SURFACE-BASED CORTICAL THICKNESS MEASUREMENT FOR IMPROVED CHARACTERIZATION OF HUNTINGTON'S DISEASE PROGRESSION

By

Kathleen E. Larson

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Engineeing

May 12, 2023

Nashville, Tennessee

Approved:

Ipek Oguz, Ph.D.

Catie Chang, Ph.D.

Benoit Dawant, Ph.D.

Adam Anderson, Ph.D.

Daniel Claassen, M.D., Ph.D.

ACKNOWLEDGMENTS

I suppose I should begin by mentioning my funding source, R01 NS094456 (NINDS), to get the boring part out of the way. The first actual humans I would like to acknowledge are those on my thesis committee: Drs. Ipek Oguz, Catie Chang, Adam Anderson, Benoit Dawant, and Daniel Claassen. I very much appreciated Dr. Claassen's eagerness to have me shadow him in the Huntington's disease clinic, although I was never able to attend due both to COVID and to my own, personal procrastination in beginning my clinical aim.

Mostly, I would like to thank Dr. Ipek Oguz (pronounced OH-z, as she neglected to inform me until I had already spent six months mispronouncing her name), who deserves far more space in my acknowledgements than I am allotting. I could not have asked for a better PhD advisor. It is mainly by sheer luck that I ended up in her lab, as I became advisor-less at the exact time she was student-less. I will never forget when I first interviewed with her over the phone and I asked her about her mentoring style, to which she responded that she tries to adapt to suit the needs of each individual student. Ipek has consistently shown me that it is possible to be both successful and kind in academia. Whenever I found a bug in my code, she would congratulate me on solving the problem rather than chiding me for making a mistake. I have learned so much from her not only regarding my research, but also about the type of PI that I hope to be one day, and for that I am so grateful.

Next, I would like to acknowledge the current and former graduate students in the Medical Image Computing Lab: Dewei, Hao, Jiacheng, Xing, Han (even though I'm not sure we have ever been in lab at the same time), Zach, Huahong, and David. There was a short time when I was the only student in FGH 304, followed by a longer time when it was just Huahong and I. It has been such a pleasure to see our lab grow to include so many intelligent and talented people. I appreciate that even though I am the only woman in the lab (and at the time I am writing this, the only student who doesn't speak a word of Mandarin), I have never felt left out by anybody. I have learned a lot from you all, and I'm looking forward to seeing where each of you end up after your time at Vanderbilt.

Before I forget, I would be remiss to not acknowledge the ACCRE support staff, particularly Fenglai Liu. As my defense date grew nearer, I realized that the sheer amount of data I had left to process would be incredibly challenging to complete by the time I needed to give my dissertation to my committee. Then, just as I had set up a large analysis to run over a weekend, the server decided it would not be running any more jobs (which I get, I've been there, too). Fenglai was kind enough to allow me to ask him to prioritize any jobs I submitted before my defense to make up for lost time. Then, when I asked for a 30% increase on computational capacities for our lab account to help me process just a bit more data, the ACCRE support team must have taken pity on me, because they instead increased everything by 1000%. So Chapter 6 truly could not have happened on time without them.

I would also like to thank the people in my non-work life who have supported me both before and during graduate school, without whom I could not have achieved this (well, I probably could have, but I would have been a lot more depressed about it). I don't want to mention any names lest I forget someone, but if you are a friend of mine who is actually reading my dissertation, then this probably includes you. I also want to acknowledge the teachers that I had growing up who were essential to my intellectual development and were always willing to provide emotional support. I will be forever grateful to the adults in my life who showed me kindness when I needed it most. I want to thank my parents for their support from the very beginning, and my sister. Lastly, and most of all, I want to thank my partner, Michael. I can't wait to enjoy a life with you where neither of us are students.

To my parents and my partner, who may actually read this,
to my sister and my friends, who probably won't,
and to my cat, who wouldn't even if she could.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AD Alzheimer's disease

ANTs Advanced Normalization Tools

CAG Cytosine-adenosine-guanine

CALC Calcarine fissure

CAP CAG-age product

CING Cingulate gyrus

CNS Central nervous system

CS Central sulcus

CSF Cerebrospinal fluid

CTh Cortical thickness

DCL Diagnostic confidence level

DiReCT Diffeomorphic Registration-based Cortical Thickness

DITE Decimation induced thickness error

DK Desikan-Killiany (atlas)

FDM Finite difference model

FLAIR Fluid attenuated inversion recovery

FOV Field of view

FS FreeSurfer

GM Gray matter

GRE Gradient echo

HC Healthy controls

HD Huntington's disease

HTT Huntintin

ICIR Inverse consistent image registration

ICV Intracranial volume

IR Inversion recovery

ITK Insight Toolkit

LSTG Left superior temporal gyrus

MPRAGE Magnetization-prepared rapid acquisition gradient echo

MRI Magnetic resonance imaging

MS Multiple sclerosis

MSDD Mean surface displacement difference

MSE Multi-spin echo

NITRC NeuroImaging Tools and Resources Collaboratory

OASIS Open Access Series of Imaging Studies

PDE Partial differential equation

PET Positron electron tomography

PNS Peripheral nervous system

PO Parietal-occipital gyrus

polyQ Polyglutamine

PVE Partial volume effect

RE Reproducibility error

RF Radio frequency

ROI Region of interest

SBL Surface-based Laplacian

SCPD Symmetric closest point distance

SCWT Stroop color word test

SDMT Single digit modality test

SDT Signed distance transform

SE Spin-echo

SF Superior frontal gyrus

SST Subject-specific template

ST Superior temporal gyrus

SYL Sylvian fissure

SyN Symmetric normalization

TE Echo time

TFC Total functional capacity

TI Inversion time

TMS Total motor score

TMT Trail making time test

TR Repetition time

UHDRS United Huntington's Disease Rating Scale

VDCRA Validation Data for Cortical Reconstruction Algorithms

VFT Verbal fluency test

VL Volumetric Laplacian

VTK Visualization Toolkit

WM White matter

CHAPTER 1

Introduction

1.1 Overview and Clinical Significance

1.1.1 Huntington's Disease

Huntington's disease (HD) is an autosomal dominant, neurodegenerative disorder that results in severe motor and cognitive decline. Although it affects approximately only three out of every 100,000 individuals, it has a 100% fatality rate and no existing cure [95]. Genetic testing can determine with certainty whether the disease will manifest throughout a patient's lifetime, but cannot indicate precisely when symptoms will manifest [Langbehn et al.]. The lack of more accurate predictive testing has led researchers to seek alternative measures, such as image-based markers for neurodegeneration, for the prediction of symptom manifestation. The use of longitudinal MRI datasets, where images are obtained from subjects and controls at multiple timepoints over a number of years, has proven to be particularly useful for studying disease progression in at-risk individuals. Most notably, research has correlated HD progression with volumetric loss in various subcortical nuclei such as the putamen and caudate [45, 127, 179, 197, 206]. These structures have been observed to exhibit atrophy approximately 10 years prior to symptom onset [19]. While subcortical changes are highly correlated with the presence of motor deficits, no significant patterns have been identified between these structures and non-motor symptoms.

In addition to subcortical changes, studies have detected widespread cortical gray matter (GM) thinning related to HD progression [15, 16, 146–148]. These thickness changes correspond more strongly to cognitive and behavioral symptoms rather than motor dysfunction, suggesting that they provide a more thorough explanation for symptom variation between patients than subcortical atrophy alone. Unfortunately, these findings vary between studies, due in part to both sample sizes and limitations of current technologies. Improved quantification of GM atrophy would provide a better understanding of the neurodegenerative mechanisms of HD progression. Moreover, more precise knowledge of which cortical areas specifically experience atrophy could provide helpful insights into developing potential treatments targeting those regions. This motivates the work presented in this dissertation to develop improved methods for quantifying cortical thickness (CTh) changes in large scale, longitudinal data to better assess the role of regional GM atrophy in HD progression.

1.1.2 Cortical Thickness

CTh has been found to be an important image-based marker in neurodevelopment [61, 171], healthy aging [61, 153, 170], and neurodegenerative disorders such as Alzheimer's disease (AD)[107, 185], non-AD demen-

tias [51, 92], Huntington's disease [15, 148], schizophrenia [221], and Parkinson's disease [87, 132]. Thus, the methods developed in this dissertation are not limited in application to HD; improved tactics to quantify GM atrophy has the potential to provide new insights into the progressive mechanisms of myriad pathologies. Aside from predictive measures, improved knowledge of specific cortical thinning patterns could assist in the development of clinical trials for drug-based treatments targeting specific neurodegenerative pathways [149].

Current approaches for CTh measurement can be classified as either volumetric or surface-based; each category presents their own set of advantages and limitations. The work in this dissertation focuses upon surface-based methods, which calculate CTh by first performing cortical segmentation and reconstructing surface representations of the gray matter and white matter tissues, and then measuring the distance between the two. The most popular method for surface-based thickness measurement is the FreeSurfer (FS) software suite [38, 141]. Although FS performs well in healthy data, it can produce inaccurate segmentations in cases where the brain has atrophied too far from standard atlas data. Further, the definition of thickness used by FS is prone to underestimation in high curvature areas [123], reducing its potential to identify subtle thickness changes such as those possibly experienced by pre-symptomatic HD patients. The development of a more accurate tool to quantify CTh would greatly enhance the ability to observe longitudinal GM changes related to disease progression.

1.2 Specific Aims

In this dissertation, we present improved cortical thickness measurement techniques to better assess the role of GM atrophy in HD progression. We achieve this through three specific aims:

Aim 1: Generate a synthetic ground truth dataset for accuracy validation of longitudinal surface-based cortical analyses. Accuracy validation of longitudinal CTh measurement and cortical segmentation techniques remains difficult due to the lack of ground truth data. In this aim, we develop a method to induce localized, synthetic atrophy using image morphology and deformable registration to create a set of images with known changes in thickness at each location on the cortex [101, 102]. While synthetic atrophy methods have been previously developed for similar applications [25, 40, 58, 89, 94, 209], this work is the first to create a tool specifically designed for accuracy validation of *surface-based* methods. Further, we introduce a novel protocol to assess accuracy that is unbiased towards any specific existing CTh technique. We accomplish this by restricting our method to measure only the *change* in thickness, which we define as the displacement of the deformed cortical boundaries, rather than CTh itself. We demonstrate the use of our pipeline to measure the accuracy of CTh measurements with respect to these ground truth changes using FreeSurfer (FS) as an example. We also detail how our methods can be applied to measure the accuracy of cortical surface placement with longitudinal segmentation algorithms.

Aim 2: Develop and validate a surface-based Laplacian algorithm for cortical thickness measurement. As stated previously, the widely used FS method [53] is prone to underestimating CTh in highly curved regions of the cortex. Further, this thickness mapping does not exhibit a symmetric property desired in a distance mapping. To overcome these pitfalls, previous research [85] established an alternative, volumetric algorithm that modelled the interior of the cortical GM ribbon using the Laplacian equation. Similar methods have since been implemented a number of times in primarily volumetric contexts [2, 65, 78]; these techniques all suffer from the shortcomings of volumetric methods, including errors induced by partial volume effect (PVE) and a higher susceptibility to noise. The only surface-based implementation proposed to date [201] is founded in finite element analysis and the construction of additional, intermediate surfaces, which renders it computationally expensive even compared to other surface-based methods. Further, none of these techniques were assessed for accuracy in brain MRI data; the authors either employed synthetic phantom data, or opted to forgo accuracy validation altogether. We thus propose a surface-based Laplacian (SBL) algorithm for CTh measurement that uses finite *volume* methods to model the interior cortical GM ribbon with an irregular grid in a domain bounded by the cortical surfaces. Within this domain, we solve for the gradient of the Laplacian solution with a highly efficient framework, and measure thickness as the distance traversed across this gradient from one boundary to the other. We validate our method first using digital, annular surface phantoms with varying degrees of curvature, and then with cortical surface reconstructions. We compare thickness results obtained using our method to those measured using the distance mapping employed by FS for surface-based CTh measurement.

Aim 3: Quantify longitudinal changes in gray matter thickness between cohorts of the PREDICT-HD dataset. In this aim, we measure changes in CTh and corresponding patterns between GM atrophy and clinical symptoms. PREDICT-HD is a longitudinal HD dataset (n=417) with pre-manifest and symptomatic HD patients, as well as age-matched controls. Each subject is associated with 3-7 timepoints; for each timepoint, there exists both structural imaging data and clinical metrics to assess symptom progression. In this dataset, we measure thickness using the SBL algorithm from Aim 2 and conduct a large scale statistical analysis to identify trends associated with differences in clinical metrics between pre-manifest and manifest patients and healthy controls. We hypothesize that we will find significant patterns between cognitive test scores and atrophy within the sensorimotor and frontal lobes.

By applying our surface-based CTh technique (Chapter 5, Aim 2), validated using our synthetic atrophy methods (Chapter 4, Aim 1), to the longitudinal PREDICT-HD dataset (Chapter 6, Aim 3), we will gain a better understanding of the role of GM atrophy in HD progression.

CHAPTER 2

Background

2.1 Relevant Neuroanatomy

Before discussing the neurodegenerative consequences of HD, we first introduce the structure and function of affected brain areas to provide context. This section will describe the brain from its microscopic to macroscopic components; we will begin with the neuron, and then move towards the cerebral cortex and relevant subcortical structures.

2.1.1 Anatomy of the Neuron

The nerve cell (Figure 2.1), also called a neuron, is the functional unit of the nervous system. Neurons comprise three main parts: a soma, or cell body, that contains the nucleus and other vital organelles; dendrites, which extend out from the soma to receive input from surrounding cells; and an axon, which propagates signals called "action potentials" towards the opposite end of the cell. Lipid bilayers called myelin sheathes wrap around the axon to increase the speed of signal transmission. When the signal reaches the axon terminals at the opposite end of the cell, it is relayed through a "synapse" to the dendrites of adjacent neurons.

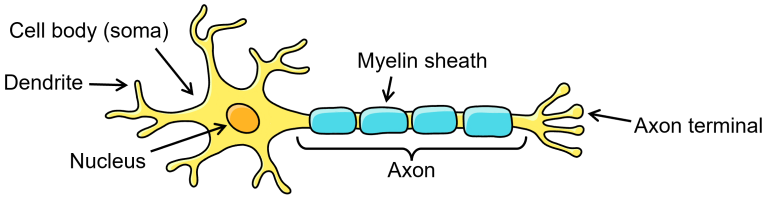

Figure 2.1: Illustration of the anatomy of a neuron.

There exist three functional classifications of neurons: afferent neurons, efferent neurons, or interneurons [204]. Afferent neurons transmit signals from sensory receptors for processing in either the brain or the spinal cord. They differ in appearance from the illustration shown in Figure 2.1 in that they have only one process connecting to an elongated axon. The soma of an afferent neuron and the majority of its axon both reside outside the central nervous system (CNS), which comprises the brain and the spinal cord, and instead lie within the peripheral nervous system (PNS). Efferent neurons, on the other hand, are very similar to the diagram in Figure 2.1. Their cell body and dendrites reside within the CNS, with a long axon extending down into the body (and the PNS) to enervate the target tissue. Afferent and efferent neurons are also referred to as sensory and motor neurons, respectively.

The third type of neuron, interneurons, account for over 99% of the total number of neurons in the human body, and exist entirely within the CNS [204]. Interneurons form integration networks with varying complexity between afferent and efferent signals. They exist in many different shapes and sizes, and can be morphologically classified by their soma (e.g. pyramidal, fusiform, round, etc.), dendritic branching (spiny or aspiny), or axon trajectory (unipolar, bipolar, or multi-polar) [183]. They can also be characterized by their electric or molecular profiles. For example, 95% of neural activity is modulated by either fast, glutaminergic excitatory neurons, which rapidly transmit signals down the axon and use glutamate as a neurotransmitter for excitatory synapses, or fast, GABAergic inhibitory neurons, which use *gamma*-aminobutyric acid for inhibitory synapses [33]. The remaining 5% of activity results from slower firing, monoaminergic (dopamine, serotonin, etc.) or non-monoaminergic (epinephrine, acetylcholine, etc.) neurons.

2.1.2 Structural anatomy of the Brain

The structural organization of the brain is best introduced in terms of its developmental components and their derivatives (see Figure 2.2). During the first stages of embryonic development, the CNS consists of a long, hollow cylinder called the "neural tube" composed of neuroepithelial cells. This eventually subdivides into the spinal cord and three distinct brain regions: the hindbrain, midbrain, and forebrain.

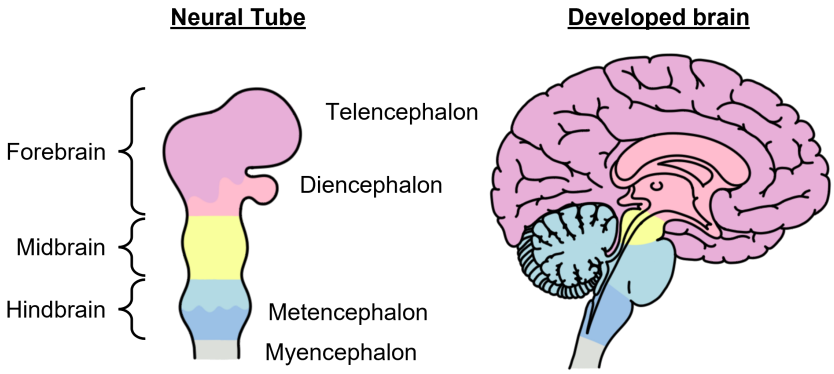

Figure 2.2: Illustration of the structural organization of the brain based on the neural tube. Forebrain structures are denoted in pink, midbrain in yellow, and hindbrain in blue. The spinal cord (not labelled) is shown in gray.

The hindbrain differentiates into the myelencephalon and the metencephalon; the former becomes the medulla oblongata, which contains the control centers for cardiac and respiratory activity, and the latter forms the pons and cerebellum [204]. The cerebellum is best known for its integral role in motor control, and the pons serves as the bridge between the cerebellum, medulla, and the rest of the brain [67]. The rest of the brainstem evolves from the midbrain and contains important nuclei involved in auditory and visual processing, as well as motor integration and homeostasis. Within the midbrain, the substantia nigra is the

only area of particular relevance to HD and is involved with motor planning, reward seeking, eye movement, and learning [169].

The forebrain develops into the remainder of the brain and is the only region from the neural tube that bifurcates into two hemispheres. Like the hindbrain, the forebrain also differentiates into two subsequent regions: the diencephalon and telencephalon. The diencephalon forms into the thalamus, hypothalamus, epithalamus, and subthalamus. All of these structures, particularly the hypothalamus, play a role in the limbic system, which is the primary network modulating emotional behavior [67]. The thalamus also contains the nuclei of most projection neurons; these are long interneurons connecting distant parts of the brain either to each other or to the brainstem, and are involved in almost every single sensory-motor pathway [204]. Lastly, the telencephalon develops into the largest and least understood portion of the brain, the cerebrum, which comprises the paleocortex, neocortex, and various subcortical structures. The paleocortex, also called the rhinencephalon, is responsible for olfaction, and is usually considered separate from the rest of the cerebrum due to differences in its molecular structure [67]. The neocortex, or the cerebral cortex, comprises the vast majority of the human brain and plays an integral role in sensory processing, motor outputs, and higher cognition. The following sections will discuss in greater detail the cerebral cortex, subcortical structures such as the putamen and caudate, and the limbic system.

The Cerebral Cortex

The cerebral cortex is the largest part of the human brain and is responsible for sensory-motor processing and cognition. It has a highly folded structure that quadruples its total surface area from that of a smooth surface, and contains approximately 100 billion neurons [67, 204]. It is composed of two primary tissue types: gray matter (GM), which consists of neuronal cell bodies, and white matter (WM), which comprises the myelinated axons that form connections between cells. Neurons within the cortex are generally described as either pyramidal or non-pyramidal based on their cell shape. Pyramidal cells are glutaminergic, excitatory neurons and have a conical soma, spiny dendrites, and a long axon [22]. These are the most common cell type in the cortex, and are responsible for sending information to the spinal cord and distant parts of the brain [22]. Non-pyramidal cells, on the other hand, are usually granular or fusiform in shape. These are usually GABAergic inhibitory neurons with shorter axons that project more locally [67]. Within cortical GM, nonpyramidal cells reside within six horizontal layers oriented such that Layer I is closest to the exterior of the brain; these are described in more detail in Table 2.1. Sensory input usually enters the cortex from the thalamus into layer IV, and departs as a motor output through Layers V and VI [3]. Signal integration primarily occurs in the upper three layers. These cortical GM cells are also arranged in vertical columns approximately 0.3-0.5 mm in diameter that span all six layers [67]. Within the somatosensory cortex, each column responds to a specific

sensory location and input type; within the motor cortex, different areas correspond to muscle movements in different parts of the body.

Layer	Primary cell type(s)	Main function
Plexiform (I)	– Horizontally projecting – Thalamocortical neuron dendrites	– Inter-columnar signalling
External granular (II)	– Granular – Small pyramidal (IT)	– Inter-columnar signalling – Inter-hemisphere signalling via corpus callosum
External pyramidal (III)	– Small pyramidal	– Inter-hemisphere signalling via corpus callosum
Internal granular (IV)	– Granular	– Sensory input from diencephalon (thalamus)
Internal pyramidal (V)	– Large pyramidal	– Project to basal ganglia, brain stem, and spinal cord
Multiform (VI)	– Fusiform polymorphic	– Project back to thalamus to modulate incoming signal response

Table 2.1: Cell types and primary functions of the six cortical GM layers. IT=intratelencephalic.

The entire cortex is organized into four basic regions: the frontal, parietal, temporal, and occipital lobes. The frontal cortex, located in the anterior most portion of the brain, contains the motor and pre-frontal cortices, where motor planning and elaboration of thought occurs. The parietal lobe, residing just posterior to the frontal, contains the somatosensory cortices. Neurons here receive signals from thalamocortical projections, those that pass through the thalamus into the cortex, for sensory processing. At the most posterior end of the brain lies the occipital lobe, also known as the visual cortex. Lastly, the temporal lobe sits inferior to the parietal lobe and lateral to the diencephalon, and contains centers for auditory and gustatory processing, as well as parts of the limbic system. These cortical regions are all depicted in the illustrations within Figure 2.3.

The processing of any sensory input generally follows a similar pipeline regardless of input type, barring a few exceptions. First, the signal travels from the sensory organ or receptor through various interneurons to the thalamus. The thalamus is often described as the "relay station" of the central nervous system, as it filters through incoming signals and sends them along to their relevant primary cortex (i.e. visual, auditory, somatosensory, or gustatory). Note that olfactory signals are the only input type that do not first pass through the thalamus, and instead pass through the olfactory bulb [187]. Within the primary cortices occur variable mechanisms of initial processing depending on the input. For example, the primary somatosensory cortex is responsible discerning the specific location on the body from where the touch signal was received, whereas the auditory cortex can identify sound frequencies as well as the spatial origin of the signal in 3D space [67].

Following the primary cortices, the signal processing pipeline complicates. In many cases, the end result

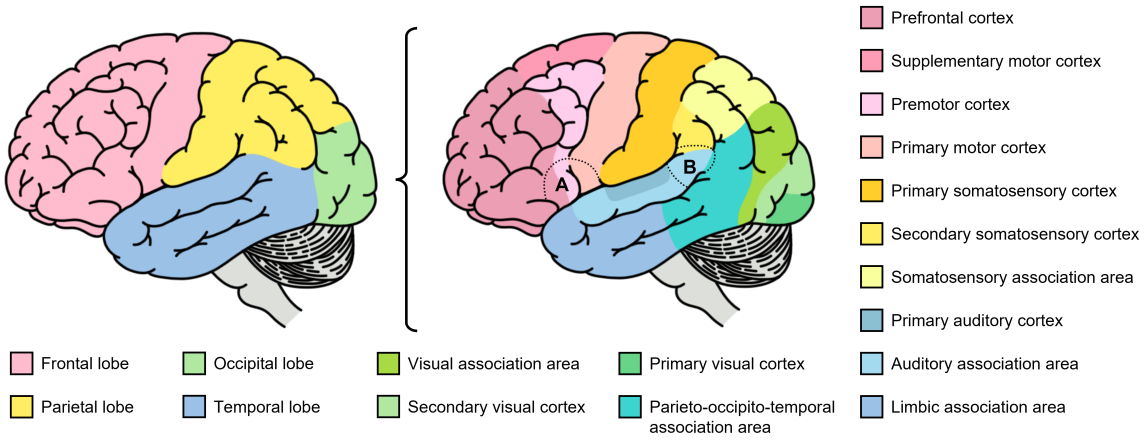

Figure 2.3: Illustration of the general organization of the human cerebral cortex. Color groupings in panel B roughly correspond to those in panel A. Note that these boundaries and classifications are approximate. For example, the pre-frontal cortex also contributes to the limbic system, which is not depicted in the illustration. The dotted regions indicate Broca's area (A) for speech production and Wernicke's area (B) for speech production.

is a motor output, where a signal passes into the primary motor cortex, and then through the spinal cord to reach a target effector for muscular innervation. These primary areas connect to myriad other cerebral regions, including the pre and supplementary motor cortices, association cortices, and the pre-frontal cortex. Additionally, signals can also be transmitted into the basal ganglia, which we discuss in the next section, or into the cerebellum and brainstem. These more complicated circuits are required for higher-level integration of multiple sensory inputs that contribute to emotional responses, working memory, and planning.

Subcortical Nuclei

Below the cerebral cortex and superior to the midbrain lie several groups of subcortical nuclei that contribute to sensory-motor signal processing and modulation. These structures, illustrated in Figure 2.4 along with parts of the limbic system (see section 2.1.2), consist of distinct clusters of somata whose axons project to different places in the cerebral cortex, cerebellum, or brainstem. Because these structures are highly connected and often contribute to multiple modulatory feedback loops, they are more easily described in terms of the neural circuits to which they contribute, rather than their individual functions.

The most notable group of structures, called the basal ganglia, consists of the striatum (caudate and putamen), globus pallidus, subthalamic nucleus, nucleus accumbens, and substantia nigra. These structures primarily contribute to motor control and planning, and play a major role in signal transmission between the cerebellum and motor cortices. For example, the putamen circuit regulates subconscious execution of learned movement. The putamen first receives input from the primary motor, premotor, supplementary motor, and prefrontal cortices. This is transmitted into the globus pallidus, onto the subthalamic nucleus and substantia

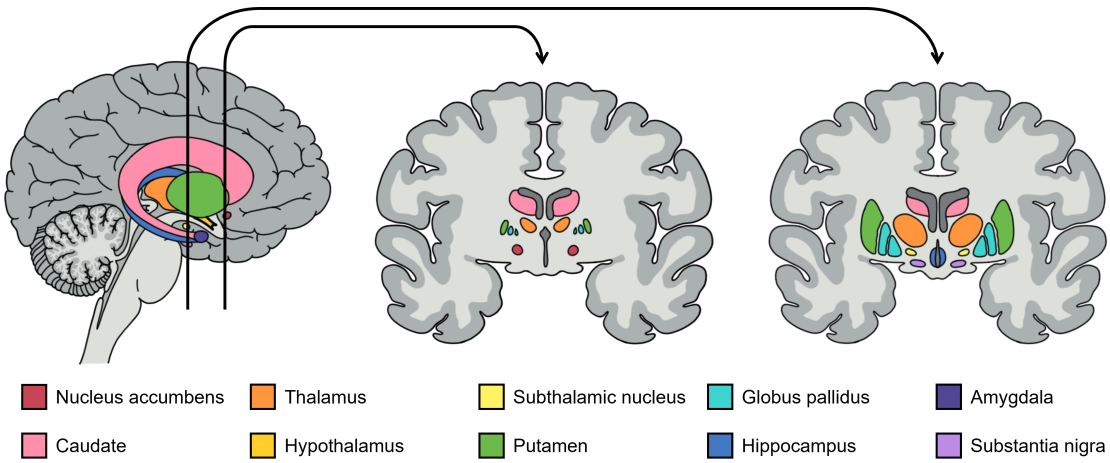

Figure 2.4: Illustration of subcortical structures, including the basal ganglia, thalamus, and limbic areas, superimposed onto the medial side of the opposite cerebral hemisphere. The left panel displays the sagittal view, and the central and right panels display an anterior and posterior coronal cross-section, respectively.

nigra, and then back into the thalamus and motor cortex. The caudate circuit, on the other hand, receives input from the pre- and supplementary motor cortices, as well as the somatosensory association cortex; information is processed in an inhibitory feedback loop with the striatum and globus pallidus, and then sent back through the thalamus and into the prefrontal and motor cortices. These two circuits both contribute to the larger, cortico-striatal-thalamus loop, which also includes modulation from the cerebellum, and ultimately transmits information to motor effector neurons for muscular innervation [67].

The Limbic System

Several subcortical nuclei, namely the amygdala and nucleus accumbens, contribute to a larger network within the telencephalon and diencephalon called the limbic system. The limbic system is a functional network responsible for emotional control, behavioral regulation, long-term memory, and olfaction [67]. Along with the amygdala and nucleus accumbens, it includes the thalamus, hypothalamus, and hippocampus (see Figure 2.4), as well as the limbic association cortex than spans regions of the temporal lobe, prefrontal cortex, and cingulate gyrus. Table 2.2 briefly decribes the function of each relevant limbic structure.

2.2 Huntington's Disease

Huntington's disease (HD) is an autosomal dominant, neurodegenerative disorder characterized by myriad symptoms including motor disorders, behavioral difficulties, and cognitive decline. The disease is both rare and severe—it affects only approximately 3 in 100,000 individuals, but has a 100% fatality rate and no existing cure [95]. Following its classification in 1872, research has been devoted to studying its genetic underlying, symptomatic expression, and mechanisms of neurodegeneration in the hopes of eventually devel-

Brain area	Structure	Function
Subcortical	Amygdala	– Behavioral awareness – Olfactory processing – Reward/punishment centers
	Nucleus accumbens	– Reward centers – Sleep regulation
	Hippocampus	– Sensory integration – Behavioral control
Cortical	Association cortex (temporal)	– Emotional regulation – Impulse control – Olfactory integration
	Association cortex (prefrontal)	– Emotional regulation – Impulse control
	Cingulate gyrus	– Emotional regulation – Impulse control
Diencephalic	Hypothalamus	– Hormonal control – Homeostasis (temperature, fluid concentration, etc.) – Circadian rhythm
	Thalamus	– Sensory processing and integration

Table 2.2: Function of cerebral structures within the limbic system

oping effective treatments [76]. In this section, we will discuss the genetic underlyings, clinical symptoms, and neurodegenerative patterns associated with HD.

2.2.1 Genetic Underlying and Patient Classification

HD is a member of a family of neurodegenerative disorders characterized by a specific, unstable mutation of a highly localized, elongated polyglutamine (polyQ) sequence [49]. These polyQ sequences are expanded repeat series of cytosine-adenosine-guanine (CAG) trinucleotides, also known as glutamine, on a gene specific to each disorder [49]. In HD, the mutation exists on the Huntingtin (HTT) gene and results in toxic aggregations of the HTT protein among widespread neuronal populations [112]. A mutation of 35 or more CAG repeats generally indicates disease presence, although this number slightly varies across studies [131, 138, 172, 199]. An individual with between 28-37 repeats may not express HD themselves, but still exhibit a high risk of passing the mutation onto their offspring [119]. Further, the instability of the mutation implies that biological children of an at-risk individual will most likely develop a longer CAG repeat expansion than the parent [138]. Genetic testing has thus proven useful for identifying disease presence for assistance with family planning, as the dominant inheritance patterns of HD already grants a 50% chance of passing the mutation onto the next generation. [200].

On a general level, HD patients are classified as “pre-manifest” or “symptomatic” based on whether or not they have begun to experience symptoms [125]. In the former category, an individual can either be pre-

symptomatic, where they have been identified as genetically at-risk but have not yet experienced any HD related symptoms, or prodromal, where they have begun to experience subtle changes in cognitive, behavioral, and motor skills [149]. An individual is considered symptomatic, or "manifest", when these neurological changes begin to affect their quality of life. Unfortunately, existing treatments are limited to symptom mitigation, and the disease will ultimately progress until the patient's death after an average of approximately 20 years prior symptom manifestation [138]. This lack of more effective treatments is thought to contribute to the low rate of genetic testing. In fact, one study found that less than 20% of the at-risk population undergo any such testing [20]. A large percentage of individuals with a positive family history therefore do not receive a diagnosis until after motor symptoms manifested, rendering any available treatments less effective [127].

Beyond the broad classifications of pre-manifest and manifest, many studies sort patients into different, systematically defined cohorts. Because research has shown that a higher number of CAG repeats correlates to both earlier disease onset and more severe symptoms [7, 131, 138, 172], almost all forms of classifications incorporate CAG-repeat length in some way. For example, Langbehn et al., 2010 [Langbehn et al.], proposed a logistic, survival model incorporating CAG length and age of onset, which can be applied to categorize subjects by an estimated time to HD onset (TTD) [18] (e.g. Far: TTD $<$ 15 yrs; Middle: 9 yrs $<$ TTD $\leq$ 15 yrs; Near: TTD $\leq$ 9 yrs). Many studies utilize a variable known as the CAG-age product (CAP) score, which is based on the universal finding that symptom severity also increases with disease progression. Zhang et al., 2011 [219], defined a patient's CAP score as

$$\text{CAP} = \text{age} \times (\text{CAG repeat length} - L) \tag{2.1}$$

where L is a constant based on the average CAG repeat length of HD positive individuals [131, 203]. Essentially, the CAP score is a variable describing the "cumulative" exposure to harmful effects caused by the genetic mutation [21]. This is most often used to model the natural history of HD, a term used to describe the progression of a disease from onset to resolution, across an entire cohort of patients. In the case of HD, resolution always refers to patient death (rather than recovery or remission) due to the lack of a cure. Zhang et al. also proposed thresholds to sort patients based on this score similar to TTD, where CAP_{low} corresponds to CAP $<$ 290, CAP_{med} to 290 $\leq$ CAP $<$ 368, and CAP_{high} to CAP $\geq$ 368. Recently, Tabrizi et al., 2022 [182], proposed the Integrated Staging System (ISS), which classifies HD patients into for distinct categories based on disease progression. Stages 0 corresponds to those with a CAP score of CAP $\geq$ 40 but no other symptoms or neurological changes, stage 1 to those with detected neurological symptoms, stage 2 to those with cognitive symptoms, and stage 3 to those with a marked loss in functional ability. Although this system is still undergoing validation and thus not as prevalent as CAP-score based studies, it presents a useful

alternative with the potential to aid in the development of clinical trials.

2.2.2 Clinical Symptoms, Diagnosis, and Evaluation

Of the wide range of symptoms currently associated with HD, the most thoroughly documented since the disease was first classified are deficits in various motor skills [137]. Most commonly, HD has been diagnosed by the presence of brief, involuntary movements known as chorea, as this is generally the first and most severe symptom to appear [77, 149]. However, clinicians have also identified several voluntary movement disorders more prevalent in earlier-onset and late-stage cases [5, 24, 44, 56]. These include bradykinesia, a decrease in movement speed [24]; akinesia, more commonly referred to as rigidity [5]; ataxia, a lack of coordination [56]; and dysarthria, which specifically affects the muscles involved with speech [44]. Patients have also been known to exhibit dystonia, another involuntary motor deficit characterized by sustained contractions and repetitive movement [172].

Currently, the Unified Huntington's Disease Rating Scale (UHDRS) exists as the primary set of clinical diagnostic criteria [77]. The UHDRS contains a numerical rating system for various symptoms, and is useful for assessing disease severity in the years following a diagnosis. It evaluates the maximal extent of full-body chorea, dystonia, bradykinesia, and dysarthria, as well as motor deficits specific to eye movement, tongue protrusion, arm rigidity, postural muscle tone, and gait. Separate ratings for each of these metrics are aggregated together to yield a total motor score (TMS). In addition, the motor assessment contains a rating for the diagnostic confidence level (DCL), which describes the certainty that a patient's motor symptoms can be attributed to HD rather than a different movement disorder such as Parkinson's disease or essential tremor [108]. The DCL ranges from 0-4, where higher values correspond to higher certainty; a DCL of 1 suggests the onset of motor impairments, while a DCL of 4 indicates a 99% likelihood that these impairments are a result of HD. A patient is generally considered to have moved from the prodromal to manifest disease stage when they first receive a DCL of 4 [109].

Historically, the diagnosis of HD and the specific definition of its manifestation time has relied solely on the presence of movement disorders [77]. However, cognitive impairment has also been identified as a major symptom [64, 75, 84, 126, 130, 167, 168, 174]. Throughout disease progression, patients will begin to experience various deficits including decreases in processing speed, visuospatial functions, attention span, and the ability to recognize emotions [64, 126, 168, 174]. These patterns differ from the more commonly known Alzheimer's dementia in that long term memory usually remains unaffected [130, 168]. There exist many different cognitive evaluations used to monitor these symptoms that report both accuracy and processing time during specific tasks. The UHDRS includes three separate assessments: a verbal fluency test (VFT); which tests the ability to orally produce words given a phonemic constraint (e.g. those starting with a specific

letter) [23]; the Symbol Digit Modality Test (SDMT), which tests the ability to recall and write symbols corresponding to specific numbers [164]; and the Stroop Color and Word Test (SCWT), which tests the ability to simultaneously processing written, incongruent color/name pairings (e.g. the word "blue" written in red ink) [175]. Other evaluations than those included in the UHDRS are Serial Response Time Tasks [207], Emotional Recognition tasks [23], and Category Learning tasks [10] to assess the ability to implicitly learn motor sequences, identify emotions based on facial expressions, or assign visual stimuli to categories based on implicit criteria. There also exist several methods that incorporate both motor and cognitive processing including the Trail Making Test (TMT) [139] and the Cued Movement Sequence Task [62], which both require the patient to draw either lines or circles based on external stimuli, and Finger Tapping Tests [62], which assess the ability to correctly time repeated finger movements. Research has shown that these test results diminish with disease progression to varying degrees compared to those in age-matched controls [75, 84, 126, 174]. Of these, the SCWT scores generally yield the most significant correlation, but this varies across cohorts and studies [130, 167].

Many patients also experience psychiatric conditions such as depression, anxiety, irritability, impulsivity, and apathy [149]. Although these symptoms are far more variable than the strictly cognitive issues, research has shown that both apathy and impulsivity are nearly ubiquitous across patients [184]. A confounding factor when considering psychiatric symptoms such as depression are whether they are caused by HD-related neurodegeneration or simply a natural response to the disease prognosis. To this end, a double-blind trial found that at-risk individuals exhibited higher levels of depression than controls despite unawareness of their diagnosis [88]. These symptoms, although less well-defined, still have dire consequences; one study found that almost 10% of deaths in HD patients were the result of suicide rather than disease progression [20]. However, similar to cognitive changes, research has also found that psychiatric abnormalities arise before motor symptoms, suggesting their significance for earlier diagnosis than previously available [120]. The UHDRS therefore also includes a behavioral assessment of relevant psychiatric symptoms, as well an evaluation for functional capacity. This evaluates the extent to which a patient's symptoms have affected their quality of life, and includes questions related to their ability to maintain employment, prepare meals, practice hygiene, etc. It measures their total level of independence and yields a score for their total functional capacity (TFC). The TFC, along with results from cognitive testing and behavioral assessment, have all been shown to worsen with disease progression to variable degrees between patients [88, 120, 184]. More significantly, research has also identified that these metrics can begin to decline even before the noticeable onset of motor symptoms, suggesting their importance for earlier clinical intervention [126, 130, 137]. The earlier presence of cognitive and behavioral symptoms also implies that causative, neurological changes may be occurring during this time, pointing researchers towards brain imaging as tool to observe HD progression.

2.2.3 Neuropathology and Image-Based Markers

The primary consequences of the underlying genetic mutation of HD are widespread aggregations of the HTT protein in neuronal populations [21]. The wild-type of this protein can be found within the cytoplasm of neurons and glial cells in every area of the brain, and is associated with myriad functions including mitochondrial processes, endocytosis, and anti-apoptotic activity [150]. In its elongated form, however, the mutation changes the configuration of expressed HTT to render it insoluble, leading to the formation of both intra- and extra-cellular HTT aggregations [49]. The relationship between these aggregates and neurodegeneration remains only partially understood. Some suggest that the mutated form of HTT possesses a toxic gain of function and directly induces harmful effects [136, 156]. Others propose that in addition, the mutated form inhibits the healthy function of wild-type HTT [29, 60, 134]. To complicate things further, even though the protein is expressed in neurons across the entire brain, certain cell populations have shown to be more vulnerable to mutated HTT than others [198]. For example, within the striatum, medium-sized spiny neurons that project to the external globus pallidus undergo more degeneration than the same type of neurons projecting to the internal pallidus [138]. Similar discrepancies are also present within the cerebral cortex, where cortical thinning is primarily limited to Layers III, V, and VI, and only in specific regions [70, 138].

Although the specific genetic mutation remained unknown until 1993, studies characterizing brain atrophy in HD date as far back as the late 1890s [8, 100, 112]. These first studies identified severe volumetric loss within the putamen [8] and globus pallidus [100] of *ex vivo* HD brains compared to healthy controls. Throughout the 20th century, researchers investigated these and numerous other structures affected by HD including the caudate [6, 48], substantia nigra [5], thalamus [48], pons and medulla oblongata [6], amygdala [48], white matter [41], cerebellum [142], and the cerebral cortex as a whole [41]. Of the areas studied, the putamen and caudate (aka the striatum) remain the most well-defined indicators of disease progression to date; an example of HD-related atrophy within these regions is depicted in Figure 2.5. In 1985, Vonsattel et al. created a standardized grading system to assess striatal atrophy in the first, large-scale *ex vivo* analysis [197]. This study established the "Vonsattel Grading System", which defines 5 tiers of neurodegeneration based on the volume and topology of the putamen and caudate. Unfortunately, due to the lack of accessible neuroimaging technology, all aforementioned studies relied on histology rather than *in vivo* data. Although this provides easier quantification of microscopic information, for macroscopic data such as volume loss, this can introduce bias caused by tissue fixation. Further, the acquisition of *ex vivo* data remained a prevalent issue, and almost every study was limited by small sample sizes ($n < 10$).

In more recent years, magnetic resonance imaging (MRI) has proven to be a powerful tool to study HD related neurodegeneration with *in vivo* data [13, 15, 19, 26, 36, 46, 63, 86, 90, 91, 113, 118, 128, 129, 144,

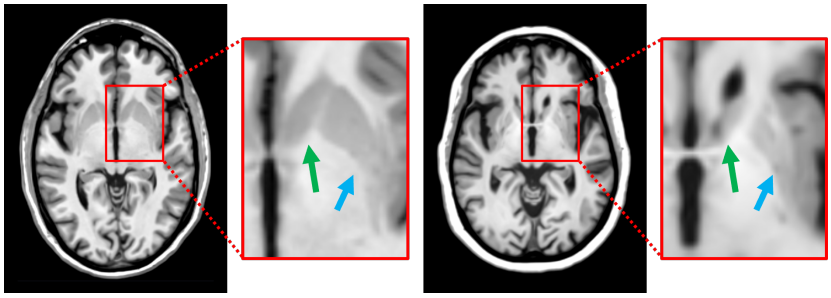

Figure 2.5: Putamen and caudate atrophy in a Huntington's disease brain (right) compared to a healthy control (left). The green arrow indicates the caudate nucleus and the blue arrow indicates the putamen.

145, 147, 148, 151, 163, 179]. A key advantage to this is the ability to observe the brains of living subjects, whereas *ex vivo* studies are limited to patients who have already deceased. While this distinction may seem obvious, it provides a broader pool of eligible subjects, thereby strengthening the power of the study. More importantly, it allows for the analysis of neuropathology over the course of disease progression rather than only after patient death. *In vivo* also provides the advantage that any observed structural changes from are not biased by the effects of tissue slicing and fixation, which are required for histological analyses.

Many imaging studies have been able to measure neuropathological changes *in vivo* that support earlier findings from histology. These include atrophy within subcortical regions such as the striatum [16, 36, 46, 104, 105, 129, 145], globus pallidus [16, 26, 46, 163], thalamus [16, 36, 46, 90], nucleus accumbens [16, 26, 36], hippocampus [26, 36], and amygdala [36, 115], as well as notable expansion of the lateral ventricles [146]. Of these, the putamen and caudate have been established as the best image-based marker for HD, as they exhibit the highest volume differential between HD cases and healthy controls [19, 113, 129]. Rosas et al., 2001, showed that atrophy of these structures increases with CAG repeat length, suggesting their correlation to age of disease onset [145]. Even more significantly, Aywlard et al., 2004, showed that the putamen and caudate begin to experience volume loss up to 9 and 11 years, respectively, prior to disease manifestation [19]. Unfortunately, these two structures alone do not account for the wide range of related symptoms associated with the disease. This has thus inspired researchers focused on other brain regions to incorporate UHDRS results (or similar metrics) into their statistical analyses in an attempt to correlate specific symptoms with specific neurodegenerative changes [26, 36, 90, 115]. For example, Kassubek et al., 2005, found that volume loss in the thalamus correlated to decreases in cognitive test scores, but not with motor symptoms [90]. Conversely, Bogaard et al., 2011, found no significant atrophy in the thalamus with respect to whole brain volume, but did observe volume loss in the nucleus accumbens, hippocampus, and pallidum that corresponded with motor and functional scores [26]. Coppen et al., 2018, found no symptomatic relationship to changes in the amygdala because they only considered motor symptoms within their analysis [36], whereas

Mason et al., 2015, reported significant correlation between the amygdala and deficits in emotional processing [115]. These discrepancies show that statistical significance of regional atrophy depends highly on the choice of clinical variables, and also suggest that subcortical changes alone do not provide a full explanation for symptom variability.

In addition to subcortical regions, studies have also used magnetic resonance imaging (MRI) to observe changes in cortical GM and their relationship to symptom expression [15, 16, 36, 46, 91, 116, 143, 146–148, 151, 178]. Aylward et al., 1998 [15] performed the first MRI analysis of cortical changes in HD, where they observed a relationship between volume decreases in the frontal lobe and cognitive symptom severity. However, this correlation was not significant after accounting for total brain volume, leading the authors to suggest that HD cortical atrophy is not unique to the frontal lobes. Rosas et al., 2002, identified the sensorimotor region as the most heavily affected, but also found changes within the occipital and temporal lobes [147]. This study found that cortical thinning in HD progresses from posterior to anterior regions, and corresponds to earlier findings regarding symptom presentation [99]. A subsequent report from Rosas et al., 2003, emphasized that GM atrophy in HD is widespread across the entire cortex, and that specific patterns of GM thinning correlate to specific symptoms in prodromal HD [146]. In contrast, in Aylward et al., 2013, the authors were surprised to find no relationships between regional thinning and any early symptoms [16]. Coppen et al., 2018, came to a similar conclusion; they studied the correlation between cortical volume loss and motor deficits, and found a relationship only between occipital lobe atrophy and eye movement [36]. This study attributed their lack of statistically significant findings in the cortex to the fact that they only considered motor symptoms, rather than cognitive or functional. Thus, although the hypothesis that widespread GM atrophy corresponds to symptom expression is generally supported, the degree and localization of thickness changes vary widely between studies, due in part to both sample sizes and limitations of current technologies [16, 36, 46, 91, 148, 178].

Until this point, all discussion of neurological changes has been limited to findings from cross-sectional analyses, those that include only one timepoint per subject. However, perhaps the most important advantage of image-based studies is the ability to obtain measurements from a single patient at multiple timepoints. Longitudinal analyses, those that include more than one timepoint per subject, not only allow for comparison between HD cohorts and healthy controls, but also enable the observation of changes within a specific subject over time [13, 14, 17, 73, 74, 86, 106, 113, 127, 128, 151, 155, 173, 179, 179–181, 205]. Moreover, these studies are particularly advantageous for examining neurological changes in pre-symptomatic individuals. For example, Wijerante et al., 2021 [205] used longitudinal structural MRI data to verify model prediction of early putamen and caudate atrophy from Aylward et al., 2004 [19] based on estimated time of disease onset. Many studies have similarly reported trends in both striatal and other subcortical atrophy alongside

clinical metrics such as cognitive testing scores, total motor score (TMS), and total functional capacity (TFC) [1, 18, 114, 128, 155, 179, 180, 206]. However, when measuring atrophy in cortical GM, most studies either quantify volume on the scale of the entire brain, rather than within individual regions, or simply exclude it altogether. In the few that examine more localized cortical GM atrophy, results differ between studies as stated previously [72, 83, 173, 206]. A more robust understanding of localized GM changes could provide a more thorough explanation of symptom variation in addition to subcortical findings. This is the primary motivation for the work in this dissertation to develop improved methods to quantify GM thickness changes.

2.3 Cortical Segmentation

Cortical segmentation is a useful tool for visualizing the human brain and quantifying important metrics such as GM thickness, surface area, and sulcal depth. It involves precisely identifying the boundaries between GM, WM, and cerebrospinal fluid (CSF) from within an MRI, and reconstructing surface representations thereof. We refer to the GM/CSF interface as the GM surface, and the GM/WM interface as the WM surface. Figure 2.6 displays an example of these surfaces both in 3D (left/center) and overlayed within a 2D slice from the corresponding image (right).

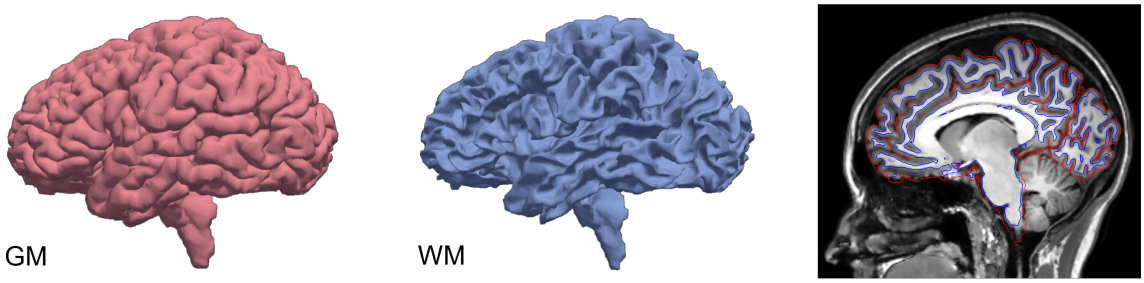

Figure 2.6: Example of GM (pink) and WM (blue) cortical surface reconstructions. Both surfaces are also displayed within the corresponding brain MRI (right).

Accurate cortical segmentations remain highly challenging to procure due to the complex geometry and folding of the human cortex. Traditional methods for surface reconstruction relied on manually tracing the cortical boundaries within 2D images, acquired first using photographs of histology, and later with 2D MRI. Early research primarily focused upon “unfolding” these contours and combining them into a 2D mapping of the entire cortex [32, 192, 193]. Throughout the 1980s and 1990s, the growing accessibility of computer-based analyses rendered it possible to reconstruct these contours into digital, 3D representations, but the initial segmentation was still founded in 2D methods reliant on manually obtained data [43, 47, 160, 191]. Other computational methods not reliant on manual contour tracing employed less complex, wire-frame surfaces for data visualization, and were almost exclusively conducted with non-human primate data [57, 158, 159].

The first automated cortical surface reconstruction protocol was proposed by Dale and Serano, 1993

[37]. In this work, they used an iterative, volumetric seed filling algorithm to perform tissue classification of GM and WM voxels, and then adapted an automatic deformable template algorithm [214] that deformed a spherical wire-frame to effectively "shrinkwrap" the voxels corresponding to the target tissue. Their method also incorporated both T1w and T2w (or proton density) images to exploit the different tissue contrasts provided by each modality (see Section 3.1 for more details). The critical insight leading to their work was that the GM surface is very difficult to segment *on its own*, particularly given the technological limitations of MRI resolution in the 1990s. Others had attempted to circumvent this by beginning with a topologically correct surface (e.g. a sphere) surrounding the entire brain, and deforming it inwards to approximate the GM/CSF interface; these methods were never able to fully resolve the deep sulci of the cortex [52]. However, Dale and Serano instead exploited the fact that WM is much easier to segment and represent with a surface. Deformation of the WM surface outwards into the cortical GM is subsequently a much more manageable problem because the initial surface already possesses both the correct topology and approximate geometry (i.e., cortical folding) of the desired result. This research was the foundation of the popular FreeSurfer (FS) toolkit for cortical segmentation by Dale et al., 1999 [38], which will be described in more detail later in this section.

Also in 1999, Xu et al. [210] introduced an alternative technique for cortical segmentation. Rather than using a hard tissue classification, this algorithm employed a probabilistic method to produce a "fuzzy segmentation", one that associates each voxel with a set of probabilities corresponding to each tissue type as opposed to a binary label [133]. Inspired by Dale and Serano, both the GM and WM surfaces were initialized by first generating a surface at the approximate GM/WM interface defined in the WM probabilistic tissue map. These initial surfaces were then deformed based on a vector field generated from the original MRI with a generalized gradient vector flow (GGVF) algorithm [211]. This research laid the groundwork for the Cortical Reconstruction Using Implicit Surface Evolution (CRUISE) algorithm [69]. The CRUISE pipeline is distinct from FS in that it generates three separate surfaces: a GM surface, a WM surface, central surface indicating the boundary between cortical layers III and V. However, it never gained the same level of popularity experienced FS.

2.3.1 The FreeSurfer Software Suite

FreeSurfer [38] is one of the most widely used tools in the field of human neuroimaging. Although several others had proposed similar methods prior to its introduction, FS was the first, fully automated pipeline for both image pre-processing and cortical segmentation in 3D MRI. Given a set of T1w volumes (and optional T2w or FLAIR volumes) for a specific subject, the pipeline will perform a series of image pre-processing, followed by segmentation and reconstruction of cortical surfaces. Over the years, its applications have in-

creased dramatically, lending to its widespread popularity. It now contains procedures for intersubject surface alignment based on cortical folding patterns [54], cortical thickness (CTh) measurement [53], volumetric segmentation of most brain structures [55], hippocampal subfield segmentation [194], WM fascicle segmentation with diffusion MRI [212], and cortical surface parcellation [42]. In this section, we will review the cortical segmentation and surface reconstruction pipeline in full, followed by a discussion on its extension to longitudinal processing.

Segmentation Pipeline

The first step of the FS pipeline is to register an input T1w image (and the optional T2w or FLAIR) into the Talairach space [35], which is a standardized coordinate space used in neuroimage analyses. Each image then undergoes bias field correction to remove intensity variation induced by the MR scanner, followed by a skull-stripping algorithm to isolate all voxels lying external to the CSF. Intensity based tissue classification is then performed to yield a binary WM voxel map used to initialize surface reconstruction. Next, the WM map is partitioned into the left and right hemispheres by defining two cutting planes: one that bisects the WM through the center the corpus callosum (CC) in the sagittal direction, and the other that horizontally bisects the pons to remove subcortical structures. The largest connected component on either side of the CC plane is extracted to yield a single, connected mass of WM voxels for each hemisphere. The remainder of the processing steps are conducted separately for each hemisphere to decrease the computational requirements for segmentation.

Once a fully connected WM volume is obtained, FS then generates an initial WM surface. This is achieved by converting the volumetric binary mask into a tessellation such that the face of each WM voxel adjacent to a background voxel is represented by two triangles. This tessellation is then smoothed with the "shrinkwrapping" algorithm used in Dale and Serano, 1993 [37]. The resulting surface is duplicated, and each copy is deformed to become either the GM or WM surface. Specifically, this step solves for an optimal deformation of each surface by minimizing a linear combination of three energy functions:

$$E = E_t + \lambda_n E_n + \lambda_I E_I \tag{2.2}$$

The first two terms, E_t and E_n, represent a spring-property of the deformation that is decomposed into tangential (t) and normal (n) components. E_t redistributes surface vertices to encourage uniform spacing and triangle size, thereby regularizing the mesh, while E_n enforces surface smoothness by penalizing vertices that travel too far in the direction of the surface normal at its neighbors. Both of these parameters are weighted equally between the GM and WM surfaces (i.e., λ_n for the GM is equal to λ_n for the WM). The E_I term

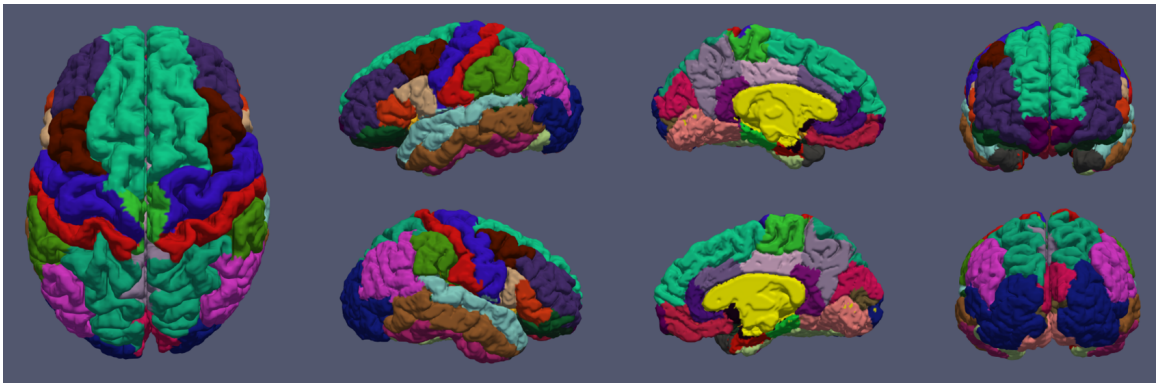

Figure 2.7: Example of the Desikan-Killiany cortical atlas [42] parcellated onto GM surface reconstructions.

allows the image intensity to influence surface placement, and this parameter *is* weighted differently between GM and WM. FS optionally uses a T2w or FLAIR modality to define E_I for the GM surface, rather than the T1w, due to the increased contrast between GM and CSF. During this stage, any self-intersections induced by vertex movement are corrected, and no holes or handles are introduced by the deformation. The end result is thus a topologically correct reconstruction of the GM and WM surface for each hemisphere. Following reconstruction, the cortical GM surface is parcellated using the Desikan-Killiany atlas to identify 33 individuals structural regions per hemisphere, an example of which is included in Figure 2.7.

Joint Initialization for Longitudinal Processing

As discussed briefly in Section 2.2.3, longitudinal analyses offer several advantages over those that employ only one timepoint per subject; this claim also holds true for cortical segmentation. Research has shown that longitudinal (4D) image processing methods provide more robust results than cross-sectional (3D) techniques because the changes detected between timepoints is more likely to be attributed to anatomical changes than intrasubject noise [11, 141, 208]. Traditional 4D methods often operate by treating the baseline image (e.g. timepoint 0) differently than the followup images (e.g. all subsequent timepoints), and use it for initialization of various processes. For example, because it is desirable to have all timepoints in the same image space for consistency, the baseline image will be employed as the "fixed image" to which all others are registered. This induces interpolation error in all moving timepoints, which can bias observations made from the data [186, 215]. Additionally, employed atlases or label maps for tissue classification are often aligned with the baseline image and then applied to all data, further introducing bias towards the selected "fixed" timepoint.

A common way to avoid inducing bias is to create a subject-specific template (SST) using all images with equal representation. The first example of this was proposed by Smith et al., 2001 and 2002 [165, 166], where the authors created such a template by registering two individual images into a "half-way space"

equidistant from their original positions prior to any analyses. This ensures that any interpolation error will be equally present in both timepoints, and any employed atlases can be registered to the template space rather than to either timepoint. In Reuter et al., 2012 [141] a similar concept was introduced into the FS pipeline that was generalizable to any number of timepoints. After conducting cross-sectional segmentation of each timepoint for a subject, FS generates an SST by iteratively aligning all timepoints to the median image using an symmetric rigid registration algorithm [140]. The median image $\hat{I}$ and corresponding set of transformations $\hat{\varphi}^i$ for each timepoint image are calculated as the solution to

$$\{\hat{I}, \hat{\varphi}^i\} = \underset{I,\varphi^i}{\operatorname{argmin}} \sum_{i=1}^{n} E_{\text{IDM}}(J^i \circ \varphi^i, \hat{I}) + D(\varphi^i)^2 \tag{2.3}$$

where $E_{\text{IDM}} = \int_\Omega |I_1(x) - I_2(x)|$ is a dissimilarity metric between a transformed image J^i and the median image from that iteration, and D the squared distance of a transform with respect to identity. Each φ^i is restricted to be an inverse consistent rigid transformation [140]. Once the SST is calculated, all images are rigidly registered to the template space using the appropriate φ^i, and all subsequent processing steps are performed in this space to minimize the amount of induced bias towards any single timepoint. The SST is used to initialize all pre-processing steps, leading to the term "joint-initialization". The template is also used to align the resulting surface reconstructions for the subject to pre-defined cortical atlases. However, the actual segmentation procedure is still conducted independently for each timepoint with the intention of avoiding over-regularization. This means there is still opportunity for the introduction of intrasubject noise that could be prevented by performing joint segmentation as well as initialization. In Chapter **??** of this dissertation, we examine the extent to which an alternative surface detection framework can jointly segment multiple timepoints without over-regularizing the results.

2.4 Cortical Thickness Measurement

Cortical thickness (CTh), the thickness of the cortical GM layer in the brain, is an important image-based marker in both healthy aging and neurodegenerative pathologies. In a healthy, adult human, the average CTh across the entire cerebral cortex is approximately 2.5 to 3 mm [71, 220]. Across the brain, this thickness varies widely. Figure 2.8 displays an example surface reconstruction of a healthy brain with CTh projected onto each point. The occipital lobe and primary somatosensory cortex are two of the thinnest regions, with values around 2 mm, while the primary motor cortex and limbic area of the temporal lobe are much thicker, with values around 4 mm.

Prior to the invention of imaging techniques to study the human brain, CTh was measured using only *ex vivo* data from either post mortem or histological analyses [28, 71, 117, 196]. In 1909, Brodmann published

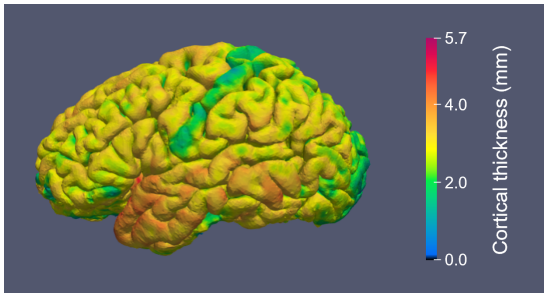

Figure 2.8: Example cortical thickness values for a healthy, adult brain projected onto a cortical surface.

the fundamental research mapping the cerebral cortex into over 50 distinct regions based on structure, location, and function; this work included metrics regarding CTh in each region [28]. This was expanded upon in von Economo and Koskinas, 1925, [196] which provided the first detailed descriptions of the cytoarchitecture of the cortical layers in different brain areas. In these studies, CTh was calculated either by inserting a depth gauge into the cortex normal to the surface, or with tissue sectioning and using caliper to measure each 2D slice. Although foundational to the study of the human cortex, histological studies cannot provide truly accurate results due to the structural changes that occur as a consequence of tissue fixation, as stated previously in Section 2.2.3. Further, by definition, post mortem data is limited to subjects who have already deceased, prohibiting observation of brain changes in healthy aging or throughout disease progression.

The introduction of MRI allowed research to pivot toward studying the human cortex *in vivo*. Numerous studies were conducted towards the end of the 20th century to observe GM in healthy brains [71, 117] and how it changes in both normal aging [80] and pathologies including Alzheimer's disease [4, 59, 80], other forms of dementia [93], Huntington's disease [68, 70, 198], and schizophrenia [96, 221]. In these studies, CTh was quantified by measuring GM thickness within 2D image slices. Although method this alleviates bias caused by tissue fixation, it does not solve the issue of attempting to quantify an inherently 3D metric within a 2D space. Figure 2.9 exemplifies this dilemma; each panel displays a CTh measurement obtained at the same point in the cortex, but measured through a different imaging slice direction. The variance between the three values indicates that CTh cannot be accurately calculated using different perpendicular slices.

Motivated by the issue presented in Figure 2.9, researchers have since invented alternative, more robust techniques to measure CTh in a 3D MRI. In general, these require two definitions: (1) the specific location of the GM/CSF and GM/WM interfaces, and (2) a distance mapping from one cortical boundary to the other. These methods can be described as either volumetric, where CTh is calculated directly within the image space, or surface-based, where GM and WM surfaces reconstructions are generated from the image prior to CTh measurement as described in Section 2.3. In the latter case, CTh is then measured as the distance between the surfaces rather than within the image itself. The rest of this section will describe several methods

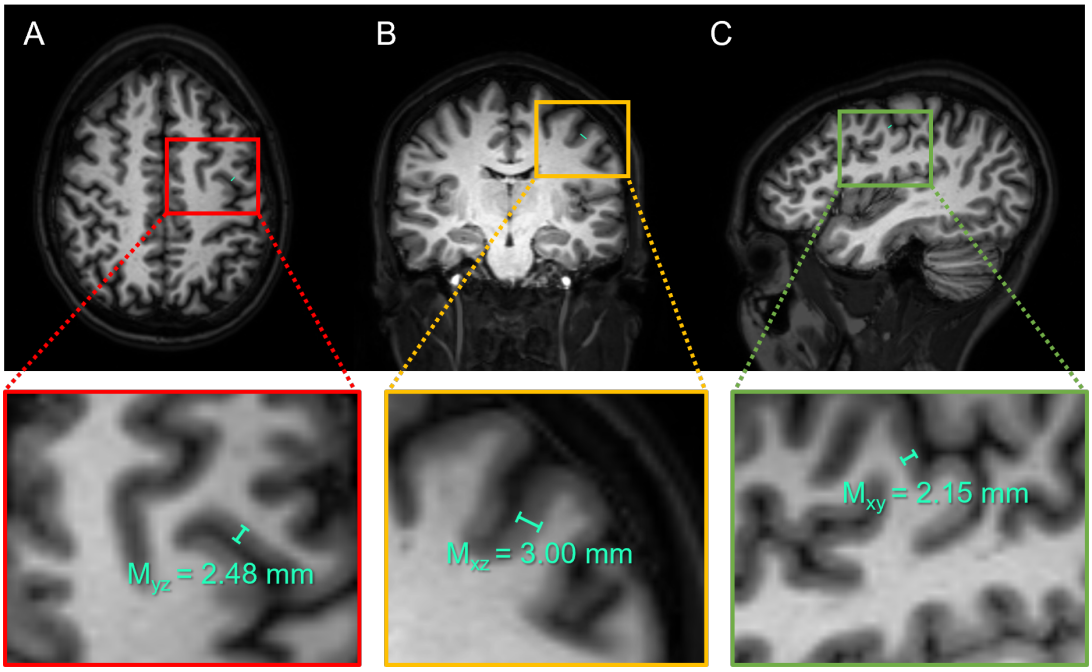

Figure 2.9: Cortical thickness measured at the same location within three different 2D slice locations.

from each category in more detail.

2.4.1 Surface-based Cortical Thickness Measurement

In general, surface-based CTh methods operate by first generating cortical surface reconstructions of the GM/CSF and GM/WM boundaries, and then calculating thickness as the distance between those two surfaces [53, 69, 122, 141, 201]. Because these operate using surfaces rather than within a voxel grid, they are less prone to error induced by image noise or partial volume effect (PVE) [34]. However, the need for surface reconstruction also renders them more computationally expensive counterparts. The actual distance mapping between these surfaces and subsequent thickness calculation steps vary in complexity between methods. This section will demonstrate the ideal properties of a CTh mapping explore several different existing methods in literature.

Let S_0 and S_1 be two non-overlapping surfaces such that S_0 is situated exterior to S_1, similar to the GM and WM surfaces. Let $D : p_0, p_1 \rightarrow \mathbb{R}$ be a function such that if p_0 is a point on S_0 and p_1 a point on S_1, then $D(p_0, p_1)$ is the distance between the two points. For example,

$$D(p_0, p_1) = \sqrt{(p_{0,x} - p_{1,x})^2 + (p_{0,y} - p_{1,y})^2 + (p_{0,z} - p_{1,z})^2} \tag{2.4}$$

is the Euclidean distance between p_0 and p_1 in 3D, Cartesian space. Next, we can define a function $f_D : S_0 \rightarrow$

S_1 that maps each point on S_0 to a point on S_1, given a specific definition for the function D. We call f_D a distance mapping. For an ideal distance mapping between surfaces, we desire a D and an f_D that satisfies the following properties:

1. For each point p_0 on S_0, there exists a unique point p_1 on S_1 such that $f_D(p_0) = p_1$ (f_D is one-to-one).
2. For each point p_1 on S_1, there exists at least one point p_0 on S_0 such that $f_D(p_0) = p_1$ (f_D is onto).
3. If $f_D(p_0) = p_1$, and $g_D : S_1 \to S_0$ is a function using the same definition D that maps each point on S_1 to a point on S_0, then $g_D(p_1) = p_0$ (D is symmetric).

If f_D is both one-to-one and onto, then we call f_D a bijection, and there exists an $f_D^{-1} : p_1 \to p_0$ such that $f_D^{-1}(f_D(p_0)) = p_0$. Further, if D is symmetric, then this implies that if a point p_0 on S_0 maps to a point p_1 on S_1, then $D(p_0, p_1) = D(p_1, p_0)$. In terms of GM thickness measurement, this means that the thickness measured from the GM surface to WM surface should be equal to that measured from WM to GM at the same location on the cortex.

Recall that prior to the rise of imaging methods, one common way to measure CTh was by inserting a depth gauge into the brain normal to the cortical surface. When analysed mathematically, it is easy to show that this distance mapping fails to satisfy any of our three desired properties. Let the distance measured by this method be called the orthonormal projection distance, denoted D_{norm}. Let S_0, S_1, T_0, and T_1 be the surfaces shown in Figure 2.10, where S_0 and S_1 are a pair of concave surfaces, and T_0 and T_1 are convex. We define functions $f_{D_{\text{norm}}} : S_0 \to S_1, T_0 \to T_1$ and $g_{D_{\text{norm}}} : S_1 \to S_0, T_1 \to T_0$ represented by the blue and green arrows in Figure 2.10, respectively. If the orthonormal projection distance met our established requirements that $f_{D_{\text{norm}}}$ be bijective and $g_{D_{\text{norm}}} = f_{D_{\text{norm}}}^{-1}$ be bijective, then $f_{D_{\text{norm}}}(p_0) = p_1$ would imply that $g_{D_{\text{norm}}}(p_1) = p_0$. Although this is true for points q_0 and q_1, this statement does not hold when the surfaces have non-parallel normals. In both cases, we have $f_{D_{\text{norm}}}(r_0) = r_1$, but $g_{D_{\text{norm}}}(r_1) = r_0'$.

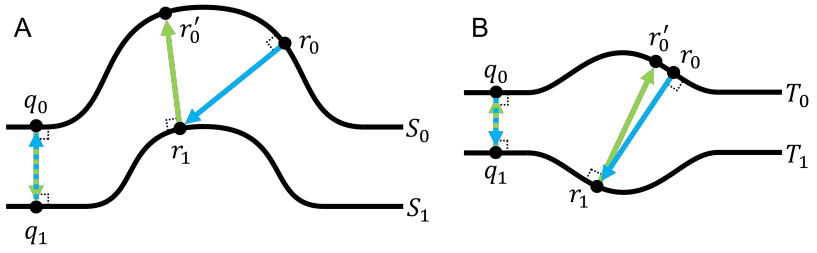

Figure 2.10: Orthonormal projection distance mapping between two sets of surfaces: S_0 and S_1 (A), and T_0 and T_1 (B). The blue arrows denote mappings from points on S_0 and T_0 to those on S_1 and T_1, and green arrows denote mappings from S_1 and T_1 to S_0 and T_0.

Another method to measure distance between surfaces, and perhaps the most logically intuitive, is by simply calculating the shortest path from each point on one surface to the other. We denote this distance

D_{min}, and the corresponding mapping between points $f_{D_{\text{min}}}$. Figure 2.11 displays this mapping between the same sets of surfaces, with concave S_0 and S_1 in panel 2.11A, and convex T_0 and T_1 in panel 2.11B. $f_{D_{\text{min}}} : S_0 \rightarrow S_1, T_0 \rightarrow T_1$ are again represented by blue arrows and $g_{D_{\text{min}}} : S_1 \rightarrow S_0, T_1 \rightarrow T_0$ by green arrows. As before, we see that while $f_{D_{\text{min}}}(q_0) = q_1$ and $g_{D_{\text{min}}}(q_1) = q_0$, we have $f_{D_{\text{norm}}}(r_0) = r_1$, and $g_{D_{\text{norm}}}(r_1) = r_0^{'}$.

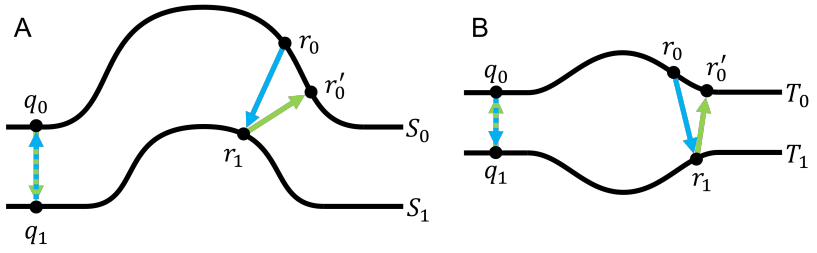

Figure 2.11: Minimum Euclidean distance mapping between two sets of surfaces: S_0 and S_1 (A), and T_0 and T_1 (B). The blue arrows denote mappings from points on S_0 and T_0 to those on S_1 and T_1, and green arrows denote mappings from S_1 and T_1 to S_0 and T_0.

In practice, creating a bijective mapping between two surfaces is highly computationally expensive, and alternative methods are often employed. Currently, the most widely used definition for surface-based CTh measurement is the symmetric closest point (SCP) distance formula, which was first introduced by Fischl and Dale, 2000 [53]. Denoted D_{SCP}, this definition builds upon the minimum Euclidean distance mapping to more closely approximate a one-to-one mapping between the GM and WM surfaces. Each point p_{GM} on the GM surface is mapped to its closest point p_{WM}. This p_{WM} is then mapped to its closest point $p_{\text{GM}}^{'}$ back on the GM surface, and thickness is calculated as the average of these two distances (equation 2.5).

$$D_{\text{SCP}}(p_{\text{GM}}) = \frac{D_{\text{min}}(p_{\text{GM}}, p_{\text{WM}}) + D_{\text{min}}(p_{\text{WM}}, p_{\text{GM}}^{'})}{2} \tag{2.5}$$

Although D_{SCP} does not lead to a bijective mapping $f_{D_{\text{SCP}}}$ between the GM and WM surfaces, it does yield more symmetric results than D_{min}. Fischl and Dale demonstrated that this CTh definition yields thickness measurements consistent with previous results, given accurate reconstructions of the cortical GM and WM surfaces [53]. Unfortunately, D_{SCP} has also been shown to underestimate CTh in areas with high curvature [123], which could prevent a study using this mapping from identifying significant differences in CTh between cohorts.

2.4.2 Volumetric Methods for Cortical Thickness

Volumetric CTh methods are those that compute thickness directly within the image space [31, 39, 65, 78, 79, 85, 188, 213]. For these algorithms, cortical boundary identification generally relies on some form of tissue classification [195, 218] to categorize voxels as either GM, WM, or CSF. Thickness is then calculated at each

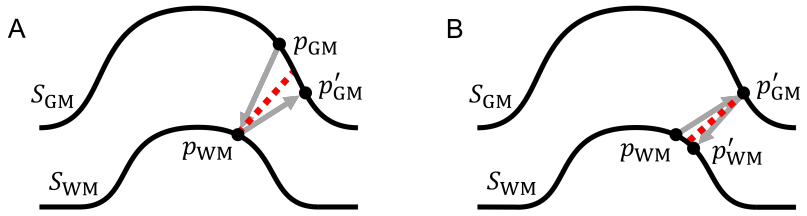

Figure 2.12: Symmetric closest point (SCP) distance mapping for cortical thickness (CT). S_{GM} and S_{WM} denote the gray matter (GM) and white matter (GM) surfaces, respectively. The gray arrows indicate mappings $f_{D_{SCP}}$ between surfaces, and the dotted red line denotes the resulting thickness D_{SCP}. **(A)** $D_{SCP}(p_{GM})$, measured from GM $\rightarrow$ WM $\rightarrow$ GM. **(B)** $D_{SCP}(p_{WM})$, measured from WM $\rightarrow$ GM $\rightarrow$ WM.

voxel labeled as GM based on a distance mapping from one side of the GM ribbon to the other.

One of the first volumetric methods for 3D CTh calculation in MRI was the Volumetric Laplacian (VL) introduced by Jones et al., 2000 [85]. This work proposed the use of the Laplacian equation

$$\nabla^2\Psi = \frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2} + \frac{\partial^2\Psi}{\partial z^2} = 0 \tag{2.6}$$

to model the interior of the cortical GM ribbon as a continuous voltage gradient. Initial conditions are set on each cortical boundary such that $\Psi = 0$ V at the GM/WM interface (B_{WM}) and $\Psi = 10{,}000$ V at the GM/CSF interface (B_{GM}). The solution Ψ can then be found for each GM voxel within the bounds using iterative methods such as those described in Appendix B. Next, the gradient $\nabla\Psi$ is calculated in each axial direction using centered differences:

$$\frac{\Delta\Psi}{\Delta x} = \frac{\Psi(x+\Delta x, y, z) - \Psi(x-\Delta x, y, z)}{2} \tag{2.7}$$

This gradient is normalized at each point to yield a vector field $\mathbf{N}$ tangent to $\nabla\Psi$. Finally, thickness (D_{VL}) is calculated at each voxel as the sum of the lengths traversed along $\mathbf{N}$ from the center of the voxel to each cortical boundary. These paths are referred to as "streamlines" along the Laplacian field. Figure 2.13 depicts this process and show the streamlines extending outward from a point p_i along $\nabla\Psi$. The thinner black lines between the GM/WM surfaces are called "equipotential surfaces" because they represent contours along which Ψ is constant. Importantly, unlike the SCP formula, this method does create a one-to-one mapping between the cortical boundaries.

The volumetric Laplacian technique has inspired many subsequent methods employing partial differential equations (PDEs) for CTh measurement [65, 78, 79, 213]. Several of these also model the cortex using the Laplacian equation while improving upon the definition of the cortical boundaries from the methods in Jones et al., 2000. One of the difficulties in measuring CTh from MRI is that deep sulci are often lost due to PVE.

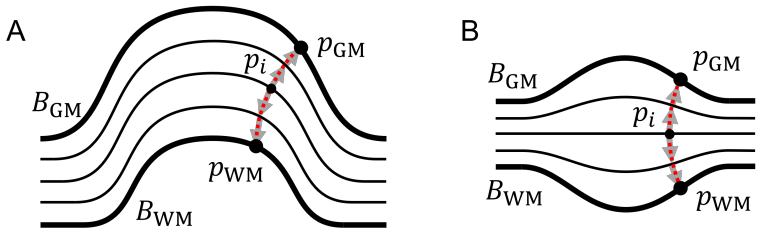

Figure 2.13: Volumetric Laplacian distance mapping for cortical thickness (CT). B_{GM} and B_{WM} denote the gray matter (GM) and white matter (GM) boundaries, respectively. The gray arrows indicate mappings $f_{D_{\mathrm{VL}}}$ between equipotential surfaces, and the dotted red line denotes the resulting thickness D_{VL}. **(A)** $D_{\mathrm{VL}}(p_i)$ in a concave region. **(B)** $D_{\mathrm{VL}}(p_i)$ in a convex region.

Jones et al. proposed a solution of inserting 1 voxel thick dilations of CSF within these deep sulcal regions to locally increase the contrast between GM and CSF. These areas are found by first generating a boxcar average image of the MRI, and then subtracting the average from the original to yield a "skeleton image" that extracts areas with high-frequency changes (such as within deep sulci). Following these corrections, the image undergoes two-threshold segmentation to classify each voxel as GM, WM, or CSF and identify cortical boundaries. Unfortunately, this restricts the accuracy of CTh measurements to the order of the image resolution.

Hutton et al., 2002 [79] and 2008 [78] expanded upon the VL by proposing a different mechanism for sulcal segmentation. To identify the locations of sulci obscured in the original MRI, this method considers sulcal voxels as bisected by the outer GM boundary and containing no CSF (i.e. half the voxel belongs to one sulcal bank, and half to another), allowing for sub-voxel resolution levels of thickness accuracy. On the other hand, Yezzi et al., 2003 [213] generalized CTh measurement beyond the Laplacian entirely for compatibility with any Eulerian PDE. Instead of calculating explicit correspondence trajectories with length equal to the GM thickness, the mathematical framework in this work yields thickness directly from the solution to the PDE. These two methods were then combined by Acosta et al., 2009 [2] to produce a method more computationally efficient than the VL, but more accurate than the Eulerian. The VL was then further improved upon in Cardoso et al., 2011 [30, 31], where the Laplacian thickness mapping was paired with the use of Khalimsky's cubic complex for topology preservation, which led to more accurate tissue segmentation and boundary definition.

In contrast to techniques founded upon PDEs, Das et al., 2009 [39] proposed the diffeomorphic, registration-based CTh (DiReCT) method that defines point-to-point correspondence between cortical boundaries using a diffeomorphic mapping (e.g. a differentiable mapping with a differentiable inverse). Similar to other methods, DiReCT uses a probabalistic tissue segmentation to define the GM/WM interface [218], but then deforms that boundary outwards based on a diffeomorphism to find the opposite edge of the GM. This method has

been implemented into the popular Advanced Normalization Tools (ANTs) pipeline [188, 190], which pairs DiReCT with an improved Atropos *n*-tissue segmentation [12] for more accurate GM and WM labels. When compared with the FreeSurfer pipeline that uses the SCP definition for surface-based thickness, ANTs was shown to yield CTh values that provided more accurate thickness-based age and gender predictions [34]. However, this method still suffers from issues characteristic of volumetric methods, motivating the work in Chapter 5 of this dissertation to develop a surface-based method exhibiting the symmetric properties of a Laplacian or diffeomorphic mapping.

CHAPTER 3

Image Datasets

3.1 Relevant MRI Modalities

This section describes the two structural image modalities referenced in this dissertation. Definitions for the MR parameters discussed are included in Table 3.1. The term "structural image" indicates that voxel intensities do not correspond to biophysical parameters as they would in quantitative or functional imaging. Traditional T1- and T2-weighted structural images are created using a spin-echo (SE) pulse sequence with the received signal S taking the general form

$$S = K \cdot [H] \cdot (1 - e^{-TR/T1}) \cdot e^{-TE/T2} \tag{3.1}$$

where K is a scaling factor and $[H]$ is the proton density within the voxel. By adjusting the repetition time (TR) and echo time (TE) based on the intrinsic T1 and T2 values of the target tissues, we can weight S to create contrast between specific tissue types. T1-weighted (T1w) brain images typically have high contrast between GM and WM with low intensity CSF, whereas T2-weighted (T2w) images generally have very bright CSF but low contrast between GM and WM. SE sequences often consist of a 90° pulse followed by a series of 180° pulses, called multi-spin-echo (MSE), or a series of pairs of 90° and 180° pulses, called turbo-spin-echo (TSE). Alternatively, one can use a gradient echo (GRE) to produce a signal of the same form by applying a single radio frequency (RF) pulse followed by a B_1 gradient reversal. The advantage of GREs is that these scans generally employ a small flip angle and a single RF pulse, and are generally much faster than SE sequences.

In certain cases, it is advantageous to apply an inversion pulse prior to the subsequent pulse sequence. The inversion pulse flips the longitudinal magnetization $\mathbf{M_z}$ to be oriented anti-parallel to the B_0 field. The result of this is that $\mathbf{M_z}$ must now relax from $-\mathbf{M_0}$ to $+\mathbf{M_0}$ in order to reach equilibrium, necessitating that at some point, $\mathbf{M_z} = 0$. If the subsequent RF pulse is applied at just the right moment, the signal from the tissue experiencing net zero magnetization will be nulled. This technique, known as "inversion recovery" (IR), is therefore highly useful for adjusting image contrast to either enhance or suppress the signal from certain tissue types. IR-prepped sequences are used in both modalities described in this section.

B_0 field	Intrinsic magnetic field of the scanner
B_1 field	Magnetic field induced by the transmitted RF pulse sequence
Gradient-echo (GRE)	Sequence of a single RF pulse followed by a gradient reversal
Inversion pulse	180° RF pulse delivered prior to a standard pulse sequence
Radio frequency (RF) pulse	Excitation pulse used in MRI in the range of 1-300 MHz
Spin-echo (SE)	Basic sequence of a 90° and a 180° RF pulse
Echo time (TE)	Duration between RF pulse and signal acquisition
Field of view (FOV)	Total dimensions of space captured within the image
Flip angle	Maximum angle of net magnetic field rotation induced by RF pulse
Inversion time (TI)	Duration between inversion pulse and SE sequence
Relaxation time, longitudinal (T1)	Elapsed time of net magnetization to re-orient parallel to B_0 field
Relaxtation time, transverse (T2)	Elapsed time of net transverse magnetization to decay back to 0
Repetition time (TR)	Duration between successive RF pulse sequences for a given slice

Table 3.1: Relevant MRI terms (top section) and parameters (bottom section), with their definitions.

3.1.1 Magnetization-Prepared Rapid Acquisition Gradient Echo

Magnetization-Prepared Rapid Acquisition Gradient Echo (MPRAGE) [121] imaging is a form of T1w imaging that utilizes an IR-prepped sequence consisting of an inversion pulse followed by a series of rapidly acquired gradient echoes (GREs). MPRAGE sequences are generally designed such that TE$<<$TI$<$TR and have small flip angles. This modality is considered highly valuable for brain imaging, as it can achieve much higher GM/WM contrast than in a typical T1w image [27]. It has also been used in scenarios such as respiratory imaging due to the ability to suppressing artifacts occurring at air-tissue boundaries [121]. For MPRAGE, TR can refer to the time between RF pulses during the gradient echo phase or to the time between each inversion pulse.

3.1.2 FLuid Attenuated Inversion Recovery

FLuid Attenuated Inversion Recovery (FLAIR) [66] imaging is similar to conventional T2w imaging with the added caveat of fluid suppression. FLAIR sequences generally consist of an IR-prepped TSE sequence with a long TR and TE, and a TI such that $\mathbf{M_z} \approx 0$ within the CSF. This results in an image similar to a T2w image, but with CSF voxels almost completely nulled rather as than the brightest tissue type. Suppressing signal from fluid provides a much higher image clarity within the tissues, leading FLAIR imaging to become one of the most ubiquitous modalities for central nervous system diagnostics [97]. While there exist both T2-FLAIR and T1-FLAIR modalities, we use only to the T2 variety in this dissertation, and will thus refer to it as FLAIR without making the distinction.

Structural	– FLuid Attenuated Inversion Recovery (FLAIR) – Magnetization-Prepared Rapid Acquisition Gradient Echo (MPRAGE)
Diffusion	– Diffusion tensor imaging (DTI) – DTI geometric reference
Functional	– Resting state functional MRI (fMRI)
Field map	– B_0, TE = 8 ms – B_0, TE = 9 ms – B_1
Perfusion	– Arterial spin labeling (ASL) – ASL magnetization (M_0) – Vascular space occupancy (VASO)
Quantitative	– qT1 (15°) – qT1 (60°) – qT2 – Magnetization transfer (qMT)

Table 3.2: Imaging sequences included in the NITRC Kirby dataset

3.2 Included datasets

3.2.1 Multi-Modal MRI Reproducibility Resource (NITRC Kirby Dataset)

The Multi-Modal MRI Reproducibility Resource is a publicly available, test-retest dataset from the NeuroImaging Tools and Resources Collaboratory (NITRC)[1] [97]. The dataset consists of n=21 healthy, adult subjects (11/10 male/female, ages 22-61). Subjects are denoted as "Kirby###", yielding the nickname "the NITRC Kirby dataset". For each subject, there are two sets of 15 MR images procured on the same day; this renders the dataset highly useful for reproducibility studies because no anatomical differences should exist between timepoints. All images were acquired at a field strength of 3 T using an Achieva Phillips Healthcare scanner, and with body coil excitation and an eight channel phased array SENSitivity Encoding (SENSE) [135] head coil for reception. Each of the 42 timepoints (two per subject) contains the image results from 15 sequences that can be divided into six separate categories. The full list of sequences are detailed in Table 3.2. From these, experiments included within this dissertation utilize only the FLAIR and MPRAGE images. FLAIR images were obtained with TR/TE/TI=8000/330/2400 ms at a $1.1\times1.1\times1.1$ mm^3 resolution over a field of view (FOV) of $242\times180\times200$ mm, and the MPRAGE images with TR/TE/TI=6.7/3.1/842 ms at a $1.0\times1.0\times1.2$ mm^3 resolution over an FOV of $240\times204\times256$ mm. This dataset is used for the analyses in Chapters 4 and 5.

[1] www.nitrc.org

3.2.2 Validation Data for Cortical Reconstruction Algorithms

The Validation Data for Cortical Reconstruction Algorithms (VDCRA)[2] [161] dataset is a publicly available resource specifically designed for assessing the accuracy of cortical segmentation algorithms. It contains 10 subjects within two separate cohorts: five healthy controls (2/3 male/female, ages 30-49), and five multiple-sclerosis (MS) subjects (1/4 male/female, ages 40-59). We refer to these cohorts as VDCRA-HC and VDCRA-MS, respectively. For each subject, the data includes an MPRAGE, FLAIR, T2w, and proton density weighted (PDw) image; we again only utilize the MPRAGE and FLAIR modalities. The MPRAGE images were obtained with an FOV of $240\times204\times256$ mm at a resolution of $1.0\times1.0\times1.2$ mm^3 (HC) and $1.1\times1.1\times1.1$ mm^3 (MS), and the FLAIR images at an FOV of $242\times180\times200$ mm at a resolution of $1.1\times1.1\times1.1$ mm^3 (HC) and $0.83\times0.83\times0.83$ mm^3 (MS). Other parameters are not included because they were not specified in the manuscript corresponding to this dataset.

In addition to the structural images, each subject is also associated with sets of manually placed fiducial landmarks denoting the GM/WM and GM/CSF boundaries in the MPRAGE image space. These landmarks were placed by two separate raters with different levels of experience, and organized into three clusters of 10 different markers within seven different brain regions per hemisphere: the calcarine fissure, cingulate gyrus, central sulcus, parieto-occipital sulcus, superior frontal gyrus, superior temporal gyrus, and Sylvian fissure. Each cluster of 10 landmarks is constrained to a single axial slice, but the three clusters within a region each lie in a separate slice. The clusters are provided in pairs for both GM and WM surfaces, such that a pair of clusters denotes the GM or WM borders of a given region within the same axial slice. In total, the data for each subject contains 56 sets (7 regions $\times$ 2 hemispheres $\times$ 2 surfaces $\times$ 2 raters) of 30 landmarks (3 clusters $\times$ 10 landmarks). An additional set of landmarks indicating MS lesion boundaries was also placed by a third rater, but these are not considered in the research contained within this dissertation. The VDCRA dataset is used for the analyses in Chapters 4 and 5.

3.2.3 PREDICT-HD

The PREDICT-HD dataset is a multimodal, longitudinal study of pre-manifest and manifest Huntington's disease patients with age-matched controls. Each subject is associated with both imaging and/or clinical testing data obtained at two to seven different timepoints, each separated by an average of 18 months. While there exists a wide range of available imaging data such as structural, diffusion, and functional data, the work in this dissertation is limited to subjects with T1 and T2 images obtained at 3 T ($n = 417$). Because these images were obtained at such a large number of different sites, acquisition parameters for each sequence are not listed in this dissertation.

[2] https://iacl.ece.jhu.edu/index.php?title=Cortical_data/

Clinical testing includes assessment for motor skills and functional capacity, symbol digit modality test (SDMT), Stroop testing, and two types of trail making test (TMT) (see Section 2.2.2 for more details). The first TMT requires the subject to draw a trail between circled numbers in ascending order (TMTa), and the second requires them to link both numbers and letters (TMTb). Tables 3.3 and 3.4 display the demographics and clinical testing scores at baseline, respectively, for the included subjects. Each HD subject is categorized based on their CAP score as either low (CAP < 290), medium ($290 \leq$ CAP ≤ 368), or high (CAP > 368), as determined by Zhang et al., 2011 [219]. This dataset is used for the analyses in Chapter 6.

CAP group	n	Male/Female	Age range (yrs)	CAP_E Score	No. diagnosed
Control	115	46/69	$23-88$	n/a	1
Low	105	32/73	$19-65$	237 ± 36	6
Medium	108	31/77	$22-83$	329 ± 23	18
High	88	40/48	$20-74$	422 ± 46	35

Table 3.3: Demographics of the subset of the PREDICT-HD dataset with 3 T MRI at first clinical visit. Subjects are categorized into groups based on CAP score at entry (CAP_E) (Low = $CAP_E < 290$; Medium = $290 \leq CAP_E \leq 368$; High = $CAP_E > 368$.

CAP group	TMS	SDMT	Stroop (color/ interference/total)	TMT (a/b)
Control	3.8 ± 4.2	53.5 ± 9.7	82.5 ± 12.8 48.2 ± 10.0 103.3 ± 17.3	23.2 ± 7.8 53.7 ± 23.0
Low	4.5 ± 6.6	55.4 ± 11.6	82.8 ± 12.5 50.5 ± 12.0 102.3 ± 17.2	22.6 ± 8.2 52.2 ± 21.2
Medium	7.7 ± 7.5	45.5 ± 10.5	76.1 ± 11.5 44.7 ± 11.5 95.7 ± 20.3	25.0 ± 9.2 62.9 ± 28.0
High	12.1 ± 13.0	44.0 ± 10.5	70.9 ± 13.3 40.2 ± 11.4 87.0 ± 17.8	32.4 ± 13.7 77.5 ± 40.8

Table 3.4: Baseline cognitive test scores for PREDICT-HD subjects. HD subjects are categorized into groups based on CAP score at entry (CAP_E) (Low = $CAP_E < 290$; Medium = $290 \leq CAP_E \leq 368$; High = $CAP_E > 368$. SDMT = single digit modality test, TMT = trail making test.

CHAPTER 4

Synthetic Atrophy for Accuracy Validation of Longitudinal Cortical Analysis

The work in this section appears in:

Larson, K. E., Oguz, I. (2022). Synthetic Atrophy for Longitudinal Cortical Surface Analyses. *Frontiers in Neuroimaging*. [102]

Larson, K. E., Oguz, I. (2021). Synthetic atrophy for longitudinal surface-based cortical thickness measurement. *SPIE Medical Imaging: Image Processing*, volume 11596, page 125. [101]

4.1 Abstract

In the fields of longitudinal cortical segmentation and surface-based CTh measurement, difficulty in assessing accuracy remains a substantial limitation due to the inability of experimental validation against ground truth. Although methods have been developed to create synthetic datasets for these purposes, none provide a robust mechanism for measuring exact thickness changes with surface-based approaches. This research developed a registration-based technique for inducing synthetic cortical atrophy to create a longitudinal ground truth dataset specifically designed to address this gap in surface-based accuracy validation techniques. Across the entire brain, our method induced up to between 0.8 and 2.5 mm of localized cortical atrophy in a given gyrus depending on the region's original thickness. By calculating the image deformation to induce this atrophy at 400% of the original resolution in each direction, we induced a sub-voxel resolution amount of atrophy while minimizing partial volume effects. We also showed that cortical segmentations of synthetically atrophied images exhibit similar segmentation error to those obtained from images of naturally atrophied brains. Importantly, our pipeline relies exclusively on publicly available software and datasets.

4.2 Introduction

CTh is an important image-based marker for measuring patterns in both healthy aging and neurodegenerative pathologies. Methods to quantify CTh are categorized as either volumetric, where thickness is measured directly from a structural brain MRI, or surface-based, where thickness is measured as the distance between surface reconstructions of the segmented GM and WM layers. Surface-based methods, while less computationally efficient than their volumetric counterparts [39, 85, 103, 188], generally yield more accurate results due to reduced errors from partial volume effects and lower susceptibility to noise and topological defects [34]. Although advancements have been made in developing robust surface-based methods for cortical segmentation and quantifying CTh [38, 53, 69, 122–124], determining the accuracy of observed measurements

still remains challenging due to the difficulty of obtaining ground truth for experimental validation. This becomes an even greater setback for validation of longitudinal studies where exact spatial correspondences across timepoints are desirable yet elusive.

Validation of CTh measurements from MRI scans was initially achieved by comparing thickness values to those obtained via histology [53, 117]; this is problematic due to differences between in vivo vs. postmortem tissue, MR vs. histology imaging differences, and 2D vs. 3D measurements. More recently, both longitudinal and cross-sectional studies often achieve validation of CTh measurements by comparing thicknesses observed by a new pipeline to those from previously existing algorithms [123, 189, 190]. However, this makes it difficult to show whether the proposed method offers any improvement in accuracy over the current state of the art in reference to ground truth. Another approach for evaluating CTh methods is through test-retest validation [53, 78, 189], which is a measure of reproducibility rather than accuracy, or by assessing the correctness of surface topology rather than anatomical accuracy [38, 141]. Finally, another method is to compare surface placement in cortical segmentation results to manual landmarks [69, 122] placed along the boundaries of the cortical ribbon within the image, such as those of the publicly available Validation Data for Cortical Reconstruction Algorithms [161] [1] (Section 3.2.2). Unfortunately, these landmark datasets are generally limited to cross-sectional analyses; even if landmarks were manually placed in a longitudinal dataset, there would not exist exact correspondence between landmarks across timepoints. A better approach would be to employ a synthetic dataset with known, ground truth changes in thickness and surface location at each point on the cortical surfaces.

Several methods for inducing synthetic deformations in brain images have been developed to validate techniques designed to measure *volumetric* cortical changes [25, 40, 58, 89, 94, 209]. For example, Freeborough and Fox, 1997 [58] implemented an image magnification technique to validate their boundary shift integral method for detecting volumetric loss. Davatzikos et al., 2001 [40] synthetically induced cortical atrophy by simulating biomechanical deformations in a localized region to validate a voxel-based morphometry approach to atrophy detection. One of the more well-known methods for inducing cortical atrophy is that of Karaçali et al., 2006 [89] where a topology preserving deformation is used to induce a predetermined amount of volumetric change at each voxel in the image. More recent approaches include Xia et al., 2019 [209] and Bernal et al., 2021 [25] which both employ deep learning to simulate atrophy in structural MRI, and Khanal et al., 2017 [94] which uses a biophysical model to generate cortical changes that are then induced with deformable registration. While these methods are all capable of generating longitudinal data with synthetically induced changes in CT, none of them also provide a robust mechanism for measuring exact thickness changes for surface-based approaches. Additionally, these volumetric approaches aside from that proposed

[1] http://iacl.ece.jhu.edu/index.php?title=Cortical_data/

by Karaçali et al. are likely to cause topological defects in the cortex.

In this aim, we present a method to induce synthetic brain atrophy in an MRI using mathematical morphology and deformable registration to create a longitudinal ground truth dataset specifically designed for accuracy validation of *surface-based* CTh measurements. We detail each step of the proposed synthetic atrophy pipeline, and demonstrate how we used cortical surface reconstructions to quantify both segmentation accuracy and the ground truth changes in CT. We then describe the experiments used to determine the degree and localization of cortical atrophy induced in different regions across the brain, and the extent to which accurate cortical segmentations were produced from our synthetically atrophied data. Finally, we present the results of these experiments and discuss their implications for our methods.

4.3 Methods

4.3.1 Materials and Data Preparation

We employed both the Kirby (Section 3.2.1) and VDCRA (Section 3.2.2) datasets to develop and validate the methods presented in this Aim. All subjects from both datasets were processed using FS (version 6) with the "-FLAIRpial" option to produce a cortical parcellation (aparc+aseg.mgz) defined by the DK atlas [42]. We used this parcellation to produce three masks required for several steps of our atrophy pipeline: a WM mask, a cortical GM ribbon mask, and a full brain mask. The full brain mask included voxels corresponding to GM, WM, subcortical structures, and the ventricles, but excluded the cerebellum and brain stem. We used the WM defined in this atlas rather than FS's volumetric WM segmentation so that the boundaries between the WM and GM labels were consistent. We also created a skull-strip mask from a thresholded, skull-stripped brain image (brainmask.mgz), which differed from the "full brain mask" because it contained voxels corresponding to CSF in addition to brain tissue. Prior to running our pipeline, we resampled all images to an isotropic resolution that was determined such that no data was lost along the highest resolution axis. In our experiments, this corresponded to $1\times1\times1$ mm^3 for the Kirby dataset and VDCRA-HC cohort, or $0.83\times0.83\times0.83$ mm^3 for the VDCRA-MS cohort.

4.3.2 Synthetic Atrophy Induction

The overall goal of our synthetic atrophy pipeline is to apply a localized deformation to a T1w image that will push the outer boundary of a GM region in towards the WM to simulate localized atrophy. This is achieved by first creating a local atrophy 'target' using a series of binary morphology operations on a region of interest (ROI) within the cortical ribbon. A restricted, deformable registration is then computed from the original ROI to the atrophied target to create a smooth deformation field, resulting in a transformation with point-to-point correspondences between timepoints along with the final atrophied image.

The process begins by generating a cortical parcellation from the target image and selecting a specific gyrus within this label map to serve as the ROI for atrophy. Although in our experiments we used FreeSurfer [38] to produce the parcellation defined by the Desikan-Killiany Atlas [42], our methods are generalizable to any parcellation. We create a binary mask of the selected ROI, which we upsample by 400% in each direction to avoid introducing partial volume artifacts. We then perform a set of binary mathematical morphology operations on the high-resolution mask using the publicly available ITK [2] library. This series of operations simulates localized atrophy by first removing a 1-voxel border around the entire ROI and then reinserting voxels adjacent to WM, thus effectively "sloughing off" the outermost layer of voxels along only the GM/CSF interface without altering the GM/WM interface. Figure 4.1 displays this entire series of binary operations, which can be repeated as many times as needed to produce the desired amount of total atrophy. Specifically, each iteration involves the following steps:

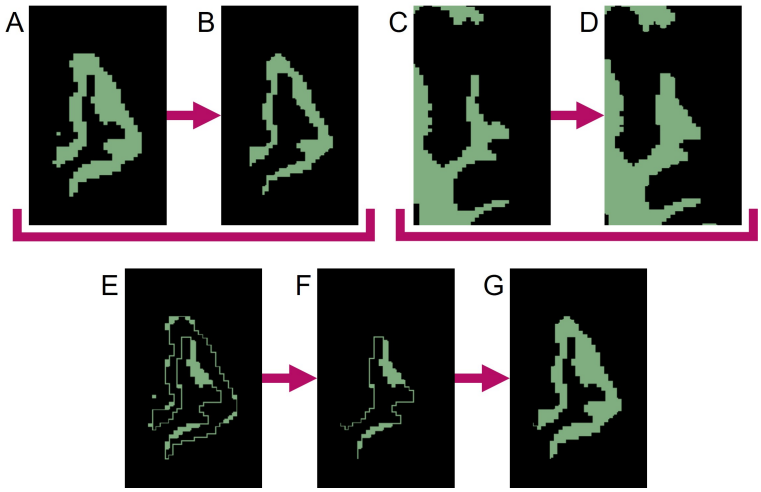

Figure 4.1: Schematic of binary mathematical morphology operations used to induce 1-voxel amount of atrophy in the left superior temporal gyrus (LSTG). **(A)** Mask of original LSTG. **(B)** Binary erosion of original LSTG mask (A) with a kernel radius of 1 voxel. Note this represent atrophy on both the inner and outer parts of the GM region, which will create undesirable holes between the WM and GM. **(C)** Mask of original WM. **(D)** Binary dilation of original WM mask (C) with a kernel radius of 1 voxel. **(E)** Original LSTG mask minus eroded LSTG mask (A−B). **(F)** Dilated WM mask times the difference between original and eroded LSTG masks (D×E). This represents the holes between the WM and GM created by the erosion operation. **(G)** Final mask of atrophied ROI, with the holes filled (B+F).

1. Apply a binary erosion filter with a kernel of radius $1\times1\times1$ voxels to the GM ROI. (Figure 4.1A $\rightarrow$ 4.1B)

2. Apply a binary dilation filter with the same kernel to the WM mask. This preserves the GM/WM interface. (Figure 4.1C $\rightarrow$ 4.1D)

[2] https://itk.org/

3. Subtract the eroded GM mask from the original to obtain a mask of the removed voxels. (Figure 4.1A − 4.1B = 4.1E)

4. Select from the mask obtained in step 3 only voxels within the dilated WM mask to extract all voxels lying on the GM/WM border. (Figure 4.1D × 4.1E = 4.1F)

5. Add back the border voxels extracted in step 3 to the eroded mask from step 1. (Figure 4.1B + 4.1F = 4.1G)

Once complete, these binary morphology operations yield a GM ROI with an unchanged WM boundary and a outer border eroded by one voxel, which in practice, also expands the CSF by one voxel. Each subsequent iteration erodes another single layer of GM voxels; this can proceed until the entire GM ribbon is removed from the ROI. Note that because we must dilate the WM to meet the eroded GM in order to preserve the boundary between the two tissues (step 2), this limits each iteration to eroding up to one half of the thickness of the original ROI. This limitation would persist even with a larger erosion kernel, and the morphology operations would fail if we used a kernel sized larger than half the tissue thickness. However, by only eroding one voxel per iteration, we bypass this issue and can atrophy deeper into the ROI. We refer to the number of iterations as the "effective erosion kernel size" because τ number of iterations removes an equal number of voxels to a single erosion operation that uses a kernel of size $\tau \times \tau \times \tau$ voxels. The end product of this series of iterations is the local atrophy target.

Once we thus obtain a binary mask for the local atrophy target, we next calculate a transformation using the steps detailed in Figure 4.2 to deform the T1w image accordingly, as well as any additional associated data such as T2w or FLAIR images. To achieve this, we first create high-resolution, full brain masks for both the original and atrophied timepoints. The original brain mask is obtained by thresholding a 400% upsampled cortical parcellation from which the ROI was selected (Figure 4.2A→4.2C), and the atrophied brain mask is obtained by substituting the eroded ROI mask for the original in the thresholded parcellation (Figure 4.2G→4.2H). We deformably register the original, full brain mask to the atrophied using the publically available Greedy software [216], yielding in a deformation field from the original to atrophied timepoint (Figure 4.2I). This registration is performed at four scales (100/100/50/100 iterations per scale, respectively) and using the mean squared difference as the similarity metric.

Next, we mask the resulting deformation with the original ROI and blur it outwards into the CSF within a custom blur mask. Creating this mask requires two inputs: a preliminary label map (4.2C) and the signed distance transform (SDT) (Figure 4.2D) of the full brain mask (Figure 4.2B). This preliminary label map parcellates the full brain mask into voxels inside the ROI (green), the rest of the GM ribbon (blue), the WM (yellow) as defined in the original DK atlas, and CSF (pink) defined using the skull-strip mask. Using these

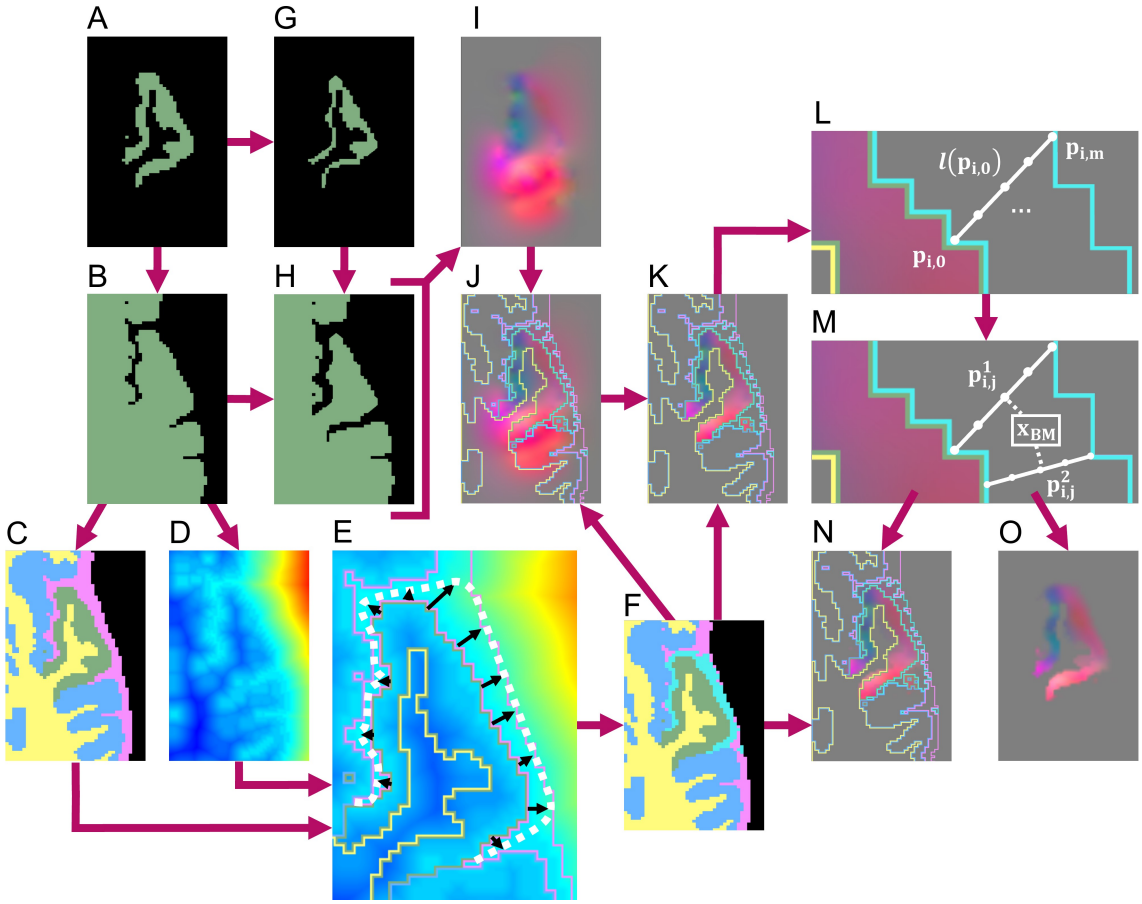

Figure 4.2: Schematic of pipeline to produce the masked and blurred deformation field for synthetic atrophy in the left superior temporal gyrus (LSTG) with 4 iterations of the binary mathematical morphology operations shown in Figure 4.1. **(A)** Original LSTG. **(B)** Original LSTG (A) with surrounding GM/WM, also referred to as the full brain mask of the original timepoint. **(C)** Full brain mask of original timepoint parcellated with 4 labels: the original ROI (green), WM (yellow), surrounding GM tissue (blue), and CSF (pink). **(D)** Signed distance transform (SDT) of original full brain mask. **(E)** Example of travelling along lines (black) normal to the GM/CSF boundary that extend to either the outer boundary of the CSF (pink), or the medial lines between neighboring gyri defined by the monotonic increase of the SDT (D) The full approximated boundary is denoted by the dotted white line. **(F)** Label map shown in (C) with added blur region (teal). **(G)** Atrophied LSTG (local atrophy target). **(H)** Atrophied LSTG (B) with surrounding GM/WM. **(I)** Deformation field obtained by registering the original LSTG + surrounding tissue mask (B) to the atrophied LSTG + surrounding tissue mask (H). **(J)** Deformation field (I) with label mask (F) as overlay. **(K)** Deformation field masked to the original ROI (A), with label mask (F) as overlay. **(L)** Example of defining m number of equally spaced points along a line $l(\mathbf{p_{i,0}})$ oriented normal to the GM/CSF boundary and extending from $\mathbf{p_{i,0}}$ to $\mathbf{p_{i,m}}$. **(M)** Example of interpolating the field value at a voxel $\mathbf{x_{BM}}$ within the blur mask using the n closest points to the center of that voxel. These points are constrained to lie on unique lines. In this 2D example, only 2 lines are used, but in 3D, the value is interpolated using the 4 closest lines. **(N)** Final deformation field after masking and outward smoothing, overlaid with label map from (F). The field inside the blur mask (teal) is a smooth transition from the original, unchanged field inside the ROI to the edge of the CSF or medial boundaries between gyri. **(O)** Final deformation field after masking and outward smoothing with no overlay.

two images (Figure 4.2C and D), we define the blur mask (teal region in Figure 4.2F) as the region bound by the GM/CSF border of the ROI, the outer edge of the CSF, and the medial lines between neighboring gyri calculated using the monotonic increase of the SDT (Figure 4.2E).

Inside this blur mask, we assign values to voxels such that we create a smooth transition from the deformation field values inside the GM ROI to zero-valued voxels at the edge of the skull-strip mask or medial lines (Figure 4.2L/M). Let $\vec{f}(p_{i,0})$ be the value of the deformation field at a point $p_{i,0}$ that lies on the interface between the ROI and the blur mask. Let $l(p_{i,0})$ be the line oriented normal to the GM/CSF interface at $p_{i,0}$ and that extends from $p_{i,0}$ to another point $p_{i,m}$ on the opposite edge of the blur mask (i.e., either at the edge of the skull-strip mask or on a medial line). Along $l(p_{i,0})$, we sample equally spaced points $p_{i,1}, \ldots, p_{i,m-1}$. Each of the $m+1$ points on $l(p_{i,0})$ is then associated with a vector $\vec{f}'(p_{i,j})$ such that

$$\vec{f}'(p_{i,0}) = \frac{m-j}{m}\vec{f}(p_{i,0}) \quad (0 \leq j \leq m) \tag{4.1}$$

This yields a series of vectors that decrease linearly from $\vec{f}'(p_{i,0}) = \vec{f}(p_{i,0})$ to $\vec{f}'(p_{i,m}) = \vec{0}$, which we calculate for each line across the entire interface between the ROI and the blur mask (Figure 4.2L). Using vectors from these series, we can interpolate field values $\vec{f}(\mathbf{x_{BM}})$ at each voxel $\mathbf{x_{BM}}$ inside the blur mask (Figure 4.2M). If we define $p^1_{i,j} \ldots p^n_{i,j}$ as the n closest points to the center of $\mathbf{x_{BM}}$, and $\vec{f}'(p^1_{i,j}) \ldots \vec{f}'(p^n_{i,j})$ as their corresponding vectors, then we can express the interpolated value $\vec{f}(\mathbf{x_{BM}})$ as

$$\vec{f}(\mathbf{x_{BM}}) = \frac{1}{n}\sum_{k=1}^{n} d(\mathbf{x_{BM}}, p^k_{i,j}) \cdot \vec{f}'(p^k_{i,j}) \tag{4.2}$$

Here, $d(x_{BM}, p^k_{i,j})$ denotes the distance between the center of $\mathbf{x_{BM}}$ and the point $p^k_{i,j}$. We require that each $p^k_{i,j}$ lie on a unique $l(p_{i,0})$; even if the center of $\mathbf{x_{BM}}$ is closest to multiple points on the same line, $\vec{f}(\mathbf{x_{BM}})$ will still be calculated using one point from n distinct $l(p_{i,0})$. In our experiments, we use $n = 4$. This entire process of obtaining the deformation field by registering the original full brain mask to its atrophied counterpart, followed by masking and smoothing the field with a custom blur mask, is detailed in Figure 4.2.

Masking and smoothing the field in this way offers several advantages over other techniques, such as a simple Gaussian blur. Firstly, we create a smooth transition between the deformed and original image regions; this ensures that the boundary of the GM is still deformed even if the edge of the ROI does not quite extend to the edge of the GM in the input T1w image (e.g., due to inaccuracies in input FreeSurfer parcellation). Constraining the blur prevents the transformation from extending into and deforming neighboring regions, which would result in the GM expanding within those areas. Finally, leaving the deformation field inside the ROI unchanged ensures that the deformation performs as expected.

After calculating the final deformation field with this custom blur operation, we apply it to the original T1w image to artificially induce localized cortical atrophy. This yields a set of two timepoints with known changes at each location in the images. Any other images, such as FLAIR or T2w, can be co-registered to the T1w image and undergo the same deformation to induce the same synthetic atrophy in these modalities.

4.3.3 Cortical Surface Thickness Change

After creating our set of images, where one timepoint is the original data and the other synthetically atrophied, our next step is to quantify the true change induced in CTh. By construction, our method induces surface erosion perpendicular to the surface, whereas the direction of thickness measurement might be at a slight angle based on the local cortical geometry. Because of this discrepancy, we expect that the true change in thickness will be less than the effective size of the erosion kernel. Thus, to determine the localized, ground truth changes in CTh, we create surface representations for each timepoint and measure the corresponding difference in thickness in the deformed region. We obtain these surfaces by performing the 3D Slicer [50] implementation of the marching cubes algorithm[3] [110] on the WM and whole brain GM masks (rather than just the ROI) of the original image. We use a smoothing factor of 10 and 0% decimation for input parameters, and then remove any topological defects in the surface such as holes or handles [81]. This resulting surfaces are warped with the same deformation field used to transform the image to yield corresponding surfaces for the atrophied timepoint. Although the marching cubes algorithm does not necessarily produce a topologically accurate mesh representation, we employ this technique rather than a specific cortical surface reconstruction pipeline (such as FreeSurfer) so that our results are not biased towards any specific reconstruction method.

Finally, we define the ground truth change in CTh as the average between the difference between the distances between the original and atrophied surfaces. Let these be denoted as $S_{i,\text{orig}}$ and $S_{i,\text{atrp}}$, respectively, where $i \in \{\text{GM}, \text{WM}\}$. For each vertex on both original surface, $v_{i,\text{orig}} \in S_{i,\text{orig}}$, we calculate the shortest, signed distance from that vertex to a point on the corresponding atrophied surface $p_{i,\text{atrp}} \in S_{i,\text{atrp}}$. Note that the point p can lie anywhere on the atrophied surface and is not necessarily coincident with one of its vertices v. Next, we average these distances across the entire ROI r to yield a mean surface displacement value, and take the difference between the mean GM displacement and the mean WM surface. We then repeat these calculations starting at each point on the atrophied surfaces and finding the shortest, signed distances to their original counterparts. The true change in thickness is thus defined as the average between the difference in surface displacements travelling from the original to atrophied surfaces, and the different in displacements travelling from the atrophied to the original. We call this definition the Mean Surface Displacement Difference (MSDD), denoted D_{MSDD} and described in equation 4.3.

[3] https://www.slicer.org/wiki/Documentation/4.3/Modules/ModelMaker

$$D_{\mathrm{MSDD}}(r) = \frac{1}{2}\left[\frac{1}{n_{r,\mathrm{GM}}}\left(\sum_j D_{\min}(v_{\mathrm{GM,orig}}, p_{\mathrm{GM,orig}}) + \sum_j D_{\min}(v_{\mathrm{GM,atrp}}, p_{\mathrm{GM,atrp}})\right) - \right.$$
$$\left. \frac{1}{n_{r,\mathrm{WM}}}\left(\sum_j D_{\min}(v_{\mathrm{WM,orig}}, p_{\mathrm{WM,orig}}) + \sum_j D_{\min}(v_{\mathrm{WM,atrp}}, p_{\mathrm{WM,atrp}})\right)\right] \quad (4.3)$$

where $1 < j < n_{r,i}$ iterates over the number of vertices on each surface $n_{r,i}$ within the ROI. Note that because the MSDD does not yield a bijective mapping between surfaces as discussed in Section 2.4.1, it is calculated over an entire region rather than at a single vertex. In this analysis, all mesh warping, topological corrections, cortical surface parcellations, and thickness measurements were obtained using VTK[4].

4.3.4 Longitudinal Cortical Segmentation Accuracy

After establishing a method for measuring the exact induced thickness change, we needed to demonstrate how data from our methods can be employed to quantify segmentation accuracy as well in the presence of cortical atrophy. We also wanted to determine whether cortical segmentation pipelines can segment our synthetic images well or whether they encounter issues such as local blurring of boundaries due to interpolation. We accomplished this by inducing synthetic atrophy in the VDCRA dataset; we deform both the images and the landmarks themselves to match the new cortical boundaries. We used the original and deformed landmarks sets to measure the accuracy of FS cortical surface reconstructions of the respective timepoints.

Because we were interested in measuring cortical accuracy only at the fiducial landmarks, we created a unique cortical parcellation for each subject based on the locations of its landmark clusters. These label maps contain ROIs that are each centered around a single cluster of GM/WM landmarks. Figure 4.3 illustrates the process of creating a single ROI for one cluster of the left superior temporal (LST) region (Figure 4.3C). The ROI was defined by first filling in the voxels within the image that lie between pairs of corresponding GM and WM landmarks (Figure 4.3D) . Because each cluster of landmarks lies within a single axial slice, this step yields a 1-voxel-thick mask with no extra voxels surrounding the landmarks. Next, we wanted to pad this initial ROI so that the deformation moving the landmarks was not affected by boundary effects. To accomplish this, we iterated between (1) dilating this initial mask outward with a $1\times1\times1$ voxel kernel and (2) applying a cortical ribbon mask to constrain the ROI to within the GM (Figure 4.3E−J). In other words, we used voxels between corresponding GM and WM landmarks as a starting seed, and flooded the ROI outwards within the cortical ribbon. This added a buffer region on either side of the landmark cluster, and of 3 axial slices on both sides of the initial slice; this also removed any holes existing within the original ROI. This

[4]https://vtk.org/

iterative dilation and masking procedure sometimes resulted in the ROI expanding into adjacent tissue; to alleviate this, each ROI was manually edited using 3D Slicer[5] to remove any voxels residing in adjacent gyri, and to ensure smoothness between axial slices (Figure 4.3J→K). Using this procedure, we generated a total of 42 ROIs (7 landmark groups × 3 clusters per group × 2 hemispheres).

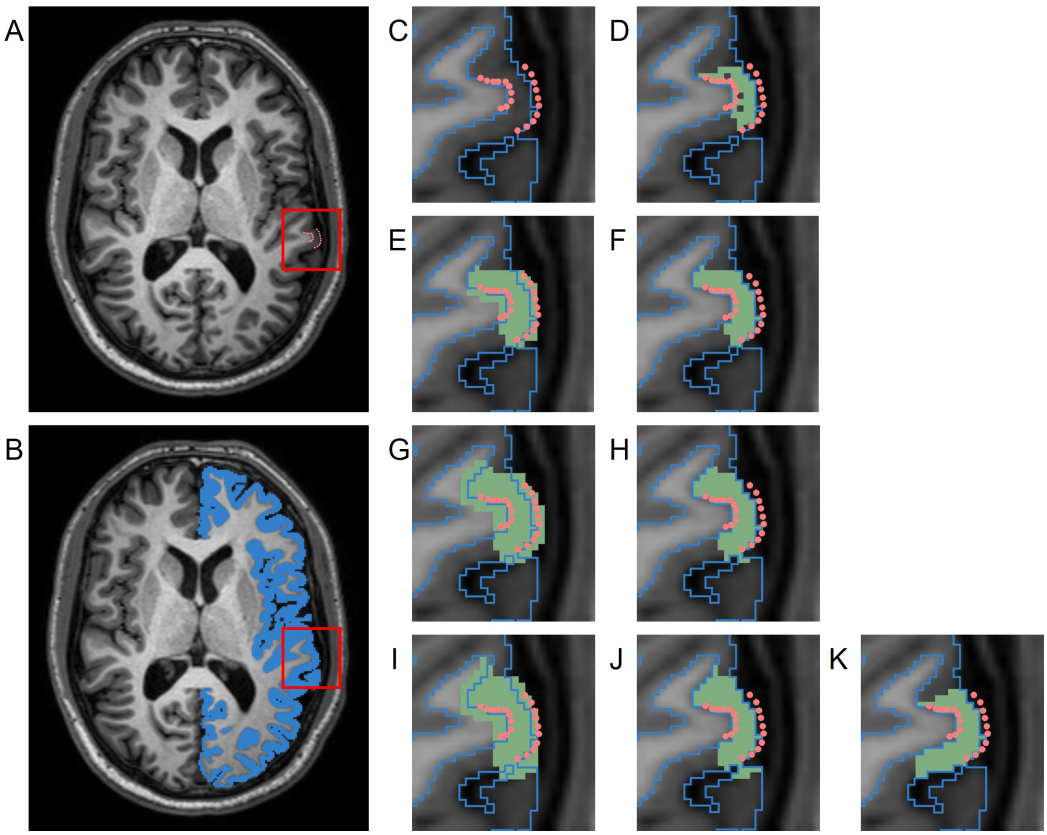

Figure 4.3: Schematic of pipeline to produce a label map specific to the left superior temporal (LST) fiducial landmarks. This data belongs to a subject from the HC cohort of the JHU cortical validation dataset. **(A)** T1w image with manual fiducial landmarks (pink). **(B)** T1w image overlayed with a mask of the left cortical GM ribbon (blue). **(C)** Close up of landmarks in T1w image overlayed with GM ribbon mask outline. **(D)** Initial ROI (green) for image and landmark deformation, obtained by filling in voxels between corresponding pairs of WM and GM landmarks. **(E)** ROI dilated with $1\times1\times1$ voxel kernel (first iteration). **(F)** ROI masked with cortical GM ribbon (first iteration). **(G)** Dilated ROI (second iteration). **(H)** Masked ROI (second iteration). **(I)** Dilated ROI (third iteration). **(J)** Masked ROI (third iteration). **(K)** Final ROI after manual cleanup of adjacent gyri.

After manual editing, we induced synthetic atrophy in each ROI with 2 iterations of the binary morphology operations shown in Figure 4.1. We used 2 iterations to induce consistent amounts of change in each region, and to make sure that at no location in any ROI the GM is fully eroded, as this would not occur during actual neurodegeneration. With the eroded atrophy target, we performed the registration step using the Greedy software with the same parameters as before (four levels, 100/100/50/100 iterations per level,

[5]https://slicer.readthedocs.io/en/latest/user_guide/modules/segmenteditor.html

and mean squared difference as the similarity metric). We calculated these individual deformation fields for all 42 ROIs, and combined them into a single deformation field. Let $T_j (1 \leq j \leq 42)$ denote these individual transformations, and let C denote the composite field. Let $\vec{v}_{x,j}$ be the value of the deformation field T_j at voxel x. There were three possible scenarios for the corresponding $\vec{v}_{x,C}$ in C:

1. If $\vec{v}_{x,j} = \vec{0}, \forall j$, then $\vec{v}_{x,C} = \vec{0}$.
2. If there exists a unique i such that $\vec{v}_{x,i} \neq \vec{0}$ and $\forall_j[(j \neq i) \rightarrow (\vec{v}_{x,j} = \vec{0})]$, which happens when x is within a single non-overlapping ROI, then $\vec{v}_{x,C} = \vec{v}_{x,i}$.
3. If there exist multiple non-zero valued $\vec{v}_{x,j}$, which can happen because the blurred regions are not necessarily non-overlapping, then, $\vec{v}_{x,C}$ is calculated as the average of all non-zero valued $\vec{v}_{x,j}$ at voxel x.

These operations were all computed using ITK.

Finally, the composite field C was applied to the T1w and FLAIR images and to the associated fiducial landmarks. Here, the images were deformed using the Greedy software and the landmarks with VTK. This created a set of images accompanied by cortical landmarks with exact correspondence between timepoints, which we could use to evaluate the accuracy of cortical surface reconstruction methods in a longitudinal setting.

4.3.5 Validation Experiments

4.3.5.1 Localization of Cortical Atrophy

The goal of our first experiment was to determine if the transformation was indeed constrained as desired, or if it bled into the surrounding voxels and deforms the GM outside the ROI. To investigate this, after calculating thickness at each vertex, we computed a mean thickness value for 3 different cortical regions: inside the atrophied ROI, within a 4×4×4 voxel neighborhood surrounding the ROI, and everywhere else on the surface. Any significant change in thickness detected within the neighborhood surrounding the ROI would have indicated that our transformation was not constrained to the target ROI as intended. We separated this surrounding neighborhood because otherwise, any measured changes would have been smoothed away if averaged together with the entire brain outside the target ROI. Figure 4.6C displays an example of this label map projected onto a cortical surface. Finally, changes in thickness were computed as the difference in mean thickness within each of the three regions. By taking the difference of the average thickness within each region instead of the average of the difference across pairs of vertices, we allowed for the application of this method to CTh pipelines that may not guarantee vertex-vise correspondence between timepoints.

4.3.5.2 Extent of Induced Localized Cortical Atrophy

Although we can fully erode the GM ROI using the series of binary morphology operations described in Figure 4.1, the regularization terms used in the subsequent deformable registration step will likely not allow for the total collapse of the GM ribbon. Thus, our next experiment aimed to determine the extent to which we could induce localized atrophy in ROIs throughout the entire cortex in practice. We tested our pipeline in each of the 62 cortical regions (33 per hemisphere) in the DK atlas by varying the amount of target (intended) atrophy induced in each ROI from 1 to 12 voxels in the upsampled mask images. Because these upsampled masks had a resolution of $0.3 \times 0.25 \times 0.25\text{mm}^3$, each iteration induced approximately $\sqrt{0.3^2 + 0.25^2 + 0.25^2} = 0.46\text{mm}$ of thickness change, resulting in a total change of about 5.5mm over 12 iterations. Note that this was intended to result in complete atrophy within the cortical region (which would not be expected to occur in a realistic dataset, except possibly in surgical removal scenarios) to test the limits of the atrophy pipeline. For each iteration of atrophy, we measured the mean change in CTh using our MSDD definition of thickness change within each cortical ROI; this allowed us to assess how the CTh changes in individual ROIs with each iteration and the maximum extent to which we could synthetically induce localized atrophy. We also measured these changes with the symmetric closest point distance formula (D_{SCP}), which defines CTh at each vertex as the average of the distances between (1) the initial GM vertex and the closest point on the WM surface, and (2) that point on the WM surface and its closest point on the GM surface (see equation 2.5, Section 2.4.1). Note that similar to the MSDD, other than the initial vertex, these points were not constrained to vertices and could fall anywhere on the GM or WM surfaces. By comparing changes in CTh measured with D_{MSDD} and D_{SCP}, we determined how our results matched those obtained with an established and widely used technique.

4.3.5.3 Effect on Longitudinal Cortical Segmentation Accuracy

In this experiment, we aimed to determine the usability of our methods in the context of validating the accuracy of the cortical segmentation methods. We explored this by inducing synthetic atrophy in the VDCRA dataset using the custom cortical parcellations generated from the landmarks associated with each subject. We used both the cross-sectional and longitudinal workflows of FreeSurfer to generate cortical surface representations of the original and atrophied timepoints, and measured the segmentation error of the resulting surfaces with respect to landmark placement. In the cross-sectional pipeline [38], each timepoint was segmented independently; in the longitudinal pipeline [141], FS first combined all timepoints to create a subject-specific template and then separately segmented each timepoint using this template for initialization. We hypothesized that if our synthetic atrophy methods were significantly corrupting the data (e.g. with blurring artifacts), this would result in more error in the jointly initialized longitudinal pipeline compared to the cross-sectional

pipeline.

After creating the synthetic dataset and surface reconstructions, we calculated the unsigned and signed segmentation errors for each landmark cluster. These errors were obtained by measuring the average minimum distance from each of the 30 landmarks within the cluster to the corresponding cortical surface. The sign of the distance was positive if the landmark resided inside the mask, and negative if outside. We expected that FS would perform worse in atrophied data than in healthy data, regardless of whether the atrophy was synthetic or naturally occurring. To test this, we compared segmentation error between the VDCRA-HC and VDCRA-MS cohorts, as the MS subjects presented varying degrees of (natural) cortical atrophy. If there was no significant difference in FS error between synthetically atrophied VDCRA-HC and the original VDCRA-MS data, then we could conclude that our method did not induce significant artifacts or error in the images or landmark placements. We used a one-way ANOVA ($\alpha = 0.05$) to test for significant differences between errors for the original and synthetic timepoints for the VDCRA-HC and VDCRA-MS groups (4 cohorts per test). These tests were conducted separately for each region (7 regions $\times$ 2 hemispheres $\times$ 2 surfaces, for a total of 28), expert (A or B), and error type (unsigned or signed). We conducted a similar set of ANOVA tests for these data to identify significant differences between the two processing types, and between the two sets of experts (4 sets total). These tests were conducted separately for each region (28 total), data type (original of synthetic), subject type (HC or MS), and error type (unsigned or signed).

4.4 Results

4.4.1 Qualitative Evaluation of Atrophied Images

Figure 4.4 displays the axial view of an example set of T1w and FLAIR images. The original images are from the NITRC Kirby dataset (Figure 4.4A/C, and atrophy was synthetically induced (Figure 4.4B/D) in the left superior temporal gyrus (LSTG) (Figure 4.4K). In this example, we used an effective erosion kernel of $\tau = 4$ (4×4×4 voxels), corresponding to 4 iterations of binary morphology operations, to compute the synthetic atrophy deformation field. We observe a noticeable decrease in CTh between the two timepoints (see Figure 4.4E through H for close-ups), with no significant changes in the GM surrounding LSTG, or at the GM/WM interface. This visually apparent localization is further supported by the difference image between the original and atrophied timepoints for each modality (Figure 4.4I/J). In these images, any zero-valued voxel is displayed as transparent. Thus, we see that the deformation only affected the ROI and the mask used during the blurring process (Figure 4.2F). We also see that the atrophied FLAIR image corresponds well to its T1w counterpart, indicating that we can apply the deformation obtained from a T1w image to atrophy additional modalities associated with the subject, as expected.

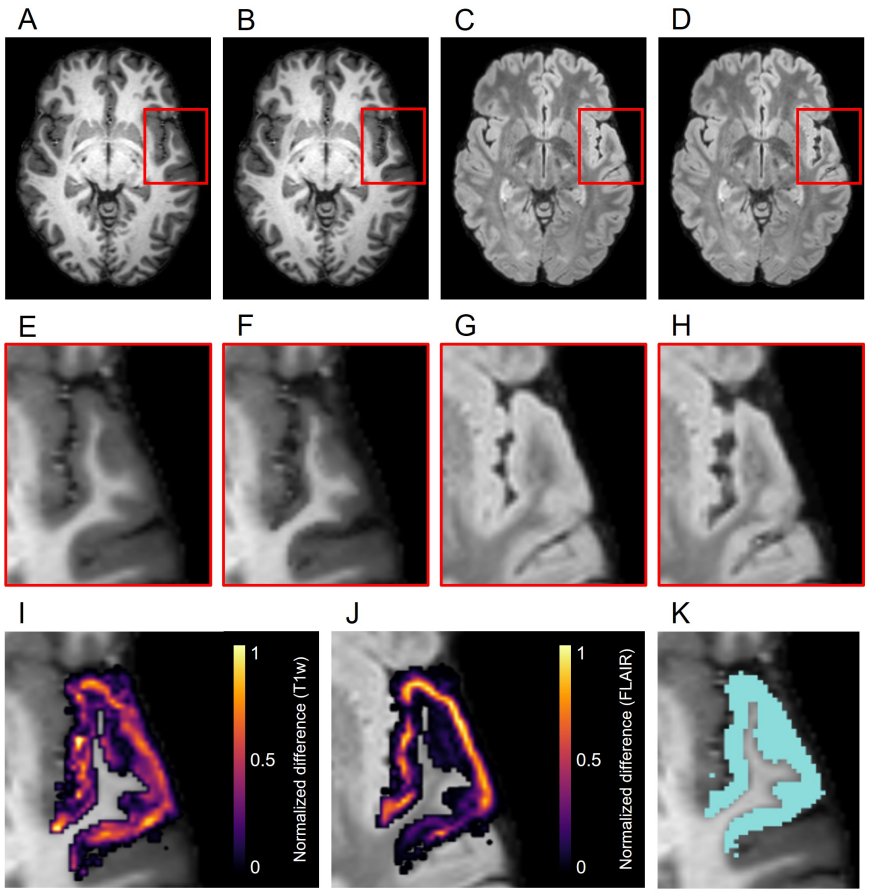

Figure 4.4: Example results of the synthetic atrophy pipeline using the NITRC Kirby dataset and four iterations of binary morphology operations to induce atrophy. **(A)** Skull-stripped original T1w image. **(B)** Skull-stripped T1w image with synthetically induced atrophy. **(C)** Skull-stripped original FLAIR image. **(D)** Skull-stripped FLAIR image with synthetically induced atrophy. **(E-H)** Close-up of ROI (LSTG) in the images of the first row. **(I)** Difference image between the original and atrophied T1w timepoints (E-F) overlayed on the original T1w image. **(J)** Difference image between the original and atrophied FLAIR timepoints (G-H) overlayed on the original FLAIR image. For both difference images, all zero-valued voxels are rendered transparent. **(K)** Close up of ROI (blue) overlayed onto original T1w image.

4.4.2 Localization and Extent of Cortical Atrophy

Figure 4.5 shows the relationship between the induced thickness changes within the LSTG of our marching cubes generated surfaces and the number of atrophy iterations. The colors of the data in panels 4.5A−C correspond to the labels overlaid as a cortical parcellation onto an example marching cubes surface in 4.5D: inside the LSTG (pink), within the surrounding $4 \times 4 \times 4$ voxel dilation neighborhood (yellow), and across the rest of the cortex (green). Panels 4.5A and 4.5B show the average thickness change across all subjects measured using the mean surface displacement difference (MSDD) and symmetric closest point (SCP) distance formulas, respectively, while panel 4.5C shows the actual SCP thickness. The MSDD yielded a higher change in thickness inside the ROI (pink) than the SCP, but a lower change in the surrounding region (yel-

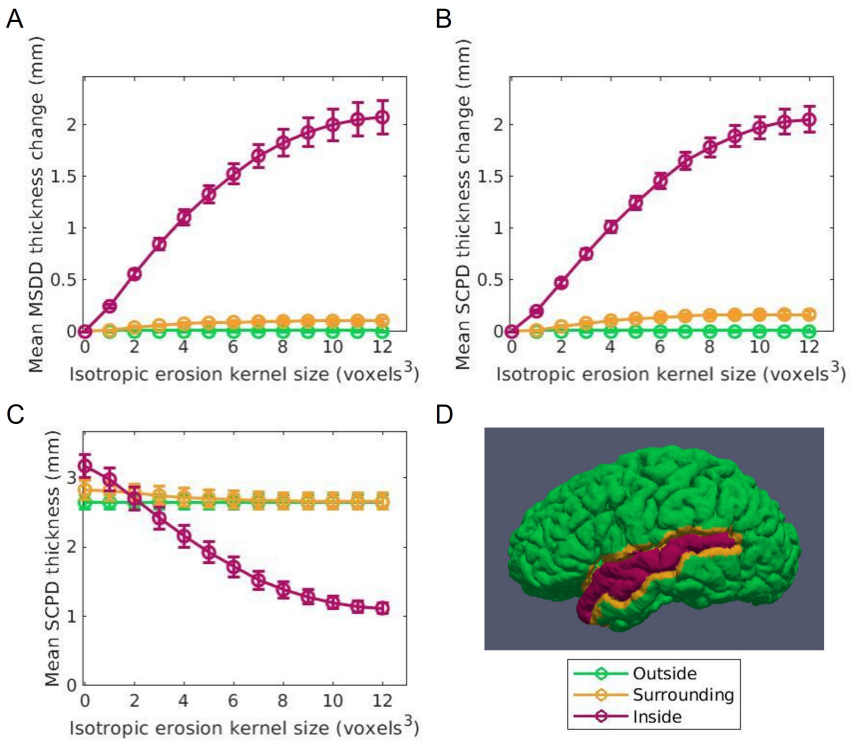

Figure 4.5: Thickness change averaged across subjects between the original and synthetic atrophied time-points as a function of isotropic erosion kernel size in the 3 regions studied for the left superior temporal gyrus (LSTG) ROI. Data shown in pink corresponds to voxels inside the ROI is shown in pink, in yellow to the $4 \times 4 \times 4$ voxel dilation neighborhood, and in green to the rest of the cortex. **(A)** Mean thickness change measured using the mean surface displacement difference (MSDD). **(B)** Mean thickness measured using the symmetric closest point (SCP) distance. **(C)** Mean change in thickness measured with SCP. These values closely match the MSDD changes measured with shown in panel A. **(D)** Example label map of a single subject in the Kirby dataset showing the vertices inside the ROI, surrounding the ROI, and outside the ROI. These labels are projected onto a cortical surface generated using the marching cubes algorithm on the full brain mask corresponding to that subject.

low). This is likely because the SCP maps a vertex on the GM surface to a vertex on the WM, and then maps that vertex to a second GM vertex. If the second GM vertex corresponds to the surrounding region while first vertex corresponds to the desired ROI, then reported thickness change will be less than if all three vertices were constrained to the ROI. Likewise, if the second corresponds to the ROI while the first corresponds to the surrounding area, the reported change will be higher. This discrepancy between measurements inside the ROI (pink) also agrees with a finding in a previous CTh study that the SCP may underestimate thickness compared to alternate surface-based methods [123]. The MSDD and SCP both found that thickness within the surrounding (yellow) region remains relatively stable as erosion kernel size increases, the MSDD more so than the SCP; this indicates that the atrophy induced by our methods is highly localized and constrained to the desired region. These thickness changes seem unaffected by the underestimation tendencies of the SCP,

most likely because they are much less pronounced than inside the ROI (pink). Finally, as expected, both methods report no change in thickness across the rest of the cortex (green). These curves shown in Figure 4.5 serve as a representative example for the atrophy trends across the entire cortex.

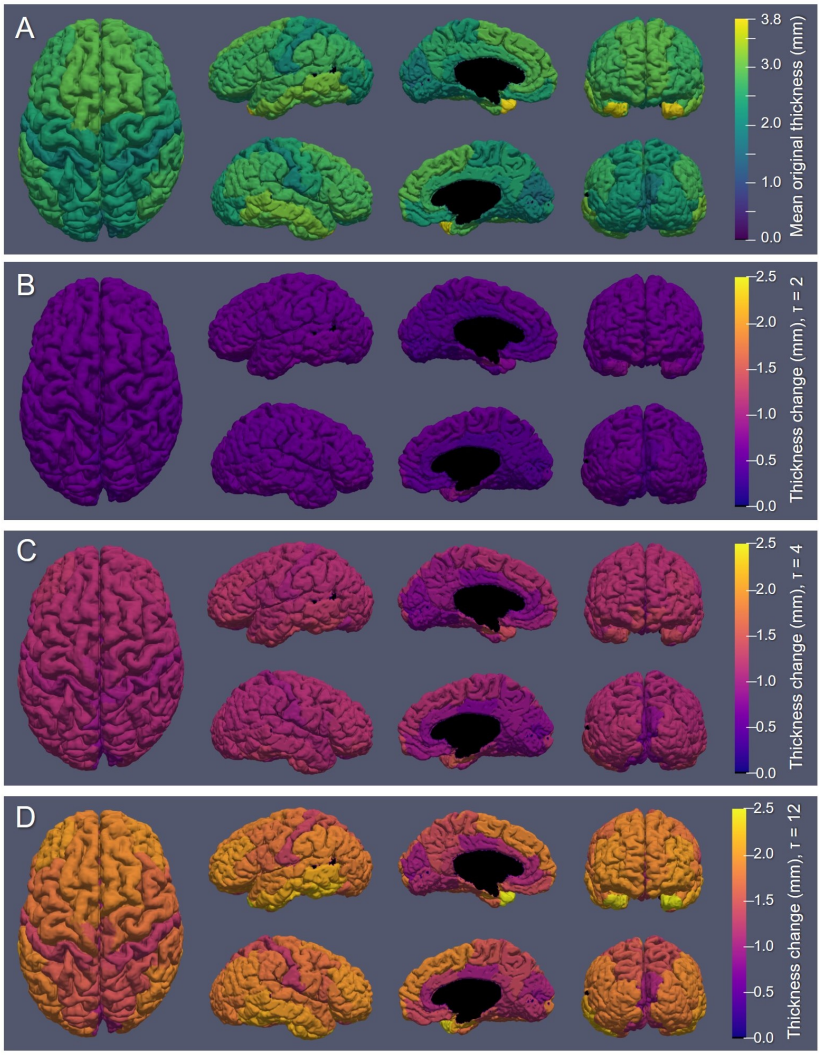

Figure 4.6: Synthetic atrophy results averaged across subjects mapped onto an example cortical GM surface obtained from the original timepoint using the marching cubes algorithm. **(A)** Mean original thickness for each region measuring using the symmetric closest point (SCP) formula. **(B)** Thickness change measured with using the mean surface displacement distances (MSDD) after $\tau = 2$ iterations of erosion. **(C)** MSDD thickness change after $\tau = 4$ iterations of erosion. **(D)** MSDD thickness change after $\tau = 12$ iterations of erosion, showing the maximum amount we can atrophy in each region.

Figure 4.6 displays the atrophy amount in each ROI within the DK atlas projected onto the cortical surface (Figure 4.6) for several different kernel sizes. Specifically, panel 4.6A displays the mean original thickness values for each region, while panels 4.6B, 4.6C, and 4.6D show the mean change in CTh for $\tau = 2$ (used to produce the data in Figure 4.7), $\tau = 4$ (used to produce the data in Figure 4.4), and $\tau = 12$, respectively. These kernel sizes correspond to approximately 0.9 mm, 1.8 mm, and 5.5 mm of intended thickness change.

We observe that for $\tau = 2$ (Figure 4.6B), there exists roughly a 0.6 mm change in thickness across the entire cortex with little variation between regions. As τ increases, more disparities in the atrophy amount occur between ROIs consistent with their original thickness. In other words, the higher the original thickness, the more iterations we can keep applying synthetic atrophy before we reach the limit. This data is also displayed numerically in Table A.1 within Appendix A, along with the SCP thicknesses of each ROI in the original and maximally atrophied ($\tau = 12$) timepoints.

4.4.3 Qualitative Evaluation of Cortical Validation Data

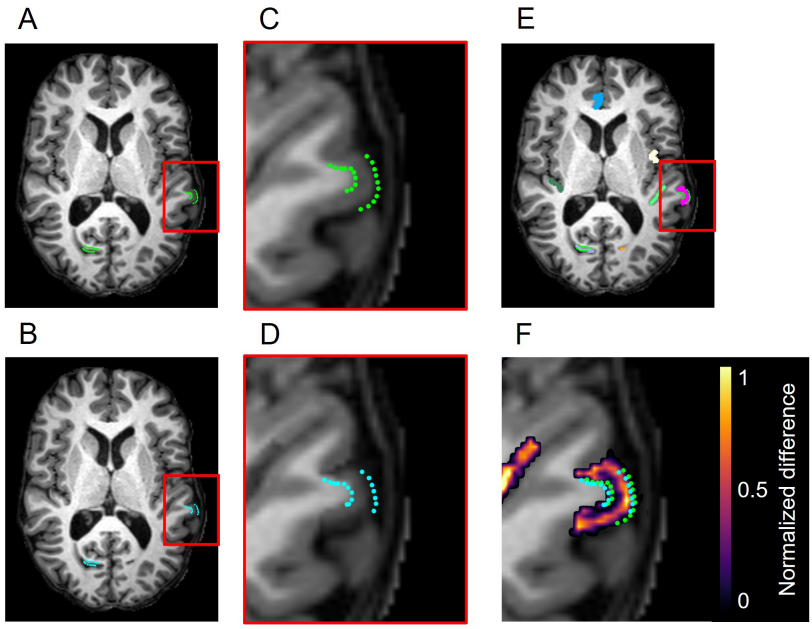

Figure 4.7: Example results from the accuracy validation for longitudinal cortical segmentation using the JHU Cortical Validation dataset and two iterations of binary morphology operations to induce atrophy. **(A)** Skull-stripped original T1w image, overlaid with original GM and WM landmarks (green). Two clusters of landmark pairs are visible in this slice: the right calcarine fissue (RCALC cluster) and the left superior temporal gyrus (LST cluster). **(B)** Skull-stripped T1w image with synthetically induced atrophy, overlaid with deformed GM and WM landmarks (blue). Again, the same RCALC and LST landmark clusters are visible in this slice. **(C)** Close-up of LST ROI selected for deformation and its landmark cluster. **(D)** Close-up of deformed LST ROI and landmark cluster. **(E)** Multi-color label map overlayed onto original T1w image, with the LST ROI in pink. **(F)** Difference image thresholded to display only non-zero voxels, as well as the original and deformed landmarks, overlaid on the original image. Note that other deformed ROIs can also be observed in this panel that are associated with additional clusters of landmarks, but that those sets are not visible in this slice.

Figure 4.7 illustrates, in addition to the conclusions stated in section 4.4.1, that the deformation field can be used to translate the landmarks from their original position to the new GM/CSF interface after synthetic atrophy. In this example, we used an effective erosion kernel of $\tau = 2$ ($2 \times 2 \times 2$ voxels), corresponding to 2 iterations of binary morphology operations, to compute the synthetic atrophy deformation field. The original

and deformed sets of left superior temporal (LST) landmarks are visible within the difference image in panel 4.7F. We note that the number of visible landmarks decreases between panels 4.7A/C and 4.7B/D because the deformation field displaces the fiducials in all 3 dimensions, thus moving some landmarks to a different axial slice.

4.4.4 Effect on Longitudinal Cortical Segmentation Accuracy

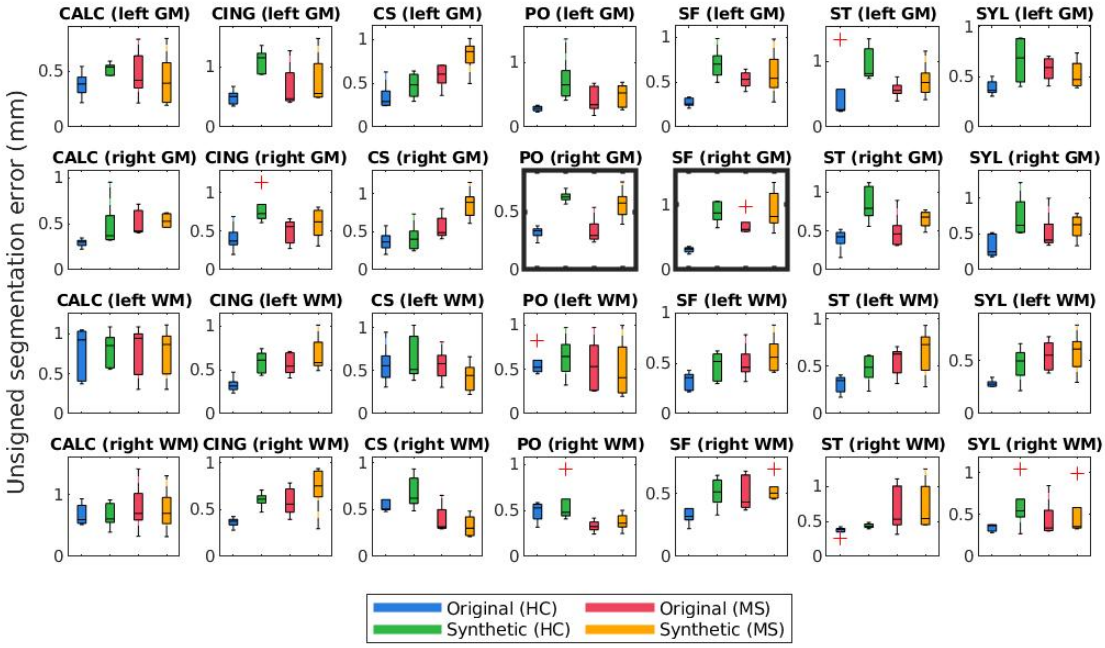

Figure 4.8: Mean unsigned segmentation errors of longitudinal FreeSurfer cortical surface reconstructions for each set of landmarks placed by expert A. Subplot rows correspond to surface and hemisphere while the subplot columns correspond to the landmarks' anatomical placements. Within each subplot, mean segmentation errors from the original images of healthy subjects are shown in blue, synthetic images of healthy subjects in green, original images of MS subjects in red, and synthetic images from MS subjects in yellow. A panel outlined in bold indicates that the statistical analysis of the associated data yielded at least one significant difference between the four groups.

The bolded panels in Figures 4.8 and 4.9 indicate the one-way ANOVA tests that yielded $p < 0.05$ after applying a Bonferroni correction of $n = 56$ (28 sets of landmarks $\times$ 2 types of errors). This analysis was also conducted on the entire dataset: first with no correction factor and then with a correction factor of $n = 224$ (2 processes (cross-sectional vs. longitudinal) $\times$ 2 experts $\times$ 28 sets of landmarks $\times$ 2 types of errors). Table 4.1 details the percent of significant tests for each comparison prior to Bonferroni correction. After correction, less than 1% of all relevant tests results are significant. In this context, we define "relevant" tests as those between the original HC and synthetic HC cohorts, the synthetic HC and original MS cohorts, and the original MS and synthetic MS cohorts. By default, the multiple comparisons analyses also compared the original HC and original MS cohorts, and the synthetic HC and synthetic MS cohorts. However, these are

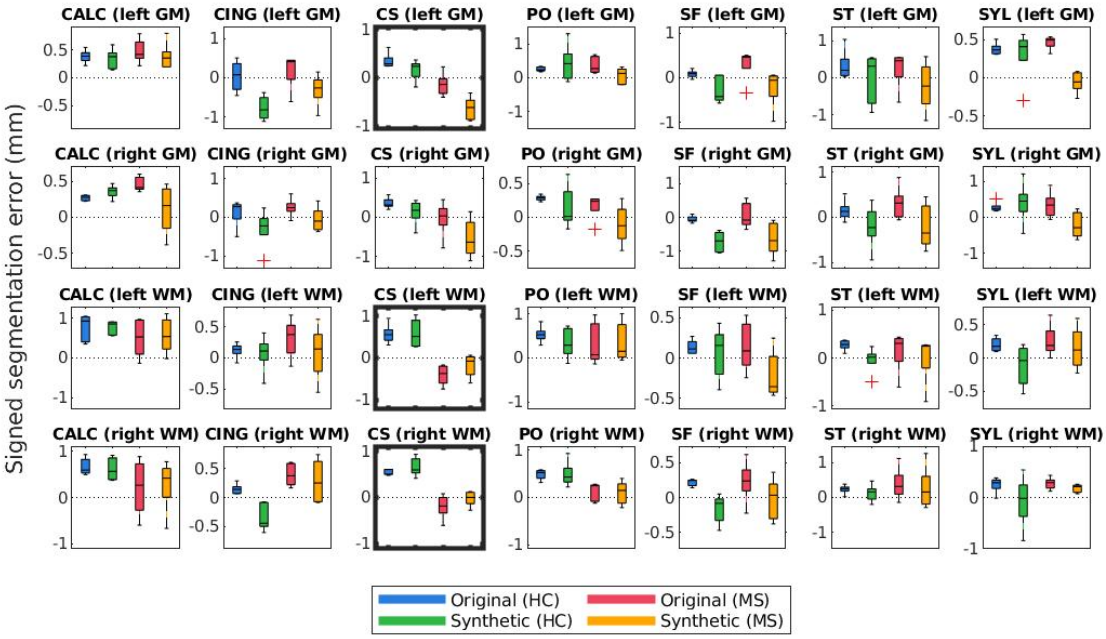

Figure 4.9: Mean signed segmentation errors of longitudinal FreeSurfer cortical surface reconstructions for each set of landmarks placed by expert A. Subplot rows correspond to surface and hemisphere while the subplot columns correspond to the landmarks' anatomical placements. Within each subplot, mean segmentation errors from the original images of healthy subjects are shown in blue, synthetic images of healthy subjects in green, original images of MS subjects in red, and synthetic images from MS subjects in yellow. A panel outlined in bold indicates that the statistical analysis of the associated data yielded at least one significant difference between the four groups.

deemed irrelevant to our study because we are not investigating the HC-MS group differences, and therefore we omitted these from the data in Table 4.1. In summary, there were a total of 672 statistical tests (2 processes $\times$ 2 experts $\times$ 28 sets of landmarks $\times$ 3 relevant comparisons $\times$ 2 types of errors).

	HC (O) vs. HC (S)	**HV (S) vs. MS (O)**	**MS (O) vs. MS (S)**	**Total**
Unsigned	16.96 % (19/112)	6.25 % (7/112)	2.79 % (2/112)	8.33 % (28/336)
Signed	8.93 % (10/112)	18.75 % (21/112)	8.93 % (10/112)	12.20 % (41/336)
Total	12.95 % (29/224)	12.5 % (28/672)	5.35 % (12/672)	10.27 % (69/672)

Table 4.1: Percent of tests resulting in significant differences for unsigned and signed errors of FreeSurfer segmentation results *before* applying Bonferroni correction for multiple comparisons. In parenthesis are the raw number of tests with $p<0.05$. A total of 336 statistical tests using a one-way ANOVA with multiple comparisons were performed for each type of error (2 processes $\times$ 2 experts $\times$ 28 landmarks $\times$ 3 columns below). In the column titles, (O) denotes the original timepoint, and (S) the synthetic.

Lastly, Figures 4.10 and 4.11 compare mean unsigned and signed errors corresponding to each expert and process type (cross-sectional vs. longitudinal FreeSurfer workflows). The row and column structures of these figures match those of Figures 4.8 and 4.9. For brevity, we show only the results for the synthetic HC data. Note that the light green data bars in Figures 4.10 and 4.11 show the same data as the darker green

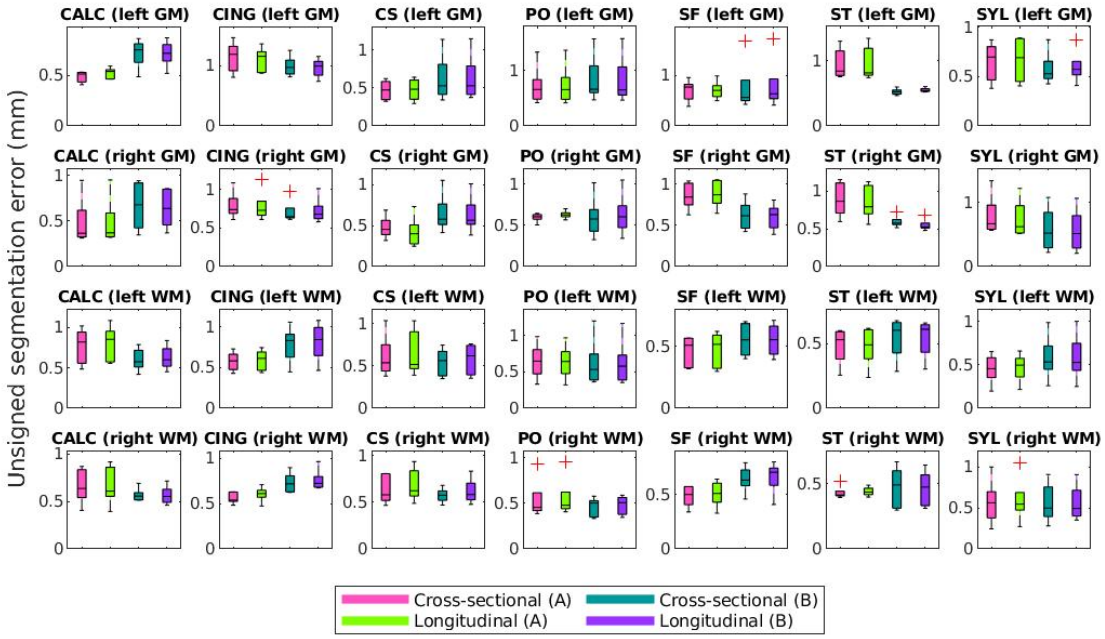

Figure 4.10: Mean unsigned segmentation errors of FreeSurfer cortical surface reconstructions of the synthetic healthy control cohort. Subplot rows correspond to surface and hemisphere while the subplot columns correspond to the landmarks' anatomical placements. Within each subplot, mean errors from the cross-sectional pipeline and measured with landmarks from expert A are in purple, the longitudinal pipeline with expert A in teal, cross-sectional with expert B in green, and longitudinal with expert B in pink. A panel outlined in bold indicates that the statistical analysis of the associated data yielded at least one significant difference between the four groups (and no bolded panels indicates no significant differences detected).

bars in Figures 4.8 and 4.9. As before, we conducted similar one-way ANOVA tests for this data, also with a Bonferroni correction of $n = 56$. We found that no significant differences exist between segmentation errors produced by different processes or experts within the same set of images.

The trends identified in the data displayed in Figures 4.8-4.11 represent those present in the full dataset. However, because each figure includes only one out of four total configurations, we include the entire set of results in Appendix A within Figures A.1-A.4. Clustering the data in these two alternative formats allows the reader to more easily observe trends across different methods and data types (e.g. data sorted by cohort and timepoint vs. by processing method and rater). This data is also shown in Tables A.2 and A.3.

4.5 Discussion

4.5.1 Qualitative Evaluation of Atrophied Images and Cortical Validation Data

The synthetic images resulting from our synthetic atrophy pipeline, shown in Figures 4.4 and 4.7, appear visually plausible compared to the original anatomical data. The FLAIR images and fiducial landmarks both deformed in a way that matches their corresponding T1w images, as expected. These results also indicate that

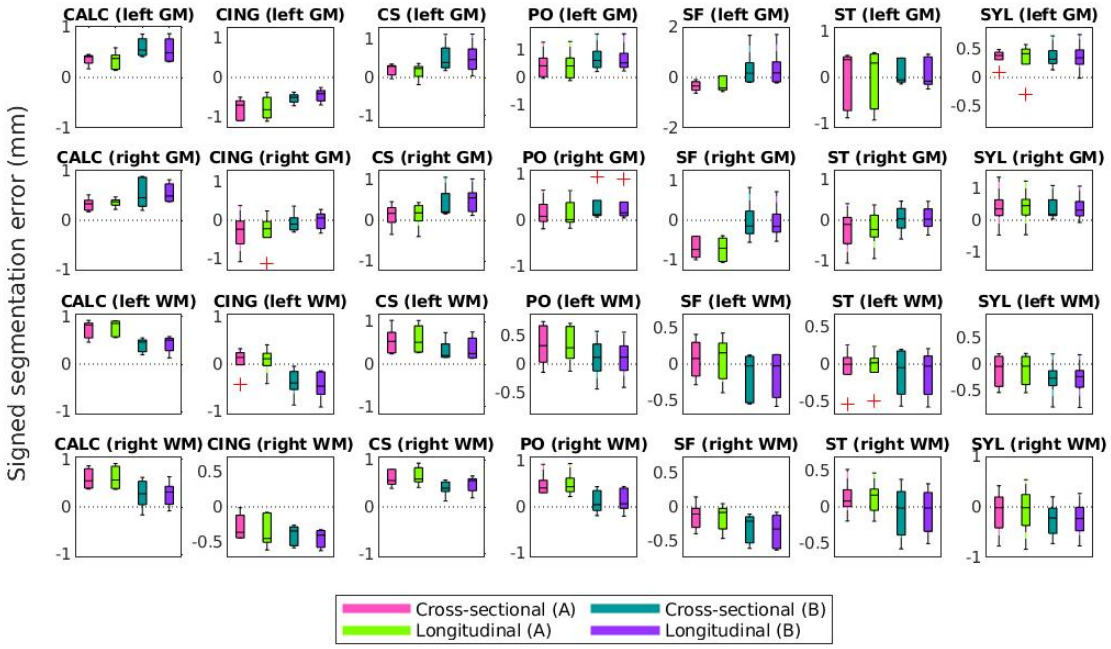

Figure 4.11: Mean signed segmentation errors of FreeSurfer cortical surface reconstructions of the synthetic healthy control cohort. Subplot rows correspond to surface and hemisphere while the subplot columns correspond to the landmarks' anatomical placements. Within each subplot, mean errors from the cross-sectional pipeline and measured with landmarks from expert A are in purple, the longitudinal pipeline with expert A in teal, cross-sectional with expert B in green, and longitudinal with expert B in pink. A panel outlined in bold indicates that the statistical analysis of the associated data yielded at least one significant difference between the four groups (and no bolded panels indicates no significant differences detected).

the presented methods can induce atrophy within a desired region without affecting the surrounding tissue. The T1w difference images in Figures 4.4J and 4.7F show that the two timepoints are identical outside the ROI, as intended.

Although Figures 4.4 and 4.7 indicate our pipeline performed as intended, it is worth noting that they do not necessarily depict cortical atrophy as it would appear naturally. Healthy aging and most neurodegenerative pathologies are often associated with changes in tissue appearance in addition to purely geometrical changes, and often a degree of both GM and WM atrophy as well as ventricular expansion are observed together rather than GM atrophy in isolation. We acknowledge that, because our methods induce changes only to the GM, we generated only an approximation of cortical atrophy. However, our goal in this work was not to create a realistic representation of naturally occurring cortical changes, but to develop a tool used specifically for accuracy validation of cortical segmentation and thickness measurement. We specifically designed our morphology-based atrophy induction to preserve the GM/WM interface so that only the GM/CSF boundary would deform. Geometrically, the WM surface is much less complex than the GM, and its segmentation is therefore an easier task. Moreover, ensuring that we alter only one of the two boundaries required for

measuring CTh changes simplified the problem compared to if both boundaries were significantly deformed between timepoints.

When testing our method in various ROIs within the DK atlas, we found that our method performs best when operating upon a cortical ROI adjacent to CSF clearly visible within the image, such as that within a wide sulcus or the sub-arachnoid space. Because of this, the LSTG proved a perfect example. When inducing atrophy on a single side of a tight sulcus, we found that our pipeline yielded less desirable results. This is because the deformation simultaneously compresses the GM layer and expands the neighboring CSF; however, if there exists no visible CSF in voxels bordering the GM of the ROI, then it simply expands the GM on the other bank of the sulcus. An example of this is shown in Figure 4.12, which displays the synthetic atrophy results in the left fusiform gyrus with four iterations of binary operations to induce atrophy (the same amount as in Figure 4.4). The sulci surrounding this ROI are too narrow and are prone to partial volume effects; because of this, the algorithm cannot expand the CSF within this region but rather fuses the GM from opposing banks of the sulcus. We thus see an artificial increase in CTh in parts of the ROI, rather than the intended decrease. Future work to address this issue involves synthetically inserting CSF voxels to improve the quality of the deformation within such regions. This research is explained in more detail within Chapter 7, and is currently under review at the Medical Imaging with Deep Learning 2023 conference.

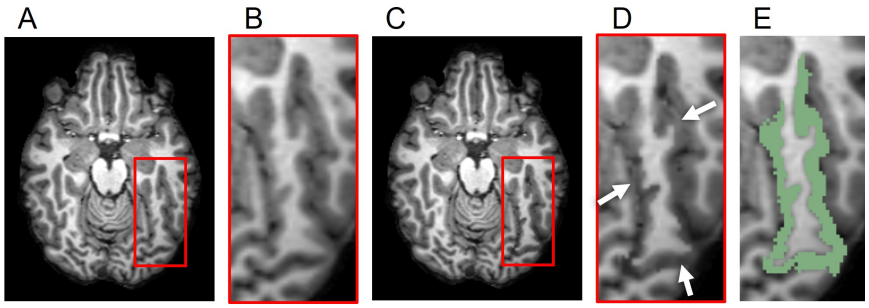

Figure 4.12: Example of less desirable results from our synthetic atrophy pipeline, where the GM surrounding the ROI is expanded in the absence of visible CSF. **(A)** Skull-stripped original T1w image. **(B)** Skull-stripped T1w image with synthetically induced atrophy. **(C)** Close-up of ROI selected for deformation. **(D)** Close-up of atrophied ROI. Arrows point to places in the image where the GM from surrounding gyri is deformed instead of CSF. **(E)** Skull-stripped original T1w image with ROI (left fusiform gyrus) overlayed in green.

4.5.2 Localization and Extent of Cortical Atrophy

Figure 4.5 confirms that, as expected, increasing the effective erosion kernel size with more iterations of binary morphology operations yields a larger change in CT. Because we used high resolution ROIs to produce the deformation, we were able to induce atrophy on a sub-voxel scale, and the change in CTh were less than the resolution of the original images. Further, the thicknesses at vertices outside the ROI boundary remained

stable. This supports what we qualitatively observed in the difference images within Figures 4.4F and 4.7F: the tissue outside the ROI remains unchanged.

In Figure 4.6B, we observe that the limit of atrophy varied between regions but is consistent within corresponding ROIs across hemispheres. That is because this limit is determined by the original thickness of each ROI, as well as the regularization term in the deformable registration, which may not allow for a total collapse of the GM ribbon. This threshold amount—the upper extent of the atrophy we were able to induce within each region—is displayed in Figure 4.6D. For reference, the SCP thickness in each region of the original timepoint (Figure 4.6A) are also included. We induced cortical thinning up to between 0.8 and 2.5 mm, which translates to approximately between 40% and 80% of a region's original thickness (as measured with the SCP formula).

The most challenging aspect of establishing ground truth is avoiding bias induced by the methods employed for its measurement. For this reason, we elected to use the marching-cubes algorithm to obtain surface representations rather than a more thoroughly validated pipeline such as FreeSurfer or CRUISE [69]. When applying these methods for the quantification of a new method for CTh measurement, one could create the synthetic atrophy dataset with our morphology-based methods, generate surface representations of each timepoint using an established segmentation pipeline, and then return back to our method to assess true change using those surfaces rather than ones obtained with marching cubes. In this case, the ground truth measurements would indeed be biased by the selected cortical segmentation method, but would use surfaces that better represent the complex geometry of cortical GM. That being said, the advantage to using marching cubes surfaces is that the resulting surface placement corresponds exactly with GM/WM and GM/CSF boundaries of the ROIs selected for deformation. The use of an established cortical segmentation pipeline may introduce additional discrepancies between the ground truth and measured thickness changes.

We also acknowledge that our definition of the true change in thickness differs from the traditional definition of CTh. In longitudinal studies of CTh, thickness change is defined as the difference between thickness measured at each timepoint. However, we define this change as the difference in surface placement. Had we elected to define thickness change in the traditional way, our results would have been biased by whatever method we employed to measure CTh. We circumvent this problem by removing the actual thickness calculation from the pipeline, and simply find the amount by which the GM and WM surfaces have been displaced by the synthetic atrophy deformation.

Finally, although the methods discussed for inducing and assessing localized cortical atrophy are presented in this chapter as a single pipeline, they are not dependent upon each other. For example, one could apply the methods presented here to obtain a synthetic longitudinal dataset with a transformation encoding the amount to which each image has been atrophied, and measure CTh volumetrically rather than with cortical

surfaces. Alternatively, one could employ a different technique to synthetically induce atrophy, such as that proposed by Karaçali et al. [89], obtain cortical surfaces corresponding to the original and synthetic timepoints, and then apply our method to assess the true change in CT. Further, our methods are not restricted to using FreeSurfer to obtain a cortical parcellation and skull-strip mask; any parcellation that contains separate labels for GM, WM, and CSF will suffice, as well as any skull-strip mask.

4.5.3 Effect on Longitudinal Cortical Segmentation Accuracy

Figures 4.8 and 4.9 illustrate the unsigned and signed errors between timepoints (original vs. atrophy). HC and MS subjects were analyzed separately rather than as a single cohort to illustrate that FreeSurfer yields surfaces with higher errors for both natural and synthetic atrophy, rather than exclusively for the synthetic atrophy induced by our methods. We observe that in general, for HC subjects, FreeSurfer has lower errors for the original timepoints than the synthetic. This may lead one to prematurely conclude that the FreeSurfer produces erroneous segmentations when processing images created by the proposed methods, perhaps due to slight blurring that results from interpolation. However, when compared to the original timepoint for the MS subjects, there exist no significant differences between the HC-synthetic (synthetic atrophy) and MS-original (natural atrophy) cohorts based on our corrected multiple comparisons analysis. Thus, it follows that these errors likely arise from the presence of any atrophy at all, rather than whether or not the image is synthetic. This is not an unexpected finding about FreeSurfer, as many algorithms have a decrease in performance as the data deviates from healthy controls. We note that the exception to this is within the right WM central sulcus cluster; the absolute signed errors are similar, but they have opposite signs, which indicates that in this location, FreeSurfer is overestimating surface placement in the HC images while underestimating the MS.

Figures 4.10 and 4.11 show that within the healthy synthetic cohort, there exist only slight differences in segmentation errors between experts, none of them statistically significant. We observe even smaller differences between images processed cross-sectionally vs. longitudinally. These results further support our previous conclusion that the data shown in Figures 4.8 and 4.9 are indeed suitable representations of the entire set. Further, the lack of discrepancy between results from the cross-sectional and longitudinal processing methods show that the inclusion of synthetic data in FreeSurfer's joint initialization steps [141] does not corrupt the final longitudinal segmentation results.

Table 4.1 shows that within the entire dataset (not simply those shown in Figures 4.8-4.11), there exists a much larger number of statistically significant mean signed segmentation errors than unsigned. This implies that FreeSurfer is overestimating surface placement in some places while underestimating in others. After applying a Bonferroni correction of $n = 224$, we found only 14 out of all 1,344 tests yielded significant differences, and only 3 out of the 672 deemed "relevant". These errors are almost entirely limited to within

the central sulcus (shown in the three bolded panels within Figure 4.9, and may be due to its higher than average thickness compared to the rest of the brain. Further, these instances, the significant differences exist between the HC and MS groups as a whole, rather than simply between the synthetic HC and original MS cohorts, which suggests the segmentation errors may arise from the original landmark placement or image quality rather than due to errors induced by our methods. Overall, these results support our observation that FreeSurfer performs somewhat worse in all cases of atrophy, regardless of whether it is synthetic or natural.

Lastly, we found that the success of our methods in this context is also dependent on fiducial landmark placement. For example, the central sulcus landmark set for one subject in the HC dataset contained two clusters in adjacent slices where one cluster existed directly on top of the other in an adjacent slice. This means that, rather then both ROIs having 3 slices of padding on either side of the landmark cluster, each ROI had 3 slices of padding on one side and none on the other. Further, in order to induce atrophy in both ROIs, the image was warped in slightly different directions in adjacent slices, so deformations at the cluster locations could have been affected by boundary effects due to the lack of padding. All this could have potentially induced image artifacts that hindered FreeSurfer's ability to yield accurate segmentations. This could be addressed by fine tuning the landmark locations such that each ROI would include an adequate buffer around the fiducials. Alternatively, instead of deforming each individual cluster separately, certain landmark clusters could be combined into a single, larger cluster and deformed as a single ROI, which would remove the issue of adjacent slices being deformed by discontinuous transformations. Our method proved to perform best when operating on isolated landmark sets rather than those placed close together.

4.6 Conclusion

In summary, we presented a registration-based method for inducing synthetic, localized cortical atrophy in MRI scans. The quantitative evaluations illustrate that this technique can be used for accuracy validation of CTh measurements, specifically those obtained using surface-based methods, by comparing experimentally measured values to the ground truth produced by our algorithm. Further, we showed that our work is also applicable to accuracy validation of cortical segmentation pipelines; the methods can be used to produce a set of longitudinal cortical landmarks with exact correspondences between the original and atrophied timepoints.

CHAPTER 5

Finite Difference Modelling of the Laplacian for Cortical Thickness Measurement

5.1 Abstract

Cortical thickness (CTh) is an important image-based marker for both healthy aging and neurodegeneration. However, the most commonly used surface-based method for its measurement, the symmetric closest point (SCP) distance mapping, is prone to underestimation in curved regions and lacks the symmetry that its name suggests. Volumetric CTh measurement techniques have provided alternative methods by employing PDEs to create a one-to-one, symmetric mapping between the GM and WM boundaries, but are prone to error from partial volume effects and susceptibility to noise. To address these shortcomings, we present a surface-based Laplacian (SBL) CTh measurement technique that circumvents the pitfalls of both the SCP and volumetric methods. Our pipeline models the interior of the cortical ribbon with an irregular grid, solves the Laplacian equation over this grid, and measures thickness by numerically integrating along the normalized solution gradient. In this work, we validate our pipeline using both digital phantoms and cortical surfaces. We compare CTh between the SBL and SCP methods, and show that the SBL produces more symmetric thickness values than the SCP. Finally, we demonstrate that the SBL yields significantly more accurate CTh values than the SCP with respect to ground truth thickness change, and that both methods are comparably reproducible.

5.2 Introduction

As described in previous sections, cortical thickness (CTh) is an important image-based marker for both healthy aging and neurodegenerative processes. In MRI scans, it can be measured using either volumetric methods [31, 39, 65, 78, 79, 85, 213], which calculate thickness directly from the image, or surface-based methods[53, 69], which calculate thickness as the distance between surface reconstructions of the gray matter (GM) and white matter (WM) outer boundaries. While surface-based methods are less computationally efficient than their volumetric counterparts due to the need to reconstruct the 3D cortical surfaces, they generally yield more accurate results due to reduced errors from partial volume effects and lower susceptibility to noise and topological defects [34]. A thorough review of methods to quantify CTh is provided in Section 2.4.

The most common surface-based method for CTh measurement is the symmetric closest point (SCP) distance mapping, which was first implemented by Dale et al. [53] as part of the FreeSurfer toolkit [38] (see Section 2.4.1). The SCP method defines thickness at each point on the cortical GM surface as the average between (1) the distance from that initial GM point to the closest point on the WM surface, and (2) the distance between that WM point and its closest point on the GM surface. Although suitable in relatively flat

regions of the cortex, the SCP formula tends to underestimate thickness in higher curvature areas such as within deep sulci and gyri [123]. Further, counter to its name, this mapping is not necessarily symmetric: using initial points from the GM surface will often yield different thickness results than using WM points.

An alternative class [65, 78, 79, 85, 201, 213] of CTh calculation methods circumvents these shortcomings by using partial differential equations (PDEs) within the cortical GM to create an bijective mapping between the GM and WM boundaries. Jones et al. [85] first proposed a PDE approach by numerically solving for the solution to the Laplacian equation $\nabla^2\Psi = 0$ within the cortical GM ribbon, and then measuring CTh as the length traversed while integrating along the solution gradient $\nabla\Psi$. This produces a one-to-one mapping between the GM and WM, and integrating in either direction will yield identical thicknesses at corresponding locations on either side of the cortical domain. Unfortunately, because these PDE methods all operate within the image domain, they are susceptible to the aforementioned sources of error associated with volumetric techniques.

In this work, we propose an alternative method for calculating CTh: the surface-based Laplacian (SBL) thickness pipeline. Rather than solving for the Laplacian within the image space, we generate a tetrahedral mesh initialized using surface reconstructions of the GM and WM boundaries, and solve for both the gradient to the Laplacian solution and the resulting thicknesses directly within the mesh space. While a similar tetrahedral approach was proposed by Zang at al., 2015 [201], this method uses a finite element model of the Laplace-Betrami heat kernel to find the Laplacian field solution Ψ. Based on this solution, isothermal surfaces are then generated within the interior of the mesh similar to those depicted in Figure 2.13, and thickness is calculated using the normals to those surfaces. Our method, on the other hand, requires no finite element analysis, and solves for the gradient of the Laplacian field $\nabla\Psi$ without solving for Ψ, rendering it theoretically more computationally efficient.

5.3 Methods

5.3.1 Theoretical approach

One of the key advantages to surface- and mesh- based methods is that unlike volumetric ones, they are not restricted to operating on a regular voxel grid. However, this presents difficulties when determining finite differences, which do not translate easily to non-uniform grids in n-dimensions when $n > 1$. An example of this is depicted in Figure 5.1. In a uniform grid, the relationship between nodes in the regular grid are well-defined, and any PDE can be easily modelled within this domain based on the spatial relationship between each nodes. In contrast, modelling even a simple equation in an irregular grid poses difficulties and requires complex frameworks to find solutions [82, 111].

For this reason, in order to employ finite different methods within our cortical annulus, we adopt the

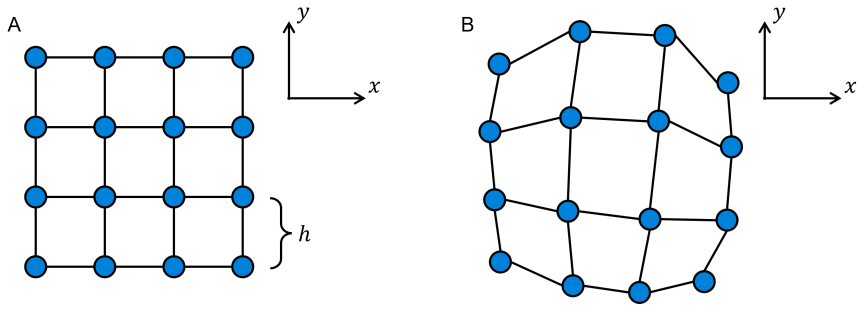

Figure 5.1: Example of a regular (left) vs. an irregular grid (right).

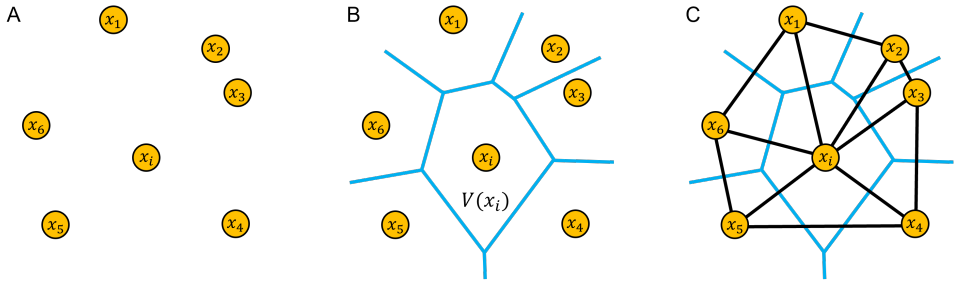

Figure 5.2: Caption

framework outlined by Sukumar et al., 2003 [176]. Rather than defining nodal connectivity based on adjacency within a uniform grid, Sukumar et al. proposed using the concept of natural neighbors as identified by the Voronoi diagram of a node set. The spatial relationship of these neighbors can then be exploited to solve for the Laplacian shape function relating each pair of adjacent nodes within the mesh, and the subsequent solution to any elliptical equation such as the Laplacian. Note that in this section, we use the terms vertex and node interchangeably.

Let Ω be the domain containing the set of n vertices $\mathbf{X} = \{x_1, ...x_i, ...x_N\}$ depicted in Figure 5.2. The Voronoi diagram of X, denoted $V(\mathbf{X})$, is a partition of Ω into distinct regions V_i such that

$$V_i = \{p \in \mathbb{R}^d : D(p,x_i) < D(p,x_j), \forall j \neq i\}, \tag{5.1}$$

where $D(\cdot,\cdot)$ is the Euclidean distance. For $d = 2$ (3), these cells are polygons (polyhedra). A natural neighbor of a node x_i is any node x_j where V_i and V_j share a $(d-1)$-dimensional boundary, called a Voronoi facet, denoted t_{ij}. In Figure 5.2, these facets are depicted by the blue lines in panels B and C. By connecting all $x_i, x_j \in V$ that are natural neighbors, we can construct specific mesh known as the Delaunay triangulation (tetrahedralization) $DT(\mathbf{X})$. An interesting fact is that a triangulation is Delaunay if and only if for every triangle $DT(x_i,x_j,x_k)$, its circumcircle contains no nodes other than x_i, x_j and x_k; this property holds for three-dimensional meshes as well.

Let $m(t_{ij})$ denote the length (area) of the facet t_{ij}, and let h_{ij} denote the length of the triangle (tetrahedron) edge connecting nodes x_i and x_j. The Laplacian shape function relating x_i and x_j is thus defined as

$$\Phi_{ij} = \frac{\alpha_{ij}}{\sum_j^n \alpha_{ij}}, \quad \alpha_{ij} = \frac{m(t_{ij})}{h_{ij}} \tag{5.2}$$

where α_{ij} the Laplacian weight function and n is the total number of natural neighbors of x_i. In other words, the Laplacian weighting is the ratio of the distance between x_i and x_j and the length (area) of the corresponding Voronoi facet, and the shape function normalizes this weight by the sum of the weights relating to x_i and all its natural neighbors. This forms a linear function, where $0 \leq \Phi_{ij} \leq 1$ and $\sum_j \Phi_{ij} = 1$.

We now show how we exploit the Laplacian shape function to numerically solve the Laplacian equation $\nabla^2 \Psi = 0$ within Ω. Let U_i be the value of the solution at the vertex x_i. Then

$$U_i = \sum_j^n \Phi_{ij} U_j = \sum_j^n \frac{\alpha_{ij} U_j}{\sum_k^n \alpha_{ik}}. \tag{5.3}$$

When solving this using the iterative Jacobian method, for each iteration l, we have

$$U_i^{l+1} = \sum_j^n \Phi_{ij} U_j^l. \tag{5.4}$$

Note that for a regular, uniform grid, this simplifies to the framework detailed in Appendix B.

5.3.2 Pipeline overview

The SBL pipeline consists of 3 general steps: (1) constructing an irregular grid within the cortical gray matter ribbon; (2) numerically solving for the gradient of solution to the Laplacian equation at each grid node; and (3) measuring CTh at each point on the cortical surface by integrating from one surface to the other along the gradient of the field solution. A full visualization of this pipeline is depicted in Figure 5.3. Unless stated otherwise, all processes reported below are implemented using the Visualization Toolkit (VTK) [1] libraries [157].

To construct our finite difference model (FDM) grid, we first generate triangulated mesh representations of the cortical GM and WM surfaces. For our validation experiments here, we use cross-sectional segmentations from FreeSurfer [38]. These surfaces act as the boundaries of the Laplacian domain, and we refer to them combined as our "cortical annulus". We construct our grid by first populating the interior of the annulus with irregularly spaced nodes, and then connecting neighboring nodes to build the final grid; we refer to these as our FDM nodes and FDM grid, respectively. We define the FDM node set by using the SPMESH

[1] https://vtk.org/

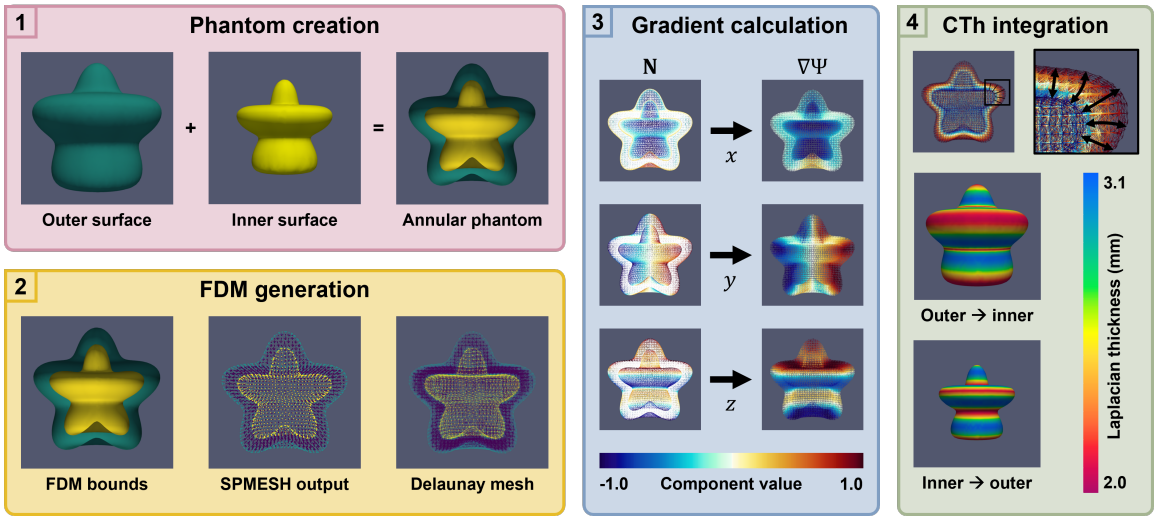

Figure 5.3: Overview of the SBL pipeline demonstrated using the star phantom. **(1)** Phantom creation: the outer and inner surfaces are generated and combined into the annular phantom. **(2)** FDM generation: SPMESH is used to populate the interior of the annulus (left) with FDM nodes (center), and the final Delaunay mesh is generated are defined using the Voronoi diagram of the FDM node set (right). Initial conditions are set at each node based on whether the node lies on the outer surface (green), inner surface (yellow), or annulus interior (purple). **(3)** Gradient calculation: using surface normals **N** (left) for initialization, the Laplacian field gradient $\nabla\Psi$ is iteratively calculated. **(4)** CTh integration: CTh is measured at each vertex (i.e. each boundary node) by iteratively travelling along $\nabla\Psi$ from one surface to the other.

tetrahedral mesh generator [177], and initializing the node spacing as the median edge length of the triangular surfaces comprising the annulus. Although this program is capable of producing a full tetrahedral mesh, we extract only the nodes from the SPMESH output for our FDM grid and rebuild the connectivity grid to meet the Delaunay requirements as described next.

Next, in order to employ the Laplacian shape function detailed in section 5.3.1, we calculate the Voronoi diagram of the FDM node set using the Voro++ library [152]. This provides the information necessary to identify the natural neighbors of each FDM node and define the connecting edges between nodes within our grid. Because calculating the Voronoi diagram of a concave domain such as the cortex proves to be highly computationally challenging, in practice, we actually define the edge set of the FDM grid based on three separate cases of node pairings. For each pairing, if an edge exists within either the SPMESH Voronoi diagram output, and the associated conditions are met, then we include the edge in the final FDM.

1. Interior $\leftrightarrow$ interior: if both nodes are non-boundary nodes and centered within adjacent Voronoi cells, and if the center of their connecting edge does not lie exterior to the cortical annulus. This preserves the integrity of the Voronoi diagram while accounting for errors induced by issues with calculating the Voronoi diagram of points in a concave domain.

2. Interior $\leftrightarrow$ boundary: if a boundary node is connected to an interior node in the initial tetrahedral

SPMESH mesh. This also corrects any additional errors induced by issues with calculating the Voronoi diagram of points in a concave domain.

3. Boundary $\leftrightarrow$ boundary: if both nodes are boundary nodes and are vertices of same triangle in the cortical annulus. This ensures the input GM and WM boundaries remain unchanged.

Once our FDM grid is fully constructed, we numerically solve for the gradient to the Laplacian equation within our cortical annulus using the Laplacian shape function framework. Due to the nature of the Laplacian function, we are able to directly calculate the gradient of the solution $\nabla\Phi$ without necessitating a solution for Φ as well. Let dU_i represent the numerical gradient solution to $\nabla\Phi$ in the direction at each FDM node x_i. We initialize our iterative process by defining Dirichlet conditions on both bounding surfaces such that for a boundary node x_{boundary}, dU_{boundary} is equal to the normal vector of the annulus at that point. We then iteratively calculate dU_i for each interior node using the Armadillo C++ library for linear algebra [154]. Finally, we calculate thickness by integrating along this gradient solution from one bounding surface to the other. During this step, we perform integration starting at the GM surface until either we reach the WM surface or the distance travelled surpasses 10.0 mm to account for any numerical errors.

5.3.3 Validation Experiments

5.3.3.1 Phantom Validation

We first evaluated the efficacy of our SBL pipeline in radially symmetric, annular digital phantoms. We constructed these by defining three sets of surfaces using the same equations as those in the volumetric, registration-based CTh pipeline developed by Das et al., 2009 [39]. In our application, we created three, surface representations of topologically spherical annuli, rather than volumetric structures, with varying degrees of curvature. Figure 5.4 displays each of the three phantoms: a sphere (left) with no undulations, a 3D "star" (center) with undulations along the one spherical axis (ϕ), and a "spore" (right) with undulations along both spherical axes (θ). Each annulus was comprised of two triangulated meshes defined by points lying on either the surface S_o or S_i, the outer and inner surfaces. These surfaces took the form

$$S_j(\gamma,\alpha,\theta,\phi) = [r_j + f(\gamma,\alpha,\theta,\phi)] \cdot (\cos\theta\cos\phi, sin\theta\cos\phi, \sin\phi) \tag{5.5}$$

where $j \in \{o,i\}$ denotes the outer or inner surface, r_j the radial distance from the center of the surface at origin, and $f(\gamma,\alpha,\theta,\phi)$ the shape of the surface. For all surfaces, we set $\gamma = 2$, $\alpha = 5$, $-\pi < \phi < \pi$, and $0 < \theta < \pi$. The radii of the outer and inner surfaces were $r_o = 10$, and $r_i = 7$, respectively. For each annulus, $f(\gamma,\alpha,\theta,\phi)$ is varied to create the sphere, star, and spore shapes such that $f_{\text{sphere}} = 0$, $f_{\text{star}} = \gamma\sin\alpha\phi$, and $f_{\text{spore}} = \gamma\sin\alpha\phi\cos\alpha\theta$. The full set of equations defining the surfaces of the annular phantoms are

$$S_{j,\text{sphere}}(\theta,\phi) = r_j \cdot (\cos\theta\cos\phi, sin\theta\cos\phi, \sin\phi) \tag{5.6}$$

$$S_{j,\text{star}}(\gamma,\alpha,\theta,\phi) = [r_j + \gamma\sin\alpha\phi] \cdot (\cos\theta\cos\phi, sin\theta\cos\phi, \sin\phi) \tag{5.7}$$

$$S_{j,\text{spore}}(\gamma,\alpha,\theta,\phi) = [r_j + \gamma\sin\alpha\phi\cos\alpha\theta] \cdot (\cos\theta\cos\phi, sin\theta\cos\phi, \sin\phi) \tag{5.8}$$

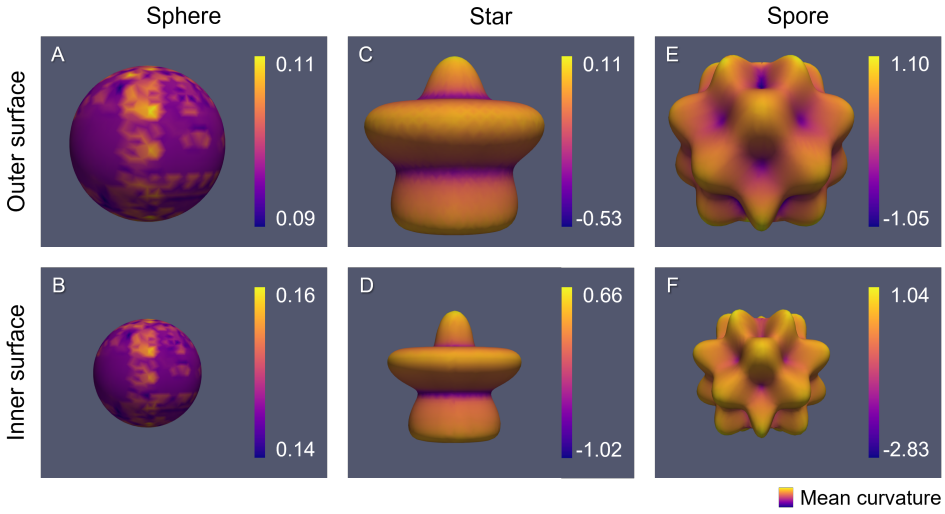

Figure 5.4: Annular phantoms used to validate the SBL pipeline. Phantoms are denoted from left to right as sphere, star, and spore. The top row contains the outer surface (with $r_o = 10$), and the bottom row the inner surface ($r = 7$). The color map on each surface depicts the value of the mean curvature at each vertex. Mean curvature was smoothed with local averaging to account for discontinuities induced by discrete triangulation.

For each annular phantom, we calculated the SBL thicknesses (D_{SBL}) at each vertex using integration step sizes (ISSs) equal to 0.1, 0.25, 0.5, and 1.0 mm, and from both the outer→inner and inner→ surfaces, denoted as $D_{\text{SBL},o\to i}$ and $D_{\text{SBL},i\to o}$. We also calculated the SCP thicknesses $D_{\text{SCP},o\to i}$ and $D_{\text{SCP},i\to o}$ for comparison. For both distance mappings, we observed the relationship between mean surface curvature and the reported thickness. Additionally, we investigated the "symmetry error" (SE) of each mapping, which we defined at each vertex as the difference between the thickness $D_{o\to i}$ measured at that vertex on one surface and the thickness $D_{i\to o}$ measured at the the corresponding point on the opposite surface. To demonstrate this, let us consider the example in Figure 5.5, which shows a vertex v of the left-hand triangle (representing the outer, GM surface) that maps to a point p lying inside the triangle $\triangle u_1u_2u_3$ on the opposite surface. The dotted red line indicates the thickness $D_{o\to i}(v)$ resulting from the mapping from v to p. The SE is then the difference between $D_{o\to i}(v)$ and $D_{o\to i}(p)$, as p should theoretically map back to v. Because p is not a surface vertex, rather than re-integrating back across the Laplacian field gradient, we instead interpolate $D_{o\to i}(p)$ using the distances $D_{o\to i}(u_i)$, $D_{o\to i}(u_2)$, and $D_{o\to i}(u_3)$.

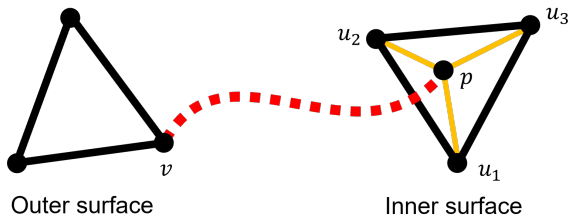

Figure 5.5: Diagram illustrating the interpolation used to calculate symmetry error.

5.3.3.2 Evaluation in Cortical Surfaces

We next applied the SBL pipeline to cross-sectional, FreeSurfer (FS) cortical surface reconstructions [38] from the NITRC Kirby dataset [97] to assess both the reproducibility and accuracy of our method. Because FS surfaces generally contain over 100,000 vertices, i.e., two orders of magnitude more than our phantoms, it is advisable to apply a decimation filter to reduce computation time. However, because over-decimating the surfaces would yield inaccurate CTh results, our first goal was to assess the extent to which decimation affects the accuracy of our thickness measurements. We accomplished this by running the SBL pipeline on the left hemisphere from a single subject with varying amounts of decimation, from 0 to 95%, induced using the vtkDecimatePro filter[2]. For each decimation factor, we measured the difference between the mean CTh in the original (non-decimated) surface and the mean CTh within the decimated surface. We refer to this difference as the decimation-induced thickness error (DITE). We conducted this analysis for both the SBL (ISS = 0.5mm) and the SCP. For all thicknesses measured in cortical surfaces, we analyzed only the thickness from the GM to WM surface.

We then assessed the reproducility and accuracy of our method, given an optimal decimation factor, compared to that of the SCP. Because the Kirby dataset consists of two same-day scans for each subject, the CTh measured in one timepoint should be equal to the CTh measured in the other. Thus, we could define the reproducibility error (RE) as the absolute change in thickness between timepoints. We calculated RE using the FreeSurfer surface reconstructions as the absolute difference in mean thickness within each of the Desikan-Killany (DK) atlas regions [42] for both the SBL and SCP methods. RE for each thickness mapping was normalized by the average thickness within each region between both timepoints and reported as a percentage rather than a raw distance value.

To assess the accuracy of SBL measurements compared to the SCP, we employed the method outlined in Chapter 4 to create a synthetic, ground truth dataset with known changes in thickness at each location on the cortical surface [101, 102]. We created this dataset using the first timepoint for each subject of the Kirby dataset and two iterations of the atrophy induction pipeline (see section 4.3.2) in each of the DK atlas regions.

[2]https://vtk.org/doc/nightly/html/classvtkDecimatePro.html

We then applied the SBL pipeline to the same marching cubes surfaces generated to measure the true change in CTh, calculated using the mean surface displacement difference (MSDD) as described in section 4.3.3. We then quantified the accuracy error of our method as the difference between the measured SBL thickness change and the MSDD in each DKT region, and compared this to the accuracy error of the SCP. Note that for both the reproducibility and accuracy studies, SCP thickness was measured in the undecimated surfaces, while SBL was measured in surfaces decimated to a target of 50% reduction in vertices.

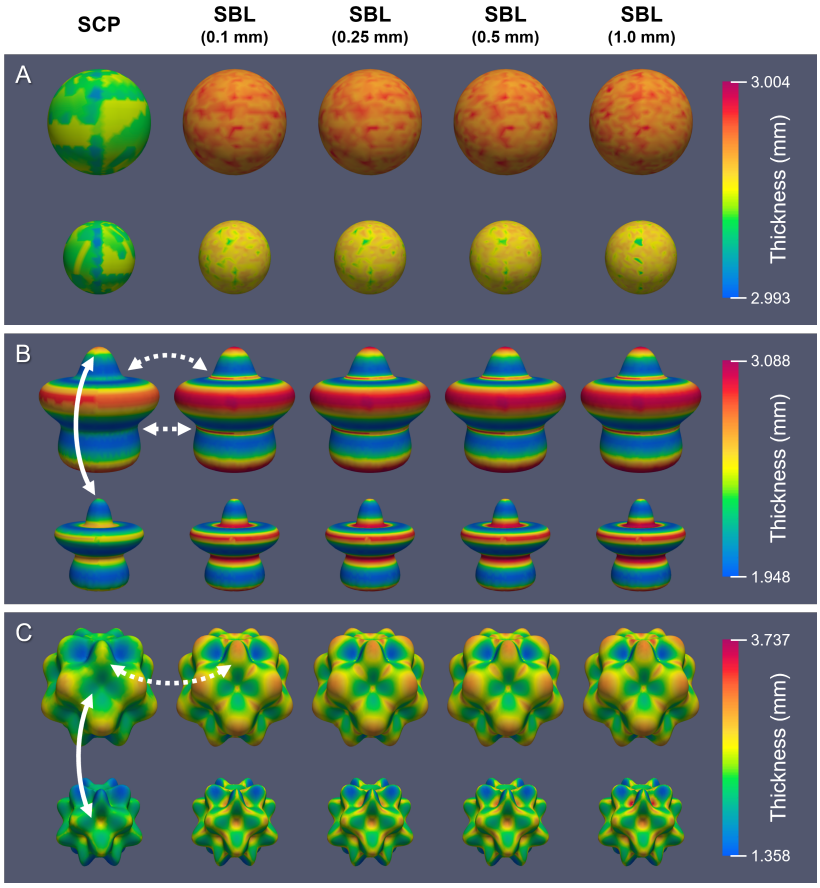

Figure 5.6: Annular phantoms with color maps indicating reported thickness. **(A)** Spherical phantom; **(B)** Star phantom; **(C)** Spore phantom. Within each panel, the top row depicts the outer surface and the bottom row the inner surface. From left to right, the thickness mapping corresponds to SCP, SBL with ISS=0.10 mm), SBL with ISS=0.25 mm) SBL with ISS=0.50 mm, and SBL with ISS=1.00 mm. The solid arrows indicate differences between the SCP and SBL, and the dotted arrows indicate differences measured with the SCP from the outer $\rightarrow$ inner vs. inner $\rightarrow$ outer surfaces.

5.4 Results

5.4.1 Phantom Validation

Figure 5.6 displays the annular thickness calculated with the SCP and the SBL for each phantom. The SBL was calculated using an integration step size (ISS) of 0.1, 0.25, 0.5, and 1.0 mm. In the spherical phantom (5.6A), where the theoretical thickness is exactly 3 mm, the SCP method reports a slightly lower than accurate thickness than the SBL. In the star and spore phantoms (5.6B and 5.6C), we observe that the SCP reports much lower thickness values in areas of high curvature, as expected. The dashed arrows in these panels indicate high curvature bands that are better identified by the SBL than the SCP, which suggests a lack of precision in the SCP method in these areas. Further, the solid arrows highlight locations with large discrepancies between SCP thickness depending on whether integration was performed starting at the outer or inner surface (i.e., symmetry errors). Note that we observe very little variation in reported SBL thickness between the different ISS values.

This observation is further supported by the data in Figure 5.7, which displays thickness for each method as a function of curvature. This figure supports our conclusions that the SBL reports higher values than the SCP. For the star and spore phantoms, the discrepancy between the two methods is more pronounced in high curvature areas. The peaks in the star and spore phantom panels (middle and right) correspond to the undulations along varying latitudes of the original surfaces. We also observe no significant difference in thickness values for different SBL step sizes.

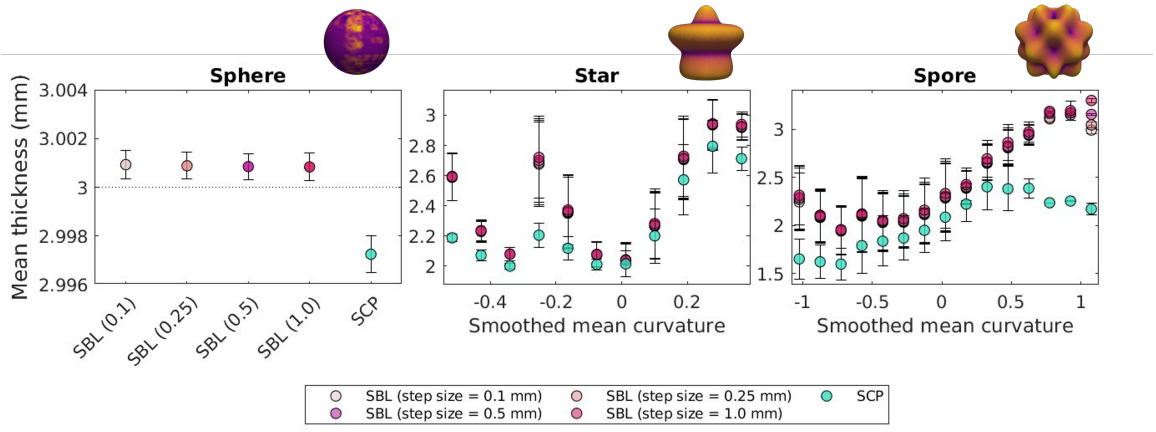

Figure 5.7: Mean SBL and SCP thicknesses measured from the outer → to inner surfaces. SBL was calculated using ISS values of 0.10, 0.25, 0.50, and 1.00 mm. Thickness for each method and ISS was averaged across vertices with similar curvature values as defining using a histogram. A picture of each phantom with curvature mappings (see Figure 5.4) is displayed beside each title.

Figure 5.8 displays the symmetry error (SE) for each phantom. For the sphere, we again see a lower error for the SCP compared to the SBL (0.000 ± 0.001 vs. 0.002 ± 0.001), although both errors are negligibly small,

likely attributable to interpolation error. For the other two phantoms, however, we see a higher symmetry error in the SCP thicknesses than the SBL (star: 0.025±0.108 vs. 0.110±0.145; spore: 0.043±0.168 vs. 0.192±0.189). The ISS for these phantoms does not affect the error substantially. For this reason, and because of the conclusions drawn from Figure 5.6, we deem it reasonable to use ISS=0.5 mm for CTh.

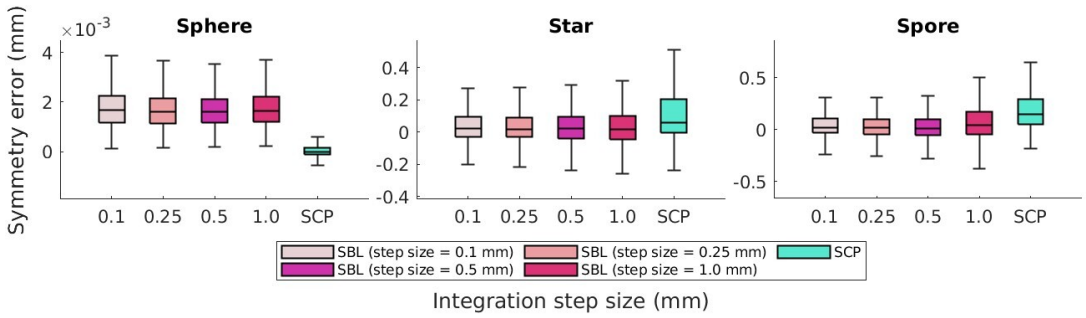

Figure 5.8: Symmetry error (SE) of SBL and SCP thicknesses averaged across the entire outer or inner surface. SBL was calculated using ISS values of 0.10, 0.25, 0.50, and 1.00 mm.

Lastly, Table 5.1 contains the run times for the SPMESH, Laplacian solver, and thickness integration stages of the SBL pipeline for each ISS value. We also report the total SBL pipeline duration, which exceeds that of the sum of the listed steps because it also is affected by the annulus generation and Voronoi diagram calculation procedures. These were excluded from Table 5.1 because their run times are negligible even when handling larger meshes such as cortical surfaces.

	ISS	**Run times (min)**		
	(mm)	**Sphere** (7311 nodes)	**Star** (12174 nodes)	**Spore** (31815 nodes)
SPMESH	n/a	0.4	1.2	7.7
Laplacian solver	n/a	0.1	0.4	1.4
Laplacian field integration	0.10	7.4	23.2	96.9
	0.25	3.0	8.9	37.1
	0.50	1.6	4.2	17.7
	1.00	0.8	1.9	8.2
Total	0.10	8.0	24.9	106.3
	0.25	3.6	10.6	46.5
	0.50	2.2	5.9	27.1
	1.00	1.4	3.6	17.5
SCP	n/a	0.0	0.0	0.1

Table 5.1: Thickness integration run time in annular phantoms.

5.4.2 Evaluation in Cortical Surfaces

Figure 5.9 shows the decimation induced thickness error (DITE) for both the SBL and SCP. As expected, more decimation yields a higher error. The DITE, number of surface vertices, and number of FDM nodes levels off after a factor of around 50% because the surface cannot be further reduced without inducing topological

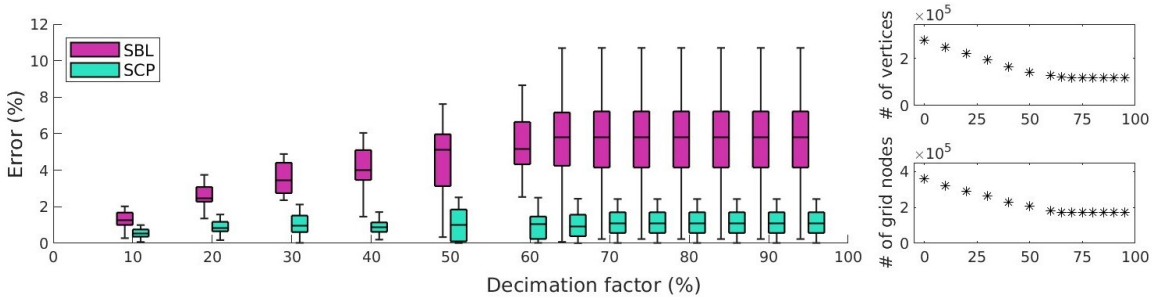

Figure 5.9: **Left:** Decimation induced thickness error (DITE) for the SBL (pink) and SCP (blue) with varying amounts of decimation. SBL was calculated using ISS values of 0.10, 0.25, 0.50, and 1.00 mm. **Right:** Number of surface vertices (top) and FDM grid nodes (bottom) for each decimation factor.

defects (i.e., regardless of the input reduction factor to the vtkDecimatePro filter, the output has the same number of nodes). In other words, an input reduction factor of 65% into the vtkDecimatePro filter yield the same output as a reduction factor of 90% because of constraints to maintain topological accuracy. In all cases, the SCP has a much lower DITE than the SBL because the decimation process does not affect the distance between surfaces used for its thickness calculation as severely as it affects the accuracy of the interpolation of the field solution during integration for SBL. Based on these results, we selected a reduction factor of 50% for the remaining validation studies.

Figure 5.10 displays the mean SBL (ISS=0.5 mm) and SCP thicknesses averaged across regions in the DK atlas[42] for a single subject. We see that the SBL yields higher thickness values as expected in highly curved surfaces, but reports similar trends corresponding regions in each hemisphere. We observe that well-known patterns such as the motor strip, the thinner visual cortex and the thicker temporal and frontal lobes are captured by both SBL and SCP. This suggests that the SBL is reporting clinically relevant thickness values, and could be used as a viable alternative to the SCP in cortical thickness analyses.

Figure 5.11 shows the mean thicknesses within each DKT region for both timepoints in the Kirby dataset measured using the SBL (pink) and SCP (green). We observe that the SBL again reports consistently higher values than the SCP, as expected, and that both CTh definitions appear to exhibit reasonably similar values between the two timepoints and within the same ROI in the left and right hemispheres. The reproducibility error (RE) is displayed in Figure 5.12, where we see that the SBL performs either comparably to or significantly better than the SCP ($p < 0.05$ with a paired t-test). Note that after applying a Bonferroni correction of $n = 66$ (the number of cortical ROIs), none of the differences maintained significance.

Lastly, Figures 5.13, 5.14, and 5.15 show the results of the accuracy validation of the SBL compared to the SCP in the synthetic atrophy experiment. The mean thickness values in each region for both timepoints are reported in Figure 5.13. We see that both distance mappings report higher thicknesses in the original

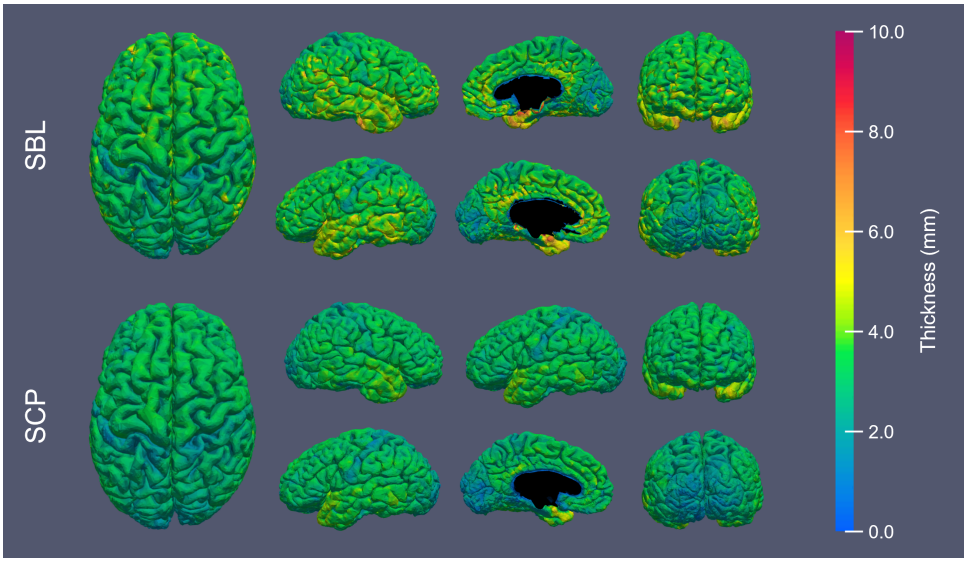

Figure 5.10: SBL (ISS=0.5 mm) and SCP thicknesses at each vertex for a single set of surfaces, decimated by 50%. The scale bar is set to 10 mm because that was the cut-off value used during SBL integration.

timepoint than the atrophied, as expected, and that the SBL reports consistently higher values than the SCP in agreement with the data presented in Figure 5.11. Note that even though the original timepoint for the synthetic atrophy data corresponds to the first timepoint from the Kirby dataset, the respective thickness values between Figures 5.13 and 5.11 are not equal. This is because the accuracy validation was assessed using marching cubes surfaces, whereas the reproducibility experiment used FreeSurfer surfaces.

Figure 5.14 shows the change in CTh measured using both pipelines alongside the mean surface displacement difference (MSDD) for comparison used to define true change. Similarly, Figure 5.15 details the subsequent error within each region. We observe that the SBL is significantly more accurate than the SCP in 54 out of 66 regions ($p < 0.05$), and in 34 out of 66 regions after applying a Bonferroni correction of $n = 66$. In only one region out of all 66 does the SBL perform worse than the SCP, significantly or otherwise. We can therefore conclude that the SBL provides significantly more accurate measures of CTh change than the SCP.

5.5 Discussion

5.5.1 General Performance

Overall, our results show that the SBL method for CTh measurement is slightly more reproducible and significantly more accurate than the SCP mapping. The only drawback to our presented methods is the run time; once the surface reconstructions are obtained, the SCP can calculate thickness at all vertices in under a minute, whereas the SBL takes several days even with surfaces decimated by 50%. The majority of this processing time consists of the thickness integration step (1-2 days), followed by the mesh generation (12-24

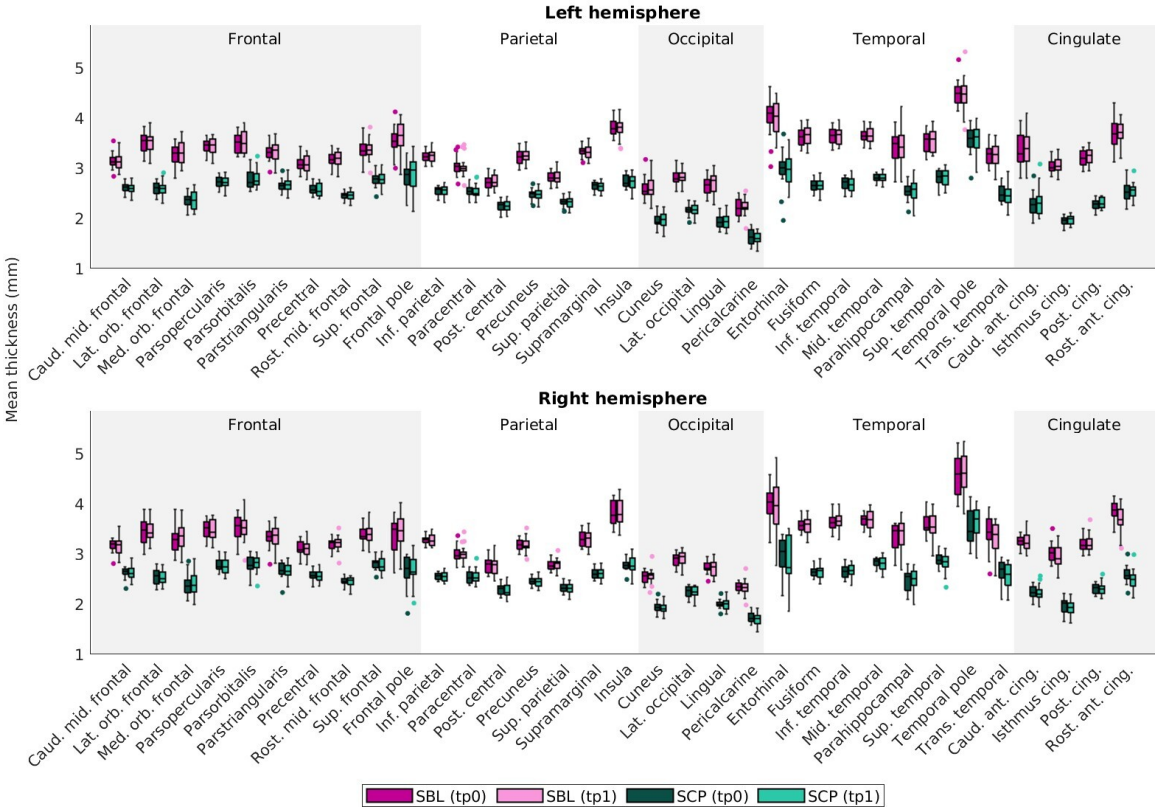

Figure 5.11: Thicknesses measured using the SBL (pink) and SCP (green) for both timepoints in the Kirby dataset. The darker and lighter colored bars differentiate between timepoints. Dots represent subjects whose mean thickness in the corresponding ROI was an outlier.

hours). The calculation of the Laplacian solution gradient is quite fast, and converges within approximately one hour per gradient dimension (e.g., about three hours total) due to the computational efficiency of both the natural neighbor adjacency framework [176] and the C++ Armadillo library [154].

A possible route to increase the efficiency of our pipeline would be to employ a more efficient mesh generation software that necessarily creates a Delaunay tessellation and also calculates the Voronoi diagram, rather than breaking this into separate steps. To this end, we initially selected an alternative mesh generation tool [162] that did exactly this; this software performed well in the annular phantoms but was unable to successfully handle boundary meshes as complex as the cortical surfaces. Further, the Voronoi diagram step is accomplished in a matter of minutes using the Voro++ library [152], and thus adds very little to the total duration. The most significant way to reduce computation time was the decimation of the input cortical surfaces; with undecimated surfaces output by FreeSurfer, which each contain over 100,000 vertices, the entire pipeline took over two weeks to run. The use of a cortical segmentation and reconstruction pipeline that could accurately recreate the GM and WM surfaces with fewer nodes would require a lower reduction factor and potentially yield more accurate results with the same computational requirements.

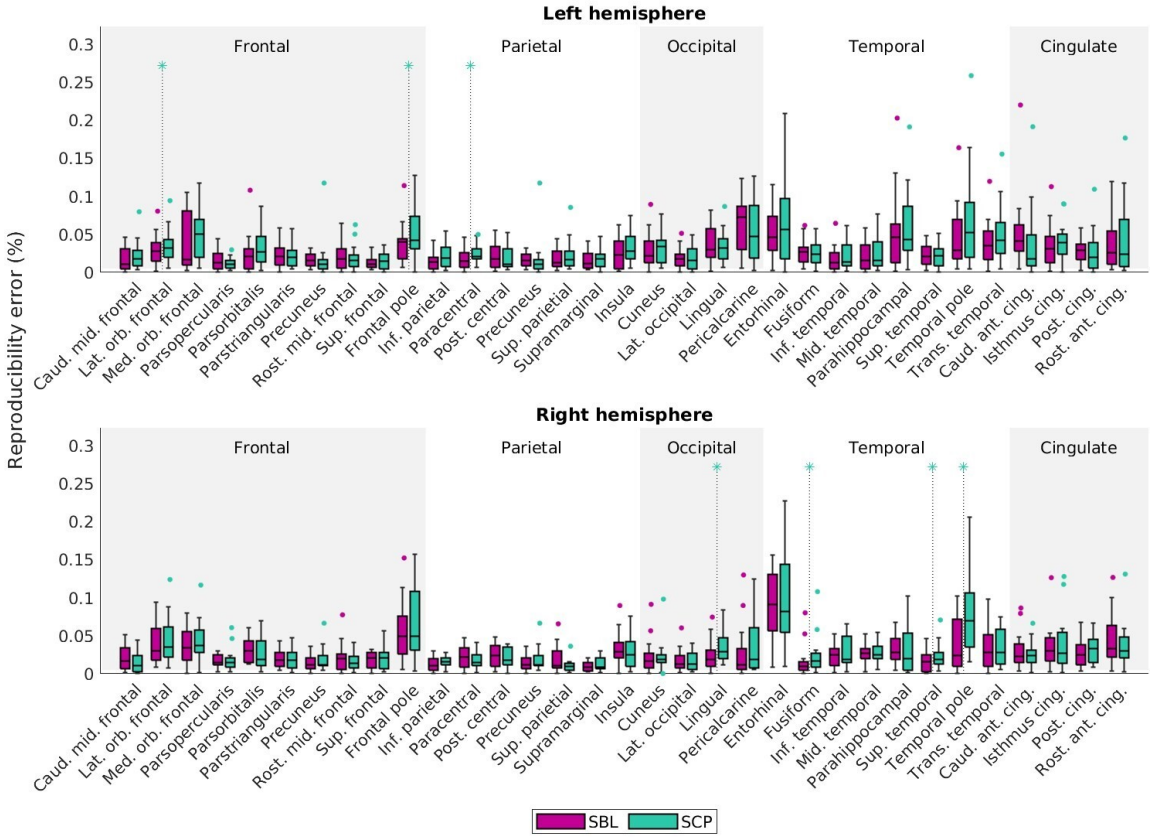

Figure 5.12: Reproducibility error for SBL (pink) and SCP (green) thicknesses measured in the Kirby dataset. Dots of the represent subjects whose mean thickness in the corresponding ROI was an outlier. An asterisk designates a significant difference ($p < 0.05$), where the corresponding color represents the thickness measurement with the higher mean error. No differences maintained significance after applying a Bonferroni correction equal to the number of ROIs compared ($n = 66$).

5.5.2 Phantom Validation

Aside from general troubleshooting within a smaller domain than the cortical ribbon, the primary purpose of our phantom validation experiments was to determine the effect of surface curvature on the SBL and SCP thickness measurements. In the sphere phantom with constant curvature, both methods yielded thickness values approximately equal to the true thickness, where the SBL slightly overestimated and the SCP slightly underestimated, and the SCP had a slightly lower symmetry error. In this case, the SBL errors result from both interpolation during numerical integration and surface triangulation that altered the true annular radii from its theoretical value. On the other hand, SCP thicknesses were only effected by triangulation. This triangulation also is the cause of the splotchy pattern observed on the spherical surfaces in Figure 5.6; this pattern is less pronounced in the other two phantoms because they required a finer triangulation to avoid topological defects.

The data from the star and spore phantoms show that the SCP results in a lower thickness estimation than

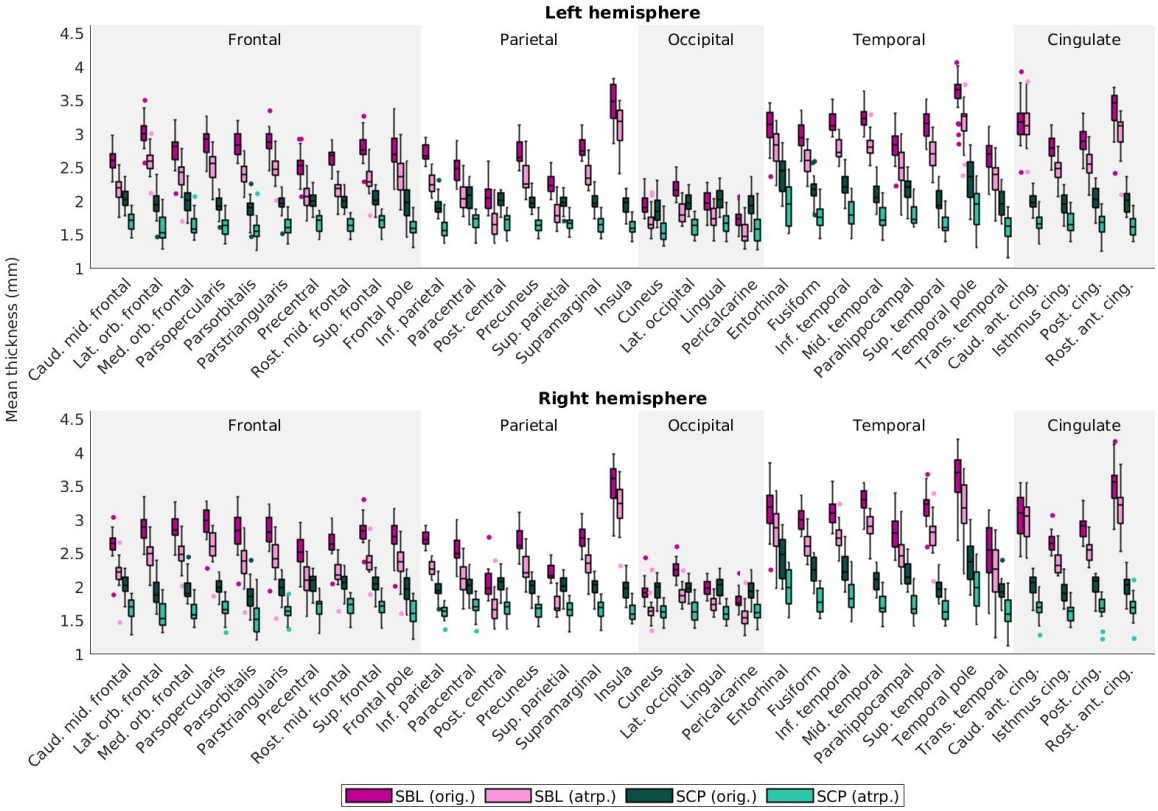

Figure 5.13: Thicknesses measured using the SBL (pink) and SCP (green) distances for both the original and synthetic atrophy timepoints. The darker colored bars correspond to surfaces from the original timepoint, and the lighter to those from the atrophied timepoint. Dots of the represent subjects whose mean thickness in the corresponding ROI was an outlier.

the SBL in higher curvature regions, as expected (Figure 5.7. Further, the SCP exhibits a higher symmetry error in these phantoms as well. These findings both support our hypothesis that a one-to-one mapping between domain boundaries offers improvements with respect to both symmetry and accurate representation of surface geometry.

5.5.3 Evaluation in Cortical Surfaces

The results from our evaluation experiments in cortical surfaces show that the SBL offers improvements in terms of both reproducibility and accuracy over the SCP. Although SBL exhibited a slightly higher mean reproducibility error (RE) in several cortical ROIs, none of these differences were significant; in contrast, the SCP performed significantly worse prior to applying a Bonferroni correction 7 out of 66 regions (Figure 5.12). However, because no differences maintained significance after correcting for multiple comparisons, we deem the two methods comparably reproducible. On the other hand, we can confidently conclude that the SBL is far more accurate than the SCP in measuring CTh changes (Figure 5.15).

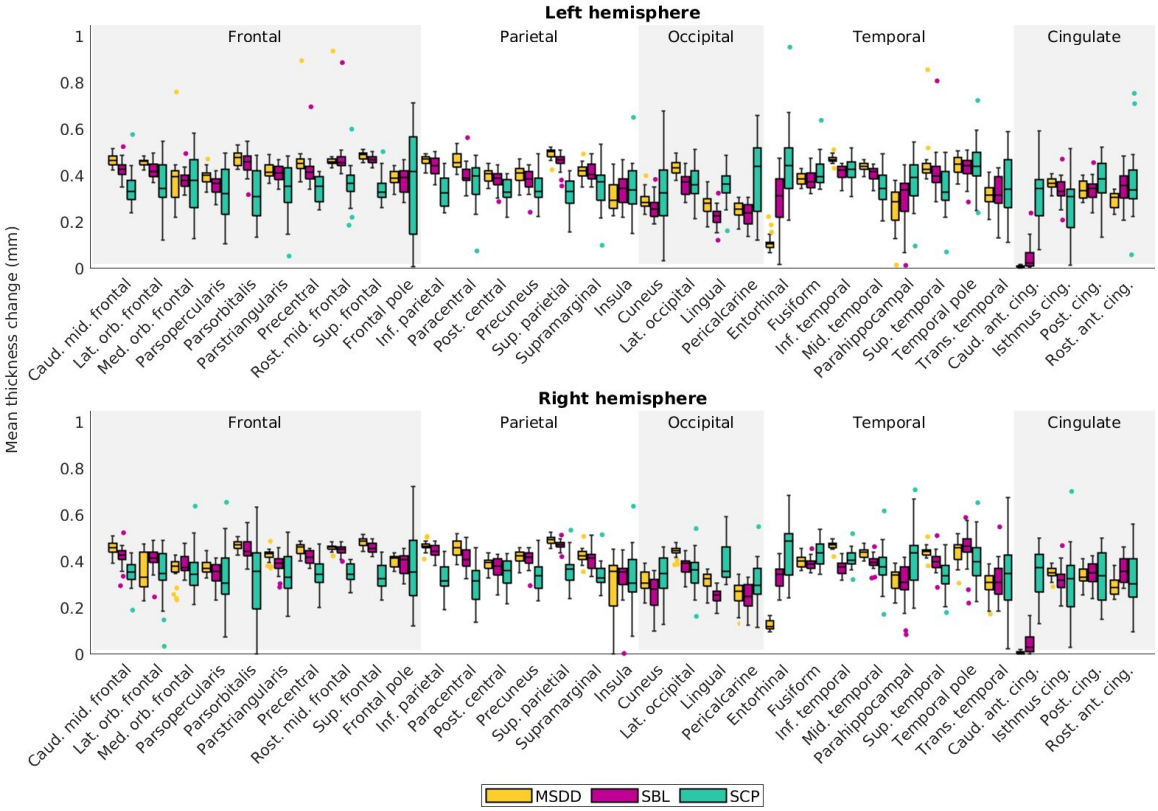

Figure 5.14: Change in thickness measured between the original and synthetic atrophy timepoints using the SBL (pink) and SCP (green) distances. MSDD (yellow) is also shown for comparison. Dots of the represent subjects whose mean thickness in the corresponding ROI was an outlier.

Figures 5.11 and 5.13 show that the SBL consistently yields a mean regional thickness value of approximately 1 mm greater than the SCP. While this may seem at first alarming, as standard CTh values in healthy data have been well documented in literature for decades, the fact that the SBL can reproduce established thickness patterns across the cortex shows indicates that this scale is of little consequence. Figure 5.10 demonstrates the similarities in thickness trends between the SBL and SCP, with lower CTh values in the post-central gyrus and occipital lobes, and higher values in the motor cortex and temporal lobes. Further, accuracy with respect to CTh *changes* is often more relevant than the thickness values themselves, particularly when longitudinal studies of neurodegeneration. The method used to assess accuracy was specifically designed to reduce bias towards any existing thickness mapping, and thus avoids calculating thickness altogether. Nonetheless, the pipeline presented in this Chapter is also applicable to cross-sectional studies with only one timepoint per subject, as long as thickness data compared between cohorts is all measured using the SBL (as opposed to comparing SBL in one group to SCP in another).

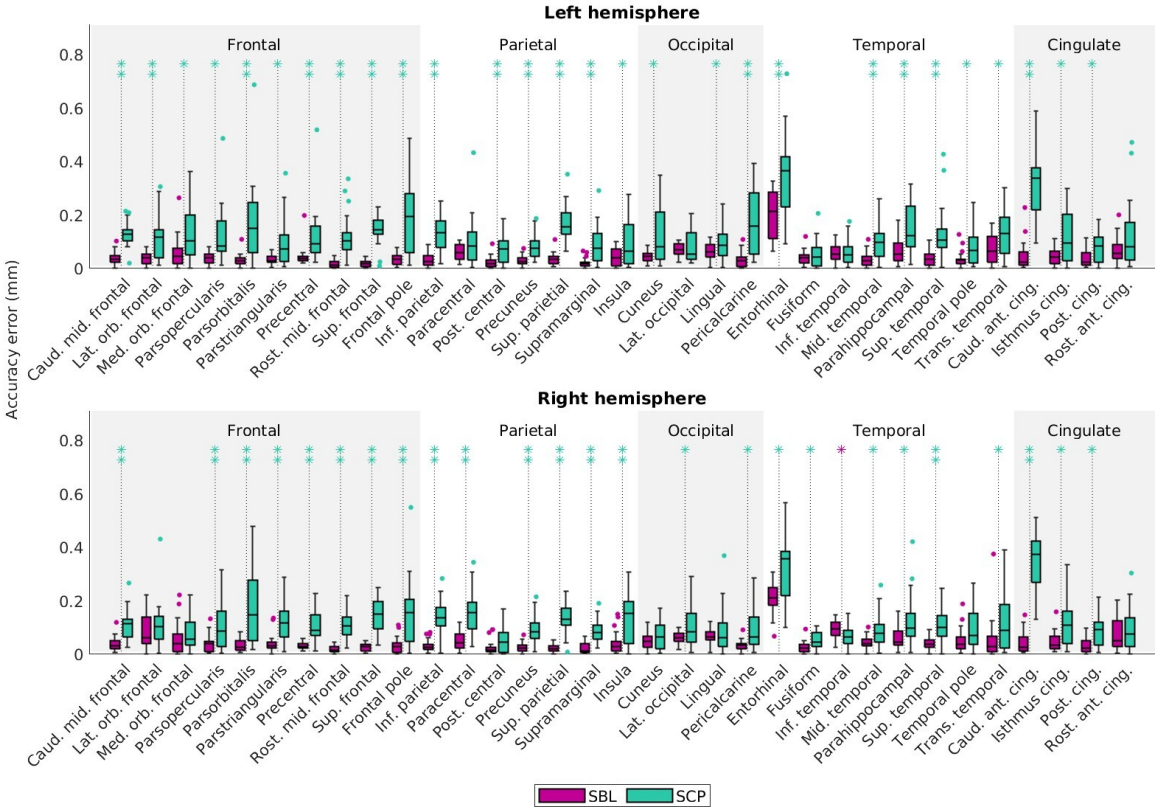

Figure 5.15: Error in thicknesses change measured using the SBL (pink) and SCP (green) compared to ground truth. Dots of the represent subjects whose mean thickness in the corresponding ROI was an outlier. An asterisk designates a significant difference ($p < 0.05$), where the corresponding color represents the thickness measurement with the higher mean error. Two asterisks designates significance after applying a Bonferroni correction equal to the number of ROIs compared ($n = 66$).

5.5.4 Comparison to Previous Methods

The experimental validation of our surface-based Laplacian method focused on comparison with only the FreeSurfer method. However, we must also contrast its merit to other existing techinques such as the volumetric Laplacian (VL) [85] (see Section 2.4.2), as well as the Laplacian-Betrami heat kernel method introduced in Wang et al. [201]. The VL, although better able to represent the complex geometry of the human cortex than the SCP distance mapping, is affected by numerous shortcomings of volumetric methods; these include errors induced by partial volume effect (PVE) and image noise, and a higher susceptibility to topological defects. Issues regarding PVE are particularly problematic within deep, narrow sulci where CSF can be difficult to resolve at a typical image resolution. While numerous approaches have been taken to overcome this [2, 78, 79, 85], they are all still limited by image resolution. Further, finite differences in volumetric domains are restricted to regular (uniform) lattices, which do not represent complex structures such as the human cortex as accurately as an irregular (non-uniform) grid.

The Laplacian-Betrami heat kernel (LBKH) approach [201] employs a similar framework to our proposed SBL in that the Laplacian equation is solved over an irregular grid bound between the GM and WM cortical surfaces, but there exist several critical differences compared to our work. Firstly, the LBKH initializes their grid from the input MRI, whereas our mesh generation relies on only the bounding surfaces and therefore is not affected by image resolution. Secondly, the LBKH uses finite *element* methods to solve for the Laplacian solution, which are far more computationally involved than the finite *volume* methods in our pipeline. Wang et al. make no mention of processing time in their manuscript. Thirdly, they calculate thickness using a heat kernel estimation over the Laplacian field, and measure thickness as the distance traversed by random-walk diffusion from one boundary to the other. This almost always yields a path much longer than the Laplacian streamlines used in both our SBL and traditional VL methods, which increases both computation time and the number of iterations to travel across the GM ribbon; because the direction travelled at each step is calculated by interpolating field values at nearby mesh points, a higher number of steps leads to a higher susceptibility to interpolation-induced error. Although Wang et al. demonstrated that the LBKH calculated CTh measurements with higher statistical power than FreeSurfer, they did not assess the accuracy of their method. Further, all experimental validation was conducted with cross-sectional data rather than longitudinal, and thus no indication of their methods sensitivity to thickness *changes* was discussed.

5.6 Conclusion

In this work, we presented a surface-based Laplacian (SBL) method to calculate cortical thickness. We validated this method in phantom data to show that our SBL method yields more symmetric and higher thickness values with better fidelity to surface geometry than the SCP. We then evaluated our method in cortical data to find an appropriate decimation factor that balances thickness accuracy and processing time. We show that the SBL pipeline produces similar thickness patterns in each hemisphere and mimics well-known thickness patterns of the human brain. Finally, we demonstrated that the SBL produces thickness values that more accurately represent the ground truth change in synthetic data, and that the SBL exhibits similar reproducibility compared to the SCP.

CHAPTER 6

Cortical Thickness in the Natural History of Huntington's Disease Progression

6.1 Abstract

Huntington's disease (HD) is an autosomal dominant, neurodegenerative disorder that results in severe motor and cognitive decline. The most common image-based marker for disease progression is subcortical atrophy in the putamen and caudate nucleus, which has been shown to correspond to the severity of motor symptoms. Studies have also identified decreases in the thickness of cortical gray matter (GM) thickness that to the progression of cognitive and behavioral symptoms, but these reports vary widely between studies, partially due to limitations of current methods to measure cortical thickness (CTh). In this aim, we quantify regional CTh changes in a longitudinal HD dataset using the surface-based Laplacian (SBL) technique presented in Chapter 5. We show that, whereas little significance was identifiable in previous studies using alternative thickness mappings, SBL thickness is a significant image-based marker for HD progression.

6.2 Introduction

Huntington's disease (HD) is a rare, autosomal dominant, genetic disorder that is characterized by widespread neurodegeneration and leads to severe motor and cognitive decline. It is caused by a distinct genetic mutation that consists of an elongated sequence of cytosine-adenosine-guanine (CAG) trinucleotide repeats on the Huntingtin (HTT) gene [112]. A sequence containing 26 or less repeats is considered healthy, whereas 35 or more repeats generally indicates disease presence, although this threshold varies between studies [131, 138, 172, 199]. Unfortunately, although genetic testing can determine whether or not someone will develop the disease, it provides no indication of which symptoms that patient will experience or when those symptoms will manifest. This has led researchers to seek alternative means, such as image-based markers with MRI, to study disease progression in at-risk individuals.

The most well quantified image-based marker for HD progression is subcortical atrophy within the striatum [14, 17, 19, 145]. One study in particular [19] found that the putamen and caudate began to exhibit volumetric loss up to nine and eleven years, respectively, prior to the onset of motor symptoms. Other subcortical areas exhibiting significant atrophy are the globus pallidus [163], substania nigra [5], thalamus [90], nucleus accumbens [26], and amygdala [115]. Of these, almost all volumetric changes have been linked only to motor symptoms, with the exception of the amygdala and thalamus. This does not fully explain the variability of cognitive and behavioral symptom expression across patient cohorts.

In addition to subcortical structures, studies have also identified significant changes in cortical GM thick-

ness with HD progression that correspond more strongly to non-motor symptoms [15, 36, 147, 148, 173]. The inclusion of cortical thickness (CTh) thus has the potential to provide a better explanation for individual symptom expression, and could be highly useful for modelling the natural history of HD progression. However, due in part to both the limitations of CTh measurement technology and small sample sizes, regional reports of GM thinning vary between studies. For example, several studies [36, 146] observed significant cortical atrophy in the occipital lobes compared to age-matched controls, whereas others found no significant GM atrophy at all [16, 91, 206].

Analyses studying regional cortical GM in HD generally quantify CTh changes using either voxel-based morphometry [36, 91, 151] or with the FreeSurfer (FS) software suite [38, 141] that measures thickness as the distance between cortical gray and white matter surface reconstructions [148, 173]. Both types of methods present their own sets of advantages and disadvantages; volumetric methods are more computationally effect, while surface-based methods are less susceptible to image noise and partial volume effect. Although the FS method has been shown to be highly robust in a number of disease applications including HD, it uses the symmetric closest point (SCP) distance mapping to measure CTh, which is prone to underestimating thickness in high curvature areas [123]. On the other hand, the surface-based Laplacian (SBL) mapping presented in Chapter 5 of this dissertation was proven to be much more accurate, suggesting that its application to study GM atrophy in HD could provide better results than previously achievable.

In this work, we apply the SBL method for CTh measurement from Chapter 5 to study GM atrophy throughout HD progression. We hypothesize that given its higher accuracy, we will be able to detect more pronounced, widespread thickness differences between CAP groups compared to previous studies that used the SCP or alternative distance mappings. These group differences will provide a more thorough understand of which specific GM regions experience atrophy throughout HD progression. We also conduct similar analyses for clinical metrics that indicate the progression of cognitive and motor symptoms; these include total motor score (TMS), symbol digit modality test (SDMT) score, trail making test (TMT) scores (a/b), and Stroop color word test (SCWT) scores (color/interference/total).

6.3 Methods

6.3.1 The PREDICT-HD Dataset

The analyses within this chapter are conducted using the PREDICT-HD dataset with individuals positive for the HTT mutation and age-matched, healthy controls. The set of HD patients includes both premanifest and manifest individuals: premanifest subjects are at risk but have not yet began to exhibit severe symptoms, while manifest subjects have already developed symptoms that significantly impact their quality of life. Both control and HD subjects are associated with up to seven timepoints acquired over a span of up to 12 years.

From the entire dataset, we selected those that had 3 T T1w images for at least two timepoints. Each timepoint contains a T1w and FLAIR image, as well as several UHDRS test scores including total motor score (TMS), symbol digit modality test (SDMT), Stroop color word test (SCWT), and trail making time (TMT), as well as a diagnostic confidence level (DCL). Subjects were sorted into cohorts based on their CAG-Age Product (CAP) score, which is calculated for this dataset as

$$\mathrm{CAP} = \mathrm{Age}_0 \times (\mathrm{CAG} - 33.66) \tag{6.1}$$

Here, Age_0 refers to baseline age for each subject, CAG indicates the number of CAG repeats, and 33.66 is a correction factor obtained using a maximum likelihood estimation of the CAG repeat length of the dataset [219]. HD subjects were divided into three cohorts based on CAP score: $\mathrm{CAP}_{\mathrm{low}}$ ($\mathrm{CAP} < 290$), $\mathrm{CAP}_{\mathrm{med}}$ ($290 \leq \mathrm{CAP} < 368$), and $\mathrm{CAP}_{\mathrm{high}}$ ($\mathrm{CAP} \geq 368$); these cutoffs were determined within the same maximum likelihood study as the correction factor. A full description of the PREDICT-HD dataset is also provided in Section 3.2.3 of this dissertation.

The analyses in this chapter utilized a subset of the PREDICT-HD dataset that consisted of 57 controls, 42 $\mathrm{CAP}_{\mathrm{low}}$ subjects, 54 $\mathrm{CAP}_{\mathrm{med}}$ subjects, and 65 $\mathrm{CAP}_{\mathrm{high}}$ ($n = 218$ total). The remaining subjects were excluded due to processing difficulties such as missing image data and FreeSurfer segmentation errors.

6.3.2 Cortical Thickness Measurement

All subjects were processed using the FreeSurfer software suite (version 7.2) [141] to perform longitudinal cortical segmentation and surface reconstruction. We applied the surface-based Laplacian (SBL) pipeline presented in Chapter 5 to measure CTh within FS surface reconstructions for each subject and timepoint. Thickness values were averaged across cortical ROIs defined by the Desikan Killiany atlas [42]; this parcellation was generated using the FS base template for each subject as part of its longitudinal segmentation algorithm.

6.3.3 Data Normalization

Our next goal was to compare two sets of observations between CAP groups: the mean CTh within each DK atlas region, and the clinical metrics described in section 6.3.1. However, these data all exists within vastly different ranges; CTh and TMS are generally an entire order of magnitude lower than the cognitive test scores (e.g. SDMT, SCW, and TMT). Thus, prior to running any statistical analyses, we normalized each set of observations Y_n, for $1 \leq n \leq 73$ total sets (66 ROIs + 7 clinical metrics), to a standardized z-score using the equation

$$Z = \frac{y - \mu_y}{\sigma_y} \tag{6.2}$$

Here, $y \in Y_n$ is an the individual observation within each set, and μ_y and σ_y are the mean and standard deviations of all $y \in Y_n$. Z-scores were computed using the means and standard deviations over all included subjects, rather than within each cohort. The standardized data for the mean CTh within each cortical ROI is displayed in Figure 6.1, and for the clinical metrics in Figure 6.2.

6.3.4 Statistical Analysis

Next, we fit each set of z-scores using a linear mixed effect (LME) model that corrected for several different fixed effects within each cohort group. Each of these n number of models took the form

$$z_{it} = \beta_0 + \sum_k \beta_k f_{itk} + \sum_j \beta_j g_{itj} + b_i + \varepsilon_{it} \tag{6.3}$$

where i indicates subject ID, t indicates timepoint, and β_0 is the model intercept. k iterates over each fixed effect f, and j iterates over each CAP group g. Control subjects are not represented in the group terms. Each β_k term corresponds to the coefficient associated with the k^{th} fixed effect, which varied depending on the specific observation z_{it}. b_i was the random error associated with each subject, and ε_{it} the residual error to account for otherwise unexplained sources of variation. Both error terms were assumed to be normally distributed and independent of each other.

The fixed effects for each model estimating cognitive test scores or CTh values were binary sex classification (male/female), education level (number of years), age at baseline (age_0), and duration within the study ($\text{age}_t - \text{age}_0$). Including both duration and baseline age allowed us to account for the fact that our inference of each metric can only begin at a subject's entrance into the study, rather than at disease onset. For the model estimating TMS,the fixed effects were binary sex, baseline age, and duration in study (education level was excluded).

For all models, our target objects of inference were the coefficients β_j corresponding to each group g_{ij}, where j indicates CAP_{low}, CAP_{med}, or CAP_{high}. We set each g_{ij} equal to either 0 or 1 depending on the cohort for a given subject. For example, for a subject in the CAP_{med}, equation 6.3 would take the form

$$\begin{aligned} z_{it} &= \beta_0 + \sum_k \beta_k f_{itk} + \beta_{\text{low}} \times 0 + \beta_{\text{med}} \times 1 + \beta_{\text{high}} \times 0 + b_i + \varepsilon_{it} \\ &= \beta_0 + \sum_k \beta_k f_{itk} + \beta_{\text{med}} + b_i + \varepsilon_{it} \end{aligned}$$

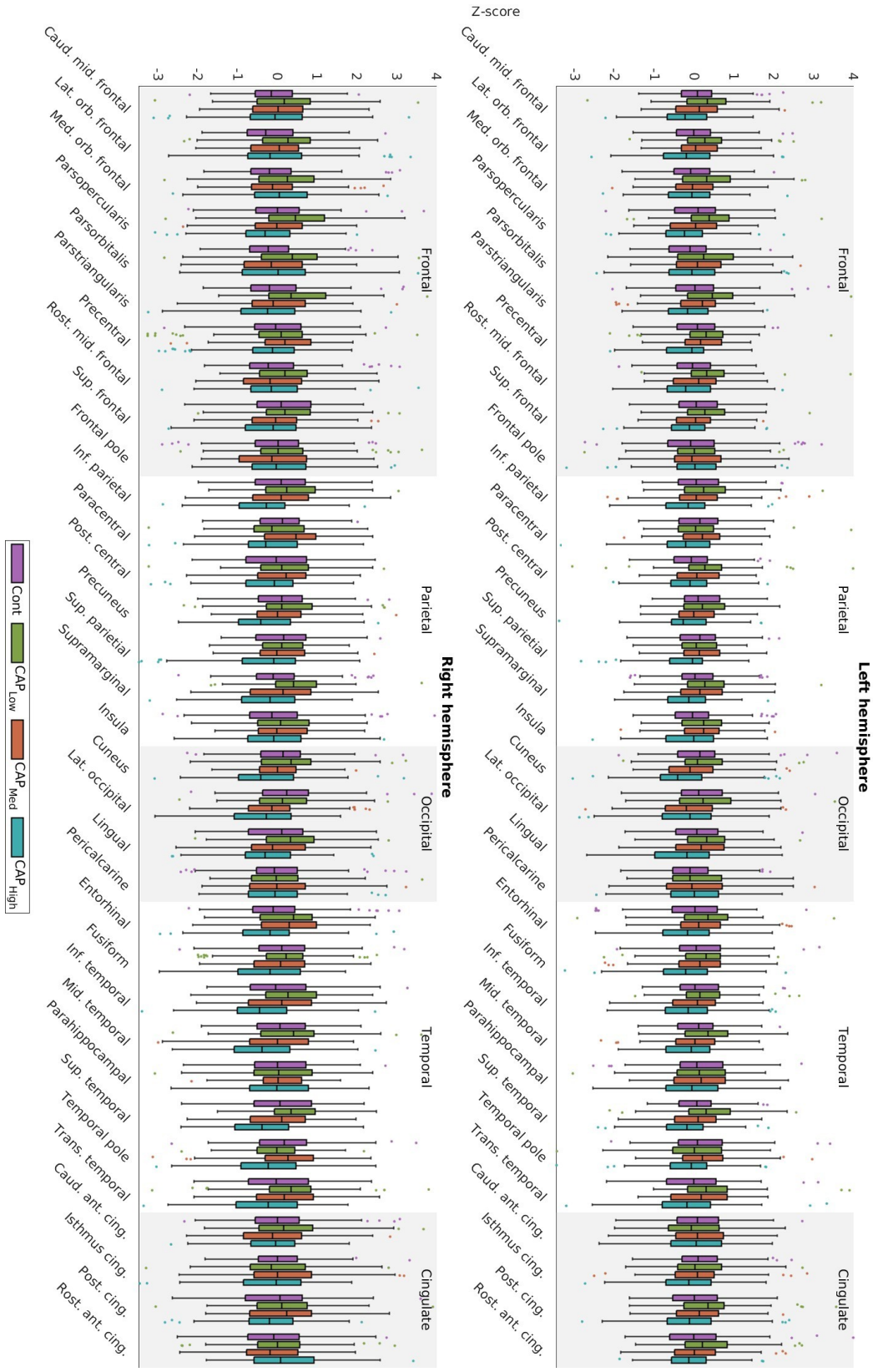

Figure 6.1: Z-scores of mean CTh values within each DK atlas cortical ROI for controls (purple), CAP_{low} (green), CAP_{med} (orange), and CAP_{high} (blue). Dots represent outliers within the corresponding group.

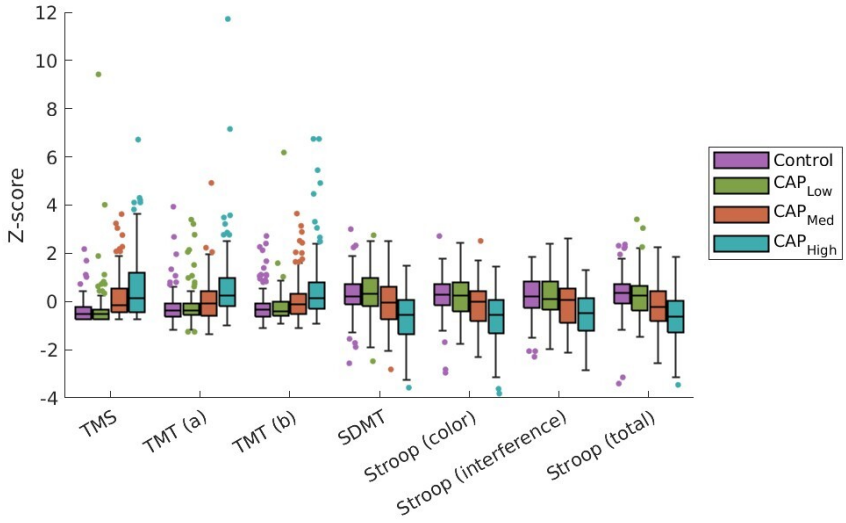

Figure 6.2: Z-scores of clinical test scores for controls (purple), CAP_{low} (green), CAP_{med} (orange), and CAP_{high} (blue). Dots represent outliers within the corresponding group.

The consequences of this are that the β_j terms fit by each model represents the difference between each corresponding CAP group and the control group. Using TMS as an example, this means that given two subjects who have the same age, duration in study, binary sex, and education level, where one subject is CAP high and the other a control, the difference in their normalized TMS values will be equal to β_{high}, within a margin of error dictated by the corresponding b_i and ε_{it} terms.

To recap, our statistical framework resulted in a total of 73 fitted models for the entire dataset: 66 that corresponded to CTh in a particular ROI (33 per hemisphere), and 7 that corresponded a clinical metric. Each model was implemented using the fitlme built-in Matlab function [1]. This analysis is based on a similar cortical shape evaluation presented by Stoeber et al., 2022[173].

6.4 Results

6.4.1 Cortical Thickness

Based on the z-score data in Figure 6.1, we see several regions with a monotonic decrease in CTh in both hemispheres across CAP groups. These include the parsopercularis and parstriangularis within the frontal lobe; the inferior parietal gyrus, precuneus, and supramarginal gyrus within the parietal lobe; the cuneus and lingual cortex within the occipital lobe; and the superior and transverse gyri within the temporal lobe. On the other hand, regions such as the left caudal middle frontal and right lateral occipital gyri decrease across CAP groups in one hemisphere, but exhibit no such trend in the other. However, because none of these data have

[1] https://www.mathworks.com/help/stats/fitlme.html

been corrected for age, sex, or duration in study, more analysis was needed to remove any confounds from the normalized data.

Figure 6.3 contains the estimated β_j intercepts from equation 6.3 where the object of inference z_{it} was the normalized mean thickness value for each subject and timepoint. Because these data account for fixed effects, the trends exhibited across CAP groups between these intercept values are better indicators of disease progression than the raw z-scores in Figure 6.1. CTh within regions such as the right superior frontal gyrus appeared relatively constant based on z-scores, but exhibit a clear trend in intercept values in 6.3. On the other hand, regions such as both entorhinal gyri exhibited the opposite trend, with z-scores decreasing between CAP groups but less change between intercepts. Note that Figure 6.3 does not display data corresponding to the control cohort due to the structure of our models.

For the group intercept data, we performed an omnibus statistical analysis to determine whether the mean thicknesses for each CAP group and region were significantly different from those of the control group. Our analysis took the form of an F-test with the null hypothesis that for each region, the group intercepts for each CAP group were equal to 0 (e.g., $\beta_j = 0$ for each j). A rejected null hypothesis indicates that $\beta_j \neq 0$ for at least one CAP, which implies that the corresponding cortical region experiences a significant change in CTh at some point during HD progression. The results of this omnibus testing are represented by the asterisks in Figure 6.3; one asterisk * indicates $p < 0.05$, two ** indicate $p < 0.01$, and three *** indicate $p < 0.05/n$, where $n = 66$ accounts for the multiple comparisons (one for each cortical ROI). Only four regions maintained significance after correcting for multiple comparisons: the parsopercularis, inferior parietal, lateral occipital, and middle temporal.

Figure 6.4 displays the ranges of the corrected, z-scored, mean thickness values for each cortical region and cohort. The trends exhibited within each region are similar to those present in the group intercepts shown in Figure 6.3. The difference in normalized CTh between the CAP_{low} and CAP_{high} groups indicates that the ROI does atrophy throughout HD progression. The proximity of the cluster corresponding to CAP_{med} to either the CAP_{low} or CAP_{high} clusters suggests when this thinning occurs throughout disease progression. For example, in the right parsopercularis and parstriangularis regions, the CAP_{low} and CAP_{med} are relatively similar compared to the CAP_{high}, whereas within the right lateral, middle, and superior gyri, the change is more evenly distributed between groups. This implies that thinning of these temporal regions occurs earlier within disease progression than thinning of the two frontal lobe regions.

Lastly, Figure 6.5 displays individual model estimates (rather than aggregated into a box plot) for the normalized mean CTh in the four regions that maintained significance following correction for multiple comparisons. CTh is plotted for each HD subject as a function of their CAP score at each timepoint. In all four regions, we observe a downward trend in thickness across CAP groups, with a larger difference between

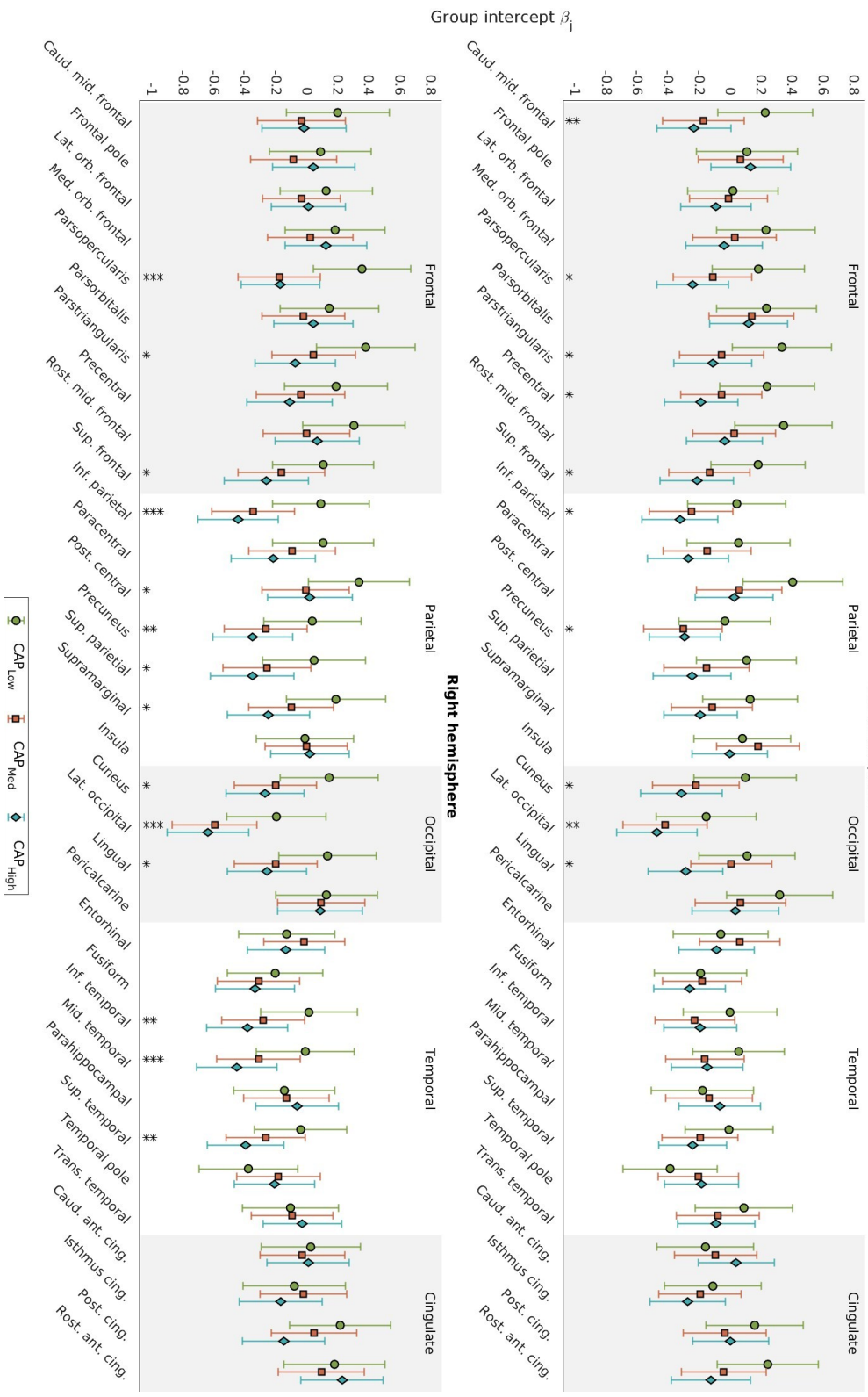

Figure 6.3: Estimated group intercepts of each normalized mean CTh for CAP_{low} (green), CAP_{med} (orange), and CAP_{high} (blue). Each marker represents the model estimated β_j value (see equation 6.3), and error bars correspond to 95% confidence intervals (CI).

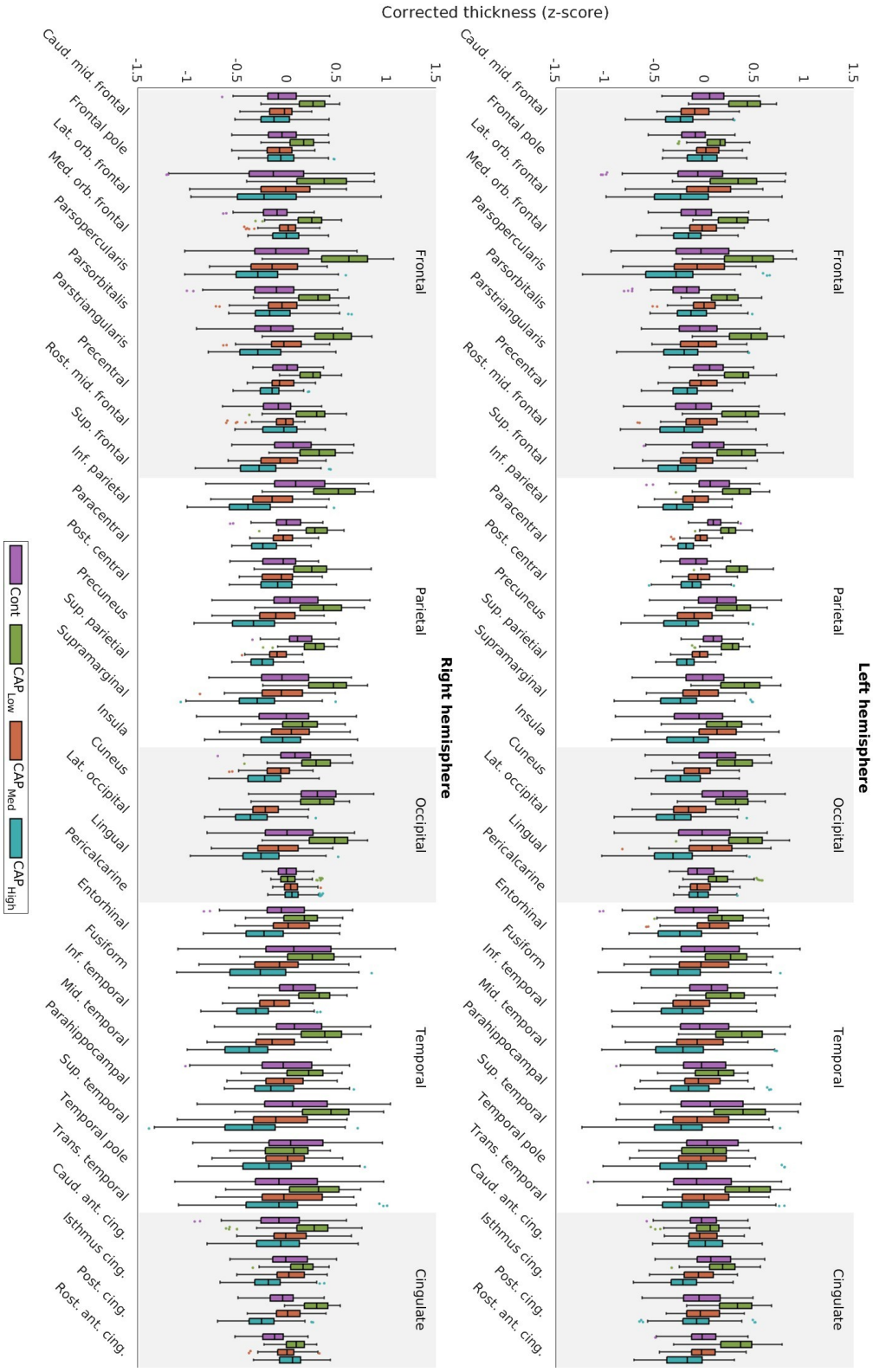

Figure 6.4: Box plots of corrected z-scores for mean CTh in controls (purple), CAP_{low} (green), CAP_{med} (orange), and CAP_{high} (blue).

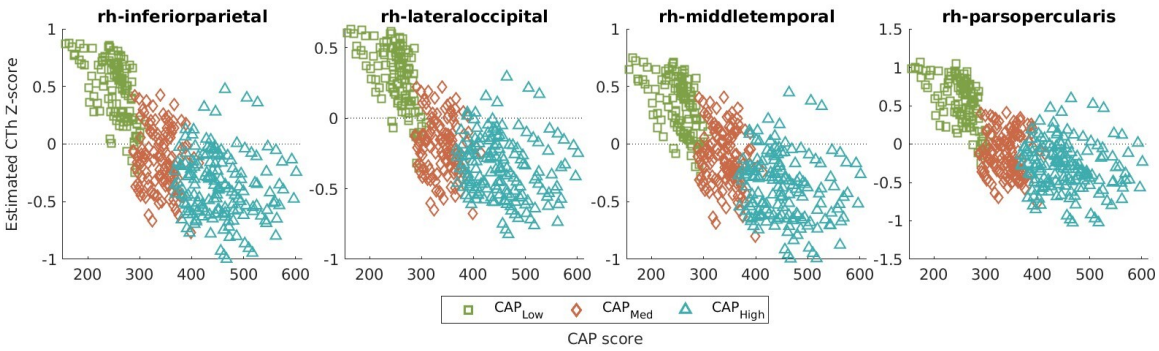

Figure 6.5: Model estimates for normalized mean thickness within select ROIs for CAP_{low} (green), CAP_{med} (orange), and CAP_{high} (blue). The horizontal dotted line indicates $z = 0$.

CAP_{low} and CAP_{med} than between CAP_{med} and CAP_{high}. Similar plots for all DK atlas regions are included in Appendix C. Controls are not included in these plots because they were not associated with a CAP score in the PREDICT-HD dataset.

6.4.2 Clinical Metrics

We also present results from a similar analysis of the normalized clinical metric scores. Figure 6.6 displays the estimated β_j intercepts from equation 6.3 where the z_{it} was one of the seven included clinical metrics. All observed variables except for total motor score (TMS) were corrected for age, binary sex, and education level; TMS was only corrected for age and sex. Similar to Figure 6.5, the differences between the CAP_{low} (green) and CAP_{high} (blue) clusters in Figure 6.7 represent the amount of change throughout disease progression, whereas the proximity of the CAP_{med} (orange) cluster to the others suggests the timing of this change. The monotonic trends are consistent between Figures 6.6 and 6.7, and similarly to the CTh data, the intercepts exhibit clearer trends than the uncorrected z-scores in Figure 6.2.

6.5 Discussion

6.5.1 Implication of Findings

Overall, the changes identified for both CTh and clinical metrics mostly agree with those reported in previous literature [109, 126, 146–148, 173, 174]. We observed bilateral atrophy in at least two ROIs within all areas of the brain except for the cingulate gyrus. However, only ROIs in the right hemisphere maintained significance after correcting for multiple comparison. One potential explanation for this is that the mean CTh values were not corrected for handedness, as that data was not contained in the PREDICT-HD demographics. Further, the low number of regions that maintained significance is likely due to a limited sample size; including the entire PREDICT-HD dataset in subsequent analyses would yield a higher number of significant regions.

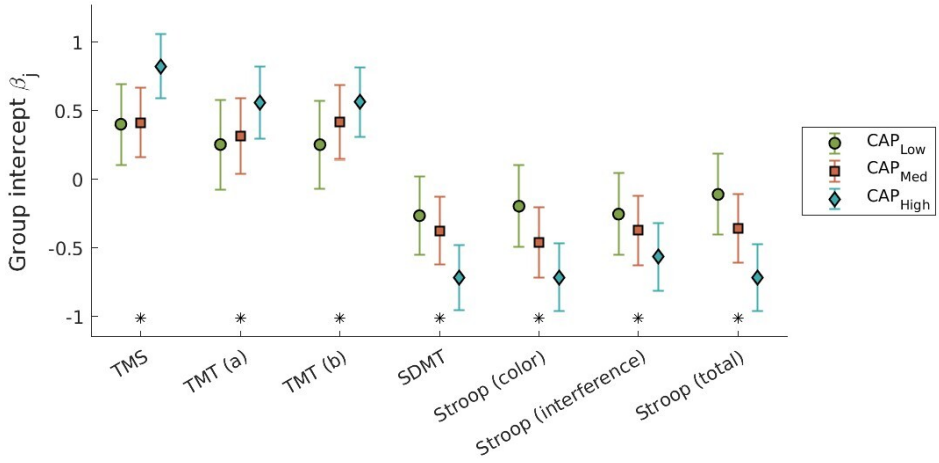

Figure 6.6: Estimated group intercepts for each normalized clinical test score for controls (purple), CAP_{low} (green), CAP_{med} (orange), and CAP_{high} (blue). Each marker represents the model estimated β_j value (see equation 6.3), and error bars correspond to 95% confidence intervals (CI). All regions expressed significant differences at p ¡ 0.05/*n* (denoted by an asterisk), for $n = 7$ comparisons.

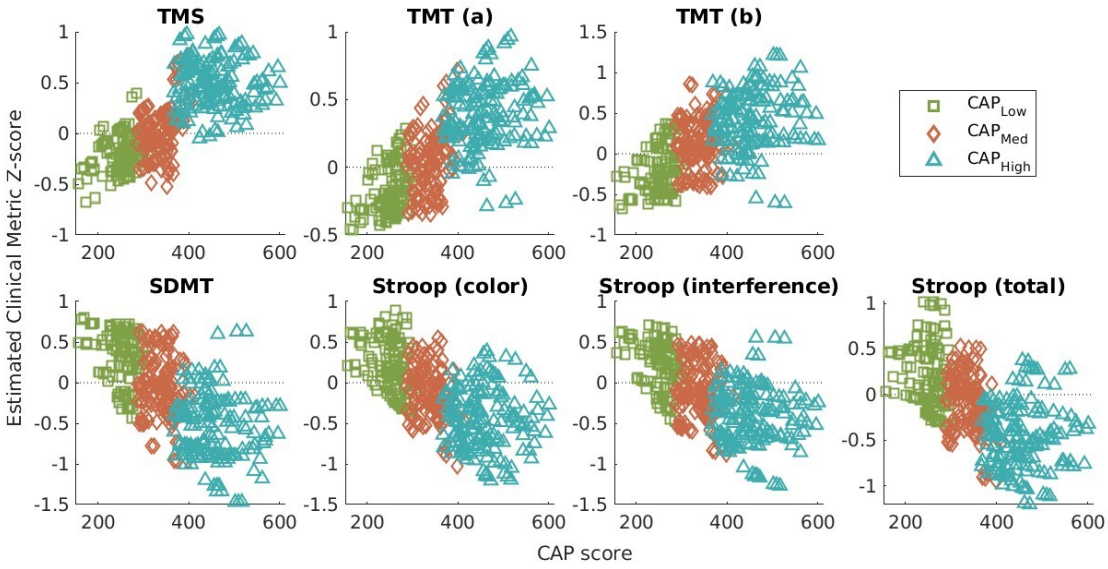

Figure 6.7: Model estimates of clinical test scores for CAP_{low} (green), CAP_{med} (orange), and CAP_{high} (blue).

In addition to bilateral atrophy, we also found that most regions that exhibited significant differences identified in the omnibus statistical analysis tended to experience greater differences between the CAP_{low} and CAP_{med} than CAP_{med} and CAP_{high} groups. The left lingual cortex was the only significant region that exhibited higher differences between the two latter CAP groups, which contradicts the findings in [148, 173] that cortical atrophy progresses from posterior to anterior regions across disease progression. This discrepancy is maybe due to the difference thickness mappings used between our analysis and theirs, as both prior studies used the SCP mapping. A deeper analysis of the full PREDICT-HD dataset beyond the subset included in this dissertation is required to better compare our findings.

We also observed that the group intercepts corresponding to the clinical metrics exhibit higher amounts of change between CAP_{med} and CAP_{high}. Because the CTh intercepts in many regions differed more between CAP_{low} and CAP_{med}, this suggests that cortical thinning could occur in those areas prior to symptom onset in a similar manner to other brain areas such as the caudate and putamen.

6.5.2 Comparison to Prior Work

One major source of the discrepancies regarding the extent of localized cortical atrophy within HD literature arises from the use of either volumetric methods or symmetric closest point (SCP) distances to measure CTh. In our analyses, we employed the surface-based Laplacian (SBL) method presented in Chapter 5. This method was shown to provide more accurate thickness measurements than the SCP, which lends itself to more statistically powerful results within a given cohort. For example, the CTh analyses conducted in Stoebner et al., 2022 [173] measured thickness using the SCP mapping in the same PREDICT-HD cohort of subjects with 3 T image data. In their work, they found very few areas with significant changes compared to our findings presented in this Chapter. This indicates that the SBL appears to find more widespread atrophy patterns than the SCP, likely due to its increased sensitivity to thickness changes.

6.5.3 Future Work

The next step for these data is to more thoroughly investigate the relationships between clinical metrics and the CTh regions that exhibit a downward trend across CAP groups. The normalization of all data into a standard range allows us to observe roughly when each variable changes along the disease progression. For example, CTh in the right pre-central gyrus (i.e., the motor cortex) and TMS both are relatively similar between CAP_{low} to CAP_{med}, but differ greatly between CAP_{med} and CAP_{high}. Although more advanced analyses are required to determine any significant correlations, as discussed in Chapter 7, these observations are promising for the use of regional CTh as an image-based marker.

6.6 Conclusion

The data presented in this chapter suggests that regional CTh could serve as a useful image-based marker for HD progression. Our use of the surface-base Laplacian to measure CTh allowed us to identify more widespread atrophy than previously capable due to its increased sensitivity to thickness changes compared to the SCP. We observed that some regions experienced thinning earlier in disease progression than others based on the differences in intercept values between different cohorts. These results, in combination with a more advanced covariate analyses as mentioned in Section 6.5 suggest that CTh measured with the SBL may provide a better explanation for cognitive symptoms, thereby leading to a better understanding of HD progression and symptom manifestation.

CHAPTER 7

Conclusions, Limitations, and Potential for Future Work

In this final chapter, we summarize the research presented in this dissertation with an emphasis on limitations and potential for future work.

7.1 Synthetic Atrophy for Ground Truth Validation

In Chapter 4, we presented a registration-based synthetic atrophy pipeline for accuracy validation of cortical segmentation and thickness measurement. Our method operates by performing a series of image morphology operations on a binary cortical GM ROI to erode a layer of voxels from the GM/CSF boundary. The eroded ROI is then registered to the original, resulting in a deformable transformation that can be applied to the anatomical images and any associated data such as cortical surfaces or landmarks. Unfortunately, methods founded in registration often induce blurring around the GM/CSF interface due to interpolation, which could cause difficulties when using the resulting image data for cortical segmentation tasks. Further, they can also struggle to induce atrophy within deep sulci where CSF is not originally visible due to partial volume effects, as this prevents deformable models from expanding CSF in the atrophied images despite GM thinning.

A potential solution to both of these shortcomings would be to apply an inpainting technique to the image data after synthetic atrophy is induced. Inpainting is often used to replace missing or corrupted regions in a volume, and been successfully implemented with a variety of different deep learning methods [9, 202, 217]. For example, Zhang et al., 2020 [217] proposed a pipeline to remove WM lesions from brain MRI and replace those voxels with inpainted, healthy WM tissue. In particular, this method employed an edge-map as additional input prior to help create more realistic boundaries between the inpainted and non-inpainted voxels. Because one proposed application for our synthetic atrophy data is cortical segmentation, the ability to exploit boundary information to generate realistic tissue interfaces renders this work potentially applicable to the task of CSF inpainting.

Another pitfall of our presented synthetic atrophy method concerns the correspondence between landmark placement and the cortical parcellation from with GM ROIs are selected. Although in our method we generated a label map specific to each landmark cluster, each ROI was thresholded to remain within the cortical ribbon as defined by a volumetric tissue labelling produced by FreeSurfer (FS) [38] regardless of whether or not the actual landmarks existed inside those voxels. Because the landmarks themselves were not restricted to a voxel grid as the label maps were, there were many instances when the landmarks lay outside the input ROI. In these cases, when applying the resulting deformable registration to induce atrophy within the MRI,

the GM/CSF interface targeted by the transformation did not correspond exactly to the boundary indicated by the landmarks; this implies that the new GM/CSF interface in the atrophied image may not correspond to the deformed landmark location. A potential solution for this would be to ensure that the label map used to define the cortical ribbon corresponds more precisely to landmark location, which would likely require a higher resolution. However, because we conducted the erosion operations and ROI deformation with a 400% increase in resolution in each dimension, this could fit easily within our established framework.

7.2 Cortical Thickness in the Natural History of Huntington's Disease

In Chapter 6, we modeled CTh within each Desikan Kiliany atlas [42] cortical ROI using a linear mixed effect (LME) model. Using the estimated group intercepts, we found that 39 out of the total 66 regions exhibit a downward trend across CAP groups, suggesting that these regions show potential to serve as image-based markers for HD progression. The next step for the data is to apply a covariance-based analysis to study the correlation between clinical metrics and CTh in areas where decreases occurred across CAP groups. This will enable us to more thoroughly examine the relationship between GM atrophy and clinical metrics.

In addition to examining the relationship between CTh and the metrics included in Chapter 6, it would be beneficial to study correlations between CTh and volumetric loss in subcortical structures. For example, the putamen and caudate circuits both connect to the premotor and supplementary motor cortices; the implications of both regions experiencing atrophy at similar times, or one experiencing atrophy but not the other, would provide more specific insight into the degenerative mechanisms of HD progression. The PREDICT-HD dataset also contains more UHDRS scores than used within this dissertation; this includes total functional capacity (TFC), as well as more specific motor scores indicating the presence and severity of a number of different motor abnormalities including chorea, bradykinesia, ataxia, akinesia, dysarthria, and dystonia. Supplementing our CTh analysis with subcortical data, which is the image-based marker most correlated with motor deficits, could a better explanation for the prevalence of specific movement disorders in HD patients.

7.3 Closing Remarks

In conclusion, the research presented within this dissertation follows a natural progression, where each specific aim supports the next. In Aim 1 (Chapter 4), we developed a registration-based, synthetic atrophy method specifically designed for the accuracy validation of surface-based cortical segmentation and thickness measurement pipelines. We used this method in Aim 2 (Chapter 5) to validate our surface-based Laplacian (SBL) CTh pipeline, which we showed produced comparably reproducible and much more accurate thickness measurements with respect to ground truth change compared to the most widely used method. Finally, we applied our SBL technique to measure CTh changes within the PREDICT-HD dataset, which allowed us to

observe more widespread thickness changes across the cortex compared to previous studies. Our work suggests that CTh measured with the SBL has the potential to provide a better understanding of HD progression and symptom manifestation.

Appendix A

Aim 1: Unabridged Data

This section of the appendix contains the unabridged data from Chapter 4 of this dissertation. The mean MSDD for each ROI reported in A.1 is depicted as a cortical parcellation in Figure 4.6. The segmentation errors contained in Tables A.2 and A.3 are the same data displayed in A.1-A.4. Representative subsets of these figures are shown in Figures 4.8-4.11. Unfortunately, LaTeX has forced the rest of this page to remain blank.

Table A.1: Synthetic atrophy limits for each ROI selected from the DK atlas. True change in thickness is measured using the mean surface displacement difference (MSDD). This data is also displayed as a cortical parcellation onto an example marching cubes surface in Figure 6. These results are shown here alongside the original thickness and maximum change in thickness measured using the SCP distance formula.

FreeSurfer Cortical Label	**Mean original SCP thickness (mm)**	**Mean SCP thickness at atrophy limit (mm)**	**Mean MSDD thickness change at atrophy limit (mm)**
lh-caudalanteriorcingulate	2.88 ± 0.20	1.43 ± 0.20	1.34 ± 0.23
rh-caudalanteriorcingulate	2.72 ± 0.22	1.24 ± 0.23	1.21 ± 0.26
lh-caudalmiddlefrontal	2.90 ± 0.12	1.90 ± 0.14	1.75 ± 0.15
rh-caudalmiddlefrontal	2.91 ± 0.17	1.82 ± 0.16	1.85 ± 0.17
lh-cuneus	2.11 ± 0.14	1.17 ± 0.10	1.16 ± 0.10
rh-cuneus	2.05 ± 0.12	1.12 ± 0.11	1.10 ± 0.10
lh-entorhinal	2.26 ± 0.22	1.47 ± 0.23	1.63 ± 0.16
rh-entorhinal	2.35 ± 0.24	1.41 ± 0.21	2.07 ± 0.31
lh-fusiform	3.06 ± 0.17	2.24 ± 0.17	2.32 ± 0.16
rh-fusiform	3.06 ± 0.18	2.22 ± 0.14	2.42 ± 0.16
lh-inferiorparietal	2.86 ± 0.10	1.90 ± 0.09	1.87 ± 0.11
rh-inferiorparietal	2.90 ± 0.10	1.90 ± 0.09	1.89 ± 0.14
lh-inferiortemporal	3.18 ± 0.16	2.22 ± 0.12	2.26 ± 0.11
rh-inferiortemporal	3.16 ± 0.19	2.09 ± 0.15	2.36 ± 0.15
lh-isthmuscingulate	2.48 ± 0.13	1.13 ± 0.21	1.30 ± 0.18
rh-isthmuscingulate	2.45 ± 0.16	1.06 ± 0.25	1.25 ± 0.21
lh-lateraloccipital	2.43 ± 0.14	1.68 ± 0.13	1.82 ± 0.13
rh-lateraloccipital	2.51 ± 0.14	1.74 ± 0.11	1.86 ± 0.14
lh-lateralorbitofrontal	2.70 ± 0.16	1.82 ± 0.12	1.98 ± 0.12
rh-lateralorbitofrontal	2.72 ± 0.14	1.84 ± 0.11	2.00 ± 0.10
lh-lingual	2.18 ± 0.19	1.43 ± 0.18	1.62 ± 0.15
rh-lingual	2.21 ± 0.13	1.47 ± 0.12	1.61 ± 0.14
lh-medialorbitofrontal	2.55 ± 0.15	1.28 ± 0.36	1.31 ± 0.37
rh-medialorbitofrontal	2.56 ± 0.22	1.30 ± 0.15	1.33 ± 0.14
lh-middletemporal	3.27 ± 0.13	2.18 ± 0.12	2.11 ± 0.14
rh-middletemporal	3.27 ± 0.18	2.07 ± 0.14	2.14 ± 0.10
lh-parahippocampal	2.44 ± 0.19	1.78 ± 0.13	1.80 ± 0.19
rh-parahippocampal	2.42 ± 0.23	1.69 ± 0.22	1.80 ± 0.20
lh-paracentral	2.84 ± 0.12	1.70 ± 0.11	1.60 ± 0.20
rh-paracentral	2.83 ± 0.19	1.58 ± 0.60	1.67 ± 0.27
lh-parsopercularis	2.91 ± 0.16	1.84 ± 0.14	1.65 ± 0.12
rh-parsopercularis	2.96 ± 0.18	1.85 ± 0.14	1.63 ± 0.15
lh-parsorbitalis	2.98 ± 0.18	1.91 ± 0.16	1.72 ± 0.14
rh-parsorbitalis	3.07 ± 0.17	1.99 ± 0.12	1.76 ± 0.09
lh-parstriangularis	2.98 ± 0.13	2.01 ± 0.11	1.84 ± 0.13
rh-parstriangularis	2.98 ± 0.16	1.95 ± 0.14	1.81 ± 0.17
lh-pericalcarine	1.67 ± 0.17	0.73 ± 0.18	0.76 ± 0.20
rh-pericalcarine	1.80 ± 0.12	0.84 ± 0.08	0.77 ± 0.18
lh-postcentral	2.34 ± 0.15	1.42 ± 0.11	1.37 ± 0.14
rh-postcentral	2.27 ± 0.12	1.35 ± 0.11	1.33 ± 0.10
lh-posteriorcingulate	2.65 ± 0.15	1.06 ± 0.19	1.12 ± 0.24
rh-posteriorcingulate	2.66 ± 0.09	1.03 ± 0.17	1.10 ± 0.16
lh-precentral	2.81 ± 0.13	1.76 ± 0.08	1.77 ± 0.11
rh-precentral	2.81 ± 0.15	1.77 ± 0.10	1.83 ± 0.12
lh-precuneus	2.73 ± 0.14	1.56 ± 0.15	1.47 ± 0.14
rh-precuneus	2.66 ± 0.14	1.48 ± 0.13	1.39 ± 0.13
lh-rostralanteriorcingulate	2.80 ± 0.20	1.06 ± 0.13	1.00 ± 0.13
rh-rostralanteriorcingulate	2.68 ± 0.23	0.93 ± 0.22	0.97 ± 0.20
lh-rostralmiddlefrontal	2.87 ± 0.12	1.90 ± 0.10	1.89 ± 0.13
rh-rostralmiddlefrontal	2.90 ± 0.14	1.91 ± 0.12	1.92 ± 0.15
lh-superiorfrontal	3.05 ± 0.14	1.95 ± 0.10	1.97 ± 0.11
rh-superiorfrontal	3.05 ± 0.15	1.92 ± 0.11	1.99 ± 0.11
lh-superiorparietal	2.54 ± 0.12	1.61 ± 0.10	1.54 ± 0.11
rh-superiorparietal	2.51 ± 0.13	1.57 ± 0.12	1.52 ± 0.12
lh-superiortemporal	3.17 ± 0.17	2.05 ± 0.12	2.07 ± 0.16
rh-superiortemporal	3.15 ± 0.23	1.98 ± 0.18	2.07 ± 0.13
lh-supramarginal	2.88 ± 0.16	1.88 ± 0.12	1.78 ± 0.14
rh-supramarginal	2.82 ± 0.20	1.82 ± 0.14	1.72 ± 0.17
lh-transversetemporal	2.26 ± 0.37	1.22 ± 0.29	1.12 ± 0.21
rh-transversetemporal	2.27 ± 0.39	1.10 ± 0.21	0.99 ± 0.24
lh-insula	3.27 ± 0.14	1.98 ± 0.10	2.05 ± 0.18
rh-insula	3.26 ± 0.24	1.97 ± 0.17	2.09 ± 0.23

Table A.2: Mean unsigned segmentation errors of FreeSurfer cortical surface reconstructions for each set of landmarks in the VDCRA dataset.

		Left GM				Right GM			
		3D (A)	**4D (A)**	**3D (B)**	**4D(B)**	**3D (A)**	**4D (A)**	**3D (B)**	**4D(B)**
CALC	**HC (0)**	0.37 ± 0.09	0.37 ± 0.09	0.37 ± 0.09	0.37 ± 0.09	0.27 ± 0.04	0.27 ± 0.04	0.27 ± 0.04	0.27 ± 0.04
	HC (1)	0.48 ± 0.05	0.48 ± 0.05	0.48 ± 0.05	0.48 ± 0.05	0.49 ± 0.24	0.49 ± 0.24	0.49 ± 0.24	0.49 ± 0.24
	MS (0)	0.49 ± 0.15	0.49 ± 0.15	0.49 ± 0.15	0.49 ± 0.15	0.51 ± 0.16	0.51 ± 0.16	0.51 ± 0.16	0.51 ± 0.16
	MS (1)	0.48 ± 0.18	0.48 ± 0.18	0.48 ± 0.18	0.48 ± 0.18	0.57 ± 0.10	0.57 ± 0.10	0.57 ± 0.10	0.57 ± 0.10
CING	**HC (0)**	0.76 ± 0.27	0.76 ± 0.27	0.76 ± 0.27	0.76 ± 0.27	0.63 ± 0.16	0.63 ± 0.16	0.63 ± 0.16	0.63 ± 0.16
	HC (1)	0.77 ± 0.20	0.77 ± 0.20	0.77 ± 0.20	0.77 ± 0.20	0.66 ± 0.17	0.66 ± 0.17	0.66 ± 0.17	0.66 ± 0.17
	MS (0)	1.11 ± 0.64	1.11 ± 0.64	1.11 ± 0.64	1.11 ± 0.64	1.02 ± 0.73	1.02 ± 0.73	1.02 ± 0.73	1.02 ± 0.73
	MS (1)	0.71 ± 0.28	0.71 ± 0.28	0.71 ± 0.28	0.71 ± 0.28	0.91 ± 0.62	0.91 ± 0.62	0.91 ± 0.62	0.91 ± 0.62
CS	**HC (0)**	0.47 ± 0.13	0.47 ± 0.13	0.47 ± 0.13	0.47 ± 0.13	0.39 ± 0.17	0.39 ± 0.17	0.39 ± 0.17	0.39 ± 0.17
	HC (1)	1.13 ± 0.23	1.13 ± 0.23	1.13 ± 0.23	1.13 ± 0.23	0.79 ± 0.16	0.79 ± 0.16	0.79 ± 0.16	0.79 ± 0.16
	MS (0)	0.53 ± 0.37	0.53 ± 0.37	0.53 ± 0.37	0.53 ± 0.37	0.48 ± 0.11	0.48 ± 0.11	0.48 ± 0.11	0.48 ± 0.11
	MS (1)	0.79 ± 0.39	0.79 ± 0.39	0.79 ± 0.39	0.79 ± 0.39	0.55 ± 0.16	0.55 ± 0.16	0.55 ± 0.16	0.55 ± 0.16
PO	**HC (0)**	0.32 ± 0.08	0.32 ± 0.08	0.32 ± 0.08	0.32 ± 0.08	0.37 ± 0.04	0.37 ± 0.04	0.37 ± 0.04	0.37 ± 0.04
	HC (1)	0.57 ± 0.11	0.57 ± 0.11	0.57 ± 0.11	0.57 ± 0.11	0.56 ± 0.06	0.56 ± 0.06	0.56 ± 0.06	0.56 ± 0.06
	MS (0)	0.54 ± 0.11	0.54 ± 0.11	0.54 ± 0.11	0.54 ± 0.11	0.61 ± 0.15	0.61 ± 0.15	0.61 ± 0.15	0.61 ± 0.15
	MS (1)	0.72 ± 0.21	0.72 ± 0.21	0.72 ± 0.21	0.72 ± 0.21	0.78 ± 0.24	0.78 ± 0.24	0.78 ± 0.24	0.78 ± 0.24
SF	**HC (0)**	0.35 ± 0.15	0.35 ± 0.15	0.35 ± 0.15	0.35 ± 0.15	0.37 ± 0.12	0.37 ± 0.12	0.37 ± 0.12	0.37 ± 0.12
	HC (1)	0.47 ± 0.12	0.47 ± 0.12	0.47 ± 0.12	0.47 ± 0.12	0.48 ± 0.12	0.48 ± 0.12	0.48 ± 0.12	0.48 ± 0.12
	MS (0)	0.72 ± 0.13	0.72 ± 0.13	0.72 ± 0.13	0.72 ± 0.13	0.54 ± 0.22	0.54 ± 0.22	0.54 ± 0.22	0.54 ± 0.22
	MS (1)	0.91 ± 0.19	0.91 ± 0.19	0.91 ± 0.19	0.91 ± 0.19	0.81 ± 0.19	0.81 ± 0.19	0.81 ± 0.19	0.81 ± 0.19
ST	**HC (0)**	0.58 ± 0.23	0.58 ± 0.23	0.58 ± 0.23	0.58 ± 0.23	0.54 ± 0.05	0.54 ± 0.05	0.54 ± 0.05	0.54 ± 0.05
	HC (1)	0.61 ± 0.23	0.61 ± 0.23	0.61 ± 0.23	0.61 ± 0.23	0.64 ± 0.14	0.64 ± 0.14	0.64 ± 0.14	0.64 ± 0.14
	MS (0)	0.53 ± 0.15	0.53 ± 0.15	0.53 ± 0.15	0.53 ± 0.15	0.39 ± 0.08	0.39 ± 0.08	0.39 ± 0.08	0.39 ± 0.08
	MS (1)	0.32 ± 0.15	0.32 ± 0.15	0.32 ± 0.15	0.32 ± 0.15	0.31 ± 0.12	0.31 ± 0.12	0.31 ± 0.12	0.31 ± 0.12
SYL	**HC (0)**	0.30 ± 0.03	0.30 ± 0.03	0.30 ± 0.03	0.30 ± 0.03	0.30 ± 0.04	0.30 ± 0.04	0.30 ± 0.04	0.30 ± 0.04
	HC (1)	0.72 ± 0.32	0.72 ± 0.32	0.72 ± 0.32	0.72 ± 0.32	0.60 ± 0.05	0.60 ± 0.05	0.60 ± 0.05	0.60 ± 0.05
	MS (0)	0.46 ± 0.08	0.46 ± 0.08	0.46 ± 0.08	0.46 ± 0.08	0.34 ± 0.08	0.34 ± 0.08	0.34 ± 0.08	0.34 ± 0.08
	MS (1)	0.61 ± 0.23	0.61 ± 0.23	0.61 ± 0.23	0.61 ± 0.23	0.62 ± 0.15	0.62 ± 0.15	0.62 ± 0.15	0.62 ± 0.15

		Left WM				Right WM			
		3D (A)	**4D (A)**	**3D (B)**	**4D(B)**	**3D (A)**	**4D (A)**	**3D (B)**	**4D(B)**
CALC	**HC (0)**	0.56 ± 0.13	0.56 ± 0.13	0.56 ± 0.13	0.56 ± 0.13	0.48 ± 0.09	0.48 ± 0.09	0.48 ± 0.09	0.48 ± 0.09
	HC (1)	0.65 ± 0.22	0.65 ± 0.22	0.65 ± 0.22	0.65 ± 0.22	0.54 ± 0.20	0.54 ± 0.20	0.54 ± 0.20	0.54 ± 0.20
	MS (0)	0.57 ± 0.30	0.57 ± 0.30	0.57 ± 0.30	0.57 ± 0.30	0.34 ± 0.07	0.34 ± 0.07	0.34 ± 0.07	0.34 ± 0.07
	MS (1)	0.49 ± 0.31	0.49 ± 0.31	0.49 ± 0.31	0.49 ± 0.31	0.42 ± 0.11	0.42 ± 0.11	0.42 ± 0.11	0.42 ± 0.11
CING	**HC (0)**	0.27 ± 0.03	0.27 ± 0.03	0.27 ± 0.03	0.27 ± 0.03	0.30 ± 0.05	0.30 ± 0.05	0.30 ± 0.05	0.30 ± 0.05
	HC (1)	0.69 ± 0.19	0.69 ± 0.19	0.69 ± 0.19	0.69 ± 0.19	0.86 ± 0.15	0.86 ± 0.15	0.86 ± 0.15	0.86 ± 0.15
	MS (0)	0.55 ± 0.13	0.55 ± 0.13	0.55 ± 0.13	0.55 ± 0.13	0.63 ± 0.14	0.63 ± 0.14	0.63 ± 0.14	0.63 ± 0.14
	MS (1)	0.65 ± 0.17	0.65 ± 0.17	0.65 ± 0.17	0.65 ± 0.17	0.89 ± 0.21	0.89 ± 0.21	0.89 ± 0.21	0.89 ± 0.21
CS	**HC (0)**	0.32 ± 0.10	0.32 ± 0.10	0.32 ± 0.10	0.32 ± 0.10	0.32 ± 0.06	0.32 ± 0.06	0.32 ± 0.06	0.32 ± 0.06
	HC (1)	0.46 ± 0.11	0.46 ± 0.11	0.46 ± 0.11	0.46 ± 0.11	0.49 ± 0.09	0.49 ± 0.09	0.49 ± 0.09	0.49 ± 0.09
	MS (0)	0.50 ± 0.11	0.50 ± 0.11	0.50 ± 0.11	0.50 ± 0.11	0.53 ± 0.12	0.53 ± 0.12	0.53 ± 0.12	0.53 ± 0.12
	MS (1)	0.58 ± 0.15	0.58 ± 0.15	0.58 ± 0.15	0.58 ± 0.15	0.59 ± 0.09	0.59 ± 0.09	0.59 ± 0.09	0.59 ± 0.09
PO	**HC (0)**	0.48 ± 0.40	0.48 ± 0.40	0.48 ± 0.40	0.48 ± 0.40	0.38 ± 0.11	0.38 ± 0.11	0.38 ± 0.11	0.38 ± 0.11
	HC (1)	0.96 ± 0.21	0.96 ± 0.21	0.96 ± 0.21	0.96 ± 0.21	0.89 ± 0.21	0.89 ± 0.21	0.89 ± 0.21	0.89 ± 0.21
	MS (0)	0.60 ± 0.15	0.60 ± 0.15	0.60 ± 0.15	0.60 ± 0.15	0.54 ± 0.23	0.54 ± 0.23	0.54 ± 0.23	0.54 ± 0.23
	MS (1)	0.75 ± 0.30	0.75 ± 0.30	0.75 ± 0.30	0.75 ± 0.30	0.57 ± 0.16	0.57 ± 0.16	0.57 ± 0.16	0.57 ± 0.16
SF	**HC (0)**	0.29 ± 0.08	0.29 ± 0.08	0.29 ± 0.08	0.29 ± 0.08	0.37 ± 0.07	0.37 ± 0.07	0.37 ± 0.07	0.37 ± 0.07
	HC (1)	0.48 ± 0.13	0.48 ± 0.13	0.48 ± 0.13	0.48 ± 0.13	0.43 ± 0.04	0.43 ± 0.04	0.43 ± 0.04	0.43 ± 0.04
	MS (0)	0.59 ± 0.17	0.59 ± 0.17	0.59 ± 0.17	0.59 ± 0.17	0.68 ± 0.26	0.68 ± 0.26	0.68 ± 0.26	0.68 ± 0.26
	MS (1)	0.62 ± 0.22	0.62 ± 0.22	0.62 ± 0.22	0.62 ± 0.22	0.70 ± 0.28	0.70 ± 0.28	0.70 ± 0.28	0.70 ± 0.28
ST	**HC (0)**	0.38 ± 0.05	0.38 ± 0.05	0.38 ± 0.05	0.38 ± 0.05	0.31 ± 0.14	0.31 ± 0.14	0.31 ± 0.14	0.31 ± 0.14
	HC (1)	0.64 ± 0.18	0.64 ± 0.18	0.64 ± 0.18	0.64 ± 0.18	0.80 ± 0.29	0.80 ± 0.29	0.80 ± 0.29	0.80 ± 0.29
	MS (0)	0.56 ± 0.11	0.56 ± 0.11	0.56 ± 0.11	0.56 ± 0.11	0.53 ± 0.25	0.53 ± 0.25	0.53 ± 0.25	0.53 ± 0.25
	MS (1)	0.56 ± 0.13	0.56 ± 0.13	0.56 ± 0.13	0.56 ± 0.13	0.53 ± 0.13	0.53 ± 0.13	0.53 ± 0.13	0.53 ± 0.13
SYL	**HC (0)**	0.28 ± 0.04	0.28 ± 0.04	0.28 ± 0.04	0.28 ± 0.04	0.34 ± 0.03	0.34 ± 0.03	0.34 ± 0.03	0.34 ± 0.03
	HC (1)	0.45 ± 0.15	0.45 ± 0.15	0.45 ± 0.15	0.45 ± 0.15	0.56 ± 0.25	0.56 ± 0.25	0.56 ± 0.25	0.56 ± 0.25
	MS (0)	0.56 ± 0.12	0.56 ± 0.12	0.56 ± 0.12	0.56 ± 0.12	0.47 ± 0.20	0.47 ± 0.20	0.47 ± 0.20	0.47 ± 0.20
	MS (1)	0.56 ± 0.17	0.56 ± 0.17	0.56 ± 0.17	0.56 ± 0.17	0.51 ± 0.27	0.51 ± 0.27	0.51 ± 0.27	0.51 ± 0.27

Table A.3: Mean signed segmentation errors of FreeSurfer cortical surface reconstructions for each set of landmarks in the VDCRA dataset.

		Left GM				Right GM			
		3D (A)	**4D (A)**	**3D (B)**	**4D(B)**	**3D (A)**	**4D (A)**	**3D (B)**	**4D(B)**
CALC	**HC (0)**	0.37 ± 0.09	0.37 ± 0.09	0.37 ± 0.09	0.37 ± 0.09	0.25 ± 0.06	0.25 ± 0.06	0.25 ± 0.06	0.25 ± 0.06
	HC (1)	0.36 ± 0.10	0.36 ± 0.10	0.36 ± 0.10	0.36 ± 0.10	0.32 ± 0.12	0.32 ± 0.12	0.32 ± 0.12	0.32 ± 0.12
	MS (0)	0.25 ± 0.27	0.25 ± 0.27	0.25 ± 0.27	0.25 ± 0.27	0.44 ± 0.22	0.44 ± 0.22	0.44 ± 0.22	0.44 ± 0.22
	MS (1)	0.22 ± 0.31	0.22 ± 0.31	0.22 ± 0.31	0.22 ± 0.31	0.19 ± 0.20	0.19 ± 0.20	0.19 ± 0.20	0.19 ± 0.20
CING	**HC (0)**	0.72 ± 0.26	0.72 ± 0.26	0.72 ± 0.26	0.72 ± 0.26	0.62 ± 0.16	0.62 ± 0.16	0.62 ± 0.16	0.62 ± 0.16
	HC (1)	0.72 ± 0.17	0.72 ± 0.17	0.72 ± 0.17	0.72 ± 0.17	0.60 ± 0.20	0.60 ± 0.20	0.60 ± 0.20	0.60 ± 0.20
	MS (0)	0.14 ± 0.97	0.14 ± 0.97	0.14 ± 0.97	0.14 ± 0.97	-0.07 ± 0.96	-0.07 ± 0.96	-0.07 ± 0.96	-0.07 ± 0.96
	MS (1)	0.58 ± 0.36	0.58 ± 0.36	0.58 ± 0.36	0.58 ± 0.36	0.06 ± 0.86	0.06 ± 0.86	0.06 ± 0.86	0.06 ± 0.86
CS	**HC (0)**	0.04 ± 0.34	0.04 ± 0.34	0.04 ± 0.34	0.04 ± 0.34	0.06 ± 0.28	0.06 ± 0.28	0.06 ± 0.28	0.06 ± 0.28
	HC (1)	-0.80 ± 0.25	-0.80 ± 0.25	-0.80 ± 0.25	-0.80 ± 0.25	-0.31 ± 0.47	-0.31 ± 0.47	-0.31 ± 0.47	-0.31 ± 0.47
	MS (0)	0.12 ± 0.36	0.12 ± 0.36	0.12 ± 0.36	0.12 ± 0.36	0.26 ± 0.13	0.26 ± 0.13	0.26 ± 0.13	0.26 ± 0.13
	MS (1)	-0.41 ± 0.38	-0.41 ± 0.38	-0.41 ± 0.38	-0.41 ± 0.38	-0.02 ± 0.24	-0.02 ± 0.24	-0.02 ± 0.24	-0.02 ± 0.24
PO	**HC (0)**	0.12 ± 0.11	0.12 ± 0.11	0.12 ± 0.11	0.12 ± 0.11	0.13 ± 0.10	0.13 ± 0.10	0.13 ± 0.10	0.13 ± 0.10
	HC (1)	0.08 ± 0.26	0.08 ± 0.26	0.08 ± 0.26	0.08 ± 0.26	-0.28 ± 0.17	-0.28 ± 0.17	-0.28 ± 0.17	-0.28 ± 0.17
	MS (0)	0.23 ± 0.24	0.23 ± 0.24	0.23 ± 0.24	0.23 ± 0.24	0.35 ± 0.25	0.35 ± 0.25	0.35 ± 0.25	0.35 ± 0.25
	MS (1)	-0.04 ± 0.38	-0.04 ± 0.38	-0.04 ± 0.38	-0.04 ± 0.38	0.24 ± 0.43	0.24 ± 0.43	0.24 ± 0.43	0.24 ± 0.43
SF	**HC (0)**	0.35 ± 0.15	0.35 ± 0.15	0.35 ± 0.15	0.35 ± 0.15	0.35 ± 0.13	0.35 ± 0.13	0.35 ± 0.13	0.35 ± 0.13
	HC (1)	0.20 ± 0.15	0.20 ± 0.15	0.20 ± 0.15	0.20 ± 0.15	0.12 ± 0.27	0.12 ± 0.27	0.12 ± 0.27	0.12 ± 0.27
	MS (0)	-0.35 ± 0.30	-0.35 ± 0.30	-0.35 ± 0.30	-0.35 ± 0.30	0.01 ± 0.47	0.01 ± 0.47	0.01 ± 0.47	0.01 ± 0.47
	MS (1)	-0.73 ± 0.29	-0.73 ± 0.29	-0.73 ± 0.29	-0.73 ± 0.29	-0.39 ± 0.41	-0.39 ± 0.41	-0.39 ± 0.41	-0.39 ± 0.41
ST	**HC (0)**	0.58 ± 0.23	0.58 ± 0.23	0.58 ± 0.23	0.58 ± 0.23	0.54 ± 0.05	0.54 ± 0.05	0.54 ± 0.05	0.54 ± 0.05
	HC (1)	0.55 ± 0.29	0.55 ± 0.29	0.55 ± 0.29	0.55 ± 0.29	0.62 ± 0.16	0.62 ± 0.16	0.62 ± 0.16	0.62 ± 0.16
	MS (0)	-0.34 ± 0.20	-0.34 ± 0.20	-0.34 ± 0.20	-0.34 ± 0.20	-0.11 ± 0.20	-0.11 ± 0.20	-0.11 ± 0.20	-0.11 ± 0.20
	MS (1)	-0.06 ± 0.21	-0.06 ± 0.21	-0.06 ± 0.21	-0.06 ± 0.21	0.04 ± 0.12	0.04 ± 0.12	0.04 ± 0.12	0.04 ± 0.12
SYL	**HC (0)**	0.27 ± 0.05	0.27 ± 0.05	0.27 ± 0.05	0.27 ± 0.05	0.26 ± 0.04	0.26 ± 0.04	0.26 ± 0.04	0.26 ± 0.04
	HC (1)	0.44 ± 0.46	0.44 ± 0.46	0.44 ± 0.46	0.44 ± 0.46	0.16 ± 0.28	0.16 ± 0.28	0.16 ± 0.28	0.16 ± 0.28
	MS (0)	0.25 ± 0.25	0.25 ± 0.25	0.25 ± 0.25	0.25 ± 0.25	0.15 ± 0.13	0.15 ± 0.13	0.15 ± 0.13	0.15 ± 0.13
	MS (1)	-0.12 ± 0.37	-0.12 ± 0.37	-0.12 ± 0.37	-0.12 ± 0.37	-0.08 ± 0.32	-0.08 ± 0.32	-0.08 ± 0.32	-0.08 ± 0.32

		Left WM				Right WM			
		3D (A)	**4D (A)**	**3D (B)**	**4D(B)**	**3D (A)**	**4D (A)**	**3D (B)**	**4D(B)**
CALC	**HC (0)**	0.52 ± 0.17	0.52 ± 0.17	0.52 ± 0.17	0.52 ± 0.17	0.45 ± 0.10	0.45 ± 0.10	0.45 ± 0.10	0.45 ± 0.10
	HC (1)	0.33 ± 0.33	0.33 ± 0.33	0.33 ± 0.33	0.33 ± 0.33	0.48 ± 0.23	0.48 ± 0.23	0.48 ± 0.23	0.48 ± 0.23
	MS (0)	0.39 ± 0.42	0.39 ± 0.42	0.39 ± 0.42	0.39 ± 0.42	0.15 ± 0.17	0.15 ± 0.17	0.15 ± 0.17	0.15 ± 0.17
	MS (1)	0.39 ± 0.39	0.39 ± 0.39	0.39 ± 0.39	0.39 ± 0.39	-0.02 ± 0.28	-0.02 ± 0.28	-0.02 ± 0.28	-0.02 ± 0.28
CING	**HC (0)**	0.07 ± 0.07	0.07 ± 0.07	0.07 ± 0.07	0.07 ± 0.07	0.00 ± 0.12	0.00 ± 0.12	0.00 ± 0.12	0.00 ± 0.12
	HC (1)	-0.34 ± 0.19	-0.34 ± 0.19	-0.34 ± 0.19	-0.34 ± 0.19	-0.68 ± 0.25	-0.68 ± 0.25	-0.68 ± 0.25	-0.68 ± 0.25
	MS (0)	0.25 ± 0.12	0.25 ± 0.12	0.25 ± 0.12	0.25 ± 0.12	0.06 ± 0.31	0.06 ± 0.31	0.06 ± 0.31	0.06 ± 0.31
	MS (1)	-0.32 ± 0.27	-0.32 ± 0.27	-0.32 ± 0.27	-0.32 ± 0.27	-0.66 ± 0.34	-0.66 ± 0.34	-0.66 ± 0.34	-0.66 ± 0.34
CS	**HC (0)**	0.10 ± 0.08	0.10 ± 0.08	0.10 ± 0.08	0.10 ± 0.08	0.21 ± 0.04	0.21 ± 0.04	0.21 ± 0.04	0.21 ± 0.04
	HC (1)	0.07 ± 0.25	0.07 ± 0.25	0.07 ± 0.25	0.07 ± 0.25	-0.14 ± 0.19	-0.14 ± 0.19	-0.14 ± 0.19	-0.14 ± 0.19
	MS (0)	0.14 ± 0.21	0.14 ± 0.21	0.14 ± 0.21	0.14 ± 0.21	0.20 ± 0.33	0.20 ± 0.33	0.20 ± 0.33	0.20 ± 0.33
	MS (1)	-0.22 ± 0.26	-0.22 ± 0.26	-0.22 ± 0.26	-0.22 ± 0.26	-0.10 ± 0.30	-0.10 ± 0.30	-0.10 ± 0.30	-0.10 ± 0.30
PO	**HC (0)**	0.34 ± 0.36	0.34 ± 0.36	0.34 ± 0.36	0.34 ± 0.36	0.12 ± 0.23	0.12 ± 0.23	0.12 ± 0.23	0.12 ± 0.23
	HC (1)	-0.05 ± 0.60	-0.05 ± 0.60	-0.05 ± 0.60	-0.05 ± 0.60	-0.24 ± 0.48	-0.24 ± 0.48	-0.24 ± 0.48	-0.24 ± 0.48
	MS (0)	0.03 ± 0.46	0.03 ± 0.46	0.03 ± 0.46	0.03 ± 0.46	0.23 ± 0.34	0.23 ± 0.34	0.23 ± 0.34	0.23 ± 0.34
	MS (1)	-0.29 ± 0.59	-0.29 ± 0.59	-0.29 ± 0.59	-0.29 ± 0.59	-0.11 ± 0.35	-0.11 ± 0.35	-0.11 ± 0.35	-0.11 ± 0.35
SF	**HC (0)**	0.23 ± 0.10	0.23 ± 0.10	0.23 ± 0.10	0.23 ± 0.10	0.23 ± 0.14	0.23 ± 0.14	0.23 ± 0.14	0.23 ± 0.14
	HC (1)	-0.05 ± 0.27	-0.05 ± 0.27	-0.05 ± 0.27	-0.05 ± 0.27	0.12 ± 0.23	0.12 ± 0.23	0.12 ± 0.23	0.12 ± 0.23
	MS (0)	0.11 ± 0.39	0.11 ± 0.39	0.11 ± 0.39	0.11 ± 0.39	0.38 ± 0.36	0.38 ± 0.36	0.38 ± 0.36	0.38 ± 0.36
	MS (1)	-0.01 ± 0.44	-0.01 ± 0.44	-0.01 ± 0.44	-0.01 ± 0.44	0.27 ± 0.49	0.27 ± 0.49	0.27 ± 0.49	0.27 ± 0.49
ST	**HC (0)**	0.37 ± 0.05	0.37 ± 0.05	0.37 ± 0.05	0.37 ± 0.05	0.27 ± 0.10	0.27 ± 0.10	0.27 ± 0.10	0.27 ± 0.10
	HC (1)	0.35 ± 0.14	0.35 ± 0.14	0.35 ± 0.14	0.35 ± 0.14	0.40 ± 0.57	0.40 ± 0.57	0.40 ± 0.57	0.40 ± 0.57
	MS (0)	0.17 ± 0.16	0.17 ± 0.16	0.17 ± 0.16	0.17 ± 0.16	0.40 ± 0.27	0.40 ± 0.27	0.40 ± 0.27	0.40 ± 0.27
	MS (1)	-0.17 ± 0.18	-0.17 ± 0.18	-0.17 ± 0.18	-0.17 ± 0.18	-0.07 ± 0.24	-0.07 ± 0.24	-0.07 ± 0.24	-0.07 ± 0.24
SYL	**HC (0)**	0.22 ± 0.09	0.22 ± 0.09	0.22 ± 0.09	0.22 ± 0.09	0.24 ± 0.15	0.24 ± 0.15	0.24 ± 0.15	0.24 ± 0.15
	HC (1)	-0.13 ± 0.29	-0.13 ± 0.29	-0.13 ± 0.29	-0.13 ± 0.29	-0.11 ± 0.41	-0.11 ± 0.41	-0.11 ± 0.41	-0.11 ± 0.41
	MS (0)	0.26 ± 0.24	0.26 ± 0.24	0.26 ± 0.24	0.26 ± 0.24	0.28 ± 0.10	0.28 ± 0.10	0.28 ± 0.10	0.28 ± 0.10

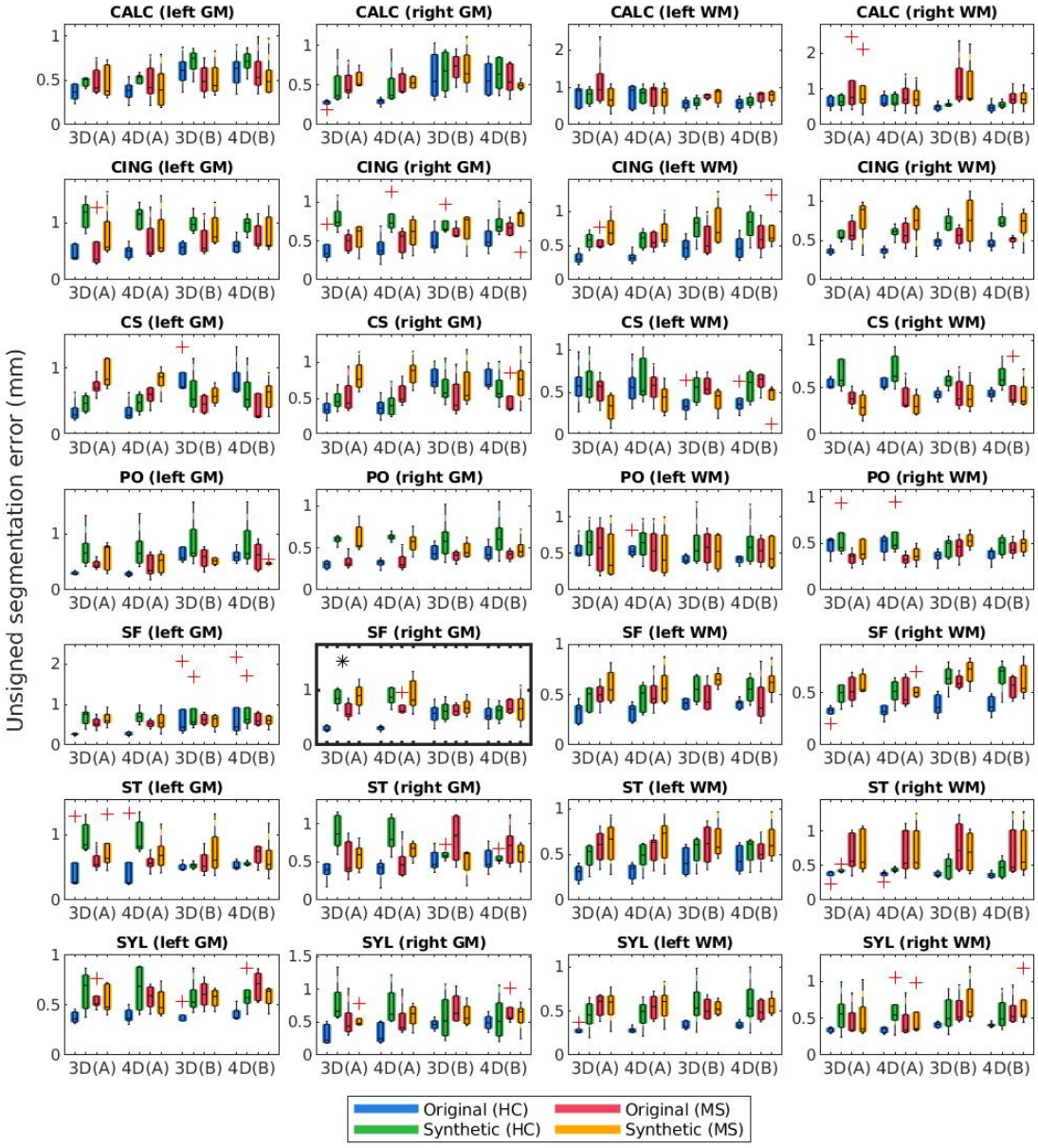

Figure A.1: Mean unsigned segmentation errors of FreeSurfer cortical surface reconstructions for each set of landmarks across healthy and MS subjects. The horizontal axis for each subplot indicates the process type (cross-sectional or longitudinal) and the expert (A or B). Columns of subplots correspond to surface and hemisphere while the subplot row corresponds to the landmarks' anatomical placements. In each panel, mean segmentation errors from the original images of healthy subjects are shown in blue, synthetic images of healthy subjects in green, original images of MS subjects in red, and synthetic images from MS subjects in yellow.

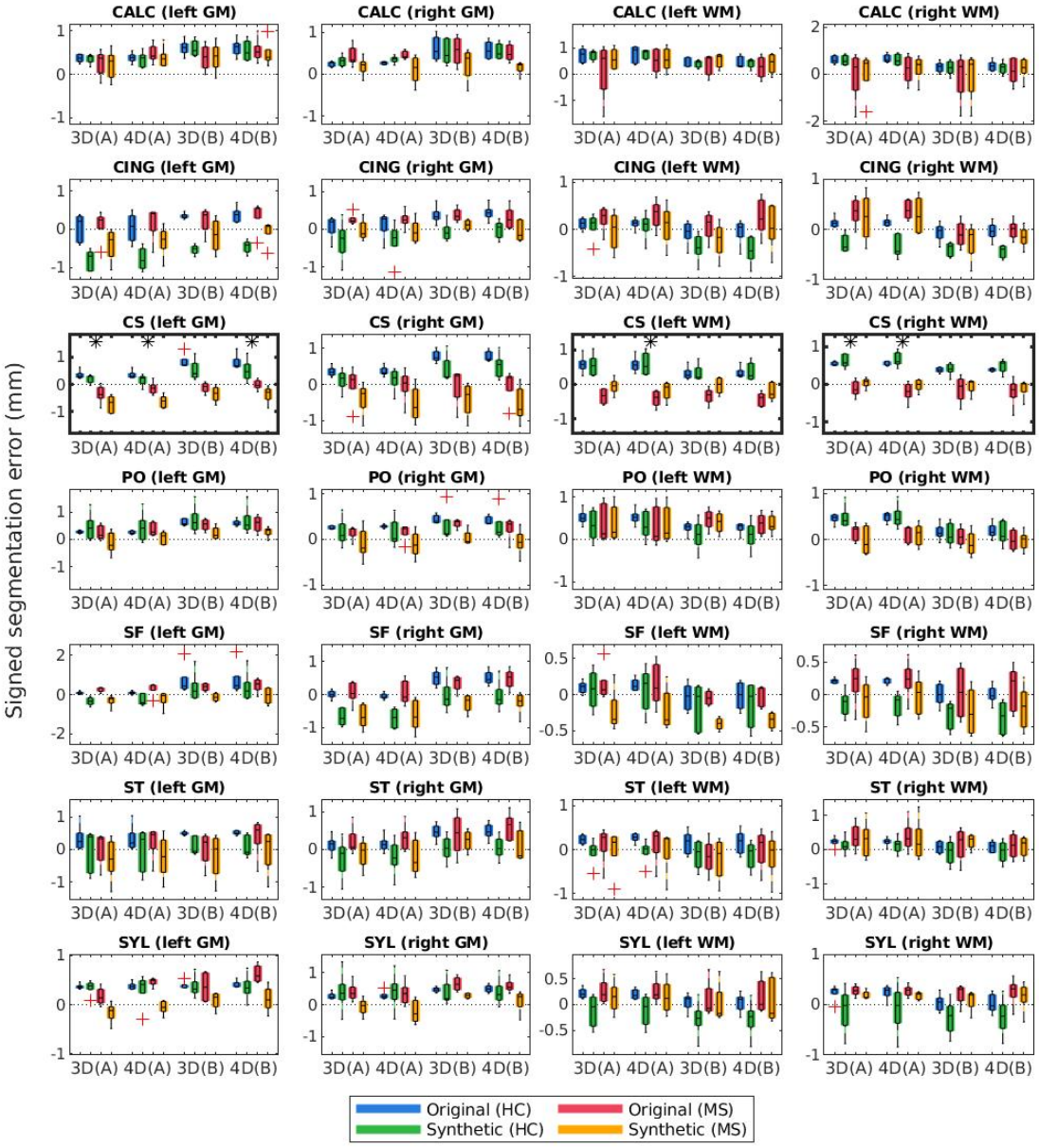

Figure A.2: Mean signed segmentation errors of FreeSurfer cortical surface reconstructions for each set of landmarks across healthy and MS subjects. The horizontal axis for each subplot indicates the process type (cross-sectional or longitudinal) and the expert (A or B). Columns of subplots correspond to surface and hemisphere while the subplot row corresponds to the landmarks' anatomical placements. In each panel, mean segmentation errors from the original images of healthy subjects are shown in blue, synthetic images of healthy subjects in green, original images of MS subjects in red, and synthetic images from MS subjects in yellow.

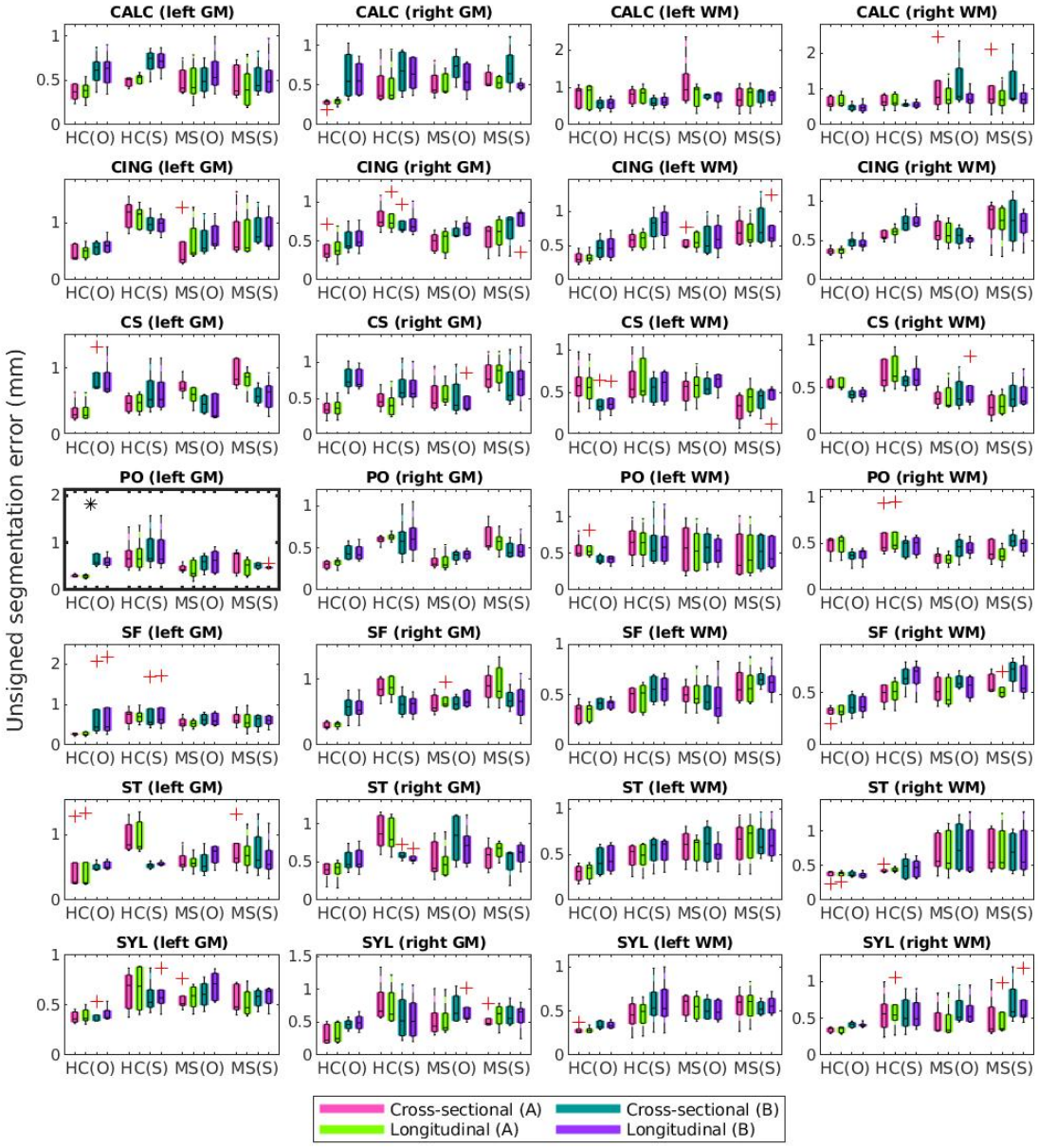

Figure A.3: Mean unsigned segmentation errors of FreeSurfer cortical surface reconstructions for each set of landmarks across healthy and MS subjects. The horizontal axis for each subplot indicates the subject group (healthy control (HC) vs. MS) and the timepoint (0=original, 1=synthetic). Columns of subplots correspond to surface and hemisphere while the subplot row corresponds to the landmarks' anatomical placement. In each panel, mean segmentation errors from the cross-sectional method using landmarks placed by expert A are shown in pink, cross-sectional with expert B in teal, longitudinal with expert A in green, and longitudinal with expert B in purple.

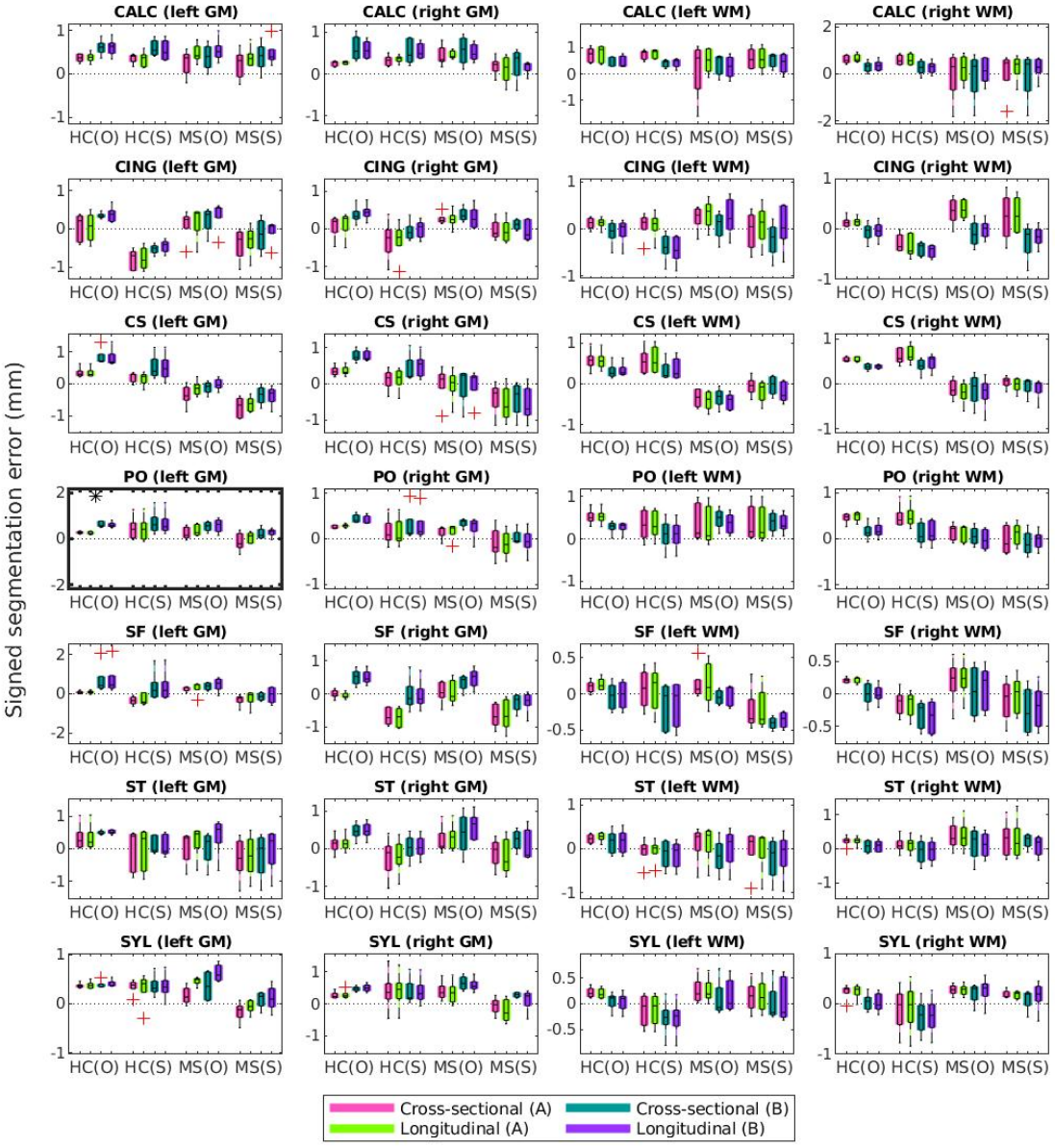

Figure A.4: Mean signed segmentation errors of FreeSurfer cortical surface reconstructions for each set of landmarks across healthy and MS subjects. The horizontal axis for each subplot indicates the subject group (healthy control (HC) vs. MS) and the timepoint (0=original, 1=synthetic). Columns of subplots correspond to surface and hemisphere while the subplot row corresponds to the landmarks' anatomical placement. In each panel, mean segmentation errors from the cross-sectional method using landmarks placed by expert A are shown in pink, cross-sectional with expert B in teal, longitudinal with expert A in green, and longitudinal with expert B in purple.

Appendix B

Aim 2: Extended Background - A Brief Overview of Finite Differences for Elliptical Equations

This section of the appendix covers the basic concepts of finite difference methods with elliptical equations, which are those that describe steady-state conditions over an enclosed domain. Elliptical equations typically are used to model physical spaces with no time dependency. Specifically, we will explore finite difference models (FDMs) of the Poisson equation $\nabla^2 U = f$ in both 1- and 2D grids. The Laplacian equation discussed in Chapter 5 is a simplified version of this where $f = 0$. Examples and explanations provided in this section are adapted from Lynch, 2005 [111].

B.1 1D Differences on Uniform Grids

Consider the partial differential equation (PDE)

$$\frac{\partial^2 U(x)}{\partial x^2} = f(x), \tag{B.1}$$

where $U(x)$ and $f(x)$ are continuous and differentiable over the entire domain x. Our goal is to develop a framework to express the solution to $U(x)$ at discretely sampled points without providing an explicit definition.

First, we explore this concept over a 1D uniform grid. Let X be the uniform grid shown in Figure B.1 with nodes $x_i \in X$, $1 \leq i \leq N$.

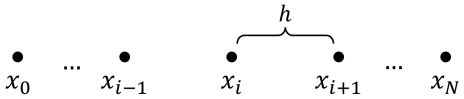

Figure B.1: A 1D grid X with nodes $x_i \in X$ and uniform spacing $h \equiv x_{i+1} - x_i$.

We can write $U(x)$ at each x_i as $U(x_i) \equiv U_i$. Given that X has uniform spacing $h \equiv x_{i+1} - x_i$, we can express U_{i+1} using the Taylor Series expansion of $U(x)$ at x_i as

$$U_{i+1} = U_i + h\frac{\partial U_i}{\partial x} + \frac{h^2}{2!}\frac{\partial^2 U_i}{\partial x^2} + O(h^3). \tag{B.2}$$

This gives us the forward difference approximation to $U'(x)$:

$$\frac{\partial U_i}{\partial x} = \frac{U_{i+1} - U_i}{h} - \frac{h}{2!}\frac{\partial^2 U_i}{\partial x^2} + O(h^2) \equiv \frac{\Delta U_i}{h} + O(h) \tag{B.3}$$

Similarly, we can express U_{i-1} using the Taylor Series expansion of $U(x)$ at x_i as

$$U_{i-1} = U_i - h\frac{\partial U_i}{\partial x} + \frac{h^2}{2!}\frac{\partial^2 U_i}{\partial x^2} + O(h^3), \tag{B.4}$$

which gives us the backwards difference approximation to $U'(x)$:

$$\frac{\partial U_i}{\partial x} = \frac{U_i - U_{i-1}}{h} + \frac{h}{2!}\frac{\partial^2 U_i}{\partial x^2} + O(h^2) \equiv \frac{\nabla U_i}{h} + O(h). \tag{B.5}$$

Combining equations B.3 and B.5 gives us the central difference expression for $U''(x)$:

$$\begin{aligned} \frac{U_{i+1} - U_i}{h} - \frac{h}{2}\frac{\partial^2 U_i}{\partial x^2} + O(h^2) &= \frac{U_i - U_{i-1}}{h} + \frac{h}{2!}\frac{\partial^2 U_i}{\partial x^2} + O(h^2) \\ \frac{\partial^2 U_i}{\partial x^2} &= \frac{U_{i+1} - 2U_i + U_{i-1}}{h^2} + O(h^2) \end{aligned} \tag{B.6}$$

We can then substitute this back into equation B.1 to find an expression for U_i:

$$\begin{aligned} \frac{U_{i+1} - 2U_i + U_{i-1}}{h^2} &= f_i \\ U_{i-1} - U_i + U_{i+1} &= h^2 f_i \end{aligned} \tag{B.7}$$

Now that we have an expression describing the value of $U(x)$ at three adjacent points, our next goal is to formulate a system of equations that will enable us to solve for U_i. To find a solution, we require the system to yield a matrix with a rank greater than or equal to the number of unknowns. We can achieve this by defining two initial conditions for $U(x)$ that establish its value along the boundaries of the domain; these are formally known as boundary conditions (BCs). For our purposes, we will be dealing with Dirichlet conditions that provide an explicit value for $U(x)$. Other types of BCs are Neumann conditions, which give an initial value for ∂U, or mixed conditions, which take the form of a linear combination of U and ∂U.

Suppose that $U_0 = f_0$ and $U_N = f_N$ are the BCs of U in X. Substituting these values into equation B.7 results in the following system of equations:

$$-2U_1 + U_2 = h^2 f_0 - U_0 \qquad i = [1] \tag{B.8}$$

$$U_{i-1} - 2U_i + U_{i+1} = h^2 f_i \qquad i = [2 : N-2] \tag{B.9}$$

$$U_{N-2} - 2U_{N-1} = h^2 f_{N-1} - U_N \qquad i = [N-1] \tag{B.10}$$

where the first and third equations represent the nodes adjacent to the boundary nodes x_0 and x_N where U_0 and U_N are known. The system of equations B.8-B.10 is equivalent to the following tridiagnonal matrix system:

$$\begin{bmatrix} 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} U_1 \\ \\ \vdots \\ \\ U_{N-1} \end{bmatrix} = \begin{bmatrix} h^2 f_1 \\ \\ \vdots \\ \\ h^2 f_{N-1} \end{bmatrix} - \begin{bmatrix} U_0 \\ 0 \\ \vdots \\ 0 \\ U_N \end{bmatrix} \tag{B.11}$$

In the left-most matrix of coefficients, each row corresponds to a specific $x_i \in X$. The left-hand side of the equation represents the unknown values of U at interior (non-boundary) nodes, and the right-hand side is a combination of the forcing conditions f_i and the BCs.

A solution to such a system can be found using either a direct or an iterative method. Direct methods involve a finite number of operations and find an exact, algebraic solution, but can be highly computationally expensive. Iterative methods, while only able to achieve an exact solution after an infinite number of operations, are much more easily implemented when solving larger systems such as the application presented in Chapter 5 of this dissertation. The simplest iterative algorithm is the Jacobian method, where at each iteration l, we calculate U_i^{l+1} based on the current values U_{i+1}^l and U_{i-1}^l at the corresponding neighboring nodes:

$$U_i^{l+1} = \frac{1}{2}\left[U_{i+1} + U_{i-1}\right]^l + h^2 f$$

Here,The full array of U_i^{l+1}s are calculated using values exclusively from the previous array of U_i^l. The Gauss-Seidel method makes a slight modification of this by updating each U_i^{l+1} instantly, rather than waiting until the end of an iteration. An example equation would be

$$U_i^{l+1} = \frac{1}{2}\left[U_{i+1}^{l+1} + U_{i-1}^l\right] + h^2 f$$

where U_i^{l+1} is calculated using an updated value of U_{i+1} but a current value of U_{i-1}.

B.2 2D Differences on Uniform Grids

We now consider the Poisson equation in 2D

$$\nabla^2 U(x,y) = f(x,y), \tag{B.12}$$

over the following domain:

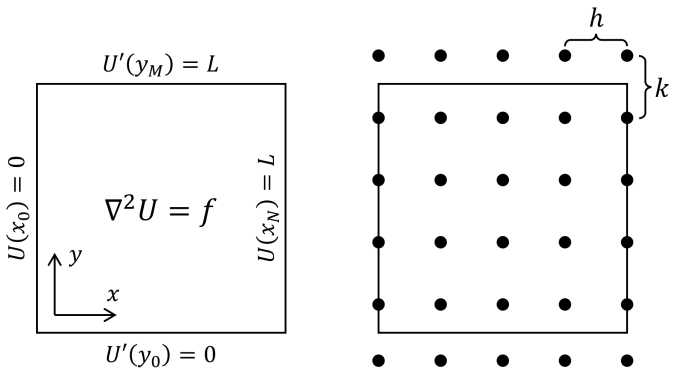

Figure B.2: Left: The domain bound by Dirichlet conditions in the horizontal direction, and Neumann in the vertical. A 2D grid with uniform spacing $h = \Delta x$ and $k = \Delta y$ that lies (mostly) inside the domain.

Given $h = \Delta x$ and $k = \Delta y$, we have

$$\nabla^2 U(x,y) = \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} \simeq \frac{\delta_x^2 U_{i,j}}{h^2} + \frac{\delta_y^2 U_{i,j}}{k^2} = f_{i,j} \tag{B.13}$$

where $0 \leq i \leq N$ and $0 \leq j \leq M$. We adapt equation B.6 to a 2D framework and write an expression for every $U_{i,j}$ in terms of its four neighboring values:

$$\frac{U_{i-1,j} - 2U_{i,j} + U_{i+1,j}}{h^2} + \frac{U_{i,j-1} - 2U_{i,j} + U_{i,j+1}}{k^2} = f_{i,j} \tag{B.14}$$

Similar to the 1D case, we can group like terms to form a system of equations based on the BCs defined in Figure B.2. For simplicity, let $\beta = {}^{h^2}/_{k^2}$. Then the equation representing each node $x_{i,j}$, for $2 \leq i \leq N-1$ and $1 \leq j \leq M-1$, is

$$\begin{aligned} U_{i-1,j} - 2U_{i,j} + U_{i+1,j} + \beta(U_{i,j-1} - 2U_{i,j} + U_{i,j+1}) &= h^2 f_{i,j} \\ U_{i-1,j} + U_{i+1,j} + \beta U_{i,j-1} + \beta U_{i,j+1} - 2(1+\beta)U_{i,j} &= h^2 f_{i,j} \end{aligned} \tag{B.15}$$

Next, we formulate equations to handle each of the BCs within our domain. Notice that we have carefully designed our grid such that nodes are placed directly on boundaries with Dirichlet conditions, but equally spaced on either side of those with Neumann conditions. At the nodes where U is known (e.g. the left and right boundaries), we have

$$U_{2,j} + \beta U_{1,j-1} + \beta U_{1,j+1} - 2(1+\beta)U_{1,j} = h^2 f_{1,j} \qquad \text{(Left, } i = 1\text{)} \qquad \text{(B.16)}$$

$$U_{N-2,j} + \beta U_{N-1,j-1} + \beta U_{N-1,j+1} - 2(1+\beta)U_{N-1,j} = h^2 f_{N-1,j} - L \qquad \text{(Right, } i = N-1\text{)} \qquad \text{(B.17)}$$

To handle the other two boundaries (e.g. the top and bottom), recall from section B.1 that we can express U' in 1D as either a forwards or backwards difference as

$$\frac{\partial U}{\partial x} \simeq \frac{U_{i+1} - U_i}{h} \simeq \frac{U_i - U_{i-1}}{h}$$

respectively. In locations where U' is known, we thus have $U_{i,1} - U_{i,0} = 0$ and $U_{i,M} - U_{i,M-1} = ka$. This allows us to solve for equations at each of the Neumann boundaries, beginning by rearranging the terms in equation B.15:

$$U_{i-1,j} - 2U_{i,j} + U_{i+1,j} + \beta(U_{i,j-1} - U_{i,j} - U_{i,j}+) = h^2 f_{i,j}$$

$$U_{i-1,j} - 2U_{i,j} + U_{i+1,j} + \beta(U_{i,j+1} - U_{i,j}) - \beta(U_{i,j} - U_{i,j-1}) = h^2 f_{i,j}$$

$$U_{i+1,1} - (2+\beta)U_{i,1} + U_{i-1,1} + \beta U_{i,2} = h^2 f_{i,1} \qquad \text{(Bottom, } j = 1\text{)} \qquad \text{(B.18)}$$

$$U_{i+1,M-1} - (2+\beta)U_{i,M-1} + U_{i-1,M-1} + \beta U_{i,M-1} = h^2 f_{i,M-1} - \beta ka \qquad \text{(Top, } j = M-1\text{)} \qquad \text{(B.19)}$$

We can further simplify equations B.15-B.19 in the corners of the domain, where both U' and U' are known. This system can then be reformulated into a pentadiagonal matrix and solved in a manner similar to our 1D example.

B.3 Non-uniform Grids

So far, we have only operated over grids with uniform spacing along each Cartesian direction. Let us now consider again the 1D Poisson example (equation B.1), but this time over the non-uniform grid X shown below:

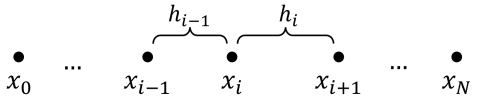

Figure B.3: A 1D grid X with nodes $x_i \in X$ and non-uniform spacing $h_i \equiv x_{i+1} - x_i$.

With non-uniform spacing, we write the forwards difference for U_i' as

$$\frac{\partial U_i}{\partial x} = \frac{U_{i+1} - U_i}{h_i} - \frac{h_i}{2!}\frac{\partial^2 U_i}{\partial x^2} + O(h_i^2) = \frac{\Delta U_i}{h_i} + O(h_i) \tag{B.20}$$

and the backwards as

$$\frac{\partial U_i}{\partial x} = \frac{U_i - U_{i-1}}{h_{i-1}} + \frac{h_{i-1}}{2!}\frac{\partial^2 U_i}{\partial x^2} + O(h_{i-1}^2) \equiv \frac{\nabla U_i}{h_{i-1}} + O(h_{i-1}). \tag{B.21}$$

Once again, we can combine these to derive an approximation for $U''(x)$

$$\begin{aligned}
\frac{U_{i+1} - U_i}{h_i} - \frac{h_i}{2!}\frac{\partial^2 U_i}{\partial x^2} + O(h_i^2) &= \frac{U_i - U_{i-1}}{h_{i-1}} + \frac{h_{i-1}}{2!}\frac{\partial^2 U_i}{\partial x^2} + O(h_{i-1}^2) \\
\frac{\partial^2 U_i}{\partial x^2}\left(\frac{h_i + h_{i-1}}{2}\right) &= \frac{h_{i-1}(U_{i+1} - U_i) - h_i(U_i - U_{i-1}}{h_i h_{i-1}} \\
\frac{\partial^2 U_i}{\partial x^2} = U_{i+1}\left[\frac{2}{h_i^2 + h_i h_{i-1}}\right] + U_{i-1}&\left[\frac{2}{h_{i-1}^2 + h_i h_{i-1}}\right] - U_i\left[\frac{2}{h_i h_{i-1}}\right]
\end{aligned} \tag{B.22}$$

This simplifies to equation B.6 when $h_i = h_{i-1}$.

As with the uniform case, we can create a system of equations in matrix form and iteratively solve for U_i at each grid node. However, unlike before, this framework does not easily translate to grids with non-uniformity in more than one dimension. This is the motivation behind adapting a system such as that introduced in Sukumar et al., 2003 [176], as detailed in Chapter 5.

Appendix C

Aim 3: Unabridged Data

This section of the appendix contains the unabridged data from Chapter 6 of this dissertation. Each figure displays the normalized mean thickness values within one of the Desikan Kiliany (DK) atlas cortical regions estimated by the linear mixed effects model. Thicknesses are plotted as a function of CAP score for the low (green), medium (orange), and high (blue) CAP groups. Control subjects are excluded, as they are not associated with a CAP score. The dotted line within each plot indicates $z = 0$.

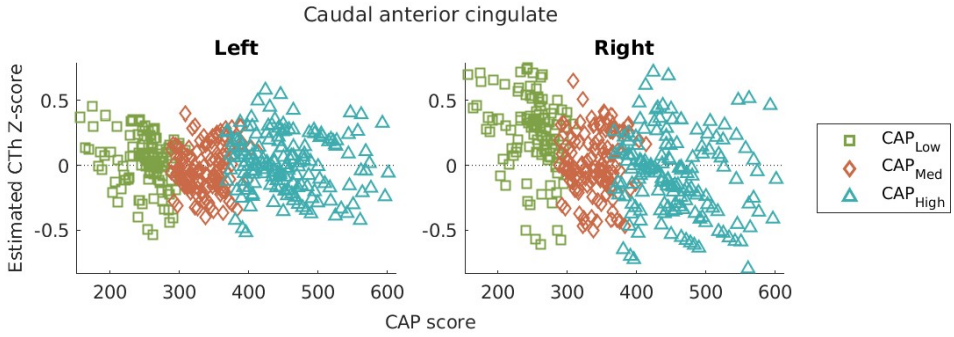

Figure C.1

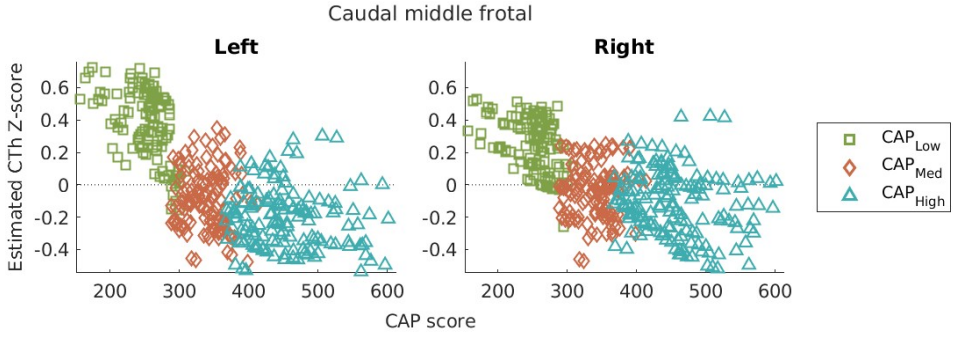

Figure C.2

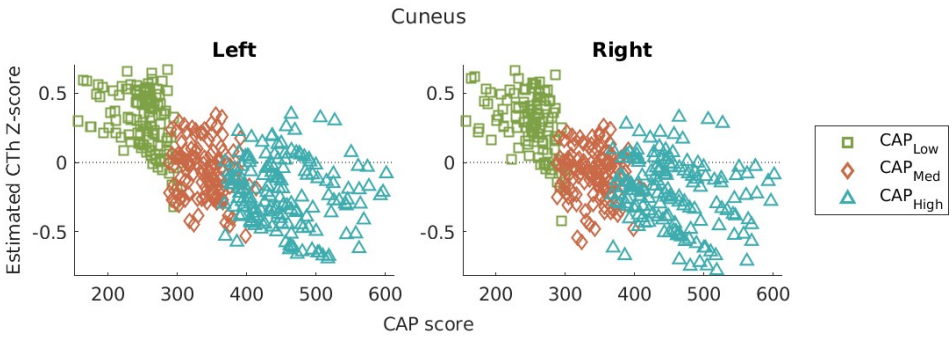

Figure C.3

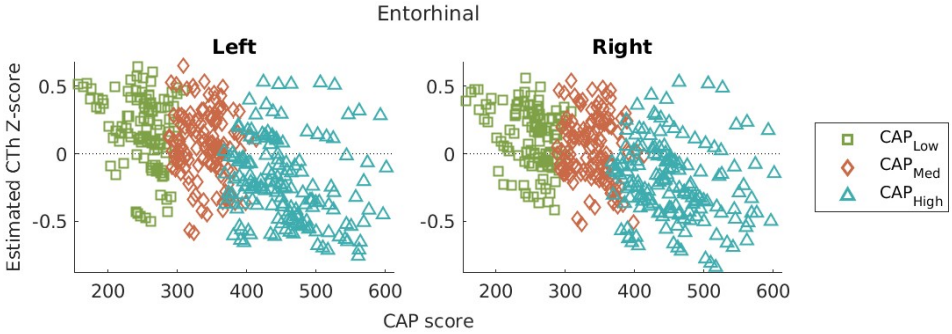

Figure C.4

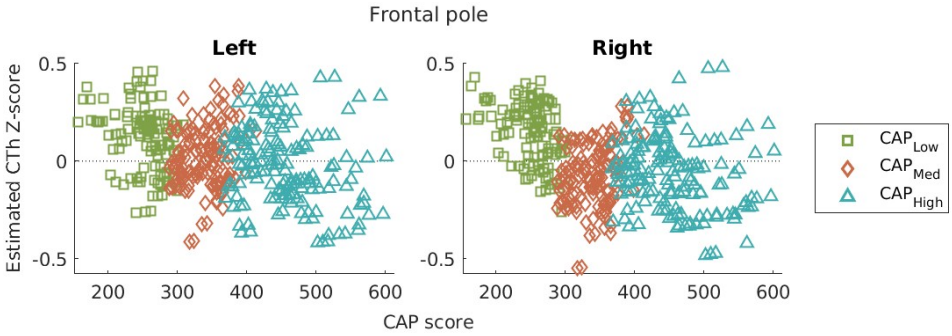

Figure C.5

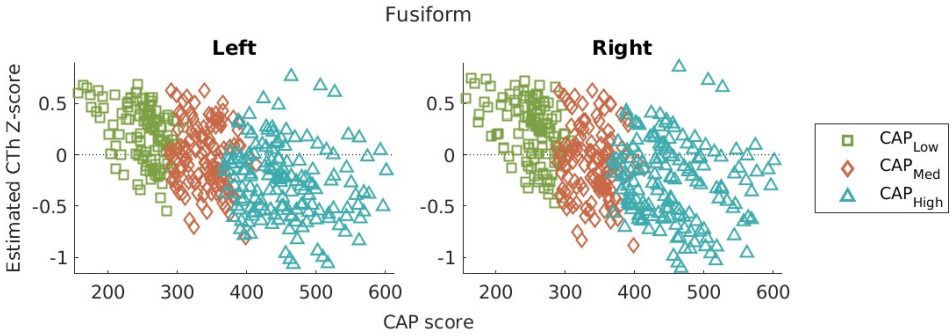

Figure C.6

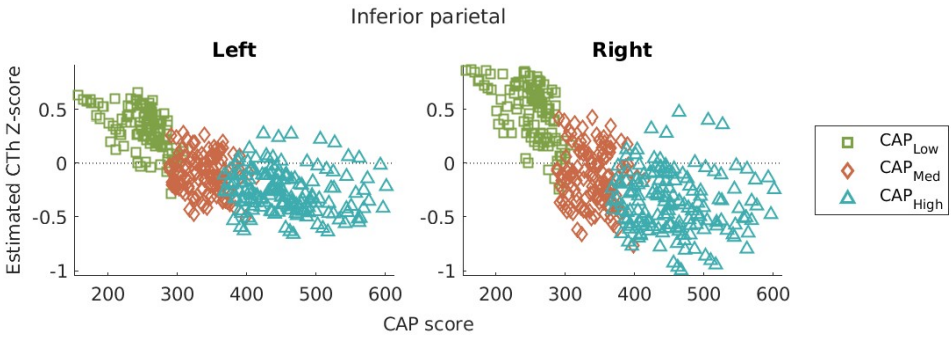

Figure C.7

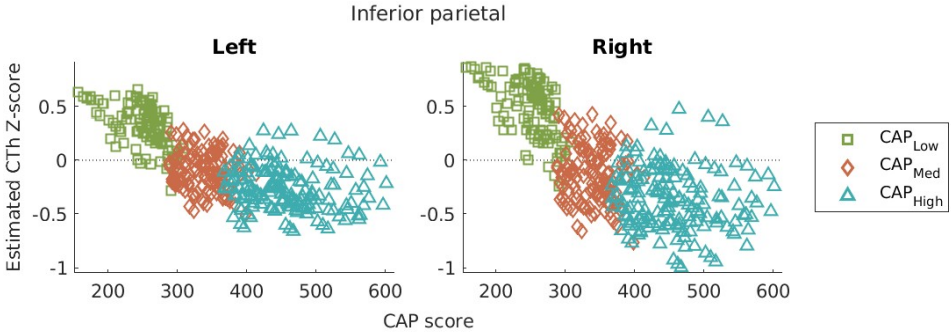

Figure C.8

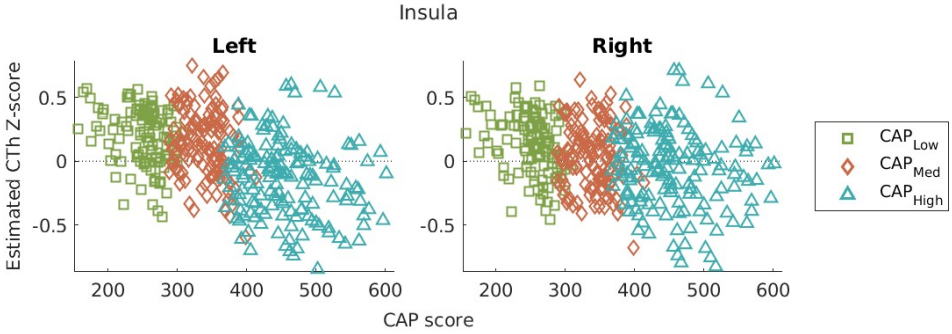

Figure C.9

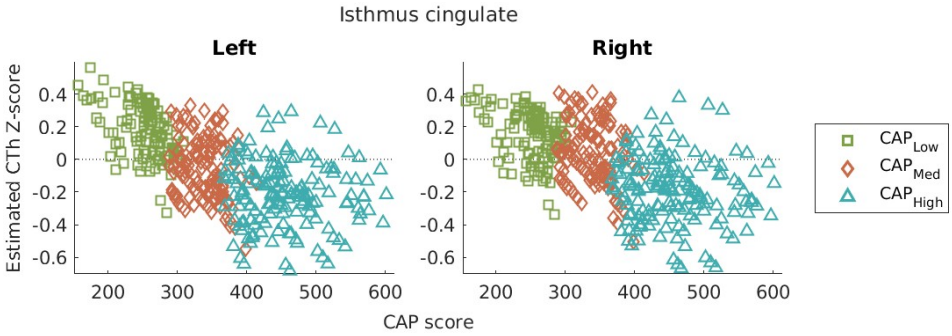

Figure C.10

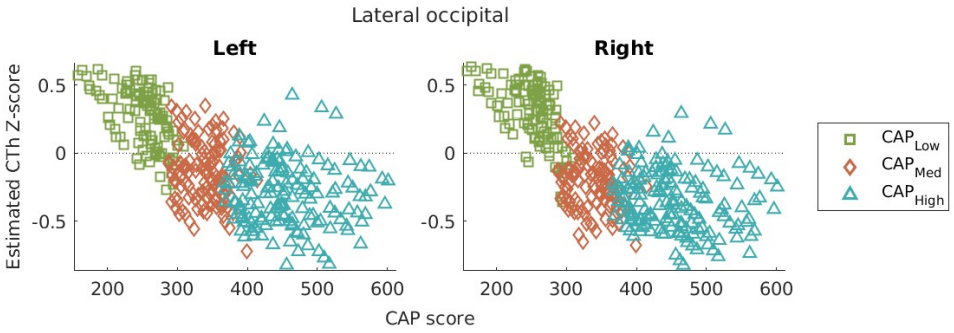

Figure C.11

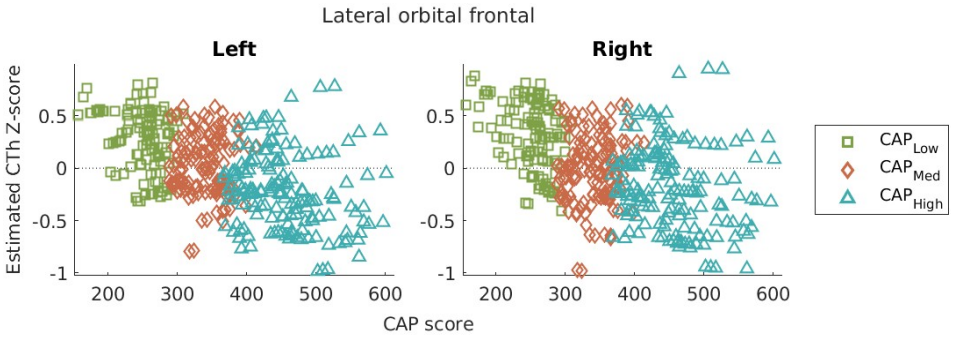

Figure C.12

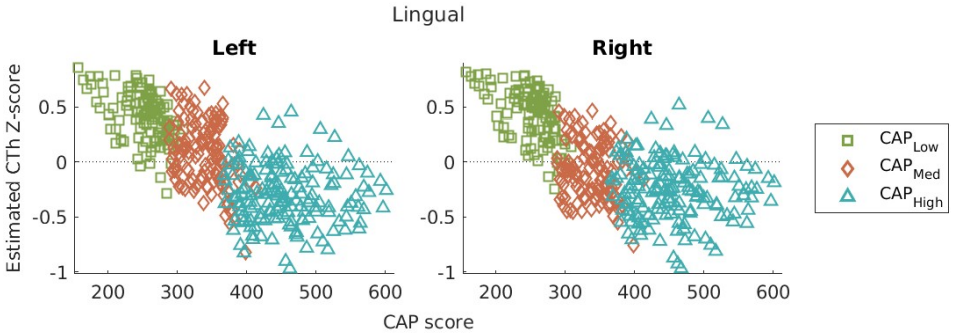

Figure C.13

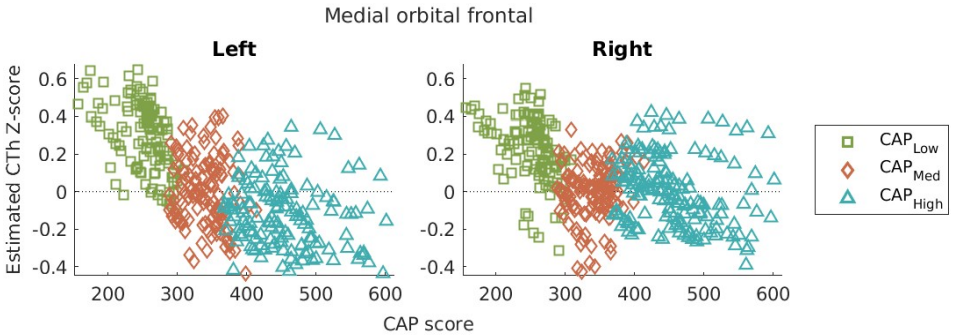

Figure C.14

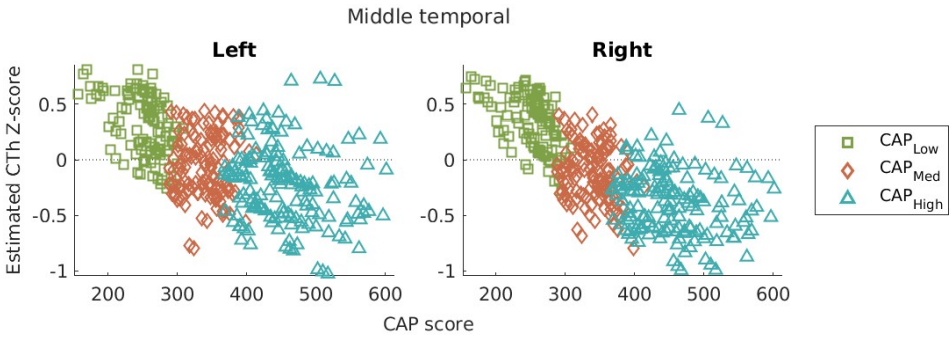

Figure C.15

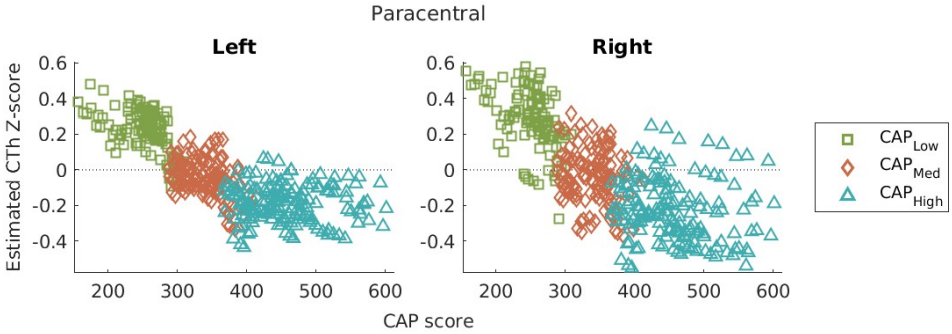

Figure C.16

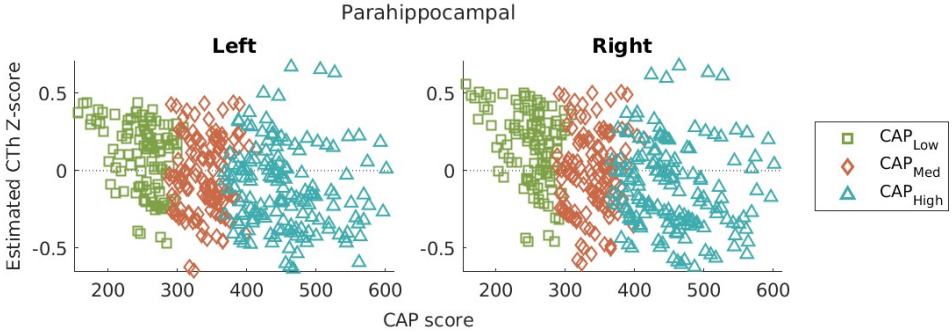

Figure C.17

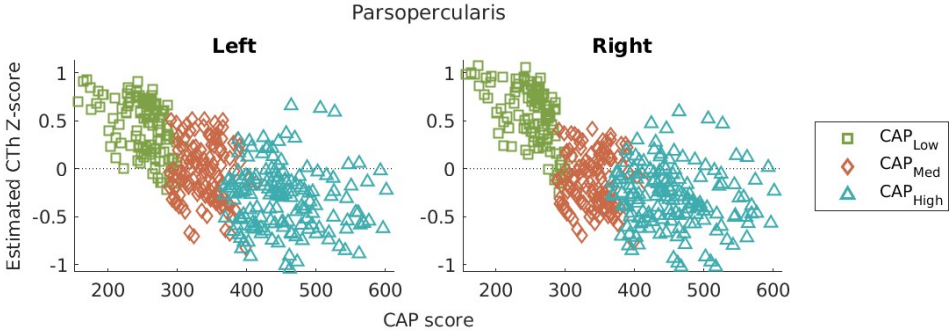

Figure C.18

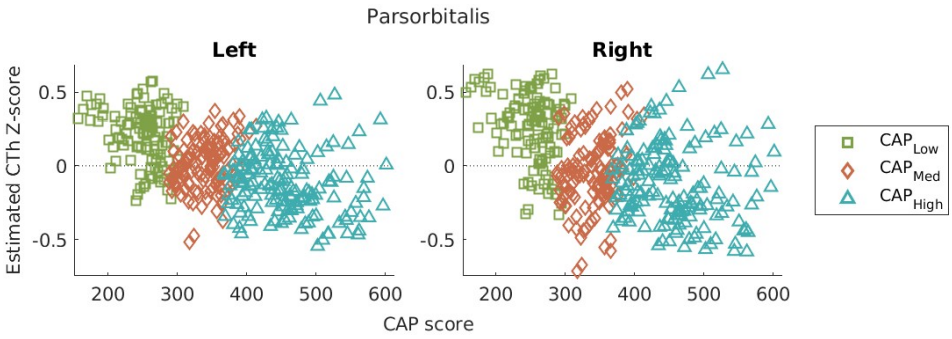

Figure C.19

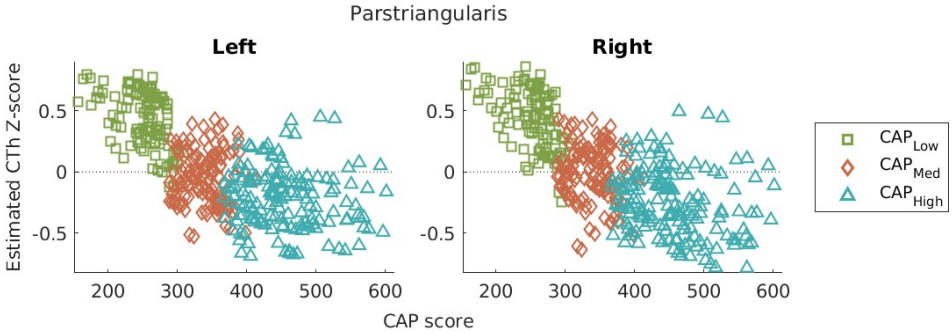

Figure C.20

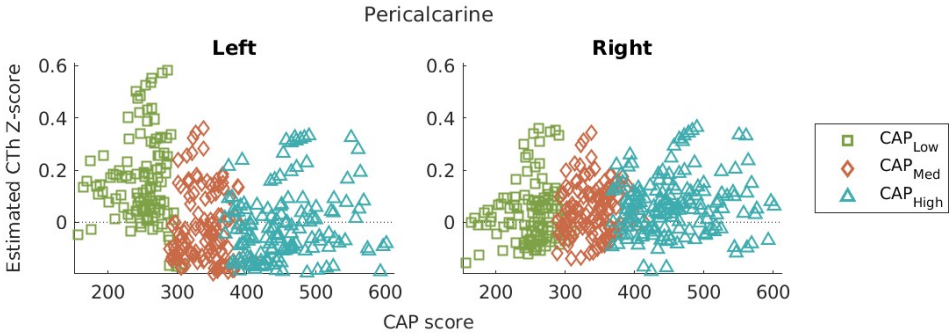

Figure C.21

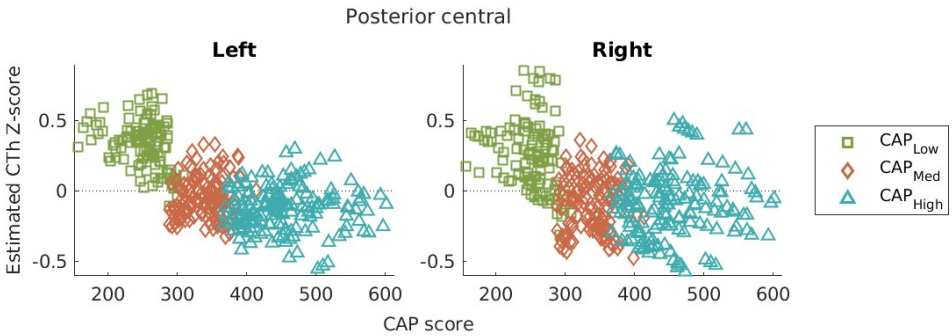

Figure C.22

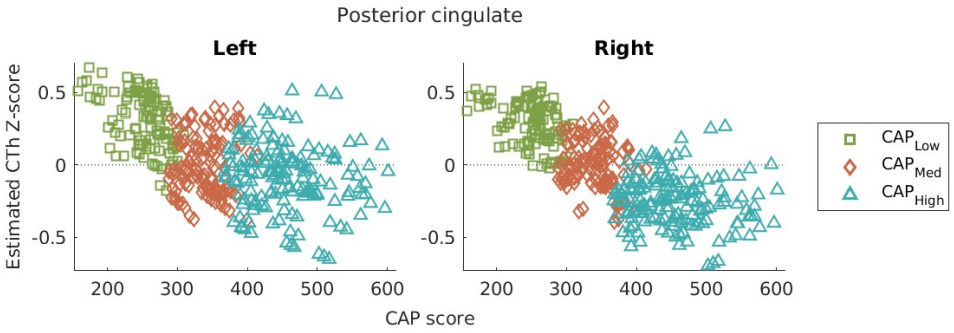

Figure C.23

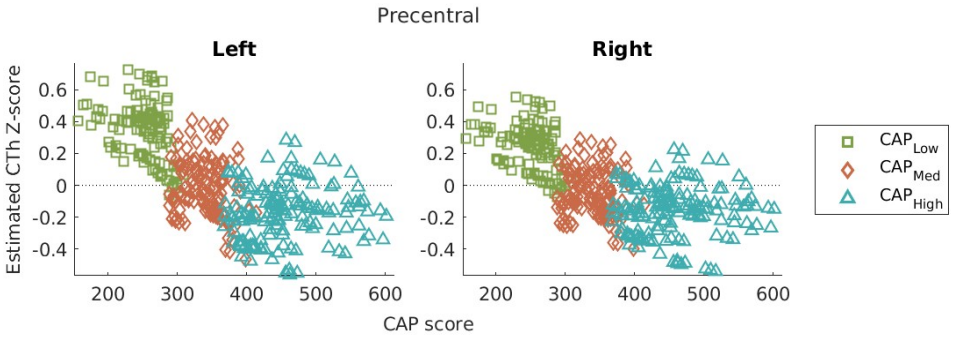

Figure C.24

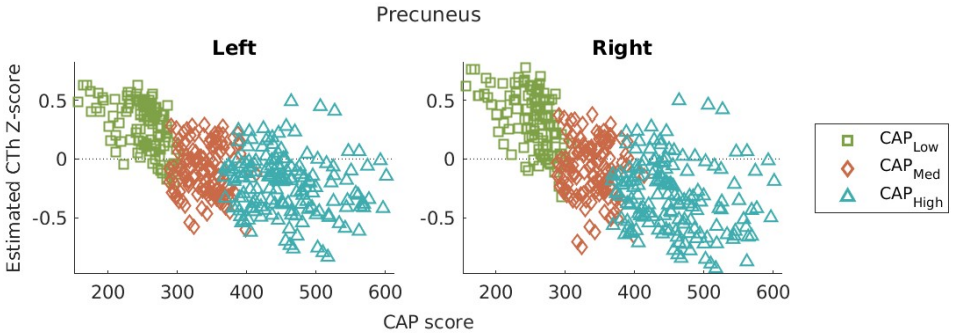

Figure C.25

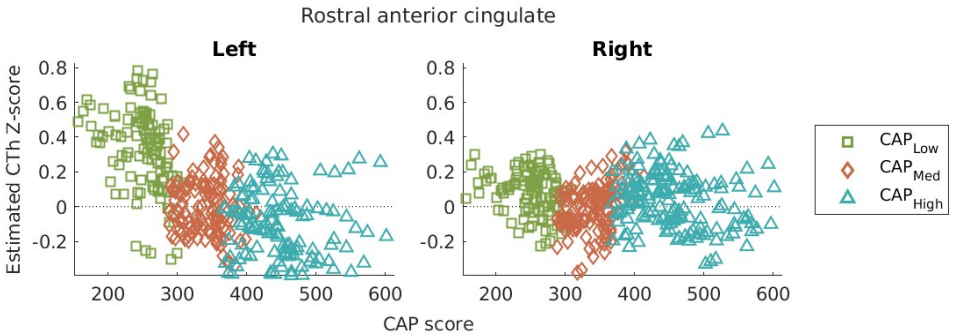

Figure C.26

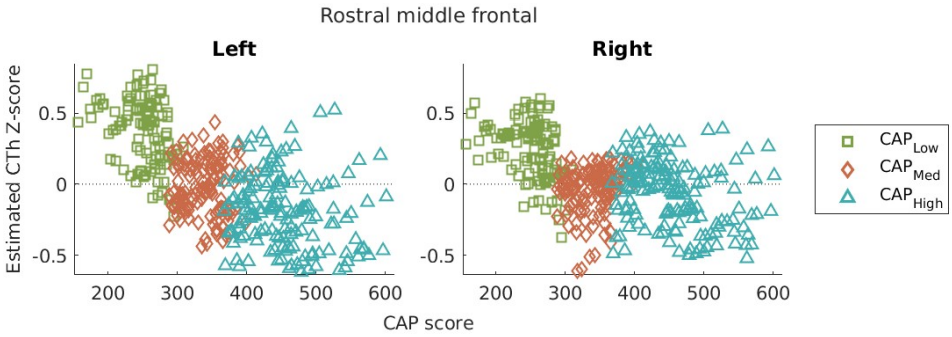

Figure C.27

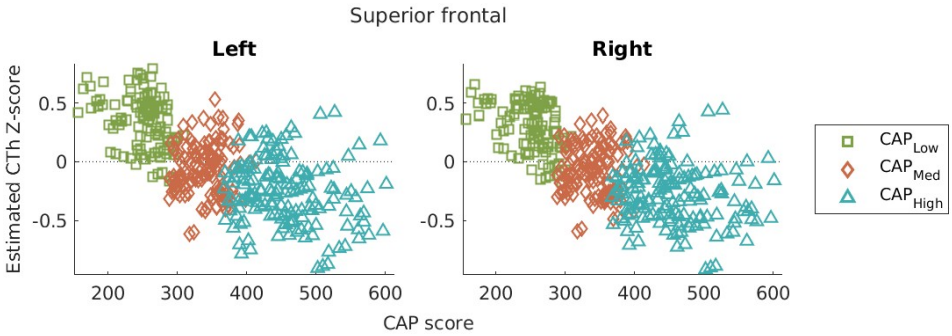

Figure C.28

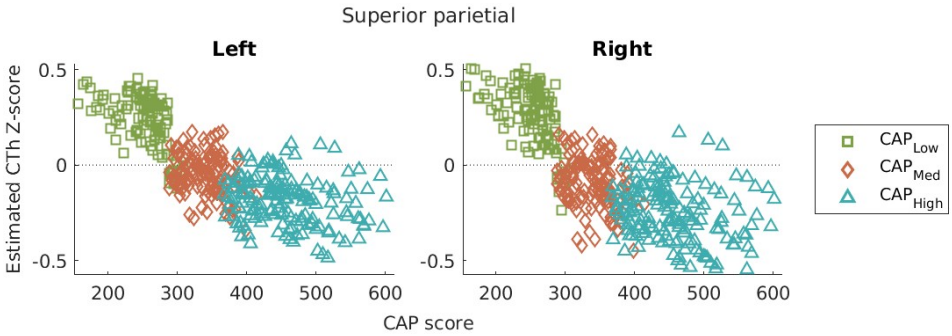

Figure C.29

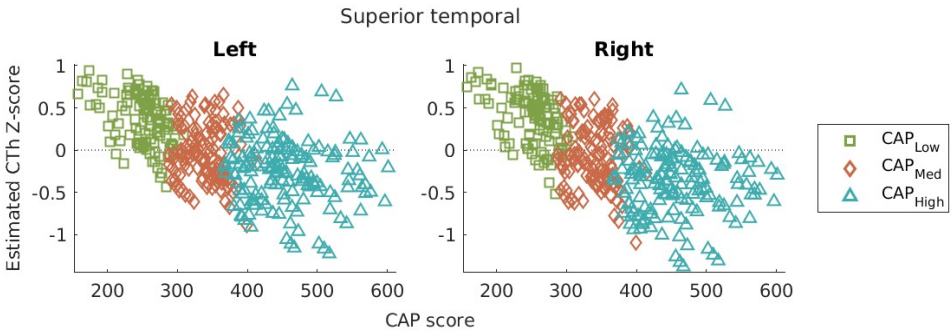

Figure C.30

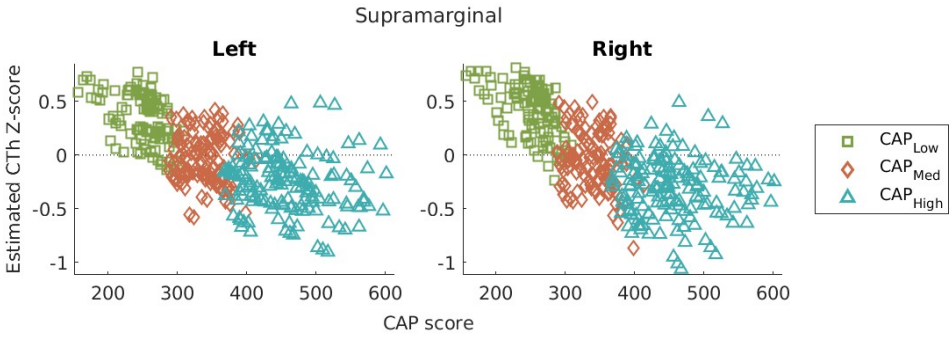

Figure C.31

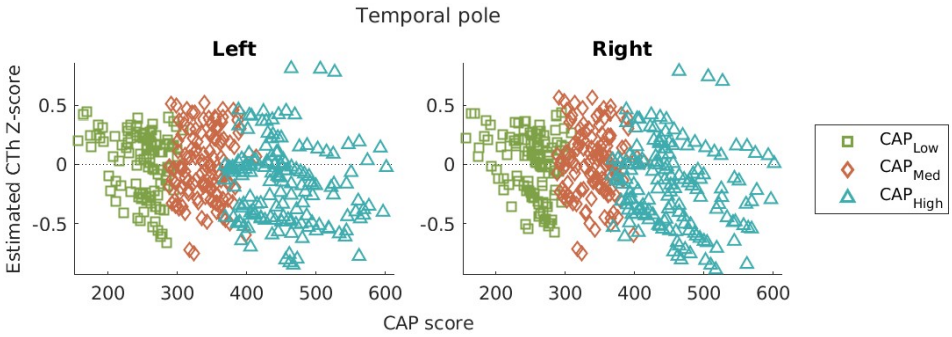

Figure C.32

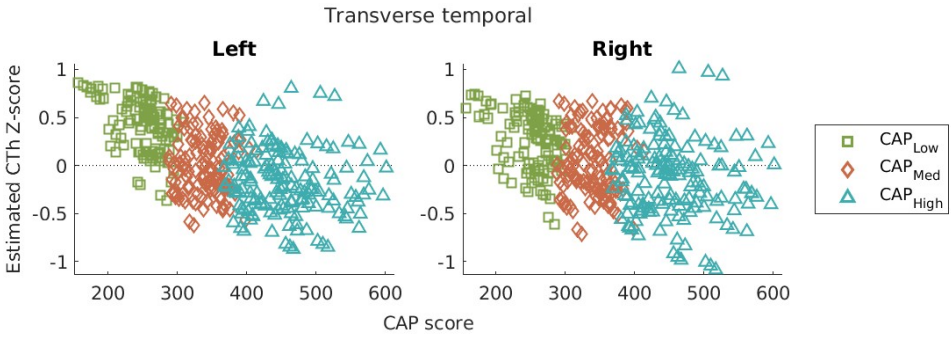

Figure C.33

References

[1] Abeyasinghe, P. M., Long, J. D., Razi, A., Pustina, D., Paulsen, J. S., Tabrizi, S. J., Poudel, G. R., and Georgiou-Karistianis, N. (2021). Tracking Huntingtons disease progression using motor, functional, cognitive and imaging markers. *Movement disorders : official journal of the Movement Disorder Society*, 36(10):2282.

[2] Acosta, O., Bourgeat, P., Zuluaga, M. A., Fripp, J., Salvado, O., and Ourselin, S. (2009). Automated voxel-based 3D cortical thickness measurement in a combined Lagrangian-Eulerian PDE approach using partial volume maps. *Medical Image Analysis*, 13(5):730–743.

[3] Adesnik, H. and Naka, A. (2018). Cracking the Function of Layers in the Sensory Cortex. *Neuron*, 100(5):1028–1043.

[4] Albert, M. S. (1996). Cognitive and neurobiologic markers of early Alzheimerdisease. *Proceedings of the National Academy of Sciences of the United States of America*, 93(24):13547.

[5] Albin, R. L., Reiner, A., Anderson, K. D., Penney, J. B., and Young, A. B. (1990). Striatal and nigral neuron subpopulations in rigid Huntington's disease: Implications for the functional anatomy of chorea and rigidity-akinesia. *Annals of Neurology*, 27(4):357–365.

[6] Alzheimer, A. (1911). Über die anatomische Grundlage der Huntingtonsche Chorea und der choreatischen Bewegungen überhaupt. *Neurologisches Centralblatt*, 30:891–892.

[7] Andrew, S. E., Goldberg, Y. P., Kremer, B., Telenius, H., Theilmann, J., Adam, S., Starr, E., Squitieri, F., Lin, B., Kalchman, M. A., Graham, R. K., and Hayden, M. R. (1993). The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nature Genetics*, 4:398403.

[8] Anton, G. (1896). Über die Beteiligung der grossen basal Gehirnganglien bei Bewegungsstörungen und insbesondere bei Chorea. *Jahrbücher für Psychiatrie und Neurologie*, 14:141–181.

[9] Armanious, K., Mecky, Y., Gatidis, S., and Yang, B. (2019). Adversarial Inpainting of Medical Image Modalities. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3267–3271.

[10] Ashby, F. G. and Maddox, W. T. (2005). Human category learning. *Annual review of psychology*, 56:149–178.

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

[12] Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A., and Gee, J. C. (2011). An open source multivariate framework for N-tissue segmentation with evaluation on public data. *Neuroinformatics*, 9(4):381–400.

[13] Aylward, E., Li, Q., Stine, O., Ranen, N., Sherr, M., Barta, P., Bylsma, F., Pearlson, G., and Ross, C. (1997). Longitudinal change in basal ganglia volume in patients with Huntington's disease. *Neurology*, 48(2):394–399.

[14] Aylward, E., Mills, J., Liu, D., Nopoulos, P., Ross, C. A., Pierson, R., and Paulsen, J. S. (2011a). Association between Age and Striatal Volume Stratified by CAG Repeat Length in Prodromal Huntington Disease. *PLoS Currents*, 3.

[15] Aylward, E. H., Anderson, N. B., Bylsma, F. W., Wagster, M. V., Barta, P. E., Sherr, M., Feeney, J., Davis, A., Rosenblatt, A., Pearlson, G. D., and Ross, C. A. (1998). Frontal lobe volume in patients with Huntington's disease. *Neurology*, 50(1):252–8.

[16] Aylward, E. H., Harrington, D. L., Mills, J. A., Nopoulos, P. C., Ross, C. A., Long, J. D., Liu, D., Westervelt, H. K., and Paulsen, J. S. (2013). Regional Atrophy Associated with Cognitive and Motor Function in Prodromal Huntington Disease. *Journal of Huntington's Disease*, 2(4):477–489.

[17] Aylward, E. H., Liu, D., Nopoulos, P. C., Ross, C. A., Pierson, R. K., Mills, J. A., Long, J. D., and Paulsen, J. S. (2012). Striatal Volume Contributes to the Prediction of Onset of Huntington Disease in Incident Cases. *Biological Psychiatry*, 71(9):822.

[18] Aylward, E. H., Nopoulos, P. C., Ross, C. A., Langbehn, D. R., Pierson, R. K., Mills, J. A., Johnson, H. J., Magnotta, V. A., Juhl, A. R., and Paulsen, J. S. (2011b). Longitudinal change in regional brain volumes in prodromal Huntington disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 82(4):405–410.

[19] Aylward, E. H., Sparks, B. F., Field, K. M., Yallapragada, V., Shpritz, B. D., Rosenblatt, A., Brandt, J., Gourley, L. M., Liang, K., Zhou, H., Margolis, R. L., and Ross, C. A. (2004). Onset and rate of striatal atrophy in preclinical Huntington disease. *Neurology*, 63(1):66–72.

[20] Baig, S. S., Strong, M., Rosser, E., Taverner, N. V., Glew, R., Miedzybrodzka, Z., Clarke, A., Craufurd, D., and Quarrell, O. W. (2016). 22 Years of predictive testing for Huntington's disease: the experience of the UK Huntington's Prediction Consortium. *European Journal of Human Genetics*, 24(10):1396.

[21] Bates, G. (2003). Huntingtin aggregation and toxicity in Huntington's disease. *Lancet*, 361(9369):1642–1644.

[22] Bekkers, J. M. (2011). Pyramidal neurons. *Current Biology*, 21(24):R975.

[23] Benton, A. L., Sivan, A. B., Hamsher, K. d., Varney, N. R., and Spreen, O. (1983). *Contributions to Neuropsychological Assessment: A Clinical Manual*. Oxford University Press, New York, NY.

[24] Berardelli, A., Noth, J., Thompson, P. D., E M Bollen, E. L., Currà, A., Deuschl, G., Gert van Dijk, J., Töpper, R., Schwarz, M., and C Roos, R. A. (1999). Pathophysiology of Chorea and Bradykinesia in Huntington's Disease. *Movement Disorders*, 14(3):398403.

[25] Bernal, J., Valverde, S., Kushibar, K., Cabezas, M., Oliver, A., and Lladó, X. (2021). Generating Longitudinal Atrophy Evaluation Datasets on Brain Magnetic Resonance Images Using Convolutional Neural Networks and Segmentation Priors. *Neuroinformatics*, 19(3):477–492.

[26] Bogaard, S. J. A., Dumas, E. M., Acharya, T. P., Johnson, H., Langbehn, D. R., Scahill, R. I., Tabrizi, S. J., Buchem, M. A., Grond, J., and Roos, R. A. C. (2011). Early atrophy of pallidum and accumbens nucleus in Huntingtons disease. *Journal of Neurology*, 258(3):412–420.

[27] Brant-Zawadzki, M., Gillan, G. D., and Nitz, W. R. (1992). MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence–initial experience in the brain. *Radiology*, 182(3):769–775.

[28] Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Groβhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues : Brodmann, K : Free Download, Borrow, and Streaming : Internet Archive*. Barth, Leipzig, Germany.

[29] Busch, A., Engemann, S., Lurz, R., Okazawa, H., Lehrach, H., and Wanker, E. E. (2003). Mutant Huntingtin Promotes the Fibrillogenesis of Wild-type Huntingtin. *Journal of Biological Chemistry*, 278(42):41452–41461.

[30] Cardoso, M. J., Clarkson, M. J., Modat, M., and Ourselin, S. (2011a). Longitudinal cortical thickness estimation using Khalimsky's cubic complex. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 6892 LNCS, pages 467–475.

[31] Cardoso, M. J., Clarkson, M. J., Modat, M., and Ourselin, S. (2011b). On the Extraction of Topologically Correct Thickness Measurements Using Khalimskys Cubic Complex. In Székely, G. and Hahn, H., editors, *IPMI 2011: Information Processing in Medical Imaging*, pages 467–475. Springer, Berlin, Heidelberg.

[32] Carman, G. J. (1990). *Mappings of the Cerebral Cortex*. PhD thesis, California Institute of Technology, Pasadena, CA.

[33] Catani, M. (2022). Neuroanatomical Bases of Human Behavior. *Encyclopedia of Behavioral Neuroscience: Second Edition*, 1-3:60–64.

[34] Clarkson, M. J., Cardoso, M. J., Ridgway, G. R., Modat, M., Leung, K. K., Rohrer, J. D., Fox, N. C., and Ourselin, S. (2011). A comparison of voxel and surface based cortical thickness estimation methods. *NeuroImage*, 57(3):856–865.

[35] Collins, D. L., Neelin, P., Peters, T. M., and Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of computer assisted tomography*, 18(2):192205.

[36] Coppen, E. M., Jacobs, M., van den Berg-Huysmans, A. A., van der Grond, J., and Roos, R. A. (2018). Grey matter volume loss is associated with specific clinical motor signs in Huntington's disease. *Parkinsonism and Related Disorders*, 46:56–61.

[37] Dale, A. and Sereno, M. (1993). Improved Localization of Cortical Activity by Combining EEG and MEG with MRI Cortical Surface Reconstruction: A Linear Approach. *Journal of Cognitive Neuroscience*, 5(2):162–176.

[38] Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical Surface-Based Analysis I. Segmentation and Surface Reconstruction. *NeuroImage*, 9(2):195–207.

[39] Das, S. R., Avants, B. B., Grossman, M., and Gee, J. C. (2009). Registration Based Cortical Thickness Measurement. *NeuroImage*, 45(3):867–879.

[40] Davatzikos, C., Genc, A., Xu, D., and Resnick, S. M. (2001). Voxel-Based Morphometry Using the RAVENS Maps: Methods and Validation Using Simulated Longitudinal Atrophy. *NeuroImage*, 14(6):1361–1369.

[41] De La Monte, S. M., Vonsattel, J. P., and Richardson, E. P. (1988). Morphometric demonstration of atrophic changes in the ceRebral cortex, white matter, and neostriatum in huntingtons disease. *Journal of Neuropathology and Experimental Neurology*, 47(5):516–525.

[42] Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., and Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980.

[43] Deyoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D., and Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences*, 93(6):2382–2386.

[44] Diehl, S. K., Mefferd, A. S., Lin, Y. C., Sellers, J., McDonell, K. E., De Riesthal, M., and Claassen, D. O. (2019). Motor speech patterns in Huntington disease. *Neurology*, 93(22):e2042–e2052.

[45] Domínguez D, J. F., Poudel, G., Stout, J. C., Gray, M., Chua, P., Borowsky, B., Egan, G. F., and Georgiou-Karistianis, N. (2017). Longitudinal changes in the fronto-striatal network are associated with executive dysfunction and behavioral dysregulation in Huntington's disease: 30 months IMAGE-HD data. *Cortex*, 92:139–149.

[46] Douaud, G., Gaura, V., Ribeiro, M.-j., Lethimonnier, F., Maroy, R., Verny, C., Krystkowiak, P., Damier, P., Bachoud-Levi, A.-c., Hantraye, P., Remy, P., and Mondor, H. (2006). Distribution of grey matter atrophy in Huntington's disease patients: A combined ROI-based and voxel-based morphometric study. *NeuroImage*, 32(4):156–1575.

[47] Drury, H. A., Van Essen, D. C., Anderson, C. H., Lee, C. W., Coogan, T. A., and Lewis, J. W. (1996). Computerized mappings of the cerebral cortex: a multiresolution flattening method and a surface-based coordinate system. *Journal of cognitive neuroscience*, 8(1):1–28.

[48] Dunlap, C. B. (1927). Pathologic changes in Huntington's chorea: with special reference to the corpus striatum. *Archives of Neurology & Psychiatry*, 18(6):867–943.

[49] Fan, H. C., Ho, L. I., Chi, C. S., Chen, S. J., Peng, G. S., Chan, T. M., Lin, S. Z., and Harn, H. J. (2014). Polyglutamine (PolyQ) Diseases: Genetics to Treatments. *http://dx.doi.org/10.3727/096368914X678454*, 23(4-5):441–458.

[50] Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J. V., Pieper, S., and Kikinis, R. (2012). 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magnetic resonance imaging*, 30(9):1323.

[51] Fein, G., Di Sclafani, V., Tanabe, P. J., Cardenas, V. A., Weiner, M. W., Jagust, W. J., Reed, B. R., Norman, D. M., Schuff, N., Kusdra, L., Greenfield, T. M., and Chui, H. (2000). Hippocampal and cortical atrophy predict dementia in subcortical ischemic vascular disease. *Neurology*, 55(11):1626–1635.

[52] Fischl, B. (2012). FreeSurfer. *NeuroImage*, 62(2):774–81.

[53] Fischl, B. and Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, 97(20):11050–11055.

[54] Fischl, B., Sereno, M. I., Tootell, R. B., and Dale, A. M. (1999). Highresolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, 8(4):272.

[55] Fischl, B., van der Kouwe, A. J., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., and Dale, A. M. (2004). Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*, 14(1):11–22.

[56] Franklin, G. L., Camargo, C. H. F., Meira, A. T., Pavanelli, G. M., Milano, S. S., Germiniani, F. B., Lima, N. S. C., Raskin, S., Barsottini, O. G. P., Pedroso, J. L., Maggi, F. A., Tumas, V., de Carvalho, P. M., de Oliveira, A. C., Braga, B., Souza, L. C., Guimarães, R. P., Piovesana, L. G., Lopes-Cendes, . T., de Azevedo, P. C., França, M. C., Martinez, A. R. M., and Teive, H. A. G. (2020). Is Ataxia an Underestimated Symptom of Huntington's Disease? *Frontiers in Neurology*, 11.

[57] Frederick, C. and Schwartz, E. L. (1990). Brain peeling: viewing the inside of a laminar three-dimensional solid. *The Visual Computer*, 6(1):37–49.

[58] Freeborough, P. A. and Fox, N. C. (1997). The boundary shift integral: An accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Transactions on Medical Imaging*, 16(5):623–629.

[59] Frisoni, G. B., Beltramello, A., Weiss, C., Geroldi, C., Bianchetti, A., and Trabucchi, M. (1996). Linear measures of atrophy in mild Alzheimer disease. *AJNR: American Journal of Neuroradiology*, 17(5):913.

[60] Gauthier, L. R., Charrin, B. C., Borrell-Pagès, M., Dompierre, J. P., Rangone, H., Cordelières, F. P., De Mey, J., MacDonald, M. E., Leßmann, V., Humbert, S., and Saudou, F. (2004). Huntingtin controls neurotrophic support and survival of neurons by enhancing BDNF vesicular transport along microtubules. *Cell*, 118(1):127–138.

[61] Gennatas, E. D., Avants, B. B., Wolf, D. H., Satterthwaite, T. D., Ruparel, K., Ciric, R., Hakonarson, H., Gur, R. E., and Gur, R. C. (2017). Age-Related Effects and Sex Differences in Gray Matter Density, Volume, Mass, and Cortical Thickness from Childhood to Young Adulthood. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 37(20):5065–5073.

[62] Georgiou, N., Bradshaw, J. L., Phillips, J. G., Chiu, E., and Bradshaw, J. A. (1995). Reliance on advance information and movement sequencing in huntington's disease. *Movement Disorders*, 10(4):472–481.

[63] Ginestroni, A., Battaglini, M., Diciotti, S., Della Nave, R., Mazzoni, L. N., Tessa, C., Giannelli, M., Piacentini, S., De Stefano, N., and Mascalchi, M. (2010). Magnetization transfer MR imaging demonstrates degeneration of the subcortical and cortical gray matter in Huntington disease. *American Journal of Neuroradiology*, 31(10):1807–1812.

[64] Gray, J. M., Young, A. W., Barker, W. A., Curtis, A., and Gibson, D. (1997). Impaired recognition of disgust in Huntington's disease gene carriers. *Brain*, 120:2029–2038.

[65] Haidar, H., Egorova, S., and Soul, J. S. (2005). New numerical solution of the laplace equation for tissue thickness measurement in three-dimensional MRI. In *Journal of Mathematical Modelling and Algorithms*, volume 4, pages 83–97.

[66] Hajnal, J. V., Bryant, D. J., Kasuboski, L., Pattany, P. M., De Coene, B., Lewis, P. D., Pennock, J. M., Oatridge, A., Young, I. R., and Bydder, G. M. (1992). Use of fluid attenuated inversion recovery (FLAIR) pulse sequences in MRI of the brain. *Journal of computer assisted tomography*, 16(6):841–844.

[67] Hall, J. E. (2016). *Guyton and Hall Textbook of Medical Physiology*. Elsevier, Philadelphia, PA, 13th edition.

[68] Halliday, G. M., McRitchie, D. A., Macdonald, V., Double, K. L., Trent, R. J., and McCusker, E. (1998). Regional Specificity of Brain Atrophy in Huntington's Disease. *Experimental Neurology*, 154(2):663–672.

[69] Han, X., Pham, D. L., Tosun, D., Rettmann, M. E., Xu, C., and Prince, J. L. (2004). CRUISE: Cortical reconstruction using implicit surface evolution. *NeuroImage*, 23(3):997–1012.

[70] Hedreen, J. C., Peyser, C. E., Folstein, S. E., and Ross, C. A. (1991). Neuronal loss in layers V and VI of cerebral cortex in Huntington's disease. *Neuroscience Letters*, 133(2):257.

[71] Henery, C. C. and Mayhew, T. M. (1989). The cerebrum and cerebellum of the fixed human brain: efficient and unbiased estimates of volumes and cortical surface areas. *Journal of Anatomy*, 167:167.

[72] Henley, S. M., Wild, E. J., Hobbs, N. Z., Scahill, R. I., Ridgway, G. R., MacManus, D. G., Barker, R. A., Fox, N. C., and Tabrizi, S. J. (2009). Relationship between CAG repeat length and brain volume in premanifest and early Huntingtons disease. *Journal of Neurology 2009 256:2*, 256(2):203–212.

[73] Hett, K., Giraud, R., Johnson, H., Paulsen, J. S., Long, J. D., and Oguz, I. (2020a). Patch-Based Abnormality Maps for Improved Deep Learning-Based Classification of Huntington's Disease. In Martel, A. L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M. A., Zhou, S. K., Racoceanu, D., and Joskowicz, L., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pages 636–645, Cham. Springer International Publishing.

[74] Hett, K., Johnson, H., Coupé, P., Paulsen, J. S., Long, J. D., and Oguz, I. (2020b). Tensor-Based Grading: A Novel Patch-Based Grading Approach for the Analysis Of Deformation Fields in Huntington's Disease. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1091–1095.

[75] Ho, A. K., Sahakian, B. J., Robbins, T. W., Barker, R. A., Rosser, A. E., and Hodges, J. R. (2002). Verbal fluency in Huntington's disease: A longitudinal analysis of phonemic and semantic clustering and switching. *Neuropsychologia*, 40(8):1277–1284.

[76] Huntington, G. (1872). On Chorea. *The Medical and Surgical Reporter: A Weekly Journal*, 26(15):317–321.

[77] Huntington Study Group (1996). Unