresholding to create binary images to clearly separate the brain and the background for computational purposes. We use Otsu thresholding to accomplish this task.

Otsu thresholding divides the contents of an image into two binary classes based on the histogram of voxel intensity values. It assumes that the classes are represented in the histogram by separable peaks. It separates the peaks by finding the separation threshold with the best separation between classes. In its original form, Otsu thresholding exhaustively searches the space of all possible thresholds. For each threshold, it calculates the within-class variance and between-class variance for the pair of classes separated by the current threshold option. The ideal threshold is the one that produces the minimal within-class variance and the maximal between-class variance. The ideal threshold is then used to convert the original image into a binary image volume where background voxels have a value of zero, and voxels in the brain have a value of one. The binarized image volumes can be compared to each other using the Dice coefficient.

In cases where a good Otsu thresholded binary image cannot be obtained, other similarity metrics such as mutual information and cross-correlation should be used instead.

#### 4.1.2 Mutual Information

The earliest description of mutual information was written in the context of mathematical theories behind networked communication (Shannon, 1948). Mutual information is a measure of the amount of information shared between two signals  $X$  and  $Y$ . Specifically, mutual information measures how the joint distribution of the two signals compares to the marginal distribution of each signal (Li, 1990). It is a more general measure of dependence than correlation, which is limited to measuring linear dependence via a comparison of the marginal distributions. In terms of information theory, mutual information is represented as

$$MI(X, Y) = H(X) + H(Y) - H(XY) \quad (4.2)$$

where  $H(X)$  is the entropy of signal  $X$

$$H(X) = - \sum_{x \in X} p_x \log(p_x) \quad (4.3)$$

where  $p_x$  is the marginal distribution of the signal  $X$ . Substituting  $y$  for  $x$  in this equation produces  $H(Y)$ , the marginal entropy of the signal  $Y$ . Similarly,  $H(XY)$  is the joint entropy

of signals  $X$  and  $Y$  given the known signals  $X$  and  $Y$

$$H(XY) = - \sum_{x \in X, y \in Y} p_{xy} \log(p_{xy}). \quad (4.4)$$

It is worth noting that since the two signals of interest are registered images,  $x$  and  $y$  refer to voxel locations in the same image space. Substituting Equations 4.3 and 4.4 in Equation 4.2 produces the following

$$\begin{aligned} MI(X, Y) &= - \sum_{x \in X} p_x \log(p_x) - \sum_{y \in Y} p_y \log(p_y) + \sum_{x \in X, y \in Y} p_{xy} \log(p_{xy}) \\ &= \sum_{x \in X, y \in Y} p_{xy} \log p_{xy} - \left( \sum_{x \in X} p_x \log(p_x) + \sum_{y \in Y} p_y \log(p_y) \right) \end{aligned} \quad (4.5)$$

In the case where signals  $X$  and  $Y$  are independent, Equation 4.5 can be simplified to

$$\begin{aligned} MI(X, Y) &= \sum_{x \in X, y \in Y} p_{xy} \log(p_{xy}) - \left( \sum_{x \in X} \sum_{y \in Y} p_{xy} \log(p_x p_y) \right) \\ &= \sum_{x \in X, y \in Y} p_{xy} \log \left( \frac{p_{xy}}{p_x p_y} \right) \end{aligned} \quad (4.6)$$

Mutual information can be used to determine how the distribution of amplitudes in one signal relates to the distribution of another signal. It is commonly used in the medical imaging domain to objectively compare images of the same tissue taken using different modalities. For example, a computed tomography (CT) scan of a patient's abdomen contains different information about each tissue types' material properties than an MRI of the same organs. Some tissue types may appear similar in one of these modalities but drastically different in the other. Combining the information about a tissue's material properties gained from both imaging modalities provides more information than could be gained from either modality independently. (In other words, the resulting information is greater than the sum of its parts.)

Even though the rs-fMRIs in our study are all obtained using the same imaging modality, the spin history effects of patient motion can impact the recorded signal such that small changes in the recorded BOLD signals are difficult to distinguish from noise. We choose

to use mutual information to quantify to BOLD signal information across the entire image sequence.

#### 4.1.3 Implementation: Tools and Libraries

To calculate metrics, we used several existing tools and libraries. When no existing tool could be found, functions were implemented manually in Python3.

For the Dice coefficient calculation, we first had to create a binary version of each image volume. To binarize the image volumes, we used Simple ITK’s Otsu thresholding function. The binary images were then passed to a manually implemented function to calculate the Dice coefficient.

The correlation ratio between each possible pair of volumes in the sequence was calculated using bash and FLIRT (FMRIBs Linear Image Registration Tool) (Jenkinson and Smith, 2001) (Jenkinson et al., 2002). We then used the average and standard deviation of the correlation ratio distribution of each image to compare the images.

The mutual information calculation was implemented in two steps. The first step was to create a function that computes the joint histogram of the voxel value distributions between the two image volumes. This histogram was fed to a second function that converts the histogram counts to probabilities and calculates the mutual information value.

## 4.2 Patient Classification Using Motion Patterns

We suggest that patient movement patterns are specific to certain age groups. For example, fetal patients live suspended in amniotic fluid and, as such, are subject to different physical constraints than patients in other age groups. Neonatal patients are often scanned using a “feed and bundle” protocol, which often results in them sleeping through the scan. However, neonatal patients sometimes wake up during the scan, and the way a baby woken from a nap moves is different from how a fidgety preadolescent moves.

There is also a chance that patients within the same age group move differently, possibly

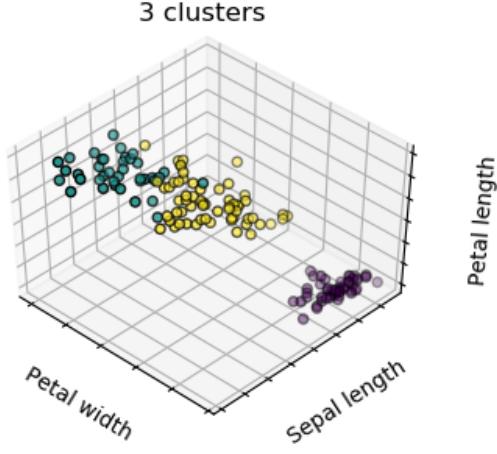


Figure 7: An example of  $k$ -means clustering performed on the Iris data set. The results of the algorithm are highly dependent on the number of clusters specified.

due to their cognitive state. Preadolescents who have ADHD likely become bored and fidgety in the MR scanner at different rates than their non-ADHD counterparts. Adults who have dementia may have more difficulty remaining still for the duration of a scan than adults from similar demographics with no dementia.

These patterns are essentially signals specific to different categories of patients. Machine learning techniques are useful for identifying patterns in signals from different sources. In addition to the motion metrics identified in the previous section, we also use demographic data as features for our machine learning models.

The goal of applying machine learning to identify population-level motion patterns lends itself well to unsupervised machine learning techniques. Unsupervised learning techniques combine samples from a population into groups based on the patterns in their features. They do not use information about any known groups in the population to inform their classification processes. In this section, we discuss several different unsupervised machine learning techniques used to measure degrees of association within subgroups of a data set.

### 4.2.1 K-means Clustering

K-means clustering divides a group of data samples with  $n$  features into  $k$  groups based on each sample's distance from the average value of the group (Hartigan and Wong, 1979), (MacQueen, 1967). In k-means clustering, the features of a set of data are viewed as the locations of each data sample in  $n$ -dimensional space. In this space,  $k$  cluster centroids are initially distributed. The distribution pattern can place the centroids either randomly between data samples or using randomly selected data points.

After the locations of the cluster centroids are initialized, the distance between each sample and each centroid is calculated. Each sample is assigned to the cluster represented by the centroid closest to it. Once the clusters are defined, the location of the centroid of each cluster is recalculated. The new centroid location is the mean of the locations of all samples in its cluster. The distance between each sample and each cluster centroid is recalculated and the samples are reassigned to their closest cluster centroid. Then the centroid of each cluster is recalculated. This process continues until a stopping criterion is fulfilled. With most unsupervised machine learning methods, the stopping criterion is that the classifications of the model do not change for a certain number of iterations. However, a maximum number of iterations is imposed on the learning process to prevent a model from running indefinitely. As a result, it is possible for a model to “time out” before reaching a stable state.

There are many variations of k-means clustering. For example, k-medians follows the same steps as k-means, but uses the median of the known data points in a cluster as the new centroid for that cluster (Juan and Vidal, 1998). Another variation called k-medoids uses the data point closest to the center of the cluster as the new cluster centroid rather than a descriptive statistic of the cluster (Kaufman and Rousseeuw, 1987).

One of the major limitations of k-means clustering is that the number of clusters must be given to the model. It is difficult to know how many clusters are needed to adequately represent subgroups within a data set. If too many clusters are used, the groups identified by the algorithm are more granular than they should be; however, using too few clusters produces large groups that mask distinct subgroups. An example of k-means clustering as applied to the Iris data set can be seen in Figure 7 (Varoquaux, 2019). The results of the

clustering are dependent on the number of clusters specified as well as the points used to initialize the algorithm.

#### 4.2.2 Spectral Clustering

While spectral clustering is related to k-means clustering, it approaches the problem of identifying associations in a group of data from a different perspective. Spectral clustering treats each data point in a sample as a node in a graph. The connections between data points are characterized by the adjacency matrix and the degree matrix of the graph. These two matrices are used to calculate the Laplacian matrix of the graph, whose properties are used to identify clusters. All three matrices are  $n \times n$  matrices, where  $n$  is the number of data points in the sample.

Herein, we discuss spectral clustering when the data can be represented using a simple graph. As such, certain mathematical shortcuts can be employed to simplify certain computations. A more general mathematical approach has been discussed by Ng, Jordan, and Weiss (Ng et al., 2002).

The adjacency matrix specifies the strength of the connections between the nodes represented by the rows and columns of the matrix. For data that does not begin in graph form, algorithms such as k-nearest neighbors can be used to generate the adjacency matrix. In the adjacency matrix, each entry  $i, j$  contains the weight of the connection between node  $i$  and node  $j$ . If the edges are unweighted, the value of the entry is either 0 or 1. If the graph is undirected, the value of entry  $i, j$  is the same as the value of entry  $j, i$ . All entries where  $i = j$  should be 0 unless node  $i$  has a self-loop.

The degree matrix is a diagonal matrix that represents the number of edges connected to each node. If the graph is directed, the directionality of the degree matrix must be specified: a directed connection from node  $a$  to node  $b$  contributes to the count for node  $a$  if the degree matrix counts the number of edges that begin at each node (outdegree), but contributes to the count for node  $b$  if the degree matrix counts the number of terminating edges at each node (indegree). In the case of an undirected graph, the connections include all edges that begin or terminate at a node. To summarize, in a directed graph, each edge contributes to

only one node count, while in an undirected graph, each edge contributes to both nodes.

The adjacency matrix and the degree matrix are used together to construct the Laplacian matrix of the graph. This calculation of the normal Laplacian for a simple graph (undirected and containing no loops) is straightforward: the adjacency matrix is subtracted from the degree matrix. The resulting matrix has the following properties:

- The diagonals are the number of connections per node less the number of self-connections
- All off-diagonal values are the negative of the weight connecting node  $i$  to node  $j$ .

It is important to note that if the graph in question contains loops or is directional, other methods must be used to calculate the Laplacian matrix.

The Laplacian matrix can be used to explore many properties of a graph. In particular, the eigenvalues of the Laplacian matrix are informative about the number of connected components in the graph. Connected components are areas of the network that are connected to each other but not anything outside that component. Each connected component is not its own cluster, though: the connected components could be large and contain smaller sets of connected nodes that are good options for clusters.

Several steps are used to determine the number of clusters in the graph. First, the eigenvalues of the Laplacian matrix are sorted in increasing order. The number of zero-valued eigenvalues is the number of connected components in the graph. Eigenvalues close to zero suggest weak edges preventing some connected components from being two separate components. Manually examining these eigenvalues before performing spectral clustering can be informative about the number of clusters to create: the number of values below the first large gap between the eigenvalues is the number of clusters,  $k$ . The eigenvectors associated with these  $k$  eigenvalues are used as a lower-dimensional representation of the data in the graph. Performing  $k$ -means clustering on this data produces the labels for the clusters within the data that are not linearly separable otherwise.

The limitations of spectral clustering are strongly related to the process of remapping the data to a lower-dimensional space. When reducing the number of features used to represent a data set, information about that data set is inherently lost. The missing information can make separating classes in the lower dimensional data significantly more difficult, if not

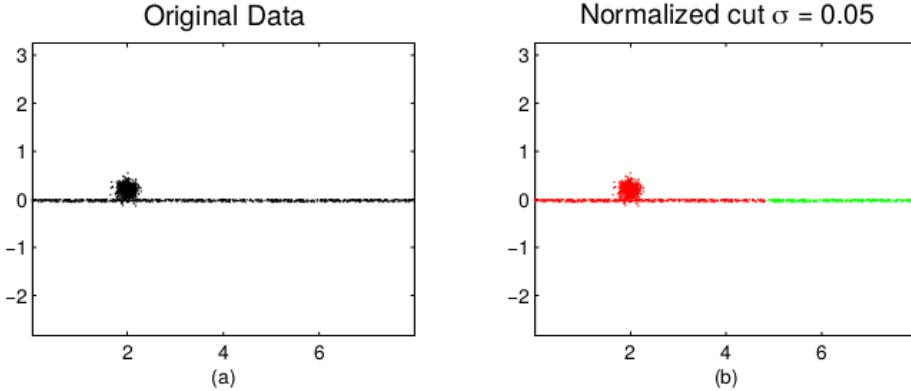


Figure 8: (a) Two different distributions, a 2D Gaussian density and a thin horizontal rectangle are difficult to separate (b) due to their overlap and the penalties built into the cost function of the spectral clustering algorithm. From (Nadler and Galun, 2007).

impossible. Consider the following two cases: a case where two classes overlap and a case where three classes are unevenly represented in the data.

In cases where two seemingly obvious clusters overlap, the clustering algorithm may be unable to accurately identify them due to the mathematical penalties imposed on separating points in the feature space. An example of this problem as described by Nadler and Galun can be seen in Figure 8 (Nadler and Galun, 2007). The two distinct groups are the small 2D Gaussian density and the horizontal rectangle. The spectral clustering algorithm (in this case, the normalized cut algorithm) is unable to separate the groups because the degree of overlap between the Gaussian density and the rectangle is greater than the height of the rectangle. It is more cost-effective for the algorithm to make a vertical cut to divide the rectangle in two than to cut the Gaussian density away from the rectangle.

Uneven distributions of data classes can severely impact spectral clustering results. Consider the case where a data set contains one highly populated class with a wide variation in its data features and two less populated classes with less variation in their data features. The scale of the large, highly populated class can overshadow the smaller classes: if the three most significant eigenvectors are more related to the large class than the smaller classes, the

algorithm cannot differentiate between the two smaller classes.

#### 4.2.3 Agglomerative Clustering

Agglomerative clustering is a specific type of hierarchical clustering that builds a tree of similarities between data samples from the “bottom up” (Ward, 1963). The data samples in agglomerative clustering are also viewed as distinct points in  $n$ -dimensional space, but the number of groups to identify is not specified.

First, the distance from every data sample to every other data sample is calculated. The two data points that are closest together in terms of a similarity metric are combined into a single cluster. In the relationship tree representing the similarities between all data samples, a node is created and the joined data points are connected to that node. That node or cluster is treated as an intermediate data sample. The distance from the new “data sample” to every other data sample is calculated, and the two closest data samples are again combined into another intermediate sample. A node representing the new cluster is added to the relationship tree and the data sample or samples merged into the cluster are connected to the node. The process of combining data points into clusters based on similarity to other data points terminates when all data points and clusters have been combined.

The results of agglomerative clustering can be interpreted by traversing the relationship tree. The relationship tree recorded the history of which nodes were merged into which clusters at each stage. Due to the nature of agglomerative clustering, these stages can be viewed as distinct levels in the tree. The granularity of the clusters can be explored beginning at the final node (the root) of the tree. At the top level, there is only one cluster, but at the second to last level of the tree, there will be two clusters; at the third level, there will be three clusters, and so on. How each cluster grew can reveal information about the relationships between the data samples within that cluster.

Agglomerative clustering also lends itself well to visualization via heatmap. The heatmap allows the researcher to see the distribution of feature values across the dataset and across visually prominent clusters. The Python library **seaborn** has a function called **clustermap**, which both performs agglomerative clustering and shows the resulting trees in a structured

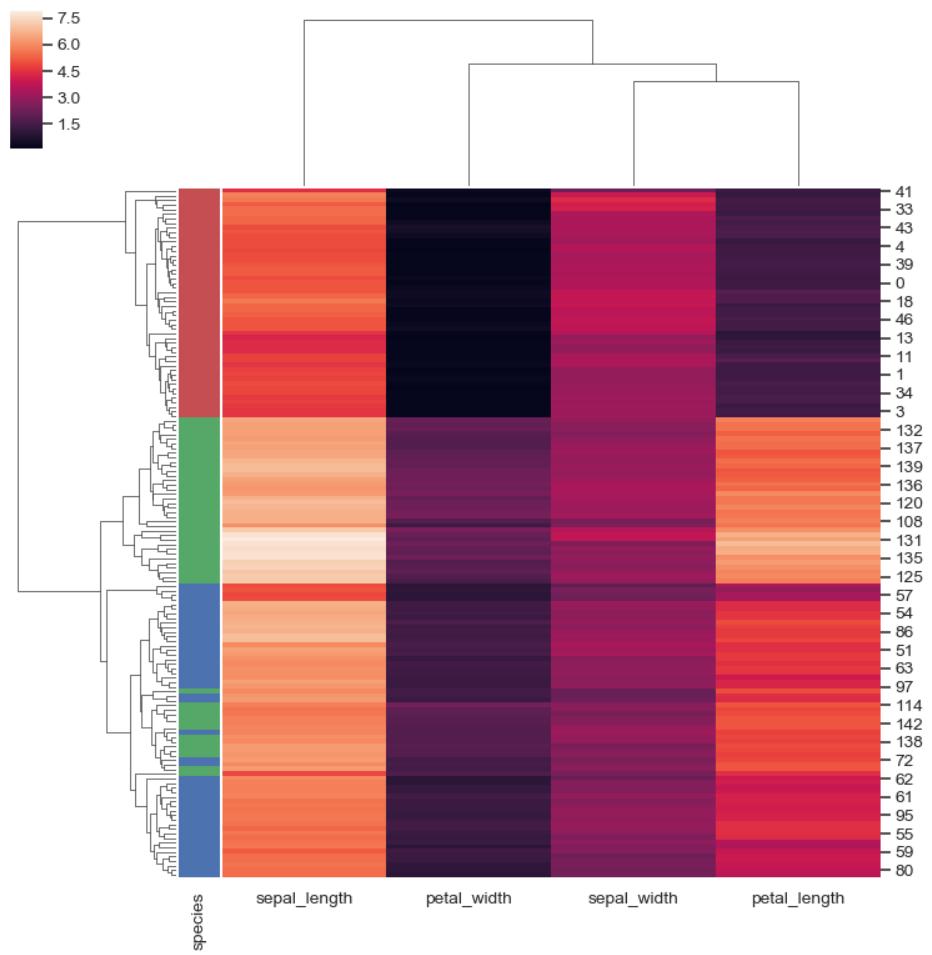


Figure 9: An example of agglomerative clustering on the Iris data set plotted using **seaborn**'s **clustermap** function.

heatmap. An example of agglomerative clustering from the **seaborn** documentation in which the clustering was applied to the Iris data set can be seen in Figure 9 (Waskom, 2018). Each column represents a feature and each row represents a single sample. The colorbar in the top left of the figure shows that lighter colors in each cell represent higher values. The dendrogram along the top of the heatmap shows the distance between each feature. The dendrogram along the left of the heatmap shows the distance between each data sample and the data sample most similar to it. The column immediately to the right of this dendrogram shows the known class (species) for each data sample. The class labels in conjunction with the dendrogram of the data samples show the distinct groups present in the data which were identified using agglomerative clustering.

#### 4.2.4 Visualizing Clustering Results

The results of unsupervised clustering algorithms can be visualized to illustrate how the computer chose each group of samples. Depending on the number of features  $n$  for each data sample, a dimensionality reduction method may be needed to transform the location of the data sample in  $n$ -dimensional feature space to a more easily visualized 2-dimensional or 3-dimensional space. The three-dimensionality reduction methods we consider are principal component analysis (PCA), T-distributed stochastic neighbor embedding (t-SNE), and uniform manifold approximation and projection (UMAP).

**PCA.** Principle component analysis is a multivariate statistical technique that can be used to transform a set of variables with some degree of intercorrelation into a set of new, independent, orthogonal variables (Abdi and Williams, 2010). These variables are called principal components of the data set. The principal components are ordered with respect to the amount of variance in the data set that can be projected onto each component. In general, PCA fits an  $p$ -dimensional ellipsoid to a data set with  $n$  features such that  $p < n$ . Each axis of the ellipsoid represents a single principal component. The first two or three principal components can be used to plot the results of a clustering algorithm in 2D or 3D space.

It is important to note that prior to the application of PCA, the data must be normalized.

Normalizing the data allows different features to be compared on the same scale.

**t-SNE.** T-distributed stochastic neighbor embedding (t-SNE) was developed by Maaten and Hilton to perform nonlinear dimensionality reduction for visualizing high-dimensional data in a 2D or 3D space (van der Maaten and Hinton, 2008). The algorithm first constructs a distribution of the high-dimensional data to measure the pairwise similarity of all data points. It also constructs a second distribution to measure the pairwise similarity of the data points in the lower dimensional space. Then, it performs gradient descent to minimize the difference between the two distributions as measured by the Kullback-Leibler divergence. At each iteration of the gradient descent, the distribution of points in the lower dimensional space is modified. The algorithm converges when the distribution of pairwise similarities between data points in the lower dimensional space most closely matches the corresponding distribution in the original high-dimensional space.

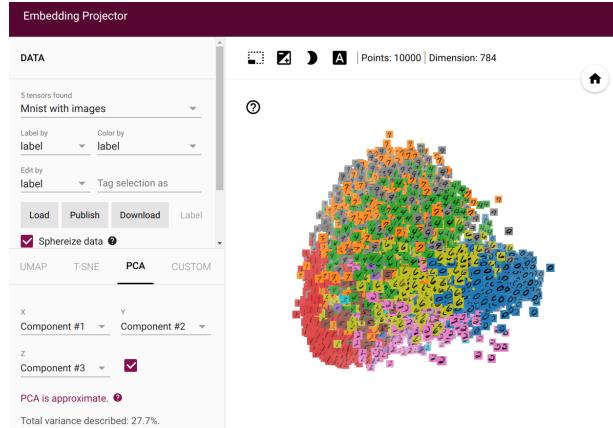
t-SNE is a computationally expensive technique with a complexity of  $O(N^2)$  where  $N$  is the number of data points. For this reason, it is recommended that data with more than 50 features undergoes another form of dimensionality reduction before t-SNE is applied to a data set. Even when this recommendation is not followed, the lower-dimensional data produced using t-SNE lacks the interpretability of PCA data: the resulting dimensions have no interpretable meaning.

**UMAP.** Uniform manifold approximation and projection (UMAP) was proposed as an alternative to t-SNE (McInnes et al., 2018). Rather than build a pairwise similarity distribution, UMAP uses topological representations of the data. For each point  $x_i$ , the distances between  $x_i$  and its  $k$  nearest neighbors are measured and normalized by the distance between  $x_i$  and the  $k$ th neighbor. These collections of distances around each point are local manifolds. The local manifolds are combined into the same global manifold using fuzzy simplicial sets. After the global of the data is known, it is used to determine where data points must lie in a lower-dimensional space so that they both adhere to the known manifold topology and retain the distance metrics from their  $k$  nearest neighbors. The topology of the lower dimensional manifold is adjusted iteratively to minimize the cross-entropy between the higher dimensional representation and the lower dimension representation. In practice, UMAP treats each data point as a node in a weighted graph. It first determines each point's

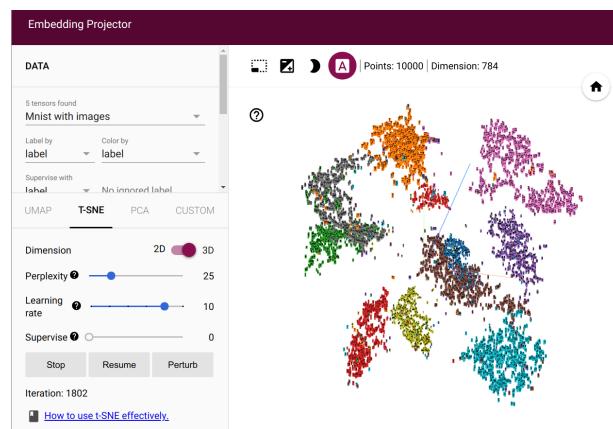
$k$  nearest neighbors, calculates the distances between them, and then computes a version of that topology in a lower-dimensional space.

UMAP was developed under the belief that local structure is more important than global structure. It learns structures in a data set, even when the local structures can only be attributed to noise. It is not suitable for use in small, noisy data sets or in large data sets with only global (not local) structure. Figures produced using UMAP-reduced data should be interpreted with care: they could contain spurious structures from the data sample and not the target population. Additionally, UMAP is similar to t-SNE in that the dimensions produced by UMAP were generated nonlinearly and contain no true meaning.

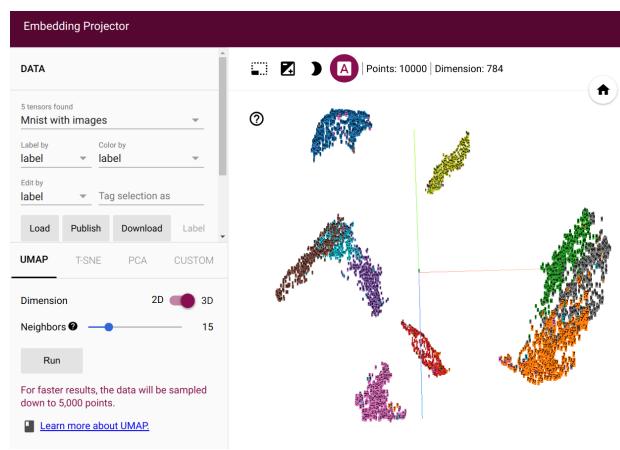
We performed a preliminary comparison of PCA, t-SNE, and UMAP for visualizing high dimensional data using TensorFlow’s Projector tool (TFP, ). This tool is a web interface for visualizing high dimensional data from either built-in data sets or data sets uploaded by the user. It supports all three of the dimensionality reduction methods discussed previously in this section as well as a customizable projection of the entire data set to two or three feature vectors. The PCA, t-SNE, and UMAP visualizations of the popular MNIST handwriting data set can be seen in Figure 10.



(a) PCA



(b) t-SNE



(c) UMAP

Figure 10: Visualization of MNIST data in 3D via TensorFlow Projector as generated using PCA, t-SNE, and UMAP dimensionality reduction techniques.

## 5.0 DATA

The data used to test the hypothesis and aims introduced in the previous chapter are drawn from a set of simulated rs-fMRI sequences and three clinical groups. In this chapter, we will first discuss the need for simulated sequences and the processes used to generate them for our work. Then, we will discuss the clinical images which were taken from several prospective studies of congenital heart defects in three age groups of pediatric patients. These groups were chosen because congenital heart disease (CHD) affects patients throughout their lifespans, and different characteristics in motion patterns have been observed in patients of different ages with and without CHD.

### 5.1 Simulated Sequences

There are two major barriers in research surrounding motion correction in rs-fMRIs: gathering data and measuring the effects of the technique.

The challenge of gathering significant amounts of data is common across the medical domain. Simulating realistic patient scans can mitigate this challenge and make rs-fMRI research more accessible.

Research on motion correction in rs-fMRIs often focuses on the amount of the positional effects of motion removed from the image. This silver standard is used because the gold standard, the amount of brain signal recovered by motion correction is unknown in clinical images. When an rs-fMRI is simulated, the amount of brain signal in the sequence can be manually specified. The known brain signal can be used to evaluate motion correction techniques with respect to the amount of brain signal recovered by the technique.

In this section, we elaborate on a simulation we designed and implemented to address both of these barriers.

### 5.1.1 Background

It is difficult to obtain enough data from a large number of subjects to perform large-scale studies in the medical domain. Collaborators can band together to create a larger and more diverse data set by participating in multicenter studies, though there are some challenges associated with multicenter studies. Each site will have a different scanner, potentially with different field strengths and from different manufacturers. Even ignoring the challenges of harmonizing data obtained using scanners from different companies, each scanner has its own set of unique inhomogeneities in the primary magnetic field. Additional scans of inanimate or human phantoms may be necessary to characterize the differences between all of the scanners involved in a multicenter study.

The second barrier to motion correction in rs-fMRI research is the complexity of identifying a gold standard metric to use when evaluating motion correction techniques. The current criteria for determining the usability of an rs-fMRI sequence developed by Power et al. evaluate motion correction techniques in terms of the reduction of positional differences and signal differences between neighboring volumes. Unfortunately, this approach does not measure the amount of signal recovered or lost through motion correction. A true gold standard evaluation of motion correction would be able to evaluate the BOLD signal present in the image before and after correction. If the BOLD signal prior to the patient motion was known, though, there would be no need for image processing in the first place: we would already have the data we are trying to obtain with motion correction.

We address these two barriers by creating a mechanism for generating simulated image sequences. The generated sequences contain simulated brain signal based in areas of the brain associated with resting-state connectivity, scanner noise, and patient motion. Our mechanism can create large quantities of unique image sequences. The simulated image sequences can also serve as a gold standard for evaluating volume registration and motion correction techniques: the signals and noise sources added to the sequence are known because they are generated as part of the simulation.

### 5.1.2 SPECTr: Simulated Phantom Emulating Cranial Transformations

Our mechanism is called Simulated Phantom Emulating Cranial Transformations (SPECTr). A phantom is an object designed to have material properties that mimic those of a specific tissue type or organ. Phantoms, either manufactured objects or healthy humans, are used in multicenter studies to obtain images of the same object or person from multiple scanners. These images are used to harmonize the data taken from different sites. We call our simulated sequence a phantom because the baseline image itself is known as are the signals added to it to simulate brain activity.

When developing the pipeline for SPECTr, it was important to consider the multiple facets of rs-fMRI sequences. The BOLD signal must be present in areas of the brain associated with resting-state neuronal networks. The effects of motion and various impacts they have on the BOLD signal must also be present. As discussed in Chapter 2, the three effects of motion are positional, spin history, and susceptibility. The positional effects of motion are straightforward to implement. For simplicity, we model the spin history and susceptibility effects as a single source of in-scanner noise.

### 5.1.3 Materials

In order to simulate the resting-state BOLD signal in an fMRI, two pieces of anatomical information are required. The first is structural information about the brain. The second is the location of functional networks associated with the resting-state BOLD signal.

While it is possible to use clinical images for the structural information, the goal of SPECTr is to be generalizable both for our use and for the use of other researchers. Additionally, clinical images inherently contain some degree of signal from some neuronal processes which could obfuscate the simulated BOLD signal.

In 1992, the International Consortium for Brain Mapping was formed to develop a set of standards for what is considered a healthy human brain. They used their criteria to develop an initial set of “average” brain structural scans based on scans from healthy volunteers. This original set of average brain scans incorporates scans from 305 subjects and is referred to as the MNI305 data. As MRI technology has evolved, the spatial resolution capabilities

of the MRI scanners have increased. In 2001, another set of 152 healthy volunteers was recruited to create a higher resolution data set. The scans from the new cohort were linearly registered to the MNI305 data to create the new MNI152 images.

Herein, we use the MNI152 data set as available at the website for the McConnell Brain Imaging Centre at McGill University (Fonov et al., 2009) (Fonov et al., 2011). This data set contains five images: a T1-weighted scan, a T2-weighted scan, a proton density-weighted scan, and two binary masks for the head and the brain. Each of the structural scans was designed to highlight the properties of different tissues in and around the brain. Proton density scans were developed for the purpose of detecting the blood-related signal. As BOLD images essentially perform the same purpose as a proton density image over a period of minutes, we choose to use the proton density-weighted scan as the structural base for the simulated images.

The functional network information is more difficult to find than structural atlases. The Functional Imaging in Neuropsychiatric Disorders Lab at Stanford has developed two sets of functional atlases. The atlases are divided into regions of interest (ROIs) associated with individual networks. The first set of atlases contains 90 functional ROIs which compose 14 networks (Shirer et al., 2012). The second set of atlases contains 499 functional ROIs and has more gray matter coverage (Altmann et al., 2015). We chose to use the original ROIs associated with the dorsal and ventral default mode networks.

#### 5.1.4 Simulation Pipeline

**5.1.4.1 Baseline Sequence** The process for generating a simulated sequence using the data discussed in the previous section has several steps. An overview of the pipeline can be seen in Figure 11. We cover these steps in detail in this section.

The MNI152 proton density image is a whole head image. To remove the skull, we apply the brain mask to the proton density-weighted image. The resulting image contains only brain tissue.

The spatial resolution of the structural brain image is  $1 \text{ mm}^3$ , but the spatial resolution for a single volume in an rs-fMRI sequence is less granular at  $4\text{mm}^3$ . In order to achieve this

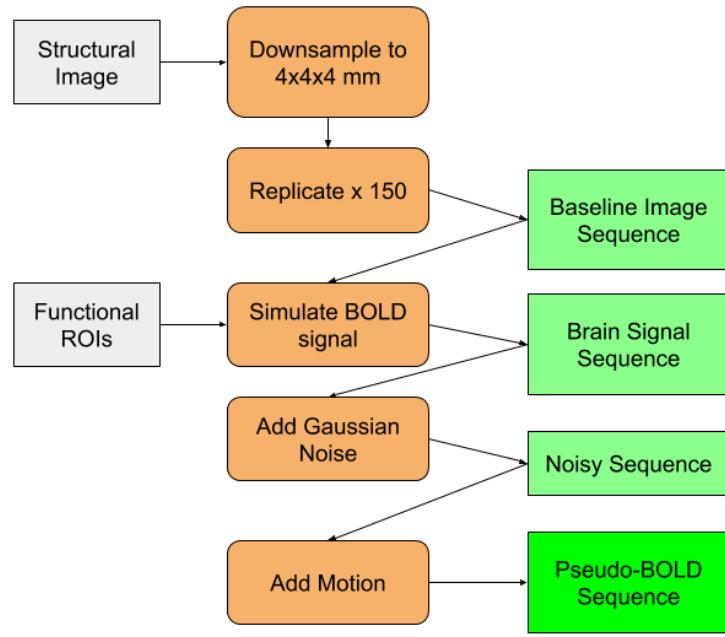


Figure 11: An overview of the SPECTR simulation pipeline. Using atlas data, a simulated phantom containing brain signal, scanner noise, and patient motion is generated.

resolution, the structural volume must be downsampled. After the downsampling, the size of the structural image is reduced from 181x217x181 voxels to 45x54x45 voxels.

Now that the structural image is the correct spatial resolution, it must be replicated to create the temporal sequence. Part of this step includes creating a new dimension in the data. The affine matrix which represents the resolution of the image has the following structure:

$$\begin{bmatrix} r_x & 0 & 0 & c_x \\ 0 & r_y & 0 & c_y \\ 0 & 0 & r_z & c_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5.1)$$

where  $r_x$ ,  $r_y$ , and  $r_z$  represent the spatial resolution in the  $x$ ,  $y$ , and  $z$  axes, respectively and  $c_x$ ,  $c_y$ , and  $c_z$  represent the location of the origin in voxels. To convert the image from a 3D volume to a 4D sequence, we add a row and a column to this matrix so that it now has the structure:

$$\begin{bmatrix} r_x & 0 & 0 & 0 & c_x \\ 0 & r_y & 0 & 0 & c_y \\ 0 & 0 & r_z & 0 & c_z \\ 0 & 0 & 0 & t & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (5.2)$$

where  $t$  is the desired temporal resolution. We choose a temporal resolution of 2 seconds for our simulated sequence. The downsampled image volume is replicated and concatenated along the new temporal dimension to create a sequence that is 150 image volumes long. This sequence is referred to as the base phantom sequence as it contains no brain signal, noise, or motion.

**5.1.4.2 Brain Signal** The next step in the pipeline is to add the BOLD signal to the sequence. We combine the functional ROIs from the dorsal and ventral default mode networks into a single binary functional ROI image. This image is referred to as the default mode network (DMN) mask.

For each nonzero voxel in the DMN mask, a temporal signal is generated. We chose to model the BOLD response as a cosine signal following the formula:

$$s(\vec{v}, t) = a * (\cos(f_0 * (t - t_{shift})) - a_{shift}). \quad (5.3)$$

This equation is for a scaled cosine function with both temporal and amplitude shifts. The temporal and amplitude shifts,  $t_{shift}$  and  $a_{shift}$  respectively, are randomly generated for each voxel from uniform distributions. Once chosen, they are consistent across the voxel's generated temporal signal. The other parameters,  $a$  and  $f_0$ , were specified using existing research.

In 2007, Biswal et al. performed a study evaluating methods to reduce changes in the BOLD signal not directly related to brain activity (i.e., vascularity). They identified a low-frequency spectral amplitude of 0.04 Hz as the highest frequency related to the BOLD signal consistent not only through scans of individual patients but in group-wise analyses. Based on their work, we chose the fundamental frequency of  $f_0$  to be 0.04 Hz (Biswal et al., 2007).

The scaling factor  $a$  was chosen based on Power et al.'s usability criteria. As they state that any signal change between volumes above 2.5% of the maximum voxel value is unlikely to be due to brain activity, we used an amplitude of 20 units on a scaled voxel value range of [0, 1000] (Power et al., 2012).

The generated BOLD signals were added to the baseline image sequence but were also saved in a separate image sequence file for reference during analysis. The generated sequence with only the BOLD signal is referred to as the brain signal sequence.

**5.1.4.3 Scanner Noise** The next step in our pipeline is to add scanner noise to the brain signal sequence. In a regular rs-fMRI of a patient, the noise in the sequence is the result of two factors: the spin history effects of motion and the susceptibility effects of motion. To simplify the simulation, we choose to model the resulting noise rather than the individual sources.

Signal acquired by MRI scanners is first recorded as a raw data matrix in k-space. K-space contains spatial frequency information. The coordinate system of k-space differs from the coordinate system of physical space. In k-space, the closer a point is to the center of

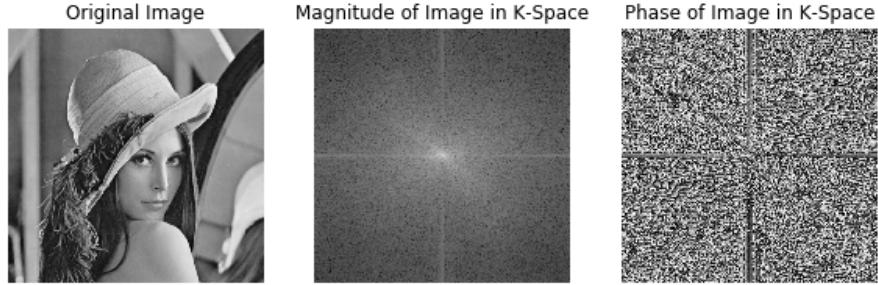


Figure 12: An example of the magnitude and phase of an image in k-space.

a zero-centered data matrix, the lower its frequency or phase. The brighter a point in the raw data matrix is, the larger its magnitude. An example of the magnitude and phase of an image in k-space can be seen in Figure 12.

Generally speaking, images contain more low-frequency information than high-frequency information. We wish to be able to control the amount of noise added to the magnitude and phase components independently, so we choose to add noise to the sequence in k-space.

When adding scanner noise to the brain signal sequence, we add noise to each image volume independently. The image volume is transformed from physical space into k-space using the fast Fourier transform (FFT). To force the image to be zero-centered, we perform an FFT shift on the Fourier space data. A matrix the same size and shape as the image volume is created. The matrix is then filled with complex Gaussian noise,  $n(\vec{l})$  where

$$|n(\vec{l})| = w_m * r_m, r_m \sim N(0, 1) \quad (5.4)$$

and

$$\angle n(\vec{l}) = w_p * r_p, r_p \sim N(0, 1) \quad (5.5)$$

In these two equations,  $\vec{l}$  is the location of the point in the matrix,  $w_*$  represents the weight assigned to the component,  $r_*$  represents the random noise variable independently generated from a standard normal distribution, and the subscripts  $m$  and  $p$  refer to the magnitude and

phase components, respectively. The complex noise is weighted so that the magnitude of the noise has a greater contribution than the phase of the noise, that is  $w_m > w_p$

The matrix of complex noise is added to the zero-centered k-space image volume. The noisy k-space volume is unshifted and undergoes an inverse Fourier transform to change the data back to physical space. This sequence containing both brain signal and scanner noise is referred to as the noisy sequence.

**5.1.4.4 Patient Movement** Now that the ground truth brain orientation and BOLD signal have been established, patient motion can be added to the BOLD phantom sequence. One of the aims of this document is to establish that patients from different populations exhibit different motion patterns. As we have not yet established what those motion patterns are, we developed a generic motion model for the simulation.

During an rs-fMRI scan, a patient theoretically has the freedom to rotate his head around three different axes, translate his head along three different axes, or some combination of translations and rotations. Realistically, once a patient has settled in the scanner, it becomes difficult for his head to only undergo translation. It is more likely that he will rotate his head. For the simplicity of the simulation, we assume no head translations occur during the scan.

A rotational transformation can be represented as a matrix. Rotations about a single axis are represented as followed:

$$R_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \quad (5.6)$$

$$R_y(\beta) = \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix} \quad (5.7)$$

$$R_z(\gamma) = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (5.8)$$

Rotational transforms are applied to the origin of the image. However, the origin of the image is not necessarily anatomically significant. Head motion naturally occurs about the base of the neck. We can approximate the location of the base of the neck by calculating the center of mass of the brain. First, the image volume is thresholded to separate the brain from the background. Then the locations of the “on” voxels are averaged to calculate the center of mass.

Motion is added to the noisy image sequence one volume at a time. No motion is applied to the first volume in the sequence, but volume is used to calculate the center of mass of the brain. For each subsequent volume, the angles of rotation about the x-, y-, and z-axes are each randomly generated from a standard normal distribution to create rotational change matrices,  $R_{*,\Delta}$  (where the subscript \* represents  $x$ ,  $y$ , or  $z$ ). The new rotations  $R_{*,\Delta}$  are added to the rotations from the previous step  $R_{*,i-1}$  to create the rotation matrices for the current volume,  $R_{*,i}$ :

$$R_{*,i} = R_{*,i-1} + R_{*,\Delta} \quad (5.9)$$

The three rotational transformations are combined into one matrix via multiplication:

$$R_i = R_{x,i}(\alpha)R_{y,i}(\beta)R_{z,i}(\gamma) \quad (5.10)$$

The compound transformation  $R_i$  is then applied to the image at the center of mass calculated from the original image volume. *Note: Since the simulation is constrained by the assumption of no head translations, the center of mass will remain consistent throughout the image sequence.*

After the motion has been added to every image volume in the sequence, the final sequence (labeled Pseudo-BOLD Sequence in Figure 11) is ready to be used in motion correction analyses.

### 5.1.5 Implementation

SPECTr is implemented in Python (version 3.7.3). The `nipy` library (version 0.4.1) was used to load images, save images, and combine the functional ROIs. The `numpy` library

(version 1.17.4) was used for matrix manipulations involved in the brain signal and noise generation steps. The image processing library `skimage` (version 0.14.2) was used to calculate the center of mass. The Python wrapper `SimpleITK` (version 1.2.4) of the Insight Toolkit library for image analysis was used to perform the rotational transformations associated with patient motion.

The SPECTr source code is available on Github: <https://github.com/jmschabdach/SPECTr>.

### 5.1.6 Simulated Sequences Experiments

We generated 90 image sequences from the MNI152 proton density image and the default mode network functional ROIs using the process described in the previous section. The image sequences were divided into three groups with different degrees of scanner noise. All simulated images underwent both registration techniques described in Chapter 3. The registered images also underwent independent component analysis (ICA) using FSL’s MELODIC (Multivariate Exploratory Linear Optimized Decomposition into Independent Components) tool.

The registered images underwent the same type of analysis as the clinical images. The ICA technique was used to determine how much brain signal was recovered by each registration technique.

This particular experiment is one of the first to investigate how much true BOLD signal is preserved through motion correction. One of the major drawbacks to existing motion correction pipelines is that they remove signals of interest along with noise. In clinical data, there is no way to know the ground truth signal contained within the image; however, simulated phantom images have a *de facto* known ground truth signal. The design for this experiment can be used to evaluate how much BOLD signal is recovered by other motion correction pipelines, and how close the recovered signal is to the signal of interest.

## 5.2 Clinical Cohorts

Our clinical cohorts are composed of healthy subjects and subjects who have congenital heart disease (CHD). The phrases congenital heart defects and congenital heart disease (CHD) both refer to defects in the heart or the vessels around the heart. These defects affect how blood moves into, through, and away from the heart. CHD has a worldwide prevalence of about 8 per 1000 live births, meaning about 1.35 million children are born with CHD every year. Many CHD patients are also affected by comorbidities, including neurodevelopmental disorders. Since the survivability of CHD has increased from 10% to 90%, the medical community is faced with a growing, aging population of CHD patients with a variety of neurological needs. Subjects in CHD studies provide a diverse set of data, particularly when it comes to studying motion in rs-fMRIs.

In this section, we provide a general overview of the impact of CHD on a global scale, the process for diagnosis, and additional risks associated with CHD. We then discuss the process of diagnosing neurodevelopmental comorbidities occurring with CHD and the impact of improved medical care on the CHD population. We end this section with a description of the three populations from which we obtained clinical rs-fMRI sequences.

### 5.2.1 CHD Background

CHD consists of a variety of defects that can affect any combination of the vessels and chambers of the heart with varying degrees of severity. The defects prevent the cardiopulmonary system as a whole from functioning correctly, but pinpointing and effectively treating the defects can be a complex process. It is important to note that each defect type has a different prevalence, a different treatment plan, and different expected outcomes. A breakdown of prevalence rates of some of the most common lesion types can be seen in Figure 13.

Different presentations of CHD are associated with a number of different genetic and environmental factors (Mozaffarian et al., 2016). Genetic conditions such as Down syndrome, Turner syndrome, 22q11 deletion syndrome, Williams syndrome, and Noonan syndrome are

**Table 15-3. Estimated Prevalence of Congenital Cardiovascular Defects and Percent Distribution by Type, United States, 2002\* (in Thousands)**

Type	Prevalence, n			Percent of Total		
	Total	Children	Adults	Total	Children	Adults
Total	994	463	526	100	100	100
VSD†	199	93	106	20.1	20.1	20.1
ASD	187	78	109	18.8	16.8	20.6
Patent ductus arteriosus	144	58	86	14.2	12.4	16.3
Valvular pulmonic stenosis	134	58	76	13.5	12.6	14.4
Coarctation of aorta	76	31	44	7.6	6.8	8.4
Valvular aortic stenosis	54	25	28	5.4	5.5	5.2
TOF	61	32	28	6.1	7	5.4
AV septal defect	31	18	13	3.1	3.9	2.5
TGA	26	17	9	2.6	3.6	1.8
Hypoplastic right heart syndrome	22	12	10	2.2	2.5	1.9
Double-outlet right ventricle	9	9	0	0.9	1.9	0.1
Single ventricle	8	6	2	0.8	1.4	0.3
Anomalous pulmonary venous connection	9	5	3	0.9	1.2	0.6
Truncus arteriosus	9	6	2	0.7	1.3	0.5
HLHS	3	3	0	0.3	0.7	0
Other	22	12	10	2.1	2.6	1.9

Average of the low and high estimates, two thirds from low estimate.<sup>24</sup>ASD indicates atrial septal defect; AV, atrioventricular; HLHS, hypoplastic left heart syndrome; TGA, transposition of the great arteries; TOF, tetralogy of Fallot; and VSD, ventricular septal defect.

\*Excludes an estimated 3 million bicuspid aortic valve prevalence (2 million in adults and 1 million in children).

†Small VSD, 117 000 (65 000 adults and 52 000 children); large VSD, 82 000 (41 000 adults and 41 000 children).

Source: Data derived from Hoffman et al.<sup>24</sup>

Figure 13: Table of prevalences of congenital heart defects as compiled by (Mozaffarian et al., 2016).

associated with certain CHD presentations. Maternal behaviors such as smoking and binge drinking are known to cause heart problems in the fetus. Other maternal risk factors are obesity, folate deficiency, and living at a high altitude. Paternal exposure to phthalates, anesthesia, sympathomimetic medications, pesticides, and solvents may increase the risk of the fetus for developing CHD. While there are quite a few factors in this list, there are many CHD cases whose causes are unknown.

Once a patient is diagnosed with one of these defects or a cause of the CHD is identified, the specific nature of his case must be clearly documented. The documentation of CHD using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) has 25 high-level codes representing various presentations of CHD, but these codes used alone are often not sufficient for describing a patient's true condition (Mozaffarian et al., 2016). Additional ICD-9-CM codes should be used to communicate the finer details of a patient's condition if they are available.

The incidence of CHD in live births vary across countries and continents. The United

States reports approximately 4-10 CHD cases per 1000 live births. Europe and Asia see about 6.9 and 9.3 CHD cases per 1000 live births, though smaller studies have been conducted in many countries to measure local prevalence (Mozaffarian et al., 2016). In China, the incidence of CHD ranges from 8.98 to 11.1 per 1000 live births (Zhao et al., 2019) (Qu et al., 2016). A pair of studies from Iran report incidences of 8.6 and 12.3 per 1000 live births, though the studies note that they were performed in different geographical locations with different populations within the country (Nikyar et al., 2011) (Rahim et al., 2008). One report from Dharan reports an incidence of 5.8 per 1000 patients admitted to a tertiary care hospital over a 12 month period (Shah et al., 2008). A study of newborns at one hospital in New Delhi, India, claims an incidence of 3.9 per 1000 live births, though this rate may be a poor estimate as there is a significant delay between patient birth and referral to a cardiac center in India (Khalil et al., 1994) (Saxena, 2005).

These incidence rates should be analyzed with some caution. In many cases, the reported rates were based on medical records. Medical records are not always correct; it is well known that human error can lead to a medical record lacking information or containing incorrect information. The only way for a person to have a medical record is for him or her to go to a medical center. Not everyone who has CHD is able to seek medical help, often because of their geographical locations or their income. Even if a patient is able to seek medical help, the availability of proper cardiac care varies between and within countries. However, it is generally expected that CHD incidence rates will increase as screening tools and treatments become more effective and more widespread, leading to earlier detection of defects.

Currently, the process of detecting and diagnosing CHD can begin before birth. A specialized ultrasound test called fetal echocardiography can detect heart abnormalities as early as the second trimester of the pregnancy. People who learn they are pregnant with a fetus who shows signs of CHD may choose to handle this information by opting for termination or pursuing a more detailed diagnosis. Additional tests such as amniocentesis and follow-up ultrasounds may be used to determine treatment options before the patient is born. Generally, severe CHD cases present and are detected at earlier stages of life, but minor defects may not become apparent until the patient is older. Tests used to diagnose CHD in post-natal patients include electro- and echo-cardiograms, chest x-rays, pulse oximetry, exercise

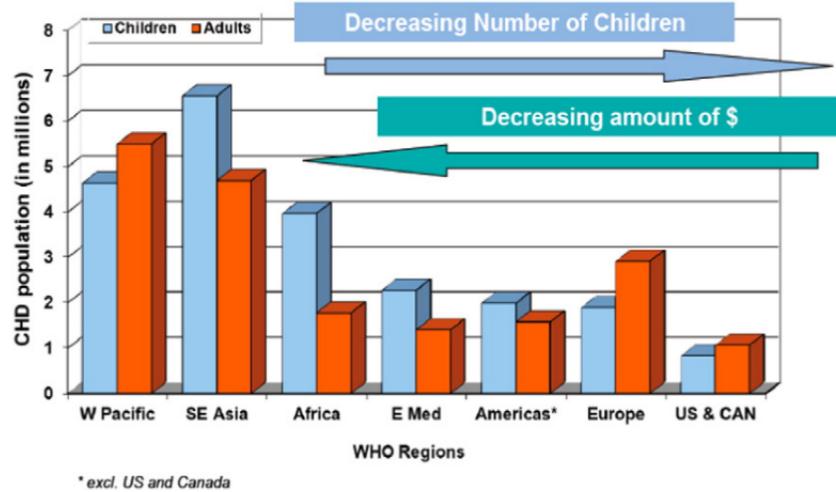


Figure 14: Estimated CHD burden in World Health Organization (WHO) regions using incidence rates of approximately 12/1000 and 4/1000 in children and adults, respectively (Webb et al., 2015).

stress tests, computed tomography or MRI scans, and cardiac catheterization. Treatment of different defects varies from monitoring and medication to surgery and cardiac implants.

The cost of diagnostic techniques and treatment plans impose different levels of the financial burden on CHD patients and their families. Certain defects require complex, expensive surgical repairs, while others can be treated with less expensive approaches (Mozaffarian et al., 2016). The burden of CHD across the globe was outlined by Webb et al. (Webb et al., 2015). Their figure illustrating the prevalence of CHD and the availability of funds with which to treat it can be seen in Figure 14. As the overall mortality of CHD declines, the burden of CHD is expected to increase (Mozaffarian et al., 2016).

Unfortunately, the cost of treating CHD is not the only burden a patient must undergo. Patients with CHD are also at increased risk for heart failure and infections (Mozaffarian et al., 2016). Children with CHD are at a 19-fold risk for stroke compared to their healthy counterparts (Fox et al., 2015). In a study of Swedish citizens born between 1970 and 1993, Giang et al. compared the prevalence of cardiac conditions in patients with and without

CHD (Giang et al., 2018). They found that patients who had a CHD diagnosis were at about eight times higher risk for intracerebral hemorrhage and subarachnoid hemorrhage than their non-CHD counterparts. The CHD patients were also more likely to suffer from arrhythmia and heart failure.

However, cardiac conditions are not the only complications CHD must deal with. Many of these patients also suffer from neurocognitive disorders that co-occur with CHD. Early research in this area focuses on the neurodevelopmental status of neonatal patients pre- and post-surgical intervention. One theory was that some factor or factors in the surgical intervention caused brain injuries in the patients. This idea proved to be inaccurate when researchers began detecting neurological malformations *in utero*.

In a systematic review of available literature regarding prenatal and postnatal presurgical CHD cases and neurodevelopmental outcomes, Mebius et al. identify two theories about the causality of neurodevelopmental delays and CHD (Mebius et al., 2017). The first theory is that abnormalities in the cardiac system prevent the developing brain from receiving enough oxygen and nutrients, which disrupts prenatal brain development. The second theory is that faulty genetic pathways used during both cardiac and brain development cause both conditions to co-occur. However, 11 of the articles Mebius et al. found during their review suggest a third theory related to blood flow through the umbilical artery. During the prenatal period, a fetus receives oxygen from the mother via the placenta. If the placenta was not functioning correctly, it could lead to the fetus receiving not enough oxygen. Lower quantities of oxygen throughout prenatal development could potentially cause problems both in neurological and cardiac growth. The 11 articles have contradictory results, but some researchers are currently investigating the role of the placenta in CHD and prenatal brain development.

The survival of CHD patients to adulthood has increased from 10% to 90% over the last several decades as CHD diagnostic tools and treatments have improved. Currently, Webb et al. estimate that at least 12 to 34 million adults have CHD, and this number is expected to increase (Webb et al., 2015). The impact of the combination of CHD and neurological conditions throughout a patient's lifetime is starting to be explored. The aging of the CHD population has also sparked interest in the relationships between CHD and adult-stage neurological disorders such as dementia and Alzheimer's.

While the purpose of this study is not to focus on CHD patients, we chose to use CHD and aging brain images in support of research being performed in the area of relationships between CHD and neurodevelopment.

### 5.2.2 Study Cohorts

The rs-fMRIs used in this study were gathered as part of ongoing studies of the relationship between CHD and neurodevelopment. Data from the CHD/neurodevelopment studies was obtained through studies approved by the IRB at the Children's Hospital of Pittsburgh of UPMC and the University of Pittsburgh. All data are stored and accessed in compliance with HIPAA policies.

The subjects included in this analysis fall into three age groups: neonatal, preadolescent, and fetal. To summarize the cohorts, there were 124 neurological rs-fMRIs obtained for the fetal cohort, 163 neurological rs-fMRIs obtained for the neonatal cohort, and 546 neurological rs-fMRIs obtained for the preadolescent cohort. The fetal cohort also contains 105 rs-fMRIs of the placenta. The race, ethnicity, and gender counts for these three cohorts can be seen in Figures 15, 16, and 17, respectively. The majority of subjects in each cohort were white and male. In the following sections, we describe each cohort in detail.

**5.2.2.1 Fetal Subject Population and Images** A fetal subject is scanned *in vivo*. He is suspended in amniotic fluid within his mother. The amniotic fluid has a buoyancy that reduces the effects of gravity and allows a fetal subject significant freedom of movement. The fetus can rotate, shift, and flip in ways that can only be accomplished when floating in a body of water. The properties of the uterus constrain the physical space in which motion could occur, but not as much as the head coil and gravity do to the other patient cohorts. A fetus is not guaranteed to be in any specific position at the start of the scan: the scan begins when the mother is ready, not when the fetus achieves a certain pose.

The fetal subjects underwent fetal echocardiography scans in a cardiac clinic to determine whether they were healthy or had a form of CHD. Images of the fetal brain and the placenta were obtained for each subject during MRI acquisitions.

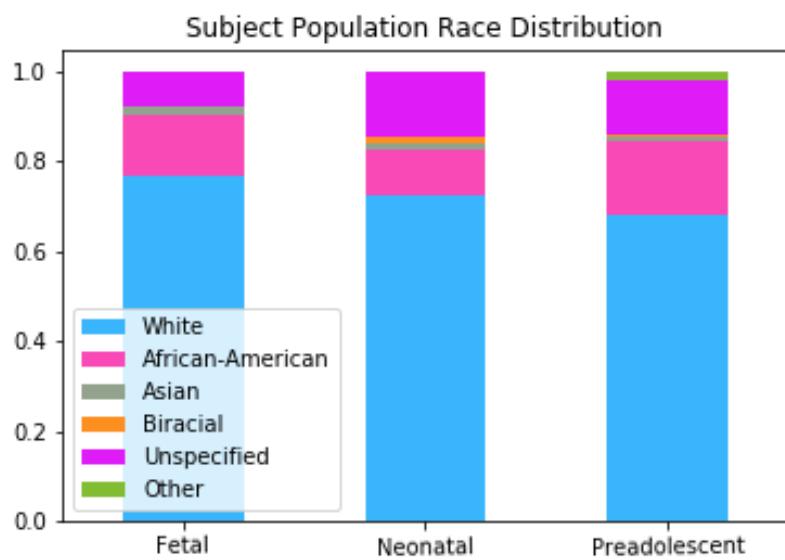


Figure 15: Distribution of subject races for all three cohorts.

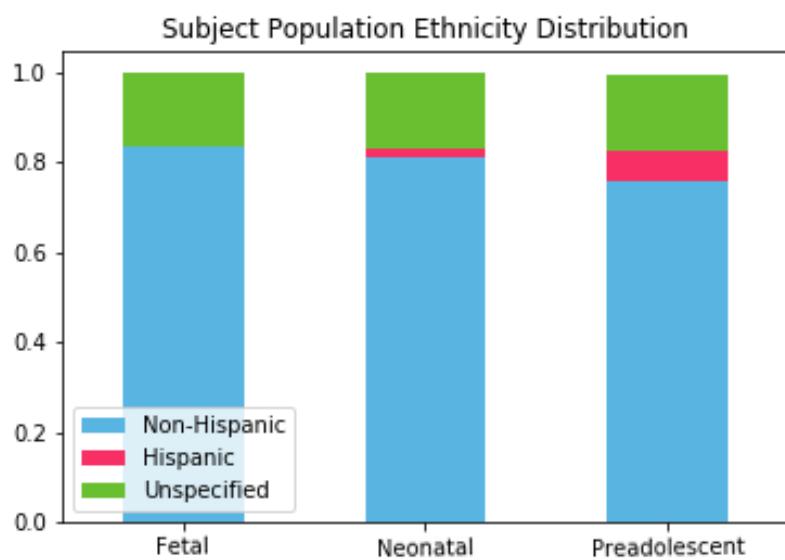


Figure 16: Distribution of subject ethnicities for all three cohorts.

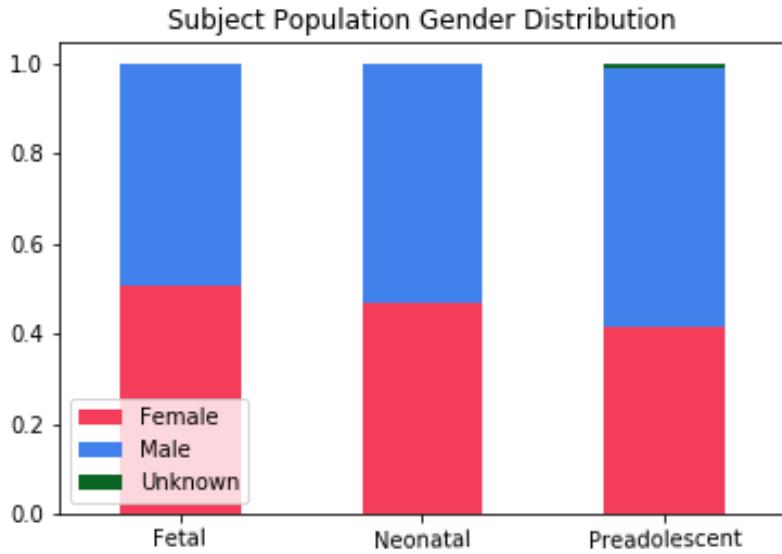


Figure 17: Distribution of subject genders for all three cohorts.

The MRI scans are performed when the mother is in her third trimester. Due to the constraints of pregnancy, the mother cannot lay directly on her side or directly on her back. The MRI scans were performed with the mother going feet first into the scanner and laying on an angle on her left or right side (left or right anterior oblique). The side chosen is the mother's preference. Her back is supported by a pillow to maintain her position during the scan. A large flexible coil is positioned on her abdomen to get coverage of the fetus and the placenta. The coil and pillow are secured with a strap. The mother is fitted with earplugs and headphones and given a button to use during the scan if she needs to get the attention of the radiology technician. No vital monitoring occurs as part of the study.

A total of 124 fetal subjects were scanned on a 3T Skyra (Siemens AG, Erlangen, Germany). The subjects were between 23 and 39 weeks old post-conception. The distribution of fetal post-conceptual age at the time of the scan can be seen in Figure 18. Axial rs-fMRI brain scans were acquired for all subjects and placental rs-fMRI scans were acquired for 105 subjects. The subjects were divided into CHD and control groups based on fetal examinations. Some subjects were moved to different groups after birth. A breakdown of these

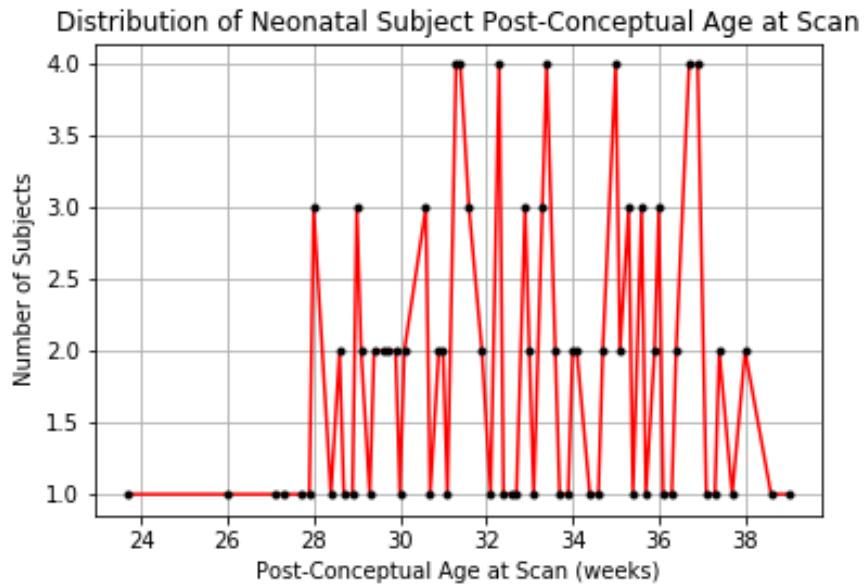


Figure 18: The distribution of post-conceptual ages at the time of the scan of all fetal subjects.

groups before and after birth can be seen in Figure 19.

The parameters for the fetal rs-fMRI protocol were FOV = 300 mm and TE/TR = 32/2280 ms with an interplane resolution of 4.7 x 4.7 mm, slice thickness of 3.0 mm and no gap between slices. The sequences were 150 volumes in length, and each volume was composed of 32 slices containing 64 x 64 voxels. Each brain or placenta image underwent manual segmentation by one of a group of four researchers to remove non-brain or non-placental tissues, respectively, from the image.

We are interested in both the fetal brain and placental images for our work partially because of the relationship between placenta and brain development, but also because these organs have very different physical properties. The fetal brain is a rigid structure floating and moving within the amniotic fluid. It undergoes translation and rotation as a single unit due to passive and active maternal and fetal motions. The placenta, on the other hand, is anchored in place on the uterine wall. It may undergo small translations or rotations due to maternal motion, but it will respond differently to fetal motion. Fetal motions cause

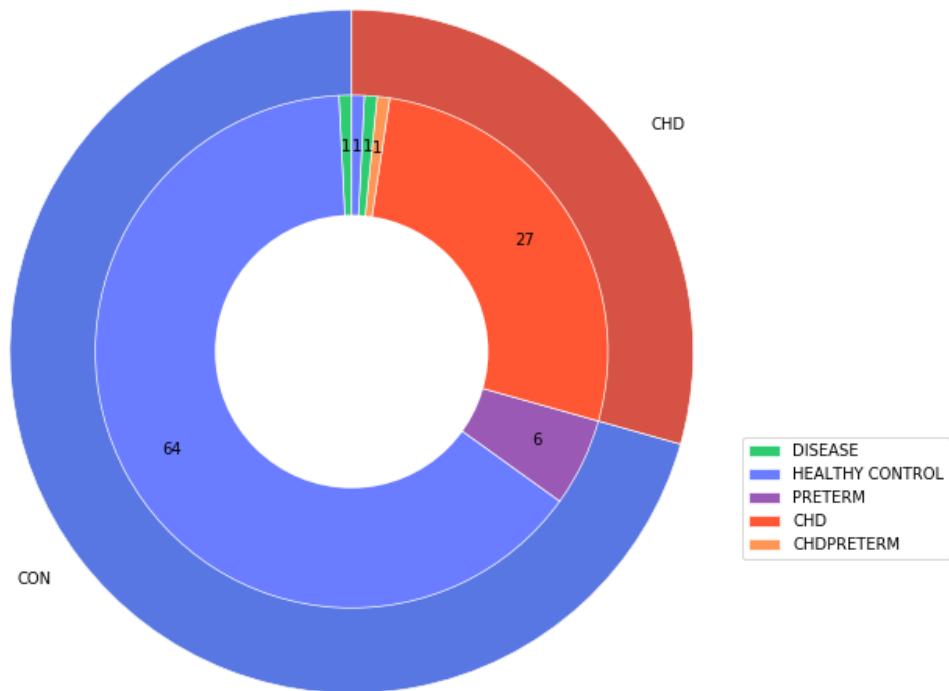


Figure 19: The outer ring represents the group the subject was assigned to based on the fetal examination while the inner ring represents the group the subject was assigned to after birth.

nonlinear deformations of the pliable placenta that can only be adequately accounted for using nonlinear registration algorithms. Nonlinear registrations have the potential to deform brain images into physically impossible shapes, so the fetal brain and placenta were manually segmented in their respective images so that each organ could undergo independent motion correction. As the fetal subjects have both brain and placenta images, their data will be used to examine the impact of volume registration on different organ types.

**5.2.2.2 Neonatal Subject Population and Images** Neonatal subjects have been recruited as part of a prospective observational study. This cohort was scanned at two sites. At Site 1, the subjects were scanned using either a 3T Skyra (Siemens AG, Erlangen, Germany) or a 3T Signa (GE Healthcare, Chicago, Illinois, United States of America). At Site 2, the subjects were scanned using a 3T Achieva (Koninklijke Philips N.V., Amsterdam, Netherlands).

The subjects were unsedated during the scans, and a “feed and bundle” protocol was used to prevent movement (Windram et al., 2011). The newborns were positioned in the coil to minimize head tilting. Newborns were fitted with earplugs (Quiet Earplugs; Sperian Hearing Protection, San Diego, CA) and neonatal ear muffs (MiniMuffs; Natus, San Carlos, CA). An MR-compatible vital signs monitoring system (Veris, MEDRAD, Inc. Indianola, PA) was used to monitor neonatal vital signs. All scans were performed using a multi-channel head coil. The parameters for the resting-state BOLD MR scans were FOV=240 mm and TE/TR=32/2020 ms with an interplane resolution of 4x4 mm, slice thickness of 4 mm, and 4 mm space between slices. The acquired images contained 150 volumes where each volume consisted of 64x64x32 voxels<sup>3</sup>.

Scans from a total of 149 patients were included in this work. The average post-conceptual age of the patients at birth was 39.08 weeks, and they were on average 42.64 weeks post-conception at the time of the scan. The distribution of post-conceptual ages of the subjects at birth and at the time of the scan can be seen in Figure 20 and Figure 21.

The subjects were a mix of control subjects and CHD subjects. A comprehensive breakdown description of the subgroups within these two cohorts can be seen in Figure 22. In this figure, the term “Control” encompasses both healthy full-term subjects, healthy pre-term

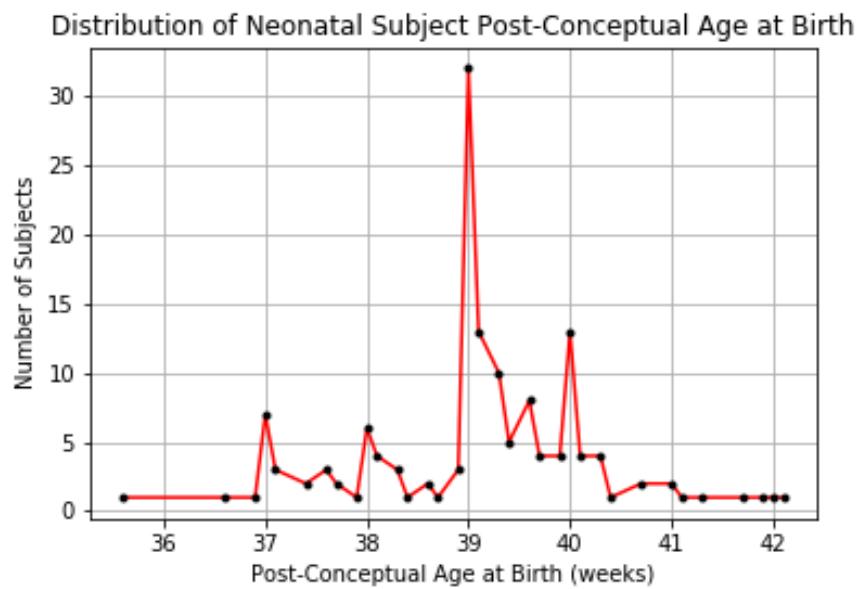


Figure 20: The distribution of post-conceptual ages at birth of all neonatal subjects.

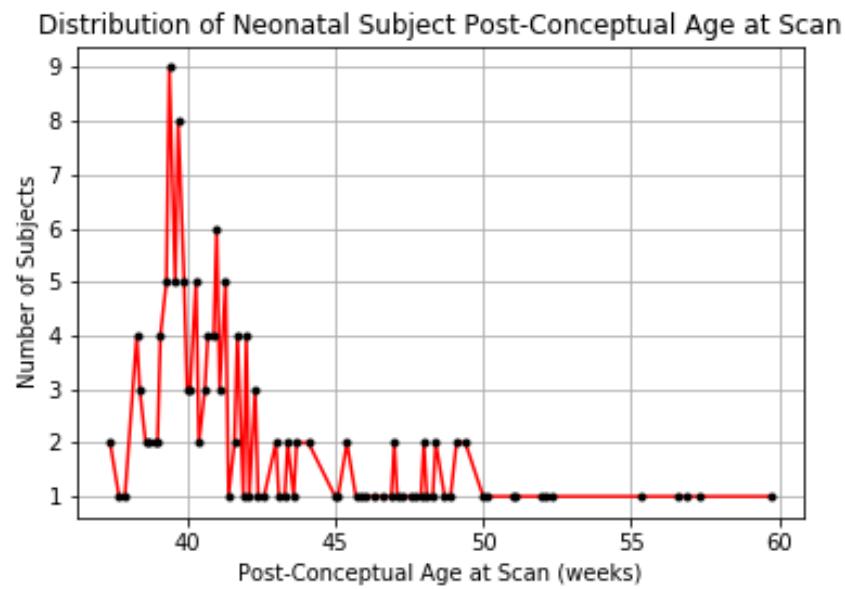


Figure 21: The distribution of post-conceptual ages at the time of the scan of all neonatal subjects.

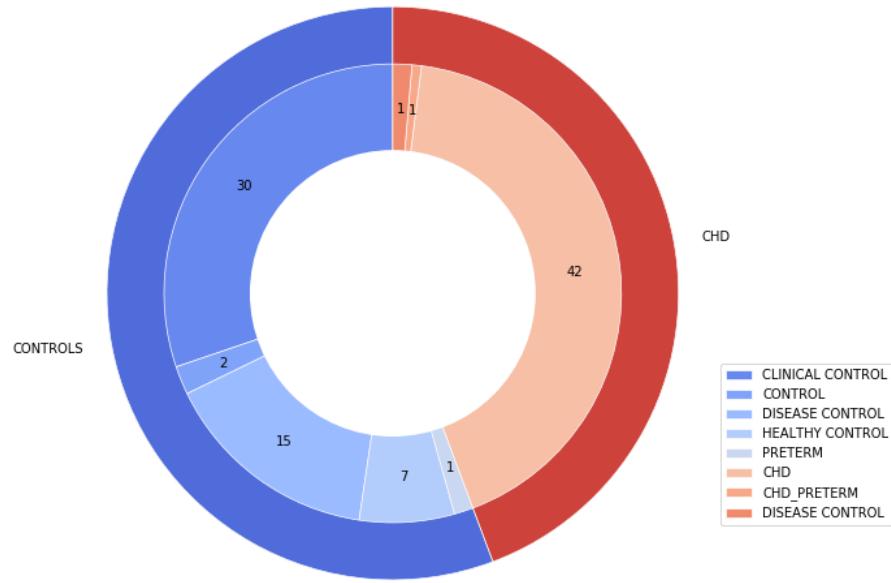


Figure 22: The breakdown of subject groups contained in the Control and CHD neonatal cohorts.

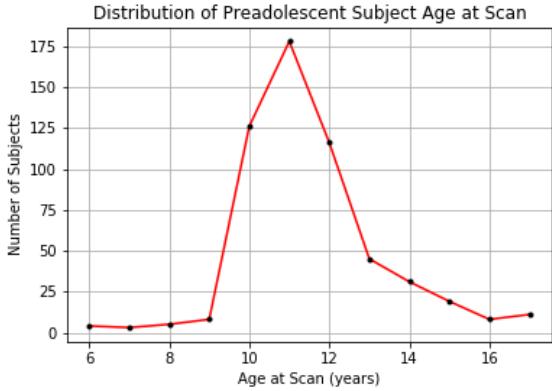


Figure 23: The distribution of preadolescent subject ages at the time of the scan.

subjects, and non-CHD clinical subjects (clinical control and disease control). The CHD group is composed of subjects diagnosed with CHD who were either born full-term or preterm. Of the entire cohort, 14 subjects underwent two scanning sessions resulting in 163 rs-fMRI scans.

As the neonates were most often asleep during the scan, they exhibit less motion overall compared to our other clinical cohorts. The high-motion neonates are an obvious exception to this concept, but many of the high-motion images contained long periods where the subject was stationary. Applying both the DAG-based framework and the traditional registration framework to these images provided the opportunity to compare the performances of both registration frameworks to each other in the context of the usability gold standard thresholds.

**5.2.2.3 Preadolescent Subject Population and Images** As part of a multicenter study of CHD in preadolescents, rs-fMRIs have been collected from twelve sites throughout the United States. The two sites from the neonatal study in the previous section also participated in this study and retained their respective labels. The subjects enrolled in this study were patients in the age range of 6 to 17 years (average age: 11 years) who either had CHD or were healthy with no neurocognitive impairments. The histogram of patient ages at the time of each scan can be seen in Figure 24.

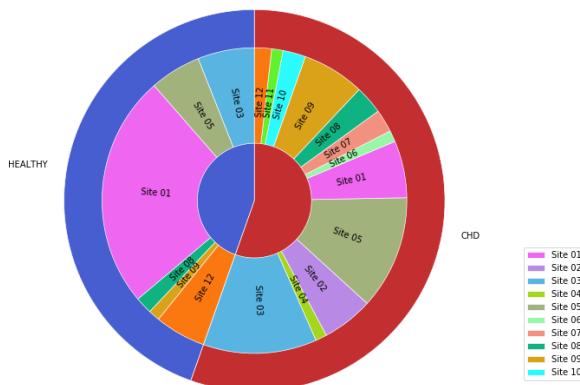


Figure 24: The distribution of CHD and healthy subjects between the 12 sites enrolled in the preadolescent study.

The majority of the subjects have CHD, though six sites also recruited healthy patients. The distribution of CHD and healthy subjects between the 12 sites can be seen in Figure 23 and in more detail in Table 1.

The subjects underwent neurodevelopmental testing prior to the neurological MRI scans. The neurodevelopmental testing was scheduled to take between 4.5 and 5 hours. The subject is given breaks to reduce study fatigue. The MRI scans were scheduled to take 60 minutes, assuming approximately 50 minutes of actual scan time and no repeat acquisitions.

As discussed earlier in this document, patient preparation and education before a scan can help reduce the amount of motion in the scan. The process used at Site 1 is partially dependent on the age of the preadolescent subjects. For younger subjects under the age of 8 years old, both the subject and the subject's family are asked whether they believe the subject could remain still for 45 minutes. The subject may also be shown the scanner to help determine his response. If the subject is comfortable being still for 45 minutes, the subject will undergo training with a mock scanner and the scan will be scheduled. For subjects older than 8 years, the subject himself determines whether or not he will undergo a brain scan. He also determines if he would like to undergo training with the mock scanner prior to the scan. This process or similar processes may be used at the other sites in the study.

Table 1: The number of CHD and healthy control preadolescent subjects scanned on each scanner type at each site.

<b>Site</b>	<b>Scanner</b>	<b>Field Strength</b>	<b>CHD</b>	<b>HEALTHY</b>
Site 01	Skyra	3T	22.0	128.0
Site 02	Unknown	Unknown	24.0	0.0
Site 03	Skyra	3T	83.0	44.0
Site 04	Unknown	Unknown	8.0	0.0
Site 05	Prisma	3T	23.0	12.0
Site 05	Skyra	3T	50.0	26.0
Site 06	Prisma	3T	6.0	0.0
Site 07	Trio	3T	14.0	0.0
Site 08	Prisma	3T	15.0	13.0
Site 09	Ingenia	3T	23.0	4.0
Site 10	Prisma	3T	14.0	0.0
Site 11	Skyra	3T	4.0	0.0
Site 12	Skyra	3T	11.0	32.0

During the scan itself, the subject is provided with ear protection in the form of earplugs and headphones. A stereotactic marker (vitamin capsule) is taped to the subject's right temple. The patient's head is positioned in a head coil so that his head and neck muscles are relaxed but not causing any head rotation. For comfort and lower body stabilization, his back or legs may be supported with an MR compatible pillow or foam pad. Additional pads or sponges may be used to stabilize and support the head. All subjects are positioned so that the centering crosshairs in the head coil are located at the subject's nasion (between the eyebrows). The patient is given a squeeze ball alarm to use if he becomes distressed during the scan and is reminded to remain still during the scan.

At Site 1, the research coordinator performs verbal checks on the subjects between every or every other scanning protocol during the same acquisition period. During the rs-fMRI scan, a star is projected onto a screen outside the scanner. The screen is reflected in the subject's line of sight via a mirror attached to the head coil. The subject is instructed to focus on the star during the rs-fMRI scan.

The parameters for the rs-fMRI scans were FOV = 256 mm and TE/TR = 32/650 ms with an interplane resolution of 4.0 x 4.0 mm, slice thickness of 4.0 mm and no gap between slices. At the sites using Siemens scanners, the sequences were 470 volumes in length while the sites using Philips scanners had sequences with a length of 380 volumes. Each volume was composed of 36 slices containing 64 x 64 voxels.

The multicenter imaging study of preadolescent subjects provides a unique opportunity to evaluate the efficacy of the DAG-based framework on a large subject cohort containing variable amounts of motion.

### 5.3 Hypothesis

At this point, we have described the methods we used to register image volumes within rs-fMRI sequences as well as the simulated data set and clinical cohorts which we used to test our methods.

We expect that the two registration frameworks will have different performances on the

different clinical age groups because we expect that the clinical age groups exhibit different types of motion.

-Overall, we expect that neonatal subjects exhibit the least amount of motion. The neonates are secured and sleeping in most cases, which limits their ranges of motion. The fetal subjects are expected to move the most because they have the fewest physical constraints out of the three age groups. The preadolescents are expected to move more than the neonatal subjects but less than the fetal subjects. The preadolescents are instructed to remain still during the scan, but are not restrained in any way.

We expect the DAG-based registration framework to be most effective at reducing the effects of motion for the fetal images. While the fetus does have full range of motion, it likely cannot move quickly around such a confined space. As a result, fetal motion is expected to be smoother than neonatal or preadolescent motion.

We also expect to see different patterns of motion between subjects who are healthy and subjects who have CHD. This expectation is based on the knowledge that neurodevelopmental disorders co-occur with CHD. Subjects with neurodevelopmental disorders are expected to have a more difficult time remaining still for the duration of a scan than the healthy subjects.

## 5.4 Summary

In this chapter, we discussed our two major data sets, a simulated data set and a clinical data set.

The simulated data were generated using an rs-fMRI simulation pipeline developed in-house and were used for the purpose of measuring ground truth signal recovered by motion correction techniques. A structural image from the ICBM Average Brain project and a set of functional ROIs associated with resting-state networks were used to generate the data. The data was given simulated scanner noise in the form of Gaussian noise injected into the k-space of the image. Finally, each image underwent rotation about the center of mass of the brain.

The pediatric data were obtained through prospective studies of CHD and neurodevelopment being conducted at the UPMC Children’s Hospital of Pittsburgh. CHD is a complex disease affecting millions of people globally. As the new treatments for CHD and its comorbidities have evolved, the life expectancy of patients with CHD has also increased.

Some of the comorbidities of CHD are related to a patient’s neurological development. As part of three studies of the relationship between CHD and brain development, rs-fMRI sequences of fetal, neonatal, and preadolescent brains have been gathered. A set of placental images was acquired alongside the fetal neurological images. The pediatric images were used to compare the two volume registration techniques on a diverse set of clinical data. We also used these images to examine patterns of motion unique to different age groups.

In the next chapter, we examine the results of the registration methods and of the machine learning techniques applied to the motion metrics for the images.

## 6.0 RESULTS

This chapter is divided into two sections. The first section focuses on the comparison of the two motion correction techniques. The second section focuses on the results of the machine learning algorithms applied to the metrics extracted from the images.

### 6.1 Volume Registration

Each of the clinical images underwent volume registration using both registration methods outlined in Chapter 3. The FD and DVARS metrics were calculated for every pair of subsequent volumes  $i$  and  $i + 1$  in the original sequences, the traditionally registered sequences, and the DAG-registered sequences.

The FD and DVARS values for each sequence were compared to the usability thresholds established by Power et al. (Power et al., 2012). These thresholds state that an image sequence is usable if at least 50% of the image volumes within it have  $FD < 2.0$  mm and  $DVARS < 2.5\%$  signal intensity change. The counts of image volumes which met these thresholds were calculated for each sequence type.

A series of independent two-sample t-tests were performed to determine the statistical significance of the differences in the counts of image volumes meeting each threshold. Three comparison pairs were used: the original sequence and the traditionally registered sequence, the original sequence and the DAG-registered sequence, and the two registered sequence types. The distributions of samples were the numbers of image volumes in each sequence meeting the usability threshold of interest. The null hypothesis of the t-tests was that the number of image volumes meeting the threshold being tested was the same for both sequence types.

The FD and DVARS metrics measure motion with respect to a pair of usability standards, but they offer a limited perspective about the motion present throughout an image sequence. Motion across the entire image was measured by calculating a similarity metric between every

image volume  $i$  and every other image volume  $j$ .

Motion across the whole sequence was measured by comparing every image volume in the sequence to every other image volume in the same sequence. The three metrics used for this comparison were the Dice metric and mutual information (MI). The calculations for a single metric on one sequence produce a two-dimensional matrix of metric values. These matrices were used to compare the quantities of motion throughout an entire sequence.

To describe the distribution of the similarity metrics matrices for the original sequence and each registered sequence, the minimum, first quartile, median, third quartile, and maximum values of each matrix were computed. The distributions of these five values were compared between registration types using two-sample t-tests.

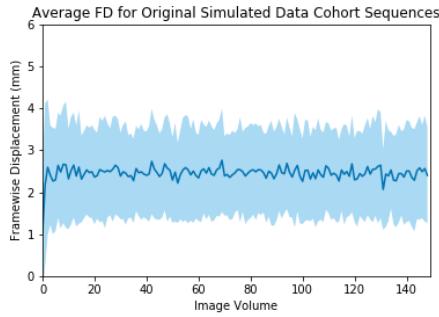
The distributions of all matrix values for the same similarity metric were compared using the Kolmogorov-Smirnov test. Three Kolmogorov-Smirnov tests were performed to compare the distributions of each metric for the original and traditionally registered sequences, the original and DAG-registered sequences, and the traditionally registered and DAG-registered sequences.

The simulated data underwent the same analyses as the clinical images, with one addition. Independent component analysis was performed on the simulated data to identify the signal components which contribute to the overall signal in the image. By correlating the components for each image with the DMN ROIs, the BOLD-related components were identified. The BOLD-related components were compared to the DMN ROIs on a voxel level to determine how many signal related voxels in the components were correctly recovered by the registrations.

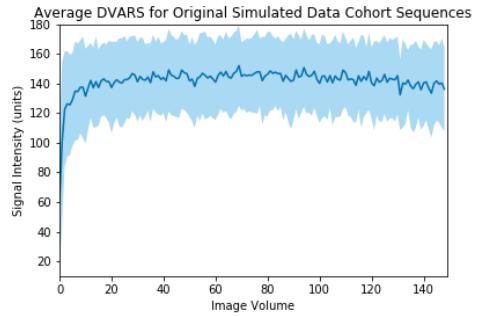
For brevity, several tables of p-values have been moved to Appendix B but are summarized here.

### 6.1.1 Simulated Data

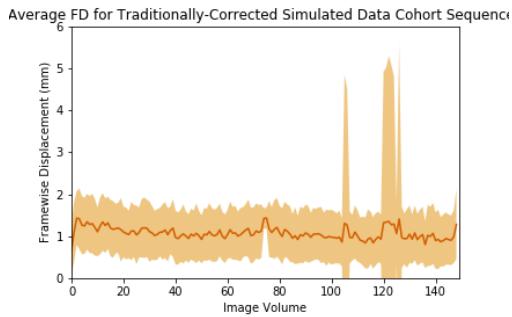
**6.1.1.1 Volume Registration: Power Thresholds** The FD and DVARS metrics were calculated between every pair of image volumes  $i$  and  $i + 1$  for the original sequences and both types of registered sequences. The mean and standard deviation of the FD and DVARS



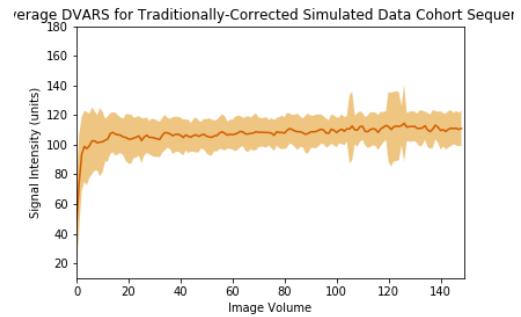
(a) FD of Original Sequences.



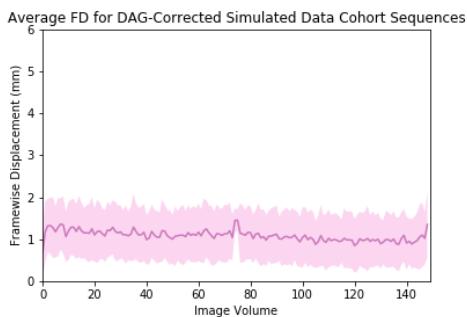
(b) DVARS of Original Sequences.



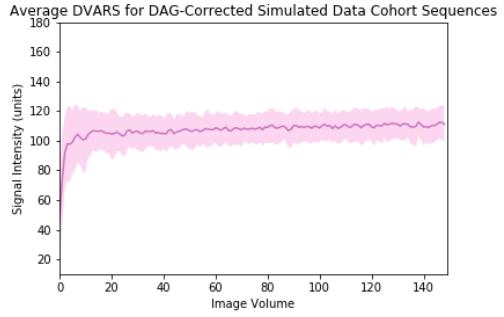
(c) FD of Traditionally Registered Sequences.



(d) DVARS of Traditionally Registered Sequences.



(e) FD of DAG-Registered Sequences.



(f) DVARS of DAG-Registered Sequences.

Figure 25: The means and standard deviations of the FD and DVARS metrics for all simulated images before and after each type of registration.

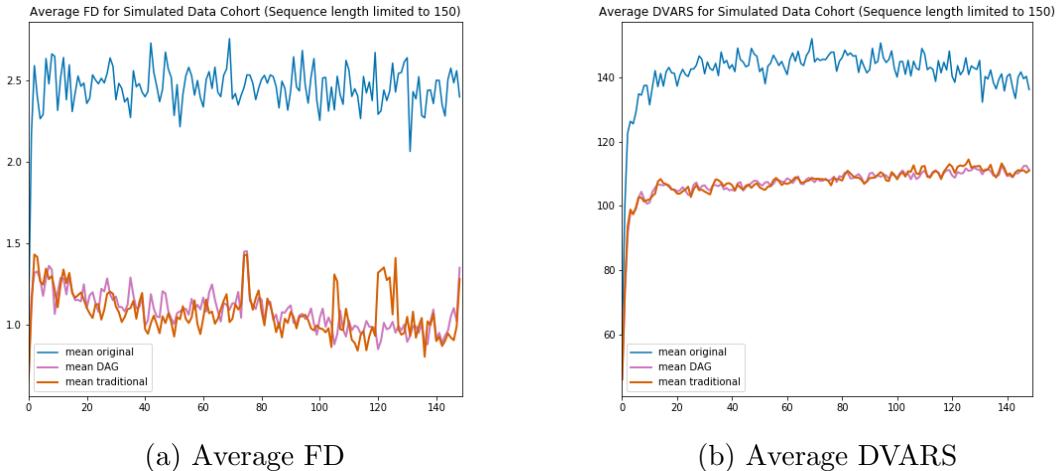


Figure 26: The means of the FD and DVARS metrics for all simulated images before and after both registrations.

metrics for each time point for each sequence type can be seen in Figure 25. Figure 26 shows the means of the FD and DVARS metrics for each time point for all sequence types on the same axes. The mean of the FD metrics decreased from 2.45 mm in the original sequences to 1.07 mm and 1.08 mm in the traditionally and DAG-registered sequences. The standard deviations of the FD slightly increased from 0.084 mm in the original sequences to 0.265 mm and 0.173 mm, respectively. Similarly, the mean DVARS of the original sequences decreased from 142 units to 107 units and 107 units in the registered images. The standard deviation of the DVARS also increased from 3.66 units in the original sequences to 4.77 units and 4.73 units in the registered sequences.

The FD and DVARS metrics for each image volume were compared to the usability criteria thresholds to determine the number of image volumes which were recovered by each registration type. The number of image volumes meeting the FD threshold, the DVARS threshold, and both thresholds were calculated. Table 2 shows the number of and percentage of image volumes that meet the specified thresholds. Each of the 90 image sequences had 150 image volumes. The total number of image volumes for the simulated data was 13410

Table 2: The counts and percentages of image volumes across each type of sequence in the simulated cohort which meet the usability thresholds of  $FD < 0.2$  mm and  $DVARS < 2.5\%$ .

Threshold	Original Sequences	Traditionally Registered Sequences	DAG-Registered Sequences
FD	98 (0.731%)	329 (2.453%)	279 (2.081%)
DVARS	54 (0.403%)	53 (0.395%)	53 (0.395%)
Both	53 (0.395%)	46 (0.343%)	44 (0.328%)

image volumes. In the original sequences, less than 1% of the image volumes met the individual thresholds, with only 0.395% of image volumes meeting both thresholds. After either registration, over 2% of the image volumes meet the FD threshold, though there was no significant change in the number of image volumes meeting the DVARS threshold.

The statistical significance of the difference in the counts of image volumes meeting each usability threshold was calculated. The only statistically significant differences in usability counts were for the number of image volumes meeting the FD threshold. The count of registered image volumes meeting the FD threshold for both registration types were significantly different from the counts for the original image sequences at  $p < 0.005$ . The difference between the counts of the two registration types was not significant ( $p = 0.127$ ). The complete set of results can be seen in Table 32.

The distributions of the FD and DVARS metrics were compared for each sequence type for each subject. The comparisons were performed using the Kolmogorov-Smirnov test. The Kolmogorov-Smirnov test evaluates the difference between a pair of probability distributions. The results of these comparisons can be found in Tables 3 and 4. The FD and DVARS distributions were significantly different between the original and traditionally registered sequences and the original and DAG-registered sequences at  $p < 0.005$ . Between the traditionally registered and DAG-registered sequences, 40 sequences (44.4%) had different FD distributions at  $p < 0.05$  and 27 sequences (30.0%) had different FD at  $p < 0.005$ . The

Table 3: The number of subjects whose sequences of types  $S_1$  and  $S_2$  had statistically different FD distributions.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	90	90
Original	DAG Registered	90	90
Traditionally Registered	DAG Registered	40	27

Table 4: The number of subjects whose sequences of types  $S_1$  and  $S_2$  had statistically different DVARS distributions.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	90	90
Original	DAG Registered	90	90
Traditionally Registered	DAG Registered	3	0

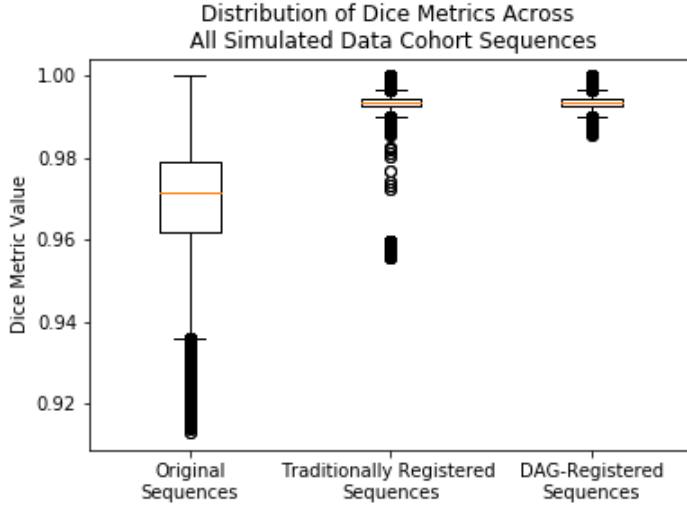


Figure 27: Boxplots of the values of all Dice matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the simulated data.

two types of registrations only had 3 sequences (3.33%) with different DVARS distributions at  $p < 0.05$  and none at  $p < 0.005$ .

Overall, the primary finding in the comparison of the registration techniques for the simulated data is the statistically significant decrease in FD by both registration types. There was also a decrease in the DVARS values for both registration types, but not a statistically significant one.

**6.1.1.2 Volume Registration: Sequence Duration Motion** The Dice and MI similarity matrices were calculated for each sequence. Box and whisker plots show the distributions of the metrics across the simulated data cohort for the Dice matrices and the mutual information matrices in Figures 27 and 28, respectively.

The Dice metric and MI matrices showed similar statistical differences, though these metrics have upper bound instead of a lower bound. The original sequences have statistically significantly different minimums, first quartiles, medians, and third quartiles from the registered sequences at  $p < 0.005$  and similar maximum values. The p-values for these

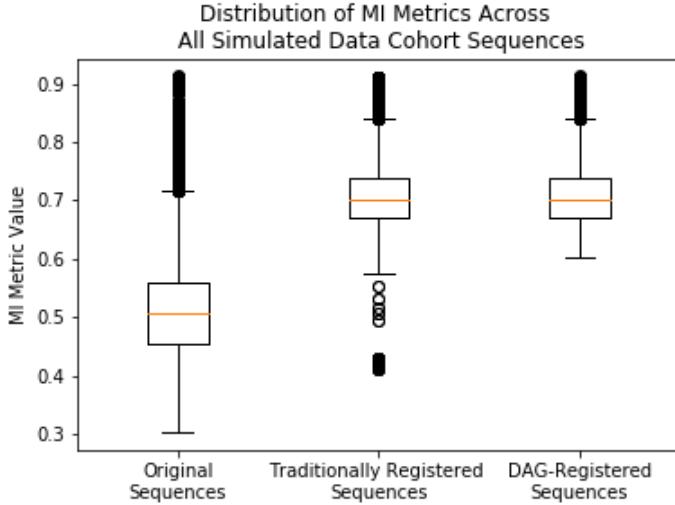


Figure 28: Boxplots of the values of all MI matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the simulated data.

calculations can be seen in Table 34 for the Dice matrices and Table 35 for the MI matrices.

The Kolmogorov-Smirnov test was used to compare the distributions for each metric. The tests showed that each pair of distributions for each metric type differed with a statistical significance of  $p < 0.005$ .

The analysis of sequence-level motion was performed using three metrics to measure the similarity of the volumes across each sequence. The matrices were compared to each other with respect to five descriptive statistics and with respect to the distribution of the entire set of similarity matrices for the original image sequences and each registration type. The distributions of the similarity metrics were found to be statistically significantly different, implying that each type of sequence contains different motion patterns.

**6.1.1.3 Volume Registration: Recovered Signal** The registered images underwent independent component analysis (ICA) to identify signal components which contribute to the image as a whole. Traditionally, ICA is used to identify signal components correlated with a patient's motion so that they can be regressed out of the image sequence. Since the

Table 5: The average rates and their confidence intervals for the classifications of component voxels as belonging to the DMN ROI for both types of registration.

Average Value	Traditionally Registered	DAG-Registered
TPR	$0.587 \pm 0.0697$	$0.623 \pm 0.0620$
FPR	$0.104 \pm 0.00188$	$0.00996 \pm 0.00180$
TNR	$0.990 \pm 0.00188$	$0.990 \pm 0.00180$
FNR	$0.413 \pm 0.0697$	$0.377 \pm 0.0620$

BOLD signal added to each simulated image is known, ICA can be used to determine how much BOLD signal is retained after volume registration.

ICA was applied to all registered sequences to identify the signal components. For each sequence, the components were compared to the DMN ROI. The components with the most significant correlation to the DMN ROI were categorized as the BOLD-related components. The average correlations between the BOLD-related components and the DMN ROI were 8122 for the traditionally registered sequences and 8040 for the DAG-registered sequences. The correlations for the two types of registered sequences were compared using both the Kolmogorov-Smirnov test and a two-sample t-test. Neither the test estimated a statistical significance between the correlation distributions for the registered sequences (KS p-value = 0.999; t-test p-value 0.807).

The accuracy of the BOLD-related components was determined by comparing the component image volumes to the DMN ROI on a voxel level. The absolute value of each BOLD-related component image volume was taken and thresholded at 10% of the maximum absolute value. The thresholded BOLD-related component image volumes were compared to DMN ROI. With the DMN ROI used as ground truth, the counts of voxels, which were true positive, false positive, true negative, and false negative rates were calculated for each registration type. These values can be seen in Table 5. The rates were compared for the traditionally registered and DAG-registered sequences. While none of the rates were statis-

tically significant at even  $p < 0.05$ , the rates were better for the DAG-registered sequences than for the traditionally registered sequences.

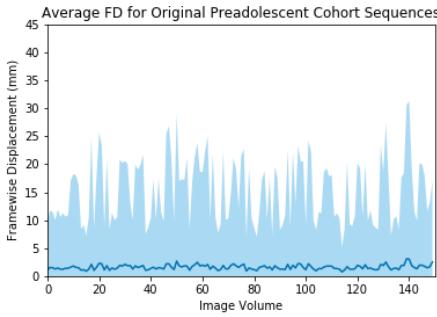
### 6.1.2 Preadolescent Cohort

**6.1.2.1 Volume Registration: Power Thresholds** The FD and DVARS were calculated for each original and registered preadolescent image sequences. The sequences varied in length from 150 volume to 450 volumes due to the differences in acquisition protocols at different sites. The number of image volumes used in this analysis was limited to the first 150 to avoid the challenges of comparing data with missing values.

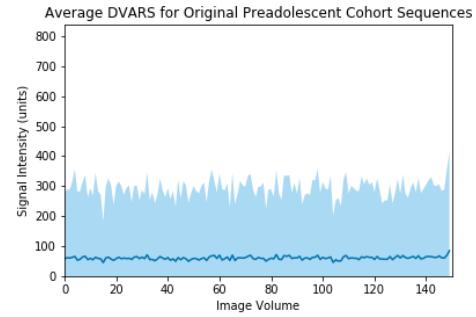
The means and stand deviations of the FD and DVARS at each timepoint for each registration type can be seen in Figure 29. The mean and standard deviation of the FD for the original sequences were 1.61 mm and 10.06 mm. The means of the FD increased to 1.83 mm for the traditionally registered sequences and 2.85 mm for the DAG-registered sequences. The standard deviation of the FD decreased to 7.74 mm for the traditionally registered sequences but increased to 15.4 mm for the DAG-registered sequences. The DVARS metrics followed a similar pattern: the mean of 60.5 units for the original sequences increased to 81.5 units and 92.5 units for the traditionally and DAG-registered sequences respectively while the standard deviation of the DVARS decreased from 197 units in the original sequences to 175 units in the traditionally registered sequences and increased to 284 units in the DAG-registered sequences.

The number of image volumes from each sequence type that met the FD and DVARS usability thresholds was calculated. The counts and percentages for each sequence type can be seen in Table 6. In the original image sequences, 60% of image volumes met the FD threshold, 49% of image volumes met the DVARS threshold, and 41% of image volumes met both thresholds. After either registration, 40% of image volumes met the FD threshold, 30% of image volumes met the DVARS threshold, and about 25% of image volumes met both thresholds.

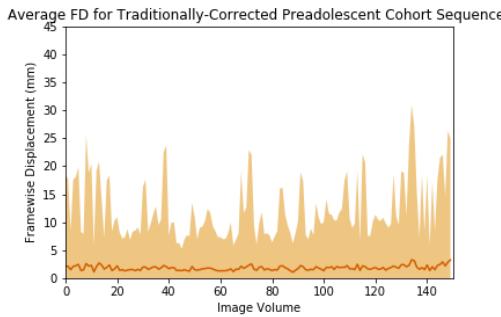
The complete set of results can be seen in Table 36. The original image sequences were statistically significantly different from both types of registered image sequences with respect



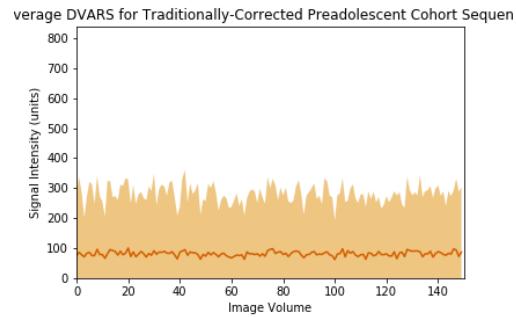
(a) FD of Original Sequences.



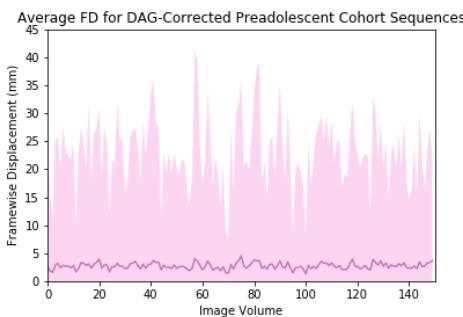
(b) DVARS of Original Sequences.



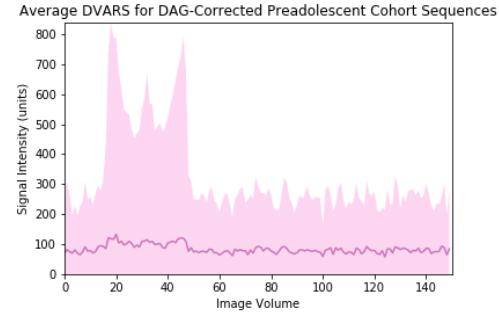
(c) FD of Traditionally Registered Sequences.



(d) DVARS of Traditionally Registered Sequences.



(e) FD of DAG-Registered Sequences.



(f) DVARS of DAG-Registered Sequences.

Figure 29: The means and standard deviations of the FD and DVARS metrics for all preadolescent images before and after each type of registration.

Table 6: The counts and percentages of image volumes across each type of sequence in the preadolescent cohort which meet the usability thresholds of  $FD < 0.2$  mm and  $DVARS < 2.5\%$ .

Threshold	Original Sequences	Traditionally Registered Sequences	DAG-Registered Sequences
FD	107879 (60.17%)	72104 (40.22%)	72114 (40.26%)
DVARS	87169 (48.62%)	54056 (30.15%)	53756 (30.01%)
Both	74107 (41.33%)	45006 (25.10%)	44682 (24.94%)

to the distributions of volumes meeting the FD threshold, the DVARS threshold, and both thresholds. The distributions for the registered image sequences were statistically similar for all three thresholds.

The distributions of the FD and DVARS values were compared for each sequence type for each subject. For each subject, the distributions of each metric for the original and traditionally registered, the original and DAG-registered, and the traditional and DAG-registered sequences were compared using the Kolmogorov-Smirnov test. The results of these tests can be seen in Tables 7 and 8. Out of 545 images, approximately 330 had FD and DVARS distributions that were statistically significantly different between the original sequences and both types of registered sequences for  $p < 0.005$ . Between the two types of registered sequences, 19 and 17 had statistically significant differences in their FD distributions at  $p < 0.05$  and  $p < 0.005$ , respectively. Similarly, 22 and 19 registered sequences had statistically significant differences in their DVARS distributions at  $p < 0.05$  and  $p < 0.005$ .

The comparison of the preadolescent sequences to the usability thresholds shows that either registration type increases the mean FD and DVARS metrics. The traditional registration method was shown to decrease the variance of the FD and DVARS metrics, but the DAG-based registration was shown to increase the variance of the FD and DVARS metrics. However, both registration types produced image sequences with similar numbers of image

Table 7: The number of preadolescent subjects whose sequences of types  $S_1$  and  $S_2$  had statistically different FD distributions.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	331	326
Original	DAG Registered	332	328
Traditionally Registered	DAG Registered	19	17

volumes meeting the FD threshold, the DVARS threshold, and both thresholds. The distributions of the FD and DVARS metrics were not statistically different for the two registration methods but were statistically significantly different between the original sequences and the registered sequences.

**6.1.2.2 Volume Registration: Sequence Duration Motion** For the preadolescent cohort, the Dice and MI matrices were calculated for the original, traditionally registered, and DAG-registered sequences. The distributions of the metric values for all matrices for each sequence can be seen in Figure 30 for the Dice matrices and in Figure 31 for the MI matrices. The minimums, first quantiles, medians, third quantiles, and maximums for each metric distribution were calculated. These descriptive statistics were compared to each other using the two-sample t-test for each sequence type. The only statistically significant difference in descriptive statistics for the Dice matrices was the difference between the third quartiles of the original sequences and both types of registered sequences at  $p < 0.005$ . There were no statistically significant differences between the descriptive statistics for the MI matrices, even at  $p < 0.05$ . The full results of these comparisons can be seen in Table

Table 8: The number of preadolescent subjects whose sequences of types  $S_1$  and  $S_2$  had statistically different DVARS distributions.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	334	331
Original	DAG Registered	334	333
Traditionally Registered	DAG Registered	22	19

B.37 for the Dice matrices and Table B.38 for the MI matrices.

The distributions of the Dice and MI matrices across all three sequence types for each subject were compared using the Kolmogorov-Smirnov test. For the Dice metrics, the original sequences and the registered sequences had approximately 190 significantly different sequences at  $p < 0.05$  and about 185 at  $p < 0.005$ . The traditionally registered and DAG-registered sequences had 84 sequences that were statistically different at  $p < 0.05$  and 69 at  $p < 0.005$ . In the comparison of MI metrics for the original sequences and the traditionally registered sequences, 188 and 185 had statistically significant differences at  $p < 0.05$  and  $p < 0.005$ . The original sequences and the DAG-registered sequences had similar differences at 187 and 185 significantly different sequences at  $p < 0.05$  and  $p < 0.005$ . The traditionally registered and DAG-registered sequences had 83 and 75 statistically significantly different sequences at  $p < 0.05$  and  $p < 0.005$ .

Overall, the motion patterns embodied by the Dice and MI matrices were fairly similar for each metric across all three sequence types. A subset of images had statistically significant differences between the original sequences and the registered sequences for both the Dice and MI matrices, and a smaller subset had statistically significant differences between the

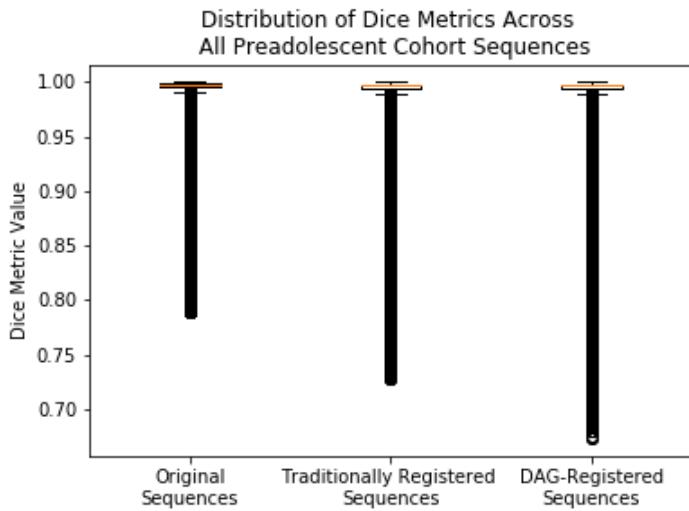


Figure 30: Boxplots of the values of all Dice matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the preadolescent cohort.

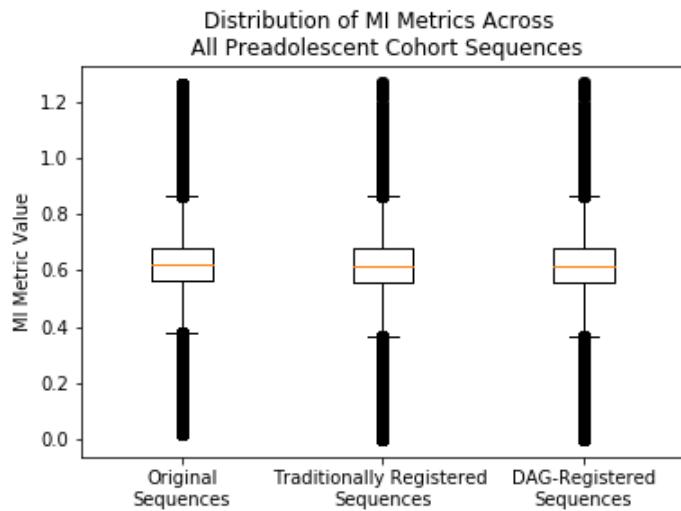


Figure 31: Boxplots of the values of all MI matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the preadolescent cohort.

Table 9: The number of preadolescent subjects whose sequences of types  $S_1$  and  $S_2$  had statistically different Dice distributions.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	188	185
Original	DAG Registered	187	185
Traditionally Registered	DAG Registered	83	75

Table 10: The number of preadolescent subjects whose sequences of types  $S_1$  and  $S_2$  had statistically different MI distributions.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	189	187
Original	DAG Registered	189	185
Traditionally Registered	DAG Registered	84	69

two registration types.

### 6.1.3 Neonatal Cohort

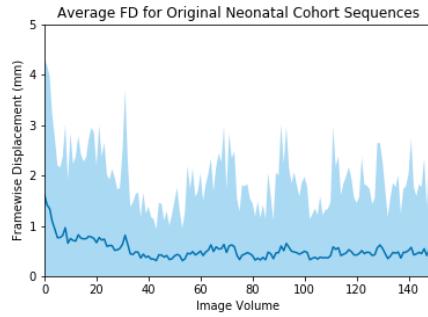
**6.1.3.1 Volume Registration: Power Thresholds** The FD and DVARS were calculated for the original and both types of registered sequences for the neonatal cohort. Each neonatal image was 150 volumes long. For each time point in the sequence, the mean and standard deviation of each metric for each image type was calculated. The means and standard deviations of these metric distributions can be seen in Figure 32.

The original sequences had an average FD of 0.522 mm with a standard deviation of 0.787 mm. After registration, these values increased to an average of 0.740 mm and a standard deviation of 1.035 mm for the traditionally registered sequences and an average of 0.736 mm and a standard deviation of 1.036 mm for the DAG-registered sequences.

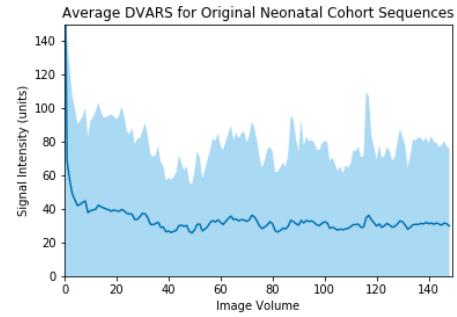
The DVARS metrics exhibited relationships between sequence types similar to the FD metrics. The average and standard deviation of the DVARS metrics for the original images were 34.1 units and 28.9 units. The traditionally registered sequences had a DVARS average and standard deviation of 47.1 units and 43.1 units, while the DAG-registered sequences had an average and standard deviation DVARS values of 46.8 units and 42.6 units.

For each sequence, the number of image volumes which met the FD, DVARS, and joint FD and DVARS thresholds were calculated. The total number of image volumes across all sequences of a single type was 23704. The counts and percentages of neonatal image volumes meeting the thresholds can be seen in Table 11. In the set of original image volumes, approximately 70% of the volumes met the FD threshold and the DVARS threshold. This percentage decreased to about 65% of the volumes when jointly considering the FD and DVARS thresholds. Both sets of registered image volumes had approximately 60% of the image volumes meeting the FD threshold, 58% of the image volumes meeting the DVARS threshold, and 54% of the image volumes meeting both thresholds.

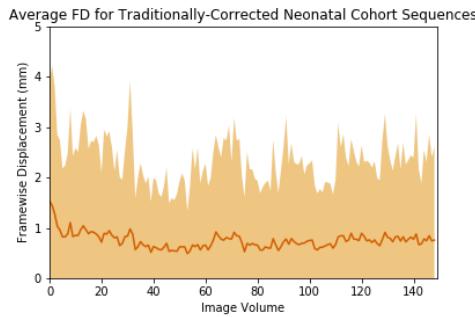
The statistical significance of the number of image volumes meeting each threshold was determined using two-sample t-tests. The three comparison pairs used were the original and traditionally registered sequences, the original and DAG-registered sequences, and the



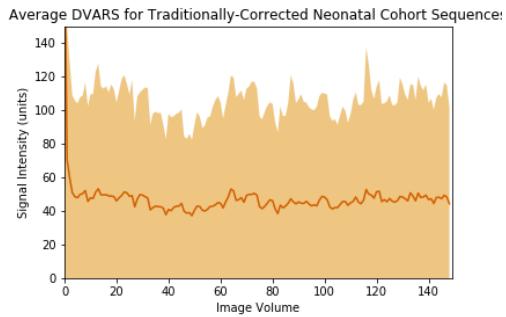
(a) FD of Original Sequences.



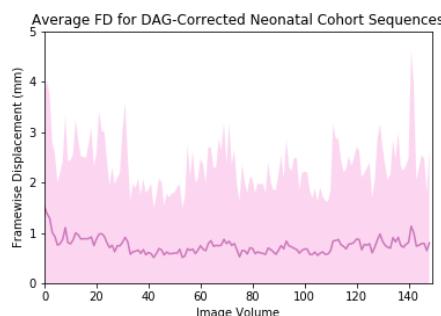
(b) DVARS of Original Sequences.



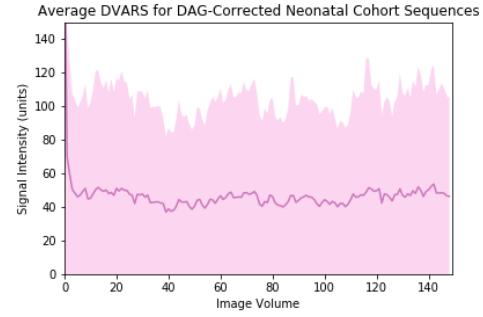
(c) FD of Traditionally Registered Sequences.



(d) DVARS of Traditionally Registered Sequences.



(e) FD of DAG-Registered Sequences.



(f) DVARS of DAG-Registered Sequences.

Figure 32: The means and standard deviations of the FD and DVARS metrics for all neonatal images before and after each type of registration.

Table 11: The counts and percentages of image volumes across each type of sequence in the neonatal cohort which meet the usability thresholds of  $FD < 0.2$  mm and  $DVARS < 2.5\%$ .

Threshold	Original Sequences	Traditionally Registered Sequences	DAG-Registered Sequences
FD	16495 (69.59%)	14264 (60.18%)	14173 (59.79%)
DVARS	16820 (70.96%)	13903 (58.65%)	13752 (58.02%)
Both	15332 (64.68%)	12837 (54.16%)	12684 (53.51%)

two types of registered sequences. The two types of registration did not have statistically significant differences in the number of image volumes meeting each threshold or the joint thresholds. The traditional registration had statistically significant differences from the number of image volumes in the original sequences in the number of image volumes meeting the FD threshold at  $p < 0.05$  and the DVARS and joint thresholds at  $p < 0.005$ . The volume counts for the DAG-registered sequences significantly differed from the volume counts for the original sequences for the individual and joint thresholds at  $p < 0.005$ . A complete set of the precise p-values can be seen in Table B.39.

For each image of all three sequence types, the distribution of the FD and DVARS metrics were compared using the Kolmogorov-Smirnov test. The number of images whose sequences of each type have different FD distributions can be seen in Table 12. The corresponding counts for the DVARS distributions can be seen in Table 13. Following the pattern of the previous results presented in this section, there were more statistically significant differences between the original sequences and either type of registered sequence than between the two types of registered sequences. The registered images had consistent counts for both the FD and DVARS metrics: 13 sequences and 11 sequences had significantly different FD and DVARS distributions at  $p < 0.05$ , respectively, and both types had 5 sequences that were significantly different at  $p < 0.005$ .

By comparing the FD and DVARS metrics of the original and registered sequences in

Table 12: The number of neonatal subjects whose sequences of types  $S_1$  and  $S_2$  had statistically different FD distributions.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	36	32
Original	DAG Registered	42	38
Traditionally Registered	DAG Registered	13	5

Table 13: The number of neonatal subjects whose sequences of types  $S_1$  and  $S_2$  had statistically different DVARS distributions.

# Sequences Type 1( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	44	39
Original	DAG Registered	45	44
Traditionally Registered	DAG Registered	11	5

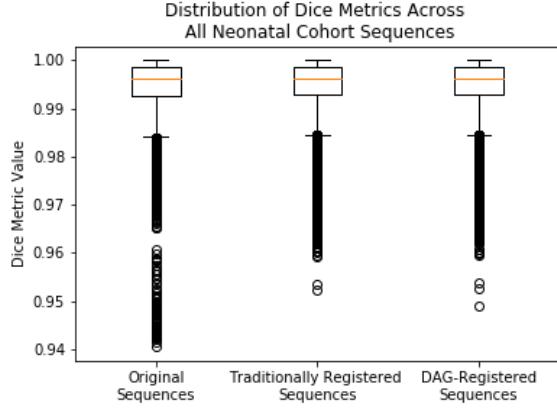


Figure 33: Boxplots of the values of all Dice matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the neonatal cohort.

several different ways, it has been established that images undergoing either type of registration differ from the original images. However, there was no statistically significant difference between the two types of registration with respect to the FD and DVARS metrics.

**6.1.3.2 Volume Registration: Sequence Duration Motion** Similarity matrices were computed to characterize patient motion throughout the duration of the image sequence. The value of each matrix at row  $i$  column  $j$  contains the similarity (either Dice or MI) between image volume  $i$  and image volume  $j$  in that sequence. The distributions of values for each metric type across the three sequence types can be seen in Figures 33 and 34. The values of the minimum, first quartile, median, third quartile, and maximum were computed for each distribution. These values were compared using two-sample t-tests. No statistically significant difference was found between the original, traditionally registered, or DAG-registered sequences for either metric (complete values can be found in Tables B.40 and B.41, respectively).

The distributions of the metrics matrices were compared to each other using the Kolmogorov-Smirnov test for each pair of sequence types. The number of image sequences with statistically significantly different distributions can be seen in Table 14 for the Dice metric and

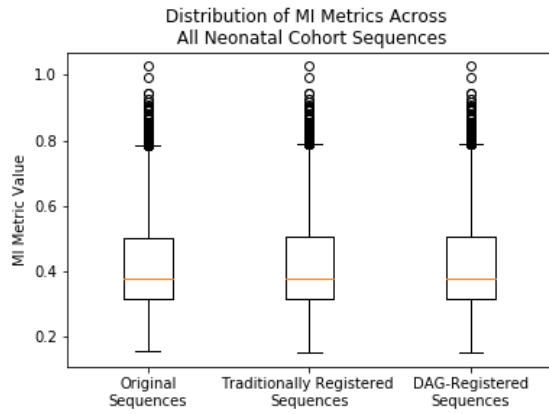


Figure 34: Boxplots of the values of all MI matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the neonatal cohort.

Table 14: The number of neonatal subjects whose sequences of types  $S_1$  and  $S_2$  had statistically different Dice distributions.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	67	66
Original	DAG Registered	71	69
Traditionally Registered	DAG Registered	68	64

Table 15: The number of neonatal subjects whose sequences of types  $S_1$  and  $S_2$  had statistically different MI distributions.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	66	64
Original	DAG Registered	71	69
Traditionally Registered	DAG Registered	64	64

Table 15 for the MI metric. For both metrics, about 70 image sequences differed at both  $p < 0.05$  and  $p < 0.005$  for all three pairs of sequence types.

The distributions of the similarity matrices were fairly similar across the whole neonatal cohort. However, the neonatal images have a greater number of statistically significantly different image sequences across all three sequence types than the simulated or the preadolescent images. There is also a higher statistical significance in the different images: at most, two image sequences had significant differences of  $p < 0.05$  and greater than  $p > 0.005$  for both metrics across all three sequence types.

#### 6.1.4 Fetal Cohort

As discussed in the previous chapter, fetal patients experience different physical constraints than any post-natal population. There is no way to prevent a fetus from moving during a scan. As a result, not all scans used in this analysis could be obtained under ideal circumstances. The images are supposed to contain 100 image volumes, but this is not the case for every sequence. Out of the 123 brain images, one contained only 81 volumes, two contained only 75 volumes, and one contained only 33 volumes. Of the 101 placenta images,

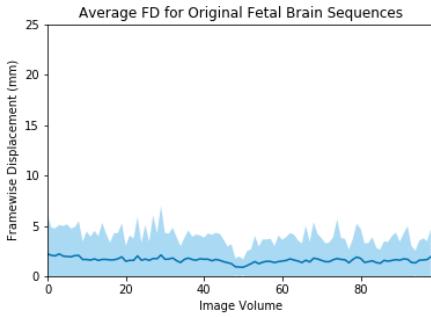
one contained only 95 volumes, one contained only 75 volumes, and one contained only 74 volumes. Additionally, not all sequences were able to undergo both types of registration successfully.

The analyses for the fetal cohort are divided into two groups: analyses for the brain images and analyses for the placenta images.

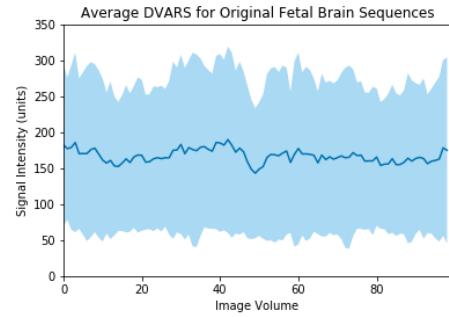
**6.1.4.1 Volume Registration: Power Thresholds, Fetal Brain** The FD and DVARS metrics were calculated for the fetal brain images. The averages and standard deviations of the FD and DVARS values across each point in time can be seen in Figure 35. The FD for the original image sequences had an average of 1.61 mm and a standard deviation of 1.33 mm. These values increased for the traditionally registered image sequences to an average of 1.81 mm and a standard deviation of 1.90 mm. However, the average and standard deviation of the DAG-registered sequences were close to those of the original image sequences at an average of 1.64 mm and a standard deviation of 1.38 mm. The DVARS metrics for all three sequence types had similar distributions with averages of 167 units, 168 units, and 169 units and standard deviations of 74 units, 75 units, and 75 units, respectively.

The FD and DVARS values for each image of each sequence type were compared to the usability thresholds. Table 16 contains the number and percentages of image volumes that meet the FD threshold, the DVARS threshold, and the joint thresholds. Of the 12041 original image volumes, only 575 (4.78%) met the FD threshold, and only 7 (0.058%) met the DVARS and joint thresholds. For the traditionally registered image volumes, of which there were 11969, 581 (4.85%) met the FD threshold, 84 (0.702%) met the DVARS threshold, and 80 (0.668%) met both thresholds. Again, the DAG-registered images closely followed the original images: of the 12041 image volumes, 561 (4.66%) met the FD threshold, and 7 (0.058%) met the DVARS and the joint thresholds. There was no statistically significant difference in the number of images meeting the thresholds between the three sequence types (complete set of p-values can be seen in Table B.42).

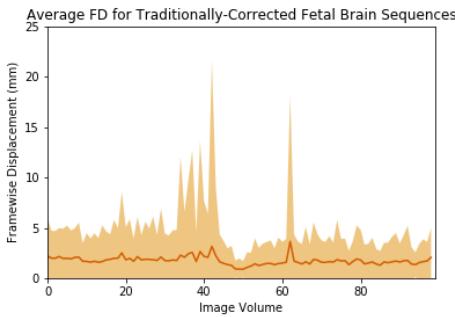
On an image level, the distributions of each metric across each type of sequence for each image were compared. Tables 17 and 18 contain the number of image sequences which were statistically different at  $p < 0.05$  and  $p < 0.005$  for the FD and DVARS metrics, respectively.



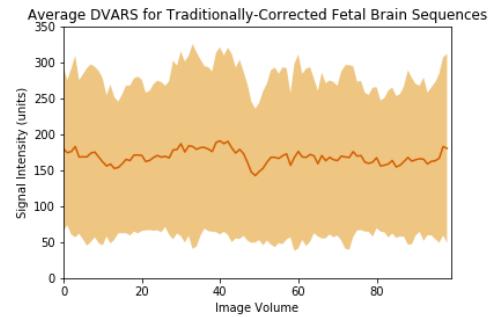
(a) FD of Original Sequences.



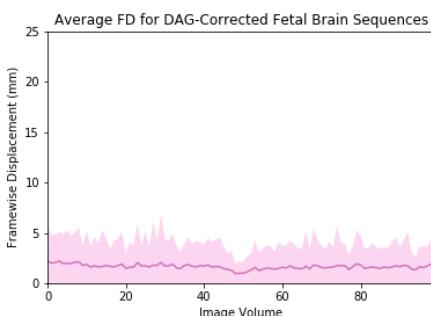
(b) DVARS of Original Sequences.



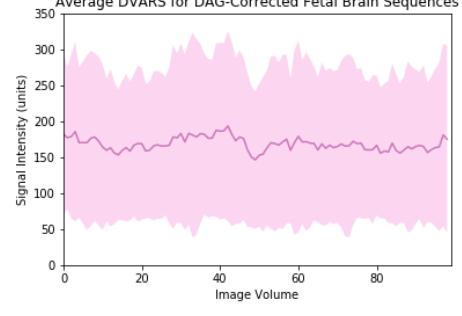
(c) FD of Traditionally Registered Sequences.



(d) DVARS of Traditionally Registered Sequences.



(e) FD of DAG-Registered Sequences.



(f) DVARS of DAG-Registered Sequences.

Figure 35: The means and standard deviations of the FD and DVARS metrics for all fetal brain images before and after each type of registration.

Table 16: The count of and percentage of image volumes across each type of sequence in the fetal brain image data set which meet the usability thresholds of  $FD < 0.2$  mm and  $DVARS < 2.5\%$ .

Threshold	Original Sequences	Traditionally Registered Sequences	DAG-Registered Sequences
FD	575 (4.775%)	581 (4.854%)	561 (4.659%)
DVARS	7 (0.058%)	84 (0.702%)	7 (0.058%)
Both	7 (0.058%)	80 (0.668%)	7 (0.058%)

Table 17: The number of subjects whose sequences of types  $S_1$  and  $S_2$  had statistically different FD distributions according to the Kolmogorov-Smirnov test.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	14	9
Original	DAG Registered	2	2
Traditionally Registered	DAG Registered	13	9

Table 18: The number of subjects whose sequences of types  $S_1$  and  $S_2$  had statistically different DVARS distributions according to the Kolmogorov-Smirnov test.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	3	3
Original	DAG Registered	2	1
Traditionally Registered	DAG Registered	2	2

These tables show that the original and the DAG-registered sequences were generally more similar than either the original or the DAG-registered sequences were to the traditionally registered sequences with respect to the FD. The distributions of the DVARS metrics were not identified as statistically different for any sequence type.

The FD and DVARS metrics suggest two concepts. First, the DAG-based registration produced image sequences that were more similar to the original sequences than the traditionally registered sequences were. Second, neither registration had a significant impact on the DVARS metrics.

**6.1.4.2 Volume Registration: Sequence Duration Motion, Fetal Brain** The changes in fetal brain position across the duration of the image sequences were measured using the similarity matrices. For each possible pair of image volumes  $i$  and  $j$ , the value of the matrix at row  $i$  column  $j$  is the similarity of the two volumes. The two similarity metrics used were the Dice metric and the MI metric. The distributions of all values for every image of each sequence type can be seen in Figure 36 for the Dice metric and in Figure 37 for the MI metric.

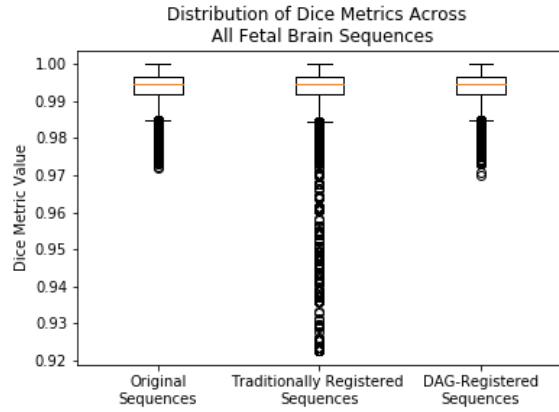


Figure 36: Boxplots of the values of all Dice matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the fetal-brain images.

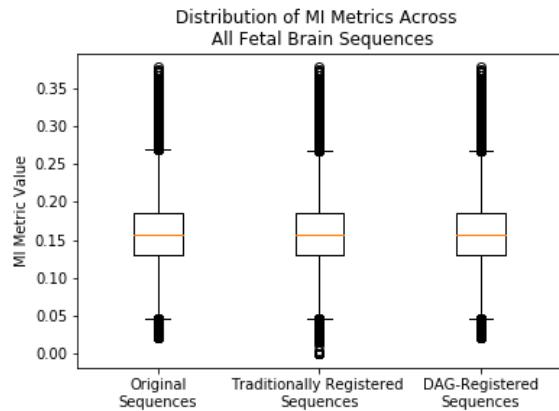


Figure 37: Boxplots of the values of all MI matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the fetal brain images.

Table 19: The number of fetal brain images whose sequences of types  $S_1$  and  $S_2$  had statistically different Dice distributions.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	11	9
Original	DAG Registered	7	7
Traditionally Registered	DAG Registered	10	8

The minimum, first quartile, median, third quartile, and maximum values for each distribution were calculated. These values were then compared using the two-sample t-test. No statistically significant difference was found between these values for any pair of sequence types for either the Dice matrices or the MI matrices. (For a complete list of p-values, see Tables B.43 and B.44.

The distributions of values in each matrix for each subject were compared to each other using the Kolmogorov-Smirnov test. The number of sequence comparisons that were statistically significant for the Dice metrics can be seen in Table 19. Very few sequences had significantly different Dice distributions: between the three sequence types, only 7 to 11 sequences were statistically different at  $p < 0.05$ , and only 7 to 9 sequences were different at  $p < 0.005$ . The number of sequence comparisons that were statistically significant for the MI metrics can be seen in Table 20. Similar to the Dice metrics, only 12 to 14 sequences were different at  $p < 0.05$  and only 9 to 10 sequences were different at  $p < 0.005$ .

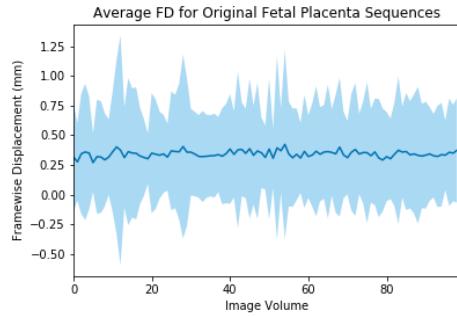
Overall, the distributions of the similarity metric matrices between the three sequences types were not statistically different.

Table 20: The number of fetal brain images whose sequences of types  $S_1$  and  $S_2$  had statistically different MI distributions.

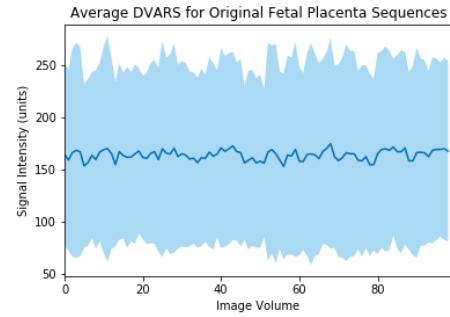
# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	13	10
Original	DAG Registered	12	9
Traditionally Registered	DAG Registered	14	9

**6.1.4.3 Volume Registration: Power Thresholds, Placenta** The FD and DVARS values for the placental images were calculated. The averages and standard deviations for the metric values at each time point in the sequence can be seen in Figure 38. The original sequences had an average FD of 0.341 mm with a standard deviation of 0.382 mm and an average DVARS of 163 units with a standard deviation of 73.8 units. The FD values increased after registration: the traditionally registered images had an average FD of 4.17 mm with a standard deviation of 8.67 mm, and the DAG-registered images had an average FD of 5.12 mm with a standard deviation of 9.14 mm. Registration had a different effect on the DVARS values. The average DVARS values decreased to 140 units for the traditionally registered sequences and to 127 units for the DAG-registered sequences. The standard deviation of the DVARS values increased, though, to 85.7 units for the traditionally registered images and to 91.9 units for the DAG-registered images.

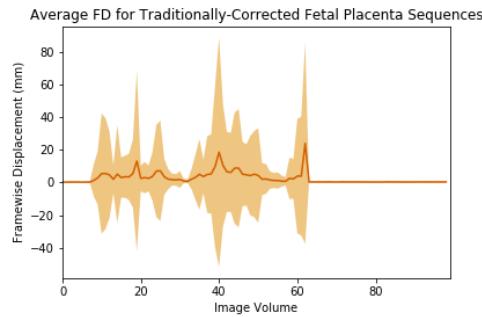
The FD and DVARS values were compared to the usability thresholds. Table 21 contains the number and percentages of image volumes from all placenta sequence that meet the FD, the DVARS, and the joint thresholds. The total number of image volumes across the original sequences was 10017, the number of traditionally registered images was 9296, and the number



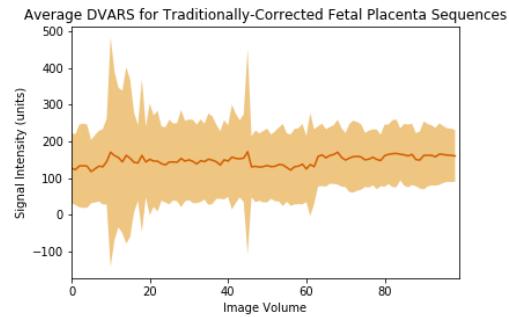
(a) FD of Original Sequences.



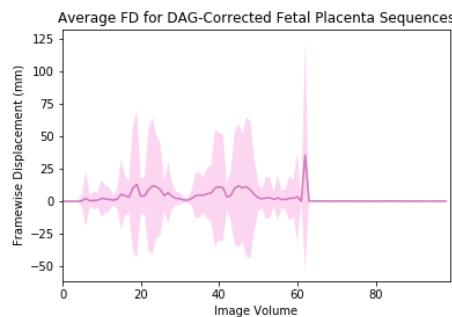
(b) DVARS of Original Sequences.



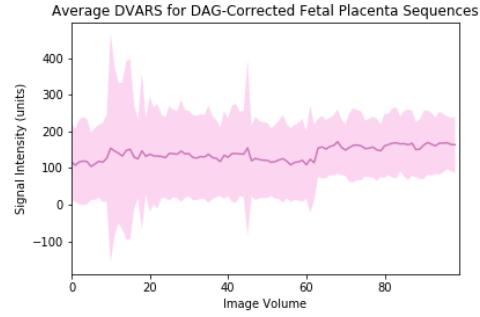
(c) FD of Traditionally Registered Sequences.



(d) DVARS of Traditionally Registered Sequences.



(e) FD of DAG-Registered Sequences.



(f) DVARS of DAG-Registered Sequences.

Figure 38: The means and standard deviations of the FD and DVARS metrics for all placenta images before and after each type of registration.

Table 21: The count of and percentage of image volumes across each type of sequence in the fetal placenta image data set which meet the usability thresholds of  $FD < 0.2$  mm and  $DVARS < 2.5\%$ .

Threshold	Original Sequences	Traditionally Registered Sequences	DAG-Registered Sequences
FD	10017 (43.646%)	4113 (44.245%)	4005 (45.020%)
DVARS	17 (0.170%)	1056 (11.36%)	1624 (18.255%)
Both	17 (0.170%)	599 (6.444%)	996 (11.196%)

of DAG-registered images was 8896.

Of the original image volumes, 10017 (43.6%) met the FD threshold. Only 17 (0.170%) original image volumes met the DVARS threshold and the FD and DVARS thresholds. Both registration techniques reduced the number of image volumes, which met the FD thresholds to 4113 (44.245%) volumes for the traditionally registered sequences and 4005 (45.1%) for the DAG-registered sequences. The registration techniques increased the number of image volumes which met the DVARS threshold and the FD and DVARS thresholds. After traditional registration, 1056 (11.4%) volumes met the DVARS threshold and 599 (6.44%) met both thresholds. After DAG-based registration, 1624 (18.3%) volumes met the DVARS threshold and 996 (11.2%) volumes met both thresholds. Two-sample t-tests, the results of which can be found in Table B.45, showed that the differences in the number of image volumes meeting the DVARS thresholds and the joint thresholds between the original sequences and the registered sequences are statistically significant at  $p < 0.005$ .

For each placenta image, the distribution of the FD and DVARS metrics for each sequence type were compared using the Kolmogorov-Smirnov test. The results of these tests can be seen in Table 22 and Table 23.

The number of sequences which had statistically different FD distributions across all three sequence types can be seen in Table 22. The original and traditionally registered sequences

Table 22: The number of placental images whose sequences of types  $S_1$  and  $S_2$  had statistically different FD distributions according to the Kolmogorov-Smirnov test.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	21	21
Original	DAG Registered	32	30
Traditionally Registered	DAG Registered	31	20

Table 23: The number of placental images whose sequences of types  $S_1$  and  $S_2$  had statistically different DVARS distributions according to the Kolmogorov-Smirnov test.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	20	20
Original	DAG Registered	29	29
Traditionally Registered	DAG Registered	19	19

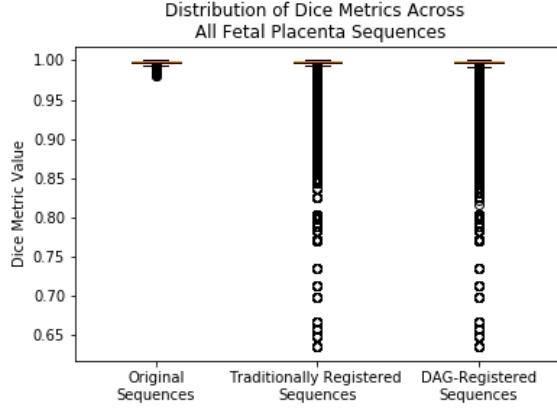


Figure 39: Boxplots of the values of all Dice matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the placenta images.

had 21 sequences with significantly different FD at both  $p < 0.05$  and  $p < 0.005$ . The original and DAG-registered sequences had 32 and 30 sequences with statistically different FD at  $p < 0.05$  and  $p < 0.005$ , respectively. The two types of registration have 31 and 20 sequences with statistically different FD at  $p < 0.05$  and  $p < 0.005$ . To summarize, approximately 20% to 30% of the images had statistically significantly different FD distributions depending on the sequence type.

With respect to the DVARS distributions, the traditionally registered sequences were more similar to the original sequences or the DAG-registered sequences. There were about 20 sequences that had different DVARS distributions between the traditionally registered sequences and the original and DAG-registered sequences at  $p < 0.005$ . The original and DAG-registered sequences were slightly less similar with 29 sequences between them which had different DVARS distributions at  $p < 0.005$ .

The FD distributions were not significantly different for the original and registered placenta sequences. However, the placenta images are the first images in our analyses which demonstrate an improvement in DVARS metrics after registration.

Table 24: Results of t-tests comparing the descriptive statistics of the Dice matrices for the fetal placenta data.

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same minimums)	6.39 E -5	3.43 E -7	0.257
P( $S_1$ and $S_2$ have same 1st quartile)	6.54 E -5	2.35 E -6	0.257
P( $S_1$ and $S_2$ have same medians)	0.000310	9.46 E -5	0.816
P( $S_1$ and $S_2$ have same 3rd quartiles)	0.096	0.104	0.902
P( $S_1$ and $S_2$ have same maximums)	1.0	1.0	1.0

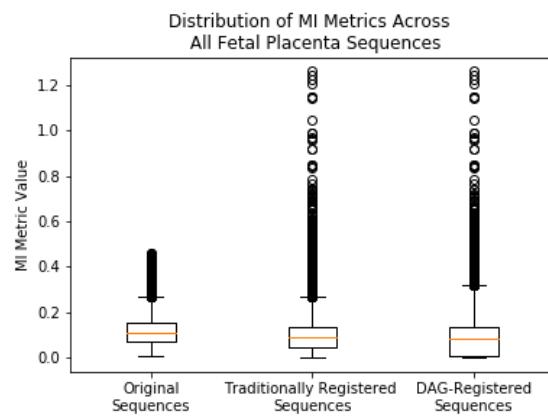


Figure 40: Boxplots of the values of all MI matrices for the original sequences, the traditionally registered sequences, and the DAG-registered sequences for the placenta images.

Table 25: Results of t-tests comparing the descriptive statistics of the MI matrices for the fetal placenta data.

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same minimums)	0.00980	0.00168	0.536
P( $S_1$ and $S_2$ have same 1st quartile)	0.00742	0.00132	0.548
P( $S_1$ and $S_2$ have same medians)	0.00713	0.00116	0.535
P( $S_1$ and $S_2$ have same 3rd quartiles)	0.00807	0.00118	0.506
P( $S_1$ and $S_2$ have same maximums)	0.00145	8.28 E -6	0.212

**6.1.4.4 Volume Registration: Sequence Duration Motion, Placenta** The image volumes across each sequence were compared by calculating the similarity between volume  $i$  to volume  $j$  and storing the results in row  $i$  column  $j$  of a 2D matrix for that sequence's similarity metric. The similarity metrics used for these comparisons were the Dice metric and the MI metric. The distributions of the values of each metric for each sequence type can be seen in Figure 39 and Figure 40 for the Dice and MI matrices, respectively. These distributions were compared by performing two-sample t-tests on the minimum, first quartile, median, third quartile, and maximum values for each sequence type.

Table 24 shows the results of the t-tests comparing the descriptive statistics for the Dice matrices. The minimum values, first quartiles, and medians were statistically different at  $p < 0.005$  for the original and traditionally registered sequences as well as for the original and DAG-registered sequences. There was no significant difference between the third quartiles and the maximum values between the original and registered sequences. Additionally, there was no significant difference between the descriptive statistics for the traditionally and DAG-registered sequences.

The results of the t-tests comparing the descriptive statistics for the MI matrices can be seen in Table 25. The differences in all five statistics between the original sequences and the traditionally registered sequences were significant at  $p < 0.05$ . The differences between the original and DAG-registered sequences for all five statistics were significant at  $p < 0.005$ . There was no significant difference in the descriptive statistics for the MI matrix distributions between the traditionally registered and DAG-registered sequences.

The distributions of the matrix values were compared on a subject level. These comparisons consisted of using the Kolmogorov-Smirnov test to determine if the distributions of the matrix values for two sequence types for the same subject were statistically different. The results of these comparisons can be seen in Table 26 for the Dice matrices and in Table 27 for the MI matrices. For both the Dice and MI comparisons, there were 22 sequences between the original and traditionally registered sequences which were statistically different at  $p < 0.005$ . The Dice and MI comparisons also both had 39 statistically different sequences between the original and DAG-registered sequences at  $p < 0.005$ . The traditionally and DAG-registered sequences had 38 statistically different sequences for the Dice comparison

Table 26: The number of subjects whose placenta sequences of types  $S_1$  and  $S_2$  had statistically different Dice distributions.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	22	22
Original	DAG Registered	39	39
Traditionally Registered	DAG Registered	30	30

Table 27: The number of subjects whose placenta sequences of types  $S_1$  and  $S_2$  had statistically different MI distributions.

# Sequences Type 1 ( $S_1$ )	# Sequences Type 2 ( $S_2$ )	# Sequences $p < 0.05$	# Sequences $p < 0.005$
Original	Traditionally Registered	22	22
Original	DAG Registered	39	39
Traditionally Registered	DAG Registered	38	38

and 30 statistically different sequences for the MI comparison.

The similarity matrices for the placenta images suggest that the two registration types both produce sequences that have similarity distributions that are different from those of the original sequences. The two registration types differ in the metric distributions for some images, but over a population these differences are not significant.

## 6.2 Characterizing Motion Across Clinical Groups

Between the preadolescent, neonatal, and fetal brain images, we have a fairly large set of images. The images come from a population that is diverse and well represented in two areas: disease status and general patient age group. In this section, we compare groups of images identified using only the metrics discussed earlier in this chapter to these two groups of interest.

As the goal of this analysis is to characterize patient motion, only the metrics for the original, unregistered sequences were used. These metrics exist in high dimensional spaces. In the high dimensional space, outlier vectors greater than three standard deviations from the mean vector were removed. The remaining feature vectors were clustered in the high dimensional spaces and then projected into 2D spaces for visualization purposes.

### 6.2.1 CHD and Control

The FD, DVARS, Dice matrices, and MI matrices for the original clinical brain images were projected into 2D space and labeled using each subject's disease status. The projections of this data can be seen in Figure 43. In this figure, each column represents a different dimensionality reduction technique, and each row represents a different metric type. The different dimensionality reduction techniques offer different views on the same data, though the relationships between these views are not always clear. It is very difficult to see any distinct groups of CHD or control labels in any of these views.

The FD, DVARS, Dice matrices, and MI matrices were then each clustered using ag-

Dimensionality Reduction: BOLD Metrics

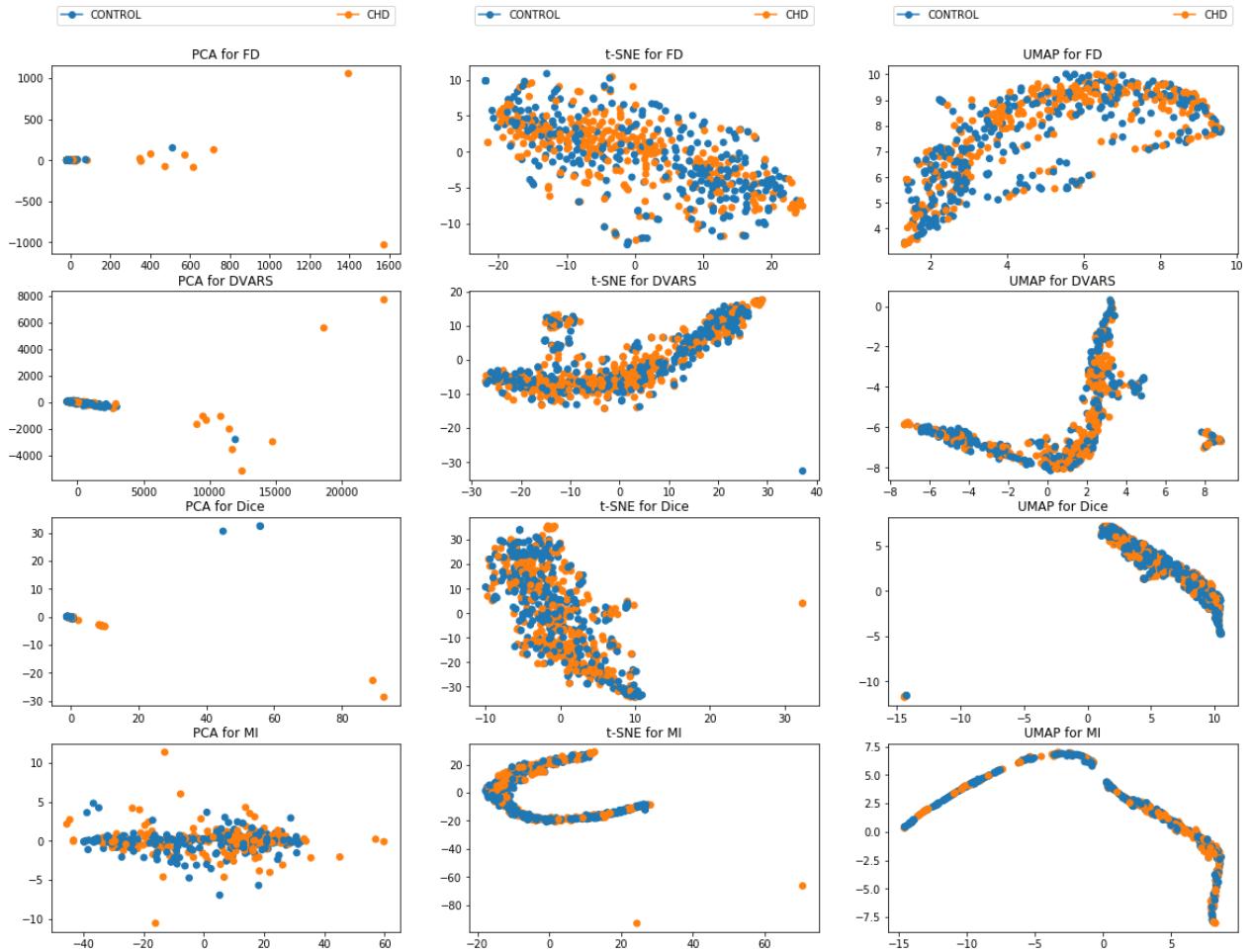


Figure 41: The data for the four metrics used to evaluate motion in the clinical brain images projected using different methods into 2D space labeled by CHD/Control status.

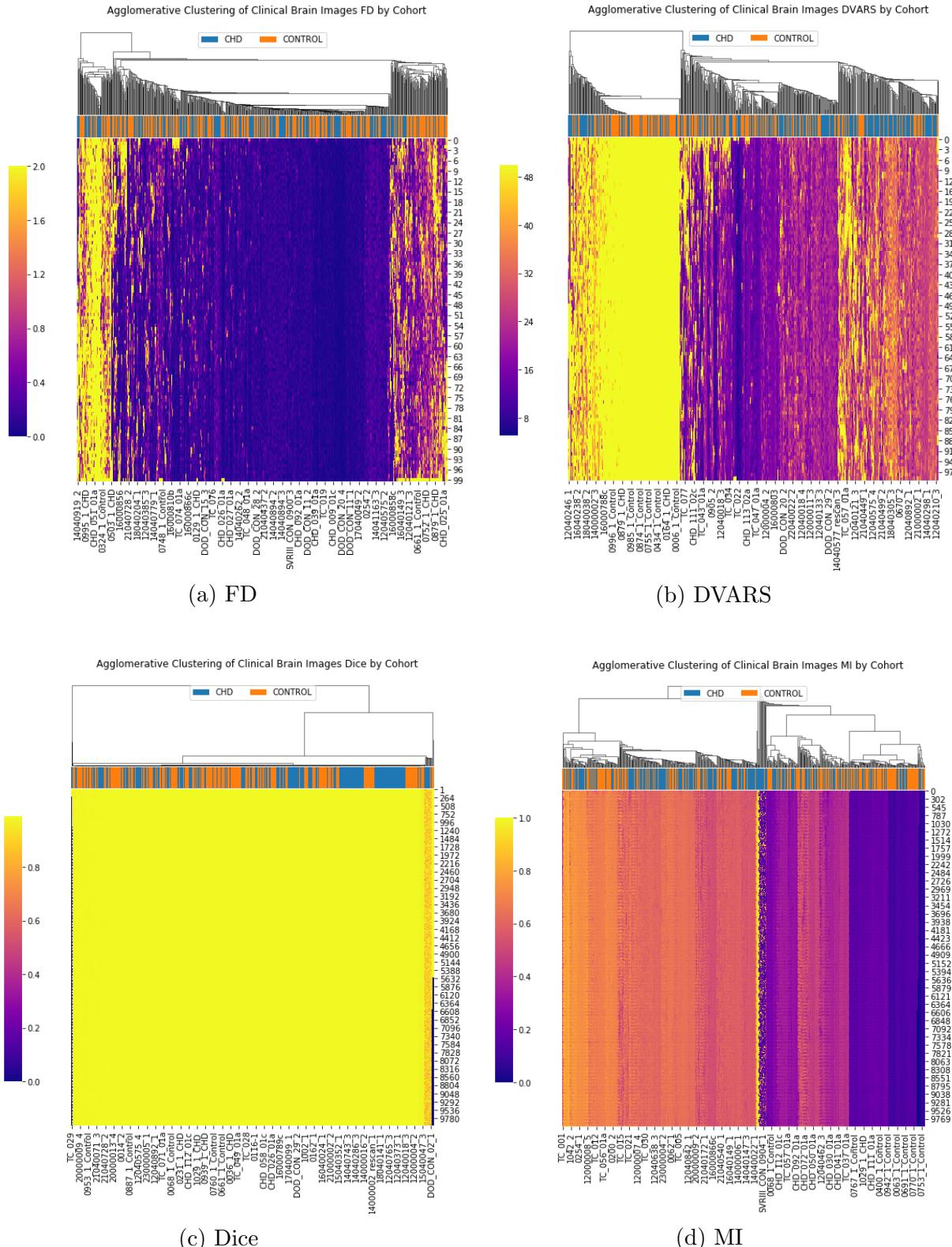


Figure 42: The original preadolescent, neonatal, and fetal images clustered by each metric using agglomerative clustering and labeled by disease status.

glomerative clustering. Heatmaps of these clustering results can be seen in Figure 42. The columns of the heatmaps show the feature vectors used to represent each image sequence. The colorbars to the left of the heatmaps show the range of values present in each heatmap. In all four heatmaps, darker colors represent lower values. At the top of each heatmap is a row of colored labels where the blue labels indicate subjects who have CHD and the orange labels indicate control subjects. Above the row of labels is a dendrogram. The dendrogram explicitly shows the distances between the feature vectors. Each vertical bar in a dendrogram represents the distance between that feature vector and the feature vector or group of feature vectors most similar to it. The horizontal bars indicate connections between feature vectors or groups of feature vectors which are the next closest to each other.

The heatmaps for the clustering results generated using the FD, DVARS, and MI metrics each show several distinct vertical bands representing clusters of similar feature vectors. The heatmap for the Dice metrics is less interpretable as most of the values in the heatmap are close to 1.0. In all four sets of clustering results, there are no distinguishable clusters which are mostly composed of either CHD or control subjects.

### 6.2.2 Age Groups

The three age groups of interest were preadolescent, neonatal, and fetal. Figure 43 shows the motion metrics projected into 2D and labeled with the correct age groups. The blue labels indicate preadolescent subjects, the orange labels indicate neonatal subjects, and the green labels indicate fetal subjects. In this figure, the first three rows of the first column (PCA of FD, DVARS, and Dice) show highly concentrated groups of subjects with some sparse, distant outliers. In the first three rows of the second and third columns (t-SNE and UMAP for FD, DVARS, and Dice), the preadolescent and fetal groups are mostly separate, and the neonatal subjects span these groups. In the last row of all three columns (PCA, t-SNE, and UMAP for MI), groups of subjects from the different age groups are clearly visible despite some overlap.

The four types of metrics were clustered using agglomerative clustering. The agglomerative clustering results labeled using the age groups can be seen in Figure 44. The same

Dimensionality Reduction: BOLD Metrics

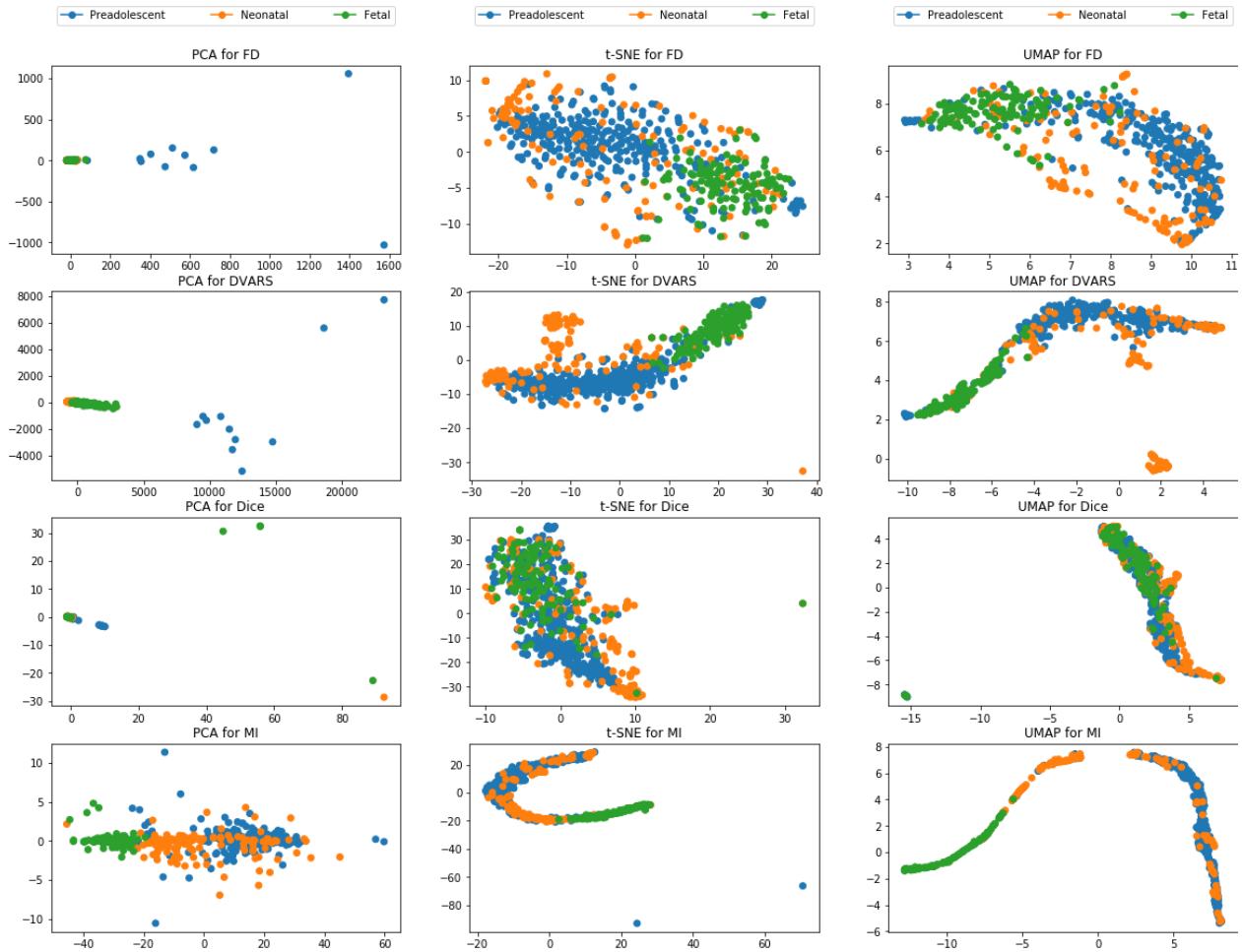


Figure 43: The data for the four metrics used to evaluate motion in the clinical brain images projected using different methods into 2D space and labeled by age group.

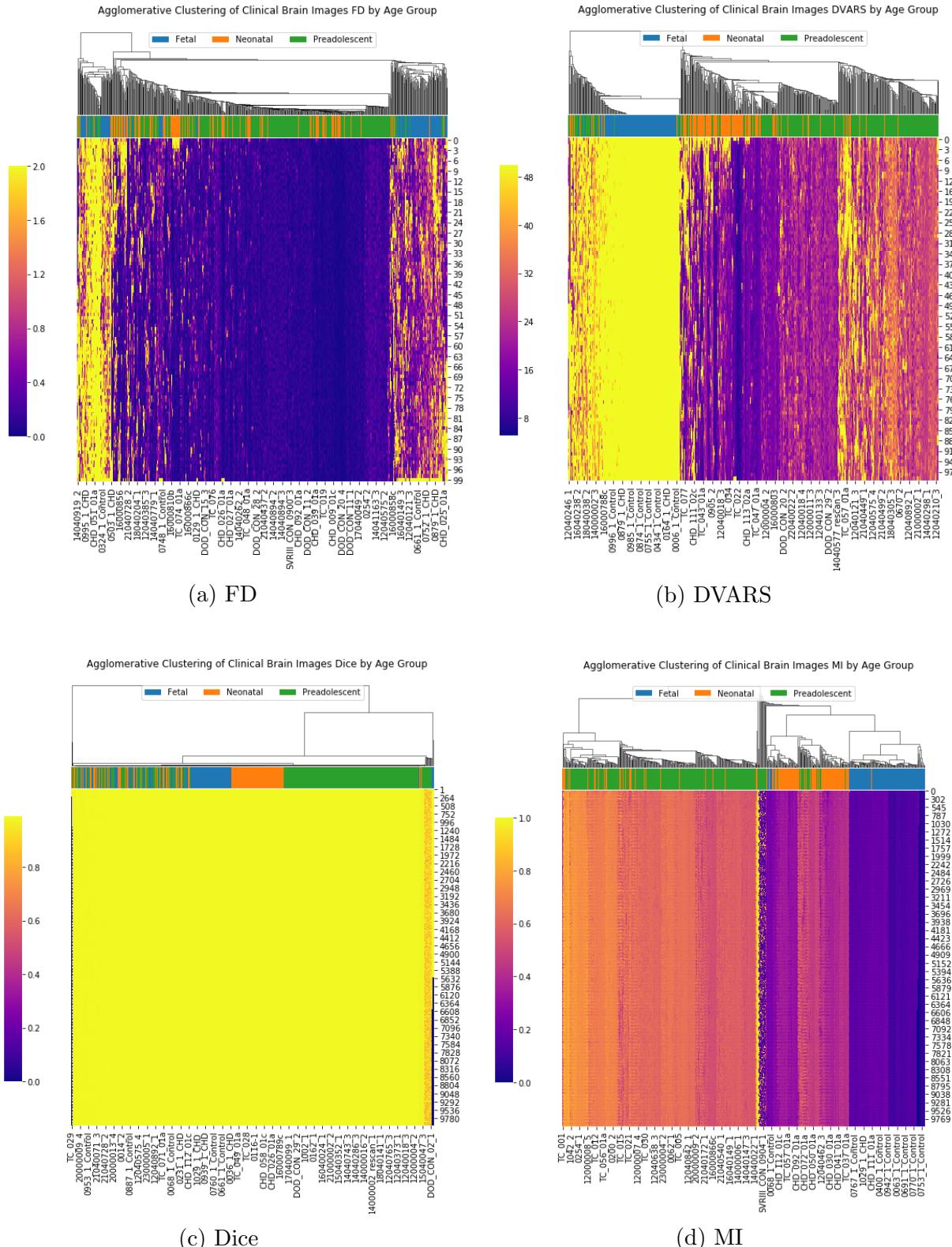


Figure 44: The preadolescent, neonatal, and fetal images clustered by each metric using agglomerative clustering and labeled by age group.

clusters present in the heatmaps of Figure 42 can be seen in these heatmaps, but the age group labels are more clearly related to the clusters in the DVARS and MI heatmaps.

In the heatmap for clustering performed with DVARS feature vectors, the brightest uniform band mostly consists of fetal subjects. The darker, more variable band next to the group of fetal subjects is primarily composed of neonatal subjects. Groups of preadolescent subjects can also be easily identified.

Clusters of subjects from the same age group can also be clearly identified in the heatmap for the MI feature vectors. The bright, wide vertical band encompassing the left half of the heatmap consists of subjects mostly from the preadolescent cohort. The darkest vertical bands on the right of the heatmap are made of almost exclusively fetal subjects. The clusters between the preadolescent group and the neonatal group comprise the neonatal subjects and some preadolescent subjects.

The MI feature vectors were used to perform k-means clustering and spectral clustering in addition to the agglomerative clustering. Figure 45 shows a grid of scatter plots of the original data and the clustering results. The feature vectors in the first column are labeled with the true age group labels of the data. The vectors in the second, third, and fourth columns are labeled using the labels produced by k-means, spectral, and agglomerative clustering, respectively, where the number of clusters is three.

Initially, the spectral clustering results essentially identified a pair of outliers in the data set. The outliers were removed and the clustering was redone, but another pair of outliers was identified in this second iteration. We performed several iterations of identifying data belonging to outlier clusters, which we defined as clusters containing less than 10% of the total data in the data set, and repeating the clustering. This process reduced the number of samples in the data set from 564 to 545 but did produce meaningful clusters. The results of spectral clustering the reduced data set can be seen in Figure 46.

The composition of the clusters produced by each type of clustering can be seen in Tables 28, 29, and 30. The frequency counts in Tables 28 and 30 were comparable despite the different clustering methods. The contents of table 29 were calculated using the reduced set of data that produced interpretable (non-outlier based) clustering results. These results mirror those in Tables 28 and 30. The results of our cluster analysis suggest that there are

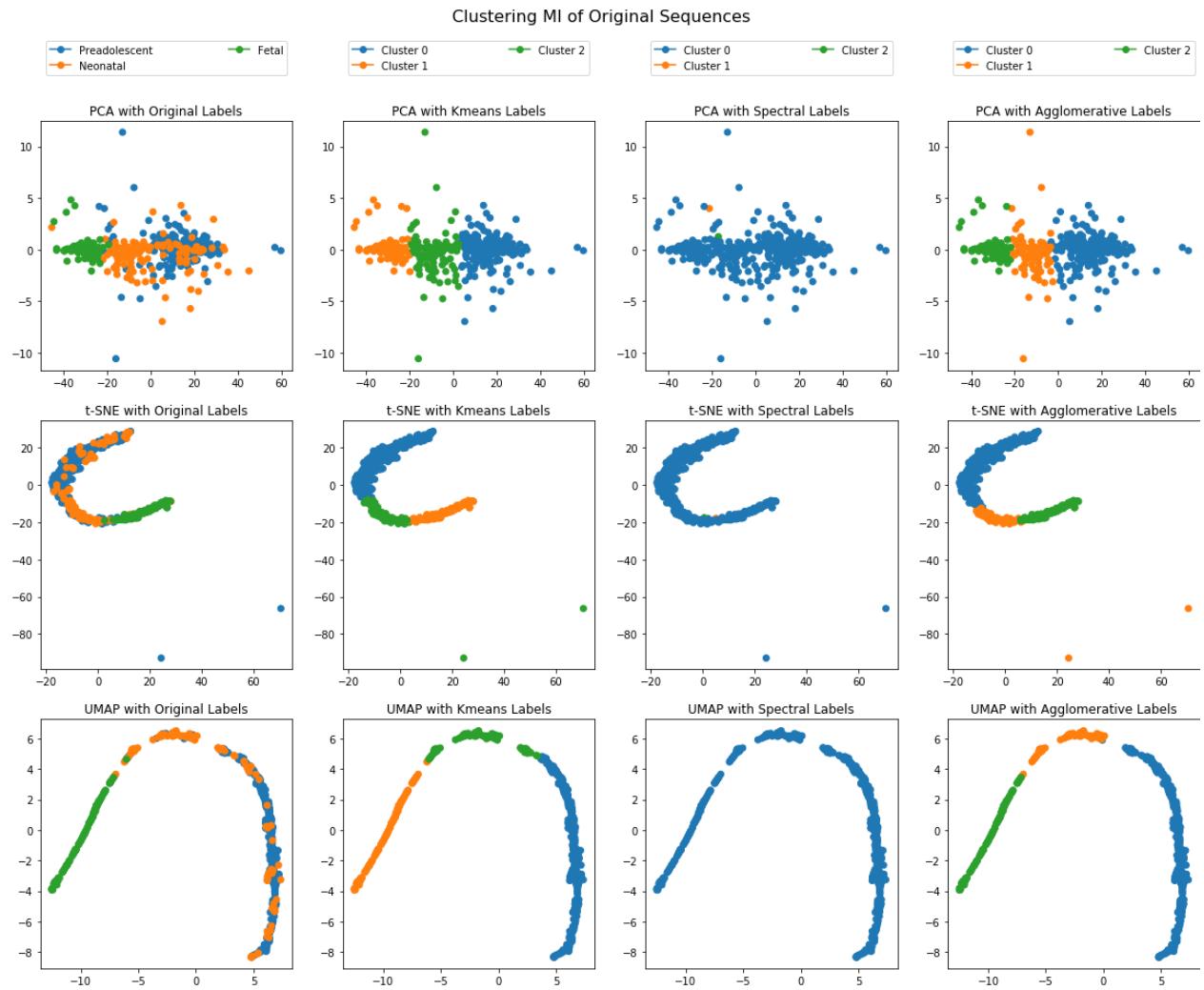


Figure 45: The results of clustering the clinical images using the MI matrices and the age group labels.

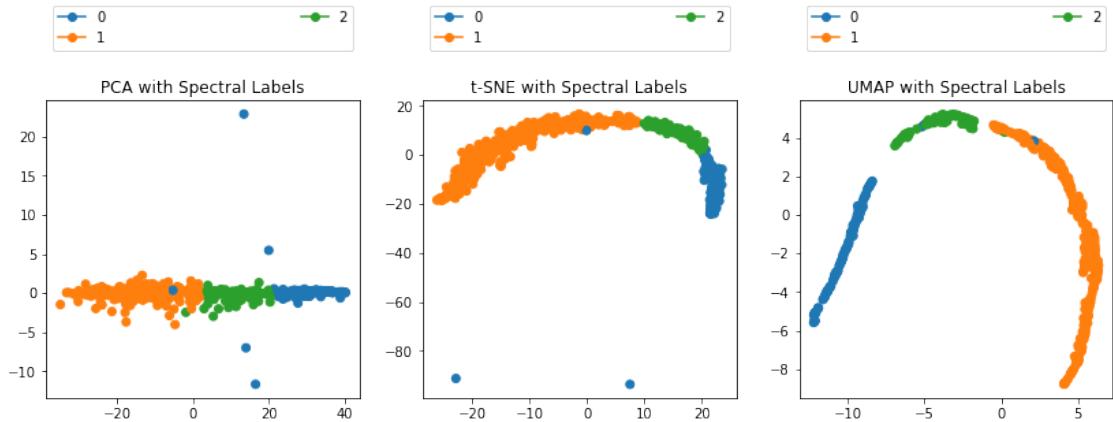


Figure 46: Iteratively removing clusters containing less than 10% of the data in the data set and repeating the spectral clustering process produced meaningful clusters similar to those produced by the k-means and agglomerative methods.

Table 28: The composition of age groups in each cluster produced using the MI feature vectors and k-means clustering.

Label	Cluster 1	Cluster 2	Cluster 3
Preadolescent	271	3	28
Neonatal	45	8	86
Fetal	0	122	1
Total	316	133	115

Table 29: The composition of age groups in each cluster produced using the MI feature vectors and spectral clustering.

<b>Label</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>
<b>Preadolescent</b>	4	278	9
<b>Neonatal</b>	6	55	76
<b>Fetal</b>	116	0	1
<b>Total</b>	126	333	86

Table 30: The composition of age groups in each cluster produced using the MI feature vectors and agglomerative clustering.

<b>Label</b>	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>
<b>Preadolescent</b>	282	19	1
<b>Neonatal</b>	58	76	5
<b>Fetal</b>	0	1	122
<b>Total</b>	340	96	128

detectable differences in the rs-fMRIs of different age groups of patients from different age groups.

### 6.3 Summary

In this chapter, we compared the rs-fMRIs before and after registration for the simulated and clinical images. These comparisons were performed using four metrics: FD, DVARS, Dice, and MI. The FD and DVARS metrics showed that both registration types had statistically significant impacts on these metrics at the population level. These impacts varied between populations. The Dice and MI metrics were used to compare all volumes in a single sequence to each other. These metrics demonstrated the impact of volume registration on the image sequences as a whole.

The motion patterns of the original image sequences characterized by the four metrics were used to characterize motion in the clinical brain images. No clear relationship between motion and disease status was found, but clustering showed a relationship between motion and age group.

## 7.0 DISCUSSION

### 7.1 Volume Registration

The comparison of the simulated and clinical brain rs-fMRIs using the local and global motion metrics showed different effects for each data set.

In the simulated images, there was a decrease in both the FD and DVARS values for both registration types. The change in the distribution of FD for both registration types was statistically significant for 40 sequences at  $p < 0.05$ . However, the difference in the DVARS distributions was only significant at  $p < 0.05$  for three sequences. The global metrics had statistically significant differences in the Dice and MI distributions for all three sequence types (original, traditionally registered, and DAG-registered), with the DAG-registered sequences having the overall highest values and smallest distributions.

For the preadolescent images, the distributions of the local motion metrics of the traditionally registered sequences were more similar to those of the original sequences than the distributions of the DAG-registered sequences. The DAG-based registration actually saw an increase in the local motion metrics. Specifically, there was a large spike in the mean and standard deviation of the DVARS values in the first 50 image volumes of the sequences. The magnitude and duration of this increase suggests that there may be a factor severely impacting a large number of preadolescent sequences in the first half of the scan. It is possible that many preadolescent subjects exhibited similar motion patterns in this part of the scan that were difficult for the DAG-based registration to correct.

The neonatal images experienced a different effect from both registrations. The FD distributions remained relatively consistent while the mean and standard deviations of the DVARS distributions increased for both registration types. These distribution changes were not expected: while registration is not designed to address the signal-related effects of motion, it was not expected to exacerbate them. As this type of change was specific to the neonatal images, it is possible that there are tissue properties specific to neonatal brains that caused these changes.

The registrations had the opposite effects on the fetal brain images than they did on the preadolescent images. The distributions for the local motion of the DAG-registered sequences were more similar to those of the original sequences than the traditionally registered distributions were. Additionally, there were several spikes in the FD of the traditionally registered sequences that were not present in the FD for the DAG-registered sequences. When comparing the numbers of usable image volumes from each sequence type, the number of image volumes meeting the FD threshold was consistent between all three sequence types. The original and DAG-based sequences had the same number of image volumes meeting the DAVRS thresholds, though the traditional registration had an increased number of volumes meeting that threshold. These results suggest that the DAG-based registration framework may be more effective at reducing the effects of motion (or at least not exacerbating them) than the traditional framework.

For all the clinical brain images, there was a decrease in the number of usable image volumes after registration. While this decrease may suggest that volume registration degrades an image, it does not mean that volume registration should be removed from the retrospective motion correction toolkit. Before registration, the relationships between the position of the patient at every time point are unknown. Registration is similar to looking in Schrodinger's box: until we do, we cannot quantify the effects of patient motion on the acquired sequence.

The changes in the similarity matrix value distributions were comparable overall for the clinical images.

In the set of placenta images, both registration types increased the mean and standard deviations of the distributions of both local motion metrics. However, even with these changes, both registration types improved the number of image volumes meeting the DVARS threshold. This result was completely unique to the placenta images and is likely related to the difference between the tissue properties of the brain and the placenta.

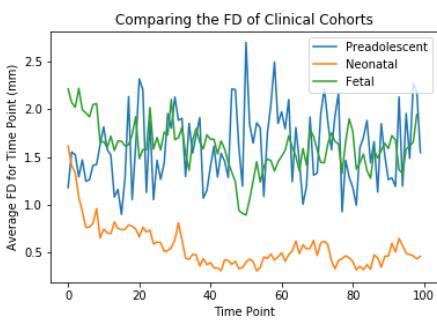
An additional analysis was performed for the simulated images. The goal of this analysis was to evaluate the effects of volume registration on the brain signal present in an rs-fMRI image. It was found that the DAG-registered sequences had correct voxel classifications when compared to the DMN ROI than the traditional registration did.

This finding has exciting potential. Subject motion during rs-MRI scans affects both the recorded position and orientation of the subject as well as the established magnetic spin gradients within the skull and the susceptibility recorded in each voxel. The primary focus of volume registration has been to reduce the positional effects of motion. Correction of the spin history effects and the susceptibility effects are considered to be a separate albeit related area of research. The results of the independent component analysis of the simulated images suggest that the DAG-based registration may contribute to the reduction of the signal-based effects of motion. It could potentially be coupled with prospective motion monitoring techniques,  $B_0$  field maps, and navigator sequences to address the other half of the motion problem.

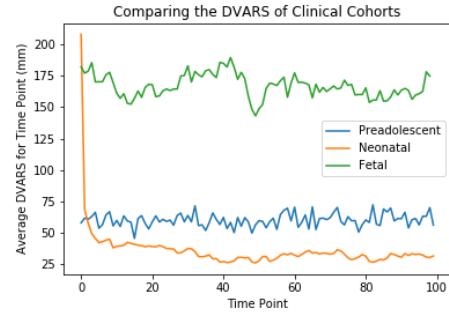
## 7.2 Characterizing Motion

In addition to the evaluation of motion within the images, clustering techniques were used to identify groups of subjects with similar motion patterns in their original sequences. Agglomerative clustering was used to confirm the existence of similar motion patterns between patients. The agglomerative clustering results consisting of a heatmap and a dendrogram were combined with two sets of labels: the disease status and age group of each subject. Examination of the dendograms and heatmaps coupled with the labels suggested that subjects in the same age group were more likely to exhibit similar motion patterns to each other than to subjects in other age groups. To reinforce this theory, k-means clustering and spectral clustering were also performed on the data. The labels produced by the clustering techniques were compared to the age group labels. The composition of the clusters reinforces the theory that patient motion patterns vary more between age groups than among age groups.

Additional analyses could be performed to evaluate the computer-detectable differences in patient groups in more detail. The models presented in the previous chapter were generated with each using a single metric type. Each metric only measures one property of the image volumes. It is possible that combinations of metrics measuring different properties could be used to better separate patient groups. For example, the combination of the FD



(a) Average FD of Original Sequences.



(b) Average DVARS of Original Sequences.

Figure 47: The average local motion metrics for all original sequences show that the neonatal subjects move the least, though differences between fetal and preadolescent motion is more clear in (b).

values and the DVARS values could more comprehensively categorize subjects based on the effects of motion, BOLD signal change, and background noise.

### 7.3 Expectations vs. Outcomes

In Chapter 5, we outlined the expected differences in motion patterns between the clinical groups, the expected effects of the registration technique on data from each clinical group, and the expected groups that could be identified via clustering.

We predicted that the neonatal subjects would demonstrate the least motion and the fetal subjects would demonstrate the most motion out of the clinical cohorts. Comparing the FD and DVARS of the original image sequences confirms that the neonatal subjects demonstrated the least motion, but the comparison between the preadolescent and fetal subject motion is slightly more complicated. Figure 47 shows a direct comparison between the average local motion metrics for the original sequences. It is only in comparing both the FD

and DVARS that we can see that the fetal subjects do move more than the preadolescent subjects.

-Due to the physical constraints on subjects from each population during the scan, we expected the registration frameworks to have different effects on the clinical cohorts. We expected that the DAG-based registration would have the best performance on the fetal subjects, but it was difficult to develop more granular hypotheses about each framework's effects on each clinical group. In general, we saw a comparable performance between the registration types for the neonatal cohort. The DAG-based framework was less effective at reducing motion in the preadolescent cohort than the traditional framework was. We also saw the DAG-based registration preserve the local motion in the fetal images while the traditional registration increased local motion. Out of these three sets of effects, the DAG-registration did indeed perform best on the fetal cohort.

Finally, we expected to see different patterns of motion dependent on clinical and age groups. In our clustering analysis of the population-wide motion patterns, we did not see distinct groups of healthy and CHD subjects based on their motion parameters. We did, however, see a relationship between age group and motion metrics. These results do not mean that there are no relationships between CHD disease status and motion: rather, they suggest that disease and motion relationships may occur on an age-group level rather than a lifespan level.

## 7.4 Relation to Existing Work

### 7.4.1 MRI Simulations

The idea of simulated MRIs originated in the 1980s. Bobman et al. suggested a process of MR image synthesis, then demonstrated its validity by creating synthetic spin-echo brain MRIs and comparing the simulated images to clinical images (Bobman et al., 1985). Since then, several MRI simulation software tools have been developed. Herein, we discuss two of these tools and compare them to our simulation tool.

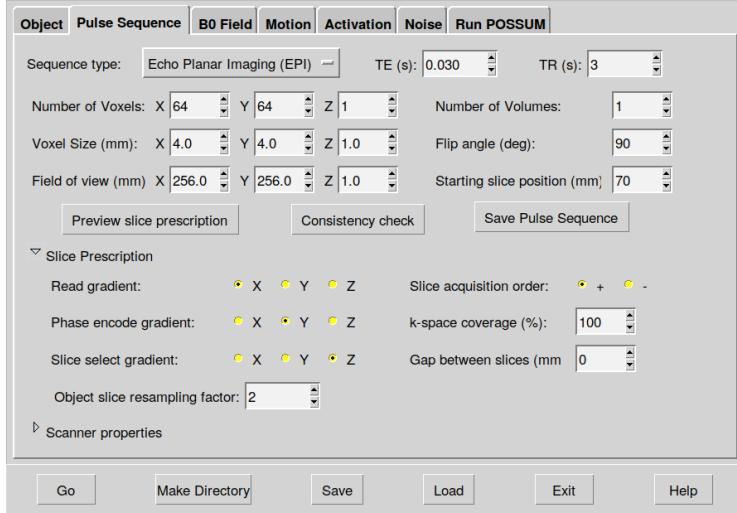


Figure 48: The “Pulse Sequence” specifications page in the POSSUM GUI

The FMRIB group developed a simulation tool called POSSUM (Physics-Oriented Simulated Scanner for Understanding MRI) (Drobnjak et al., 2006) (Drobnjak et al., 2010). POSSUM offers a realistic, physics-based simulation of structural, functional, and diffusion tensor images. It requires a gradient echo pulse sequence and a segmented object with known tissue properties as inputs for the simulation. It allows the user a high degree of control over the physical properties to be simulated. The user can specify the pulse sequence information, the method for generating brain signal, the addition of motion and noise, and the method for image reconstruction. Both a GUI and a command-line interface are available for POSSUM.

The biggest drawback to POSSUM is the degree of MR physics knowledge needed to use it effectively. For MR physicists, specifying the details of a pulse sequence may be trivial. For researchers from other fields, customizing pulse sequence parameters, an example of which can be seen in Figure 48, can be a challenging task. These customizations may even be unnecessary depending on the goals a researcher hopes to achieve using the simulated data.

If a researcher’s goal is to test the impact of a new MRI processing technique on known signals in an image, the BrainWeb MRI simulator might be a better option than POSSUM. BrainWeb was created and is actively developed by the McConnell Brain Imaging Centre at

McGill University to assist in validation computer-aided MRI analysis tools (Kwan et al., 1999) (Collins et al., 1998) (Cocosco et al., 1997) (Kwan et al., 1996). A set of simulated images have been generated using BrainWeb and can be found in the BrainWeb Simulated Brain Database (SBD). The SBD contains simulated images for healthy subjects and for subjects who have lesions due to multiple sclerosis.

Custom simulations can be generated on the BrainWeb server by submitting a request via a browser. The simulation request form has three areas that can be customized: the type of brain to simulate, the MR pulse sequence, and the imaging artifacts. The simulation is run on the server and the user is notified via email when the simulated images are ready to download.

BrainWeb’s simulator is slightly more approachable than POSSUM: a limited number of pulse sequence parameters are available to customize, and a description is listed next to each parameter in the pulse sequence and imaging artifacts sections. However, the user has slightly less control over the simulated image. The brain models used by BrainWeb are healthy brains or brains with MS lesions. It is not an option for the user to upload an image to use as the structural information in the sequence. For our purposes, the most significant limitation of BrainWeb is that it only simulates structural images.

Our simulation tool, SPECTR, is one of the few tools which simulates resting-state fMRIs. It offers the opportunity for researchers to explore the effects of their motion correction techniques on the BOLD signal, background noise, and patient motion using a lightweight simulation that can be run on a personal computer. It should be noted that SPECTR is not a substitute for the physics-based models in POSSUM and BrainWeb: it is intended to evaluate signal changes in rs-fMRI sequences as a result of post-acquisition image processing.

#### 7.4.2 Volume Registration

To the best of our knowledge, the only other study that has used a variant of the DAG-based method was performed by Liao et al. (Liao et al., 2016). Liao et als dataset consisted of 10 fetal rs-fMRIs. In each of these sequences, the fetal brain, fetal liver, and placenta were manually segmented in the first volume of the sequence as well as in five other randomly

chosen volumes. These overlap of these manual segmentations before and after registration, as measured using the Dice coefficient, was used to quantify the amount of motion in each sequence. Even though the Dice coefficients increase more in each sequence after Liao et al.s registration than after traditional registration, their measure of positional change fails to quantify any changes in position between any other pairs of volumes that do not have manual segmentations.

The fetal images used in Liao et al.’s study included images of singleton, twin, and triplet pregnancies. There is significant potential value in the study of fetal motion in multifetal pregnancies. Considering the motion patterns alone, it would be expected that the singleton pregnancies have different motion patterns than the multifetal pregnancies.

#### **7.4.3 Age Group Specific Motion**

It has been established that motion is often correlated with patient age in the adolescent population. Satterthwaite et al. specifically designed an imaging study of adolescents ages 8-23 such that patient age and motion were uncorrelated (Satterthwaite et al., 2012) In our study, patient age was described only as fetal, neonatal, or preadolescent despite a range of ages in each group. We establish that there are characteristics of patient motion specific to each group, but we did not consider more granular ranges of patient age. Additional analyses would need to be considered to determine whether the post-conceptual age for fetal and neonatal subjects impacts motion characteristics.

In the fetal cohort, there may be motion characteristics linked to post-conceptual age. As a fetus grows, the amount of room in the uterus in which it can move decreases. However, as the fetal brain develops, the fetus may begin to move in different ways to test its biological systems.

Studies involving neonatal cohorts often track two ages for the subjects: the post-conceptual age and the age since birth. The relationship between these two ages and a neonate’s brain development could impact the amount of motion exhibited during a scan.

## 7.5 Limitations

### 7.5.1 Spin History and Susceptibility Effects of Motion

As discussed in Chapter 2, patient motion has three effects on the acquired images. In this work, we have provided an analysis of how one step of retrospective motion correction can be used to reduce the positional effects of motion. While some of the clinical images also saw changes in the signal-related effects of motion (spin history effects and susceptibility effects), registration is not designed to alleviate these effects of motion. Comprehensive motion correction pipelines can be used to filter and denoise sequences registered using the DAG-based framework. Furthermore, additional scan types such as  $B_0$  field maps or navigator sequences contain valuable information that can be used to correct the spin history and susceptibility effects of motion directly.

### 7.5.2 Runtime

One topic we have not yet addressed in this work is the amount of time needed to perform the registrations and motion measurements. Herein, we compare the amount of time needed to run each registration framework. Assuming each operation for the same initial sequence is performed on the same computing hardware, the operations are limited by the number of image volumes in the rs-fMRI,  $N$ . We define  $r$  as the maximum amount of time to perform one pairwise volume registration.

In traditional registration, the contents of a 4D image sequence containing  $N$  image volumes are all aligned to the contents of a single image volume. At face value, the runtime  $R$  of the registration is  $r * N$ . However, there are no dependencies of the registration of one volume in the registration of another volume. Therefore, all registrations can theoretically be performed at the same time in parallel. Parallelizing the registrations for a single sequence means that the expected runtime of the sequence registration should decrease to

$$R_{trad} = \frac{hrN}{p} \quad (7.1)$$

where  $h$  is a factor specific to the hardware properties of the computer being used and  $p$

Table 31: The statistics for the runtimes of the registrations of 17 healthy neonatal subjects demonstrate that runtime of registration is not a trivial concern.

Registration Framework	Traditional	DAG-based
Average (HH:MM:SS)	00:41:55.25	01:38:20.02
Median (HH:MM:SS)	00:44:44.20	01:38:42.63
Maximum (HH:MM:SS)	01:21:46.95	03:02:36.10

is the degree to which the registration can be parallelized. In a scenario where the computer has enough power and resources to run all  $N$  pairwise volume registrations in parallel, the runtime decreases to  $hr$ .

-This decrease due to parallelization is only an option for the traditional registration framework. For the DAG-based registration framework, each pairwise volume registration is dependent on the affine transformation produced by the previous volume's registration. While this dependency accounts for the important temporal relationships between subsequent image volumes, it means that the image volume registrations must be performed in series, not in parallel. As a result, the runtime for the DAG-based registration is

$$R_{dag} = hrN. \quad (7.2)$$

An example of the power of parallelization can be seen in Table 31. This table shows the average, median, and maximum runtimes for both registration types for a subset of high-motion sequences belonging to healthy neonatal subjects on a desktop computer with 8 cores. Each of the runtime statistics for the DAG-based registration is almost twice as much as the corresponding runtime statistic for the traditional registration.

There are at least two occasions where the runtime of the registration frameworks may suggest the use of the traditional registration over the DAG-based registration. The first occasion is when the registered image is needed urgently. The second is when there are many registrations to perform. Access to computational clusters may help decrease the registration

time, but this decrease could be expected for both frameworks.

### 7.5.3 Manual Segmentations

All images used in these analyses consisted of brain tissue against an empty background. The images had undergone processing to remove tissues outside the skull, either using automated tools or manual segmentation. The use of manual segmentation with multiple annotators has the potential to confound the results of motion correction experiments. The segmentation process may not remove all non-brain tissue from the image. Those images would then contain brain tissue, non-brain tissue, and dark background.

The combination of these three tissue types has the potential to cause problems during the registration of the manually segmented images. The registration process optimizes the alignment of values in a pair of images. In some cases, the registration reaches a state where the lowest-cost alignment aligns anything not part of the background together, not specifically brain tissue to brain tissue. While these alignments do have the lowest cost, the results are not physically correct. This problem is difficult to detect in the metrics extracted from the image sequences: the metrics only measure the properties of the voxel values in the sequence, not of the physiological information it contains.

Specifically, this limitation pertains to our fetal scans. The masks generated during the segmentation process were created to be uniform across the whole sequence. However, fetal motion is highly variable. The subject may drastically change position in the middle of the scan, possibly several times. The manually created masks were developed using a software tool that allows 3D image masks to be applied to an entire 4D image sequence. The masks were required to be created to ensure the fetal brain or placenta would be inside the masked area at all times and therefore, may not have removed all tissues that were not of interest.

This limitation can be resolved by incorporating computer vision principles into the anatomical aspects of image segmentation. Filters used in computer vision applications to identify edges, smooth areas, and track objects have great potential when applied to the complex challenge of segmenting fetal brain tissue in the presence of motion.

## 7.6 Future Work

### 7.6.1 Theory-based Work

As seen in our results, there are different cases where one registration framework outperforms the other due to the characteristics of the patient's motion. These characteristics include the magnitude and frequency of the motion. We can use the SPECTr simulation to investigate the impact of these factors on the results of both registration frameworks. A set of simulated sequences can be generated, each with a different pair of magnitude and frequency properties. Each sequence will undergo both registration frameworks. The results of the registrations will be compared with respect to how much motion was removed and how much signal was preserved for each set of values for each factor. These results will be condensed into a curve dividing the space of motion magnitude and frequency combinations in areas where each registration method performs best.

Another area of registration research to consider is the development of a hybrid registration framework. In this hybrid framework, a sequence would undergo both traditional and DAG-based registration independently. Then, the corresponding traditional and DAG-based results would be compared for each image volume. The registered version of the image volume that either is more similar to the stationary image volume or has less motion compared to the previous image volume will be used in the sequence of registered image volumes. The hybrid framework could be tested on cases falling close to the cutoff curve produced in the results of the work suggested in the previous paragraph.

### 7.6.2 Application-based Work

This study focused on the motion of CHD and healthy pediatric subjects. As the prognosis for patients with CHD improves, their life expectancy also increases. The aging CHD population presents new questions about the connection between CHD and neurocognitive challenges associated with aging. As patients age, there is an expectation that their images will contain less motion for a time. If a patient begins to show signs of cognitive impairment due to aging, it can be expected that their images will begin to contain more motion as their

neurocognitive state deteriorates.

We would like to evaluate the impact of the two types of registration on a cohort of older adult subjects. The Alzheimer's Disease Neuroimaging Initiative has been collecting medical imaging, cognitive testing, genetic, and other types of data from adult subjects enrolled in their multisite study of Alzheimer's disease. Their subjects range from elderly controls with healthy brains, subjects with mild cognitive impairments, and Alzheimer's disease patients. Of these groups, we expect that healthy controls would exhibit different motion patterns than subjects with different degrees of cognitive impairments. The effects of different motion patterns of the adult cohorts may be more easily recovered using the DAG-based registration than the traditional registration.

In the current data set, there may be groups of subjects within each pediatric population who exhibit more similar motion patterns than other members of that population. Specifically, we are interested in the differences in motion between subjects who have neurodevelopmental concerns and subjects who do not. The distribution of neurodevelopmental problems is varied: it includes autism, ADHD, and neurodevelopmental delays. If motion patterns could be identified for subjects with similar cognitive abilities, the development of motion prevention protocols and post-acquisition motion correction techniques could be targeted more precisely to each group. The ability to either successfully complete scans or successfully recover scans for subjects in these groups would significantly aid in neurodevelopmental research.

## 8.0 Conclusions

The two primary goals of this work were to compare two volume registration techniques and to characterize motion in different clinical groups. Five sets of images were used for these experiments: a simulated data set, brain images from three clinical cohorts, and a set of placenta images.

An rs-fMRI simulation tool was developed to generate data with a known BOLD signal, background signal noise, and patient motion. The base for the structural information in the simulated sequences was the proton density image from the MNI average brain data, while the regions of interest used to generate the simulated BOLD signal were taken from the 90 functional ROI atlases created by Shirer et al. (Shirer et al., 2012).

The clinical images were obtained as part of prospective, long-term studies of CHD and brain development. The neonatal subjects were recruited from a single site while the fetal subjects were recruited at one of two sites, and the preadolescent subjects were recruited at one of 12 sites in the United States. The sequences differed in length depending on the age group and scanning site. In addition to the images, basic demographic information was gathered for each subject.

All rs-fMRIs from the simulated and clinical data sets underwent both the traditional registration and the DAG-based registration. The original sequences and both types of registered sequences were compared. Four metrics were used for these comparisons. The FD and DVARS metrics were calculated between every image volume  $i$  and  $i + 1$  in each sequence. The Dice and MI similarities were calculated for every possible pair of image volumes  $i$  and  $j$  in a single sequence.

The FD and DVARS metrics were compared to the usability thresholds set forth by Power et al. (Power et al., 2012). These thresholds state that an rs-fMRI has sufficiently low quantities of motion if more than 50% of the volumes in the sequence have  $FD < 0.2$  mm and  $DVARS < 2.5\%$  signal intensity units. The number of image volumes meeting these thresholds for each sequence type (original, traditionally registered, and DAG-registered) were calculated and compared using two-sample t-tests.

Comparing the FD and DVARS metrics for the simulated, preadolescent, neonatal, and fetal brain images to the usability threshold showed that the two registration types had comparable effects with respect to recovering image volumes to the usability standards. Between these four data sets, there were slightly different impacts of each registration type on each cohort, but the registration techniques did not recover statistically significantly different numbers of image volumes. The placenta images, however, demonstrated a statistically significant increase in the number of image volumes meeting the DVARS threshold and the pair of thresholds after registration.

The rs-fMRIs were also compared using matrices of similarity metrics where the value at every row  $i$  and column  $j$  was the similarity between volumes  $i$  and  $j$  in the same image sequence. The purpose of these matrices was to better characterize the positional and signal changes due to motion throughout the image sequence. Analysis of the sequence duration motion in the simulated data set showed mixed results. For the simulated sequences, the DAG-based registration was better at improving the range and quartile values of the distribution of the original similarity metrics than the traditional registration was.

One additional analysis on the simulated images was used to compare the traditional and DAG-based registration techniques. The registered images underwent independent component analysis. The components for each image were compared to the DMN ROI to identify the component which correlated best with the simulated BOLD signal. The best matching components were compared voxelwise to the DMN ROI to determine the number of true positive, false positive, true negative, and false negative identifications of component voxels as belonging to the DMN ROI. The DAG-based registration had a higher true positive rate and a lower false positive rate (0.623 and 0.00996) than the traditional registration did (0.587 and 0.104).

After analyzing the volume registration techniques, the patterns of motion in the clinical brain images were characterized into two categories of interest. No detectable groups were identified when the metrics for the original image sequences were clustered and labeled according to disease status. However, analysis of the clusters labeled according to age group suggests that there are detectable differences in the patterns of motion between subjects from different age groups.

## Appendix A

### Volume Registration Parameters

The parameters used for the registration of pairs of image volumes can be seen below.

```
1 ##  
2 # Register a pair of image volumes  
3 #  
4 # Effects: save a copy of the registered image and the registration param  
5 #  
6 # @param fixedImgFn The filename of the fixed image as a string  
7 # @param movinImgFn The filename of the moving image as a string  
8 # @param regImgOutFn The filename as a string specifying where to save the  
9 # @param transformPrefix  
10 # @param initialize Optional parameter to specify the location of the tra  
11 # @param regType Optional parameter to specify the type of registration t  
12 def registerVolumes(fixedImgFn, movinImgFn, regImgOutFn, transformPrefix,  
13     # Registration set up: for both Affine and SyN transforms  
14     reg = Registration()  
15     reg.inputs.fixed_image = fixedImgFn  
16     reg.inputs.moving_image = movinImgFn  
17     reg.inputs.output_transform_prefix = transformPrefix  
18     reg.inputs.interpolation = 'NearestNeighbor'  
19     reg.inputs.dimension = 3  
20     reg.inputs.write_composite_transform = False  
21     reg.inputs.collapse_output_transforms = False  
22     reg.inputs.initialize_transforms_per_stage = False
```

```

23     reg.inputs.num_threads = 100
24     reg.inputs.output_warped_image = regImgOutFn
25
26     # Registration set up: Specify certain parameters for the Affine registration
27     reg.inputs.transforms = ['Affine']
28     reg.inputs.transform_parameters = [(2.0,)]
29     reg.inputs.number_of_iterations = [[1500, 200]]
30     reg.inputs.metric = ['CC']
31     reg.inputs.metric_weight = [1]
32     reg.inputs.radius_or_number_of_bins = [5]
33     reg.inputs.convergence_threshold = [1.e-8]
34     reg.inputs.convergence_window_size = [20]
35     reg.inputs.smoothing_sigmas = [[1,0]]
36     reg.inputs.sigma_units = ['vox']
37     reg.inputs.shrink_factors = [[2,1]]
38     reg.inputs.use_estimate_learning_rate_once = [True]
39     reg.inputs.use_histogram_matching = [True] # This is the default, but can be turned off
40
41     # Registration set up: nonlinear transforms only
42     if regtype == 'nonlinear':
43         reg.inputs.transforms.append('SyN')
44         reg.inputs.transform_parameters.append((0.25, 3.0, 0.0))
45         reg.inputs.number_of_iterations.append([100, 50, 30])
46         reg.inputs.metric.append('CC')
47         reg.inputs.metric_weight.append(1)
48         reg.inputs.radius_or_number_of_bins.append(5)
49         reg.inputs.convergence_threshold.append(1.e-9)
50         reg.inputs.convergence_window_size.append(20)
51         reg.inputs.smoothing_sigmas.append([2,1,0])
52         reg.inputs.sigma_units.append('vox')

```

```
53     reg.inputs.shrink_factors.append([3,2,1])
54     reg.inputs.use_estimate_learning_rate_once.append(True)
55     reg.inputs.use_histogram_matching.append(True) # This is the defa
56
57 # If the registration is initialized, set a few more parameters
58 if initialize is not None:
59     reg.inputs.initial_moving_transform = initialize
60     reg.inputs.invert_initial_moving_transform = False
61
62 # Keep the user updated with the status of the registration
63 print("Starting", regtype, "registration for", regImgOutFn)
64
65 # Run the registration
66 reg.run()
67
68 # Keep the user updated with the status of the registration
69 print("Finished", regtype, "registration for", regImgOutFn)
```

## Appendix B

### Supplemental Statistical Tables

#### B.1 Simulated Data

Table 32: Results from the t-tests comparing the counts for the numbers of images meeting the FD, DVARS, and FD and DVARS thresholds for sequence types  $S_1$  and  $S_2$ .

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same FD counts)	1.05 E -16	4.49 E -11	0.127
P( $S_1$ and $S_2$ have same DVARS counts)	0.941	0.941	1.0
P( $S_1$ and $S_2$ have same FD and DVARS counts)	0.590	0.486	0.872

Table 33: Results of t-tests comparing the descriptive statistics of the correlation ratio matrices for the simulated data.

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same minimums)	0.3487	0.3407	0.9821
P( $S_1$ and $S_2$ have same 1st quartile)	9.750 E -113	1.246 E -112	0.8019
P( $S_1$ and $S_2$ have same medians)	5.288 E -88	5.409 E -88	0.9997
P( $S_1$ and $S_2$ have same 3rd quartiles)	6.534 E -81	6.730 E -81	0.9577
P( $S_1$ and $S_2$ have same maximums)	2.536 E -98	6.180 E -103	0.4068

Table 34: Results of t-tests comparing the descriptive statistics of the Dice matrices for the simulated data.

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same minimums)	9.976 E -105	2.520 E -110	0.3778
P( $S_1$ and $S_2$ have same 1st quartile)	5.225 E -93	5.582 E -93	0.931
P( $S_1$ and $S_2$ have same medians)	1.988 E -104	2.158 E -104	0.9578
P( $S_1$ and $S_2$ have same 3rd quartiles)	1.679 E -131	2.190 E -131	0.842
P( $S_1$ and $S_2$ have same maximums)	1.0	1.0	1.0

Table 35: Results of t-tests comparing the descriptive statistics of the MI matrices for the simulated data.

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same minimums)	5.016 E -114	6.328 E -126	0.5397
P( $S_1$ and $S_2$ have same 1st quartile)	4.68 E -105	7.90 E -105	0.995
P( $S_1$ and $S_2$ have same medians)	1.65 E -97	3.57 E -97	0.994
P( $S_1$ and $S_2$ have same 3rd quartiles)	1.065 E -84	2.374 E -84	0.974
P( $S_1$ and $S_2$ have same maximums)	0.00473	0.00794	0.8761

## B.2 Preadolescent Cohort

Table 36: Results from the t-tests comparing the counts for the numbers of images meeting the FD, DVARS, and FD and DVARS thresholds for sequence type  $S_1$  and sequence type  $S_2$ .

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same FD counts)	2.81 E -16	2.35 E -16	0.998
P( $S_1$ and $S_2$ have same DVARS counts)	9.43 E -12	5.30 E -12	0.950
P( $S_1$ and $S_2$ have same FD and DVARS counts)	1.12 E -11	5.60 E -12	0.938

Table 37: Results of t-tests comparing the descriptive statistics of the Dice matrices for the preadolescent data.

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same minimums)	0.770	0.695	0.916
P( $S_1$ and $S_2$ have same 1st quartile)	0.976	0.906	0.880
P( $S_1$ and $S_2$ have same medians)	0.883	0.562	0.643
P( $S_1$ and $S_2$ have same 3rd quartiles)	0.000343	0.000586	0.390
P( $S_1$ and $S_2$ have same maximums)	1.0	1.0	1.0

Table 38: Results of t-tests comparing the descriptive statistics of the MI matrices for the preadolescent data.

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same minimums)	0.624	0.718	0.896
P( $S_1$ and $S_2$ have same 1st quartile)	0.489	0.497	0.992
P( $S_1$ and $S_2$ have same medians)	0.364	0.324	0.928
P( $S_1$ and $S_2$ have same 3rd quartiles)	0.121	0.0882	0.851
P( $S_1$ and $S_2$ have same maximums)	0.946	0.932	0.987

### B.3 Neonatal Cohort

Table 39: Results from the t-tests comparing the counts for the numbers of images meeting the FD, DVARS, and FD and DVARS thresholds for sequence type  $S_1$  and sequence type  $S_2$ .

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same FD counts)	0.0110	0.00813	0.924
P( $S_1$ and $S_2$ have same DVARS counts)	0.00163	0.000942	0.880
P( $S_1$ and $S_2$ have same FD and DVARS counts)	0.00779	0.00475	0.879

Table 40: Results of t-tests comparing the descriptive statistics of the Dice matrices for the neonatal cohort.

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same minimums)	0.523	0.542	0.977
P( $S_1$ and $S_2$ have same 1st quartile)	0.468	0.515	0.933
P( $S_1$ and $S_2$ have same medians)	0.329	0.292	0.937
P( $S_1$ and $S_2$ have same 3rd quartiles)	0.149	0.115	0.890
P( $S_1$ and $S_2$ have same maximums)	1.0	1.0	1.0

Table 41: Results of t-tests comparing the descriptive statistics of the MI matrices for the neonatal data.

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same minimums)	0.853	0.874	0.978
P( $S_1$ and $S_2$ have same 1st quartile)	0.794	0.809	0.985
P( $S_1$ and $S_2$ have same medians)	0.762	0.758	0.996
P( $S_1$ and $S_2$ have same 3rd quartiles)	0.755	0.743	0.987
P( $S_1$ and $S_2$ have same maximums)	0.956	0.938	0.982

## B.4 Fetal Cohort

### B.4.1 Brain

Table 42: Results from the t-tests comparing the counts for the numbers of images meeting the FD, DVARS, and FD and DVARS thresholds for fetal brain sequence type  $S_1$  and sequence type  $S_2$ .

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same FD counts)	0.811	0.926	0.883
P( $S_1$ and $S_2$ have same DVARS counts)	0.159	1.0	0.159
P( $S_1$ and $S_2$ have same FD and DVARS counts)	0.159	1.0	0.159

Table 43: Results of t-tests comparing the descriptive statistics of the Dice matrices for the fetal brain data.

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same minimums)	0.259	0.932	0.289
P( $S_1$ and $S_2$ have same 1st quartile)	0.254	0.996	0.252
P( $S_1$ and $S_2$ have same medians)	0.658	0.970	0.686
P( $S_1$ and $S_2$ have same 3rd quartiles)	0.973	0.921	0.896
P( $S_1$ and $S_2$ have same maximums)	1.0	1.0	1.0

Table 44: Results of t-tests comparing the descriptive statistics of the MI matrices for the fetal brain data.

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same minimums)	0.673	0.845	0.816
P( $S_1$ and $S_2$ have same 1st quartile)	0.765	0.963	0.798
P( $S_1$ and $S_2$ have same medians)	0.764	0.963	0.798
P( $S_1$ and $S_2$ have same 3rd quartiles)	0.760	0.959	0.797
P( $S_1$ and $S_2$ have same maximums)	0.539	0.999	0.539

### B.4.2 Placenta

Table 45: Results from the t-tests comparing the counts for the numbers of images meeting the FD, DVARS, and FD and DVARS thresholds for fetal placenta sequence type  $S_1$  and sequence type  $S_2$ .

Sequence Type 1 ( $S_1$ )	Original	Original	Traditionally Registered
Sequence Type 2 ( $S_2$ )	Traditionally Registered	DAG Registered	DAG Registered
P( $S_1$ and $S_2$ have same FD counts)	0.519	0.350	0.775
P( $S_1$ and $S_2$ have same DVARS counts)	5.38 E -6	5.65 E -9	0.101
P( $S_1$ and $S_2$ have same FD and DVARS counts)	5.18 E -6	1.92 E -8	0.0571

## Bibliography

TensorFlow Projector.

Abdi, H. and Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, 2:433–459.

Afacan, O., Erem, B., Roby, D. P., Roth, N., Roth, A., Prabhu, S. P., and Warfield, S. K. (2016). Evaluation of motion and its effect on brain magnetic resonance image quality in children. *Pediatric Radiology*, 46(12):1728–1735.

Aksoy, M., Newbould, R., Straka, M., Holdsworth, S. J., Skare, S. T., Santos, J. M., and Bammer, R. (2008). A Real Time Optical Motion Correction System Using a Single Camera and 2D Marker. *Proceedings of the 16th Scientific Meeting of ISMRM*, 16:3120.

Alexander, M. (2012). Managing Patient Stress in Pediatric Radiology. *Radiologic Technology*, 83(6):549–560.

Altmann, A., Ng, B., Landau, S. M., Jagust, W. J., Greicius, M. D., and Alzheimer’s Disease Neuroimaging Initiative (2015). Regional brain hypometabolism is unrelated to regional amyloid plaque burden. *Brain : a journal of neurology*, 138(Pt 12):3734–46.

Assaf, M., Jagannathan, K., Calhoun, V. D., Miller, L., Stevens, M. C., Sahl, R., O’Boyle, J. G., Schultz, R. T., and Pearlson, G. D. (2010). Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients. *NeuroImage*, 53(1):247–256.

Avants, B. B., Tustison, N. J., Stauffer, M., Song, G., Wu, B., and Gee, J. C. (2014). The Insight ToolKit image registration framework. *Frontiers in Neuroinformatics*, 8(April).

Barnea-Goraly, N., Weinzimer, S. A., Ruedy, K. J., Mauras, N., Beck, R. W., Marzelli, M. J., Mazaika, P. K., Aye, T., White, N. H., Tsalikian, E., Fox, L., Kollman, C., Cheng, P., and Reiss, A. L. (2014). High success rates of sedation-free brain MRI scanning in young children using simple subject preparation protocols with and without a commercial mock scanner - the Diabetes Research in Children Network (DirecNet) experience. *Pediatric Radiology*, 44(2):181–186.

Barnes, K. A., Nelson, S. M., Cohen, A. L., Power, J. D., Coalson, R. S., Miezin, F. M., Vogel, A. C., Dubis, J. W., Church, J. A., Petersen, S. E., and Schlaggar, B. L. (2011). Parcellation in left lateral parietal cortex is similar in adults and children. *Cerebral Cortex*, 22(5):1148–1158.

Beckmann, C. F., Deluca, M., Devlin, J. T., and Smith, S. M. Investigations into Resting-state Connectivity using Independent Component Analysis (FMRIB Technical Report TR05CB1).

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.

Bharatha, A., Hirose, M., Hata, N., Warfield, S. K., Ferrant, M., Zou, K. H., Suarez-Santana, E., Ruiz-Alzola, J., D'Amico, A., Cormack, R. A., Kikinis, R., Jolesz, F. A., and Tempany, C. M. (2001). Evaluation of three-dimensional finite element-based deformable registration of pre- and intraoperative prostate imaging. *Medical Physics*, 28(12):2551–60.

Biswal, B. B., Kannurpatti, S. S., and Rypma, B. (2007). Hemodynamic scaling of fMRI-BOLD signal: validation of low-frequency spectral amplitude as a scalability factor. *Magnetic Resonance Imaging*, 25(10):1358–1369.

Bobman, S. A., Riederer, S. J., Lee, J. N., Suddarth, S. A., Wang, H. Z., Drayer, B. P., and MacFall, J. R. (1985). Cerebral magnetic resonance image synthesis. *American Journal of Neuroradiology*, 6(2):265–269.

Cocosco, C. A., Kollokian, V., Kwan, R. K.-S., Pike, G. B., and Evans, A. C. (1997). Brainweb: Online interface to a 3d mri simulated brain database. In *NeuroImage*. Citeseer.

Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G., Kabani, N. J., Holmes, C. J., and Evans, A. C. (1998). Design and construction of a realistic digital brain phantom. *IEEE transactions on medical imaging*, 17(3):463–468.

Coté, C. J. and Wilson, S. (2016). Guidelines for Monitoring and Management of Pediatric Patients Before, During, and After Sedation for Diagnostic and Therapeutic Procedures: Update 2016. *Pediatrics*.

Dewey, M., Schink, T., and Dewey, C. F. (2007). Claustrophobia during magnetic resonance imaging: Cohort study in over 55,000 patients. *Journal of Magnetic Resonance Imaging*, 26(5):1322–1327.

Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302.

Dosenbach, N. U., Koller, J. M., Earl, E. A., Miranda-Dominguez, O., Klein, R. L., Van, A. N., Snyder, A. Z., Nagel, B. J., Nigg, J. T., Nguyen, A. L., Wesevich, V., Greene, D. J., and Fair, D. A. (2017). Real-time motion analytics during brain MRI improve data quality and reduce costs. *NeuroImage*, 161(January):80–93.

Drobnjak, I., Gavaghan, D., Suli, E., Pitt-Francis, J., and Jenkinson, M. (2006). Development of a functional magnetic resonance imaging simulator for modelling realistic rigid-body motion artifacts. *Magnetic Resonance in Medicine*, 56(2):364–380.

Drobnjak, I., Pell, G. S., and Jenkinson, M. (2010). Simulating the effects of time-varying magnetic fields with a realistic simulated scanner. *Magnetic Resonance Imaging*, 28(7):1014–1021.

Fair, D. A., Dosenbach, N. U. F., Church, J. A., Cohen, A. L., Brahmbhatt, S., Miezin, F. M., Barch, D. M., Raichle, M. E., Petersen, S. E., and Schlaggar, B. L. (2007). Development of distinct control networks through segregation and integration. *Proceedings of the National Academy of Sciences of the United States of America*, 104(33):13507–12.

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

Fonov, V., Evans, A. C., Botteron, K., Almlí, C. R., McKinstry, R. C., and Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54(1):313–327.

Forman, C., Aksoy, M., Hornegger, J., and Bammer, R. (2011). Self-encoded marker for optical prospective head motion correction in MRI. *Medical Image Analysis*, 15(5):708–719.

Fox, C. K., Sidney, S., and Fullerton, H. J. (2015). Community-based case-control study of childhood stroke risk associated with congenital heart disease. *Stroke*, 46(2):336–340.

Fransson, P., Skiöld, B., Horsch, S., Nordell, A., Blennow, M., Lagercrantz, H., and Aden, U. (2007). Resting-state networks in the infant brain. *Proceedings of the National Academy of Sciences of the United States of America*, 104(39):15531–6.

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J. ., Heather, J. D., and Frackowiak, R. S. (1995). Spatial registration and normalization of images. *Human Brain Mapping*, 3(3):165–189.

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., and Turner, R. (1996). Movement-Related effects in fMRI time-series. *Magnetic Resonance in Medicine*, 35(3):346–355.

Gale, C., Jeffries, S., Logan, K. M., Chappell, K. E., Uthaya, S. N., and Modi, N. (2013). Avoiding sedation in research MRI and spectroscopy in infants: our approach, success rate and prevalence of incidental finding. *Arch Dis Child Fetal Neonatal Ed*, 98:F267–F268.

Garcia-Palacios, A., Hoffman, H. G., Richards, T. R., Seibel, E. J., and Sharar, S. R. (2007). Use of Virtual Reality Distraction to Reduce Claustrophobia Symptoms during a Mock Magnetic Resonance Imaging Brain Scan: A Case Report. *CyberPsychology and Behavior*, 10(3):485–488.

Giang, K. W., Mandalenakis, Z., Dellborg, M., Lappas, G., Eriksson, P., Hansson, P.-O., and Rosengren, A. (2018). Long-Term Risk of Hemorrhagic Stroke in Young Patients With Congenital Heart Disease. *Stroke*, pages 1155–1162.

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

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 of the United States of America*, 100(1):253–258.

Harned, R. K. and Strain, J. D. (2001). MRI-compatible audio / visual system : impact on pediatric sedation. *Pediatric Radiology*, pages 247–250.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm AS 136 A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society*, 28(1):100–108.

- Hofman, M. B., Paschal, C. B., Li, D., Haacke, E. M., Van Rossum, A. C., and Sprenger, M. (1995). Mri of coronary arteries: 2d breath-hold vs 3d respiratory-gated acquisition. *Journal of Computer Assisted Tomography*, 19(1):56–62.
- Hu, X., Le, T. H., Parrish, T., and Erhard, P. (1995). Retrospective estimation and correction of physiological fluctuation in functional MRI. *Magnetic Resonance in Medicine*, 34(2):201–212.
- 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*, 17:825–841.
- Jenkinson, M. and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5:143–156.
- Jo, H. J., Gotts, S. J., Reynolds, R. C., Bandettini, P. A., Martin, A., Cox, R. W., and Saad, Z. S. (2013). Effective preprocessing procedures virtually eliminate distance-dependent motion artifacts in resting state FMRI. *Journal of Applied Mathematics*, 2013.
- Jo, H. J., Saad, Z. S., Simmons, W. K., Milbury, L. A., and Cox, R. W. (2010). Mapping sources of correlation in resting state FMRI, with artifact detection and removal. *NeuroImage*, 52(2):571–582.
- Jones, T. B., Bandettini, P. A., Kenworthy, L., Case, L. K., Milleville, S. C., Martin, A., and Birn, R. M. (2010). Sources of group differences in functional connectivity: an investigation applied to autism spectrum disorder. *NeuroImage*, 49(1):401–14.
- Juan, A. and Vidal, E. (1998). Fast Median Search in Metric Spaces. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 905–912. Springer.
- Kaufman, L. and Rousseeuw, P. J. (1987). Clustering By Means Of Mediods. In *Statistical Data Analysis Based on the  $\{ \backslash displaystyle L_1 \}$  Norm and Related Methods*, pages 405–416. North Holland/Elsevier, Amsterdam.
- Kennedy, D. P. and Courchesne, E. (2008). The intrinsic functional organization of the brain is altered in autism. *NeuroImage*, 39(4):1877–85.

Khalil, A., Aggarwal, R., Thirupuram, S., and Arora, R. (1994). Incidence of Congenital Heart Disease Among Hospital Live Births in India. *Journal of Indian Pediatrics*, 31:519–527.

Khan, J. J., Donnelly, L. F., Koch, B. L., Curtwright, L. A., Dickerson, J. M., Hardin, J. L., Hutchinson, S., Wright, J., and Gessner, K. E. (2007). A program to decrease the need for pediatric sedation for CT and MRI. *Applied Radiology*, April:30–33.

KinetiCor Biometric Intelligence (2019). The kineticor platform.

Kwan, R. K.-S., Evans, A. C., and Pike, G. B. (1996). An extensible mri simulator for post-processing evaluation. In *International Conference on Visualization in Biomedical Computing*, pages 135–140. Springer.

Kwan, R.-S., Evans, A. C., and Pike, G. B. (1999). Mri simulation-based evaluation of image-processing and classification methods. *IEEE transactions on medical imaging*, 18(11):1085–1097.

Laumann, T. O., Gordon, E. M., Adeyemo, B., Snyder, A. Z., Joo, S. J., Chen, M.-Y., Gilmore, A. W., McDermott, K. B., Nelson, S. M., Dosenbach, N. U., Schlaggar, B. L., Mumford, J. A., Poldrack, R. A., and Petersen, S. E. (2015). Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron*, 87(3):657–670.

Li, W. (1990). Mutual Information Functions versus Correlation Functions. *Journal of Statistical Physics*, 60(5/6):823–837.

Liao, R., Krolik, J. L., and McKeown, M. J. (2005). An information-theoretic criterion for intrasubject alignment of fMRI time series: Motion corrected independent component analysis. *IEEE Transactions on Medical Imaging*, 24(1):29–44.

Liao, R., Turk, E. A., Zhang, M., Luo, J., Ellen Grant, P., Adalsteinsson, E., and Golland, P. (2016). Temporal registration in in-utero volumetric MRI time series. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9902 LNCS:54–62.

MacLaren, J., Herbst, M., Speck, O., and Zaitsev, M. (2013). Prospective motion correction in brain imaging: A review. *Magnetic Resonance in Medicine*, 69(3):621–636.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA.

Malviya, S., Voepel-Lewis, T., Prochaska, G., and Tait, A. R. (2000). Prolonged Recovery and Delayed Side Effects of Sedation for Diagnostic Imaging Studies in Children. *Pediatrics*, 105(3).

Mathur, A. M., Neil, J. J., McKinstry, R. C., and Inder, T. E. (2008). Transport, monitoring, and successful brain MR imaging in unsedated neonates. *Pediatric Radiology*, 38(3):260–264.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv preprint arXiv:1802.03426*.

Mebius, M. J., Kooi, E. M., Bilardo, C. M., and Bos, A. F. (2017). Brain Injury and Neurodevelopmental Outcome in Congenital Heart Disease: A Systematic Review. *Pediatrics*, 140(1):e20164055.

Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., Das, S. R., Ferranti, S. D., Després, J.-p., Fullerton, H. J., Howard, V. J., Huffman, M. D., Isasi, C. R., Jiménez, M. C., Judd, S. E., Kissela, B. M., Lichtman, J. H., Lisabeth, L. D., Liu, S., Mackey, R. H., Magid, D. J., McGuire, D. K., Mohler, E. R., Moy, C. S., Muntner, P., Mussolino, M. E., Nasir, K., Neumar, R. W., Nichol, G., Palaniappan, L., Pandey, D. K., Reeves, M. J., Rodriguez, C. J., Rosamond, W., Sorlie, P. D., Stein, J., Towfighi, A., Turan, T. N., Virani, S. S., Woo, D., Yeh, R. W., and Turner, M. B. (2016). *Heart Disease and Stroke Statistics 2016 Update A Report From the American Heart Association*. American Heart Association.

Murphy, K. J. and Brunberg, J. A. (1997). Adult claustrophobia, anxiety and sedation in MRI. *Magnetic Resonance Imaging*, 15(1):51–54.

Nadler, B. and Galun, M. (2007). Fundamental limitations of spectral clustering. *Advances in Neural Information Processing Systems*, pages 1017–1024.

Ng, A. Y., Jordan, M. I., and Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856.

Nikyar, B., Sedehi, M., Mirfazeli, A., Qorbani, M., and Golalipour, M.-J. (2011). Prevalence

and Pattern of Congenital Heart Disease among Neonates in Gorgan, Northern Iran (2007-2008). *Iranian Journal of Pediatrics*, 21(3):307–12.

Parkes, L., Fulcher, B. D., Yucel, M., and Fornito, A. (2017). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *bioRxiv*.

Patel, A. X., Kundu, P., Rubinov, M., Jones, P. S., Vértes, P. E., Ersche, K. D., Suckling, J., and Bullmore, E. T. (2014). A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. *NeuroImage*, 95:287–304.

Patriat, R., Reynolds, R. C., and Birn, R. M. (2017). An improved model of motion-related signal changes in fMRI. *NeuroImage*, 144:74–82.

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

Power, J. D., Schlaggar, B. L., and Petersen, S. E. (2015). Recent progress and outstanding issues in motion correction resting state fmri. *NeuroImage*, 105:536–551.

Pruim, R. H., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., and Beckmann, C. F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*, 112:267–277.

Qu, Y., Liu, X., Zhuang, J., Chen, G., Mai, J., Guo, X., Ou, Y., Chen, J., Gong, W., Gao, X., Wu, Y., and Nie, Z. (2016). Incidence of Congenital Heart Disease: The 9-Year Experience of the Guangdong Registry of Congenital Heart Disease, China. *PLOS ONE*, 11(7):e0159257.

Rahim, F., Ebadi, A., Saki, G., and Remazani, A. (2008). Prevalence of Congenital Heart Disease in Iran: A Clinical Study. *Journal of Medical Sciences(Faisalabad)*, 8(6):547–552.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(2):676–682.

Rutman, M. S. (2009). Sedation for emergent diagnostic imaging studies in pediatric patients. *Current Opinion in Pediatrics*, 21(3):306–312.

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., and Smith, S. M. (2014). Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90:449–468.

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. (2013). 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., Wolf, D. H., Loughead, J., Ruparel, K., Elliott, M. A., Hakonarson, H., Gur, R. C., and Gur, R. E. (2012). Impact of In-Scanner Head Motion on Multiple Measures of Functional Connectivity: Relevance for Studies of Neurodevelopment in Youth. *NeuroImage*, 60(1):623–632.

Saxena, A. (2005). Congenital Heart Disease in India: A Status Report. *Indian Journal of Pediatrics*, 72(7):595–598.

Shah, G., Singh, M., Pandey, T., Kalakheti, B., and Bhandari, G. (2008). Incidence of congenital heart disease in tertiary care hospital. *Kathmandu University Medical Journal*, 6(1):33–36.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423.

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., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-berg, H., Bannister, P. R., Luca, M. D., Drobniak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., Stefano, N. D., Brady, J. M., and Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23:208–219.

Smyser, C. D., Inder, T. E., Shimony, J. S., Hill, J. E., Degnan, A. J., Snyder, A. Z., and Neil,

- J. J. (2010). Longitudinal analysis of neural network development in preterm infants. *Cerebral Cortex*, 20(12):2852–2862.
- Smyser, C. D. and Neil, J. J. (2015). Use of resting state functional MRI to study brain development and injury in neonates. *Seminars in Perinatology*, 39(2):130–140.
- Smyser, C. D., Snyder, A. Z., and Neil, J. J. (2011). Functional connectivity MRI in infants: exploration of the functional organization of the developing brain. *NeuroImage*, 56(3):1437–52.
- Togo, H., Rokicki, J., Yoshinaga, K., Hisatsune, T., Matsuda, H., Haga, N., and Hanakawa, T. (2017). Effects of field-map distortion correction on resting state functional connectivity MRI. *Frontiers in Neuroscience*, 11(DEC):1–10.
- TracInnovations (2019). Tracinnovations mri markerless motion tracker and monitor system.
- United States Food and Drug Administration (2016). Drug Safety and Availability - FDA Drug Safety Communication: FDA review results in new warnings about using general anesthetics and sedation drugs in young children and pregnant women. Technical report, United States Food and Drug Administration.
- van der Linde, D., Konings, E. E., Slager, M. A., Witsenburg, M., Helbing, W. A., Takkenberg, J. J., and Roos-Hesselink, J. W. (2011). Birth Prevalence of Congenital Heart Disease Worldwide. *Journal of the American College of Cardiology*, 58(21):2241–2247.
- van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579—2605.
- van Dijk, K. R., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., and Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human connectomics: Theory, properties, and optimization. *Journal of Neurophysiology*, 103(1):297–321.
- 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.
- Varoquaux, G. (2019). K-means Clustering.
- Ward, J. H. . (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301):236–244.

Waskom, M. L. (2018). `seaborn.clustermap`.

Webb, G., Mulder, B. J., Aboulhosn, J., Daniels, C. J., Elizari, M. A., Hong, G., Horlick, E., Landzberg, M. J., Marelli, A. J., O'Donnell, C. P., Oechslin, E. N., Pearson, D. D., Pieper, E. P., Saxena, A., Scherzer, M., Stout, K. K., Warnes, C. A., and Khairy, P. (2015). The care of adults with congenital heart disease across the globe: Current assessment and future perspective: A position statement from the International Society for Adult Congenital Heart Disease (ISACHD). *International Journal of Cardiology*, 195:326–333.

Windram, J. D., Grosse-Wortmann, L., Shariat, M., Greer, M.-L., and Yoo, S.-J. (2011). The Feed and Sleep method: how to perform a cardiac MRI in the 1st year of life without the need for General Anesthesia. *Journal of Cardiovascular Magnetic Resonance*, 13(Suppl 1):P224.

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2004). Multilevel linear modelling for fMRI group analysis using Bayesian inference. *NeuroImage*, 21:1732–1747.

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. (2013a). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *NeuroImage*, 76:183–201.

Yan, C.-G., Craddock, R. C., He, Y., and Milham, M. P. (2013b). Addressing head motion dependencies for small-world topologies in functional connectomics. *Frontiers in Human Neuroscience*, 7(December):1–19.

Zaitsev, M., Burak, A., LeVan, P., and Knowles, B. R. (2017). Prospective Motion Correction in Functional MRI. *NeuroImage*, pages 33–42.

Zaitsev, M., Dold, C., Sakas, G., Hennig, J., and Speck, O. (2006). Magnetic resonance imaging of freely moving objects: prospective real-time motion correction using an external optical motion tracking system. *NeuroImage*, 31(3):1038–1050.

Zang, Y. F., Yong, H., Chao-Zhe, Z., Qing-Jiu, C., Man-Qiu, S., Meng, L., Li-Xia, T., Tian-Zi, J., and Yu-Feng, W. (2007). Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain and Development*, 29(2):83–91.

Zhao, Q. M., Liu, F., Wu, L., Ma, X. J., Niu, C., and Huang, G. Y. (2019). Prevalence of Congenital Heart Disease at Live Birth in China. *Journal of Pediatrics*, 204:53–58.

Zijdenbos, A. P., Dawant, B. M., Margolin, R. A., and Palmer, A. C. (1994). Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation. *IEEE Transactions on Medical Imaging*, 13(4):716–724.

Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M., Kaus, M. R., Haker, S. J., Wells, W. M., Jolesz, F. A., and Kikinis, R. (2004). Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index. *Academic Radiology*, 11(2):178–189.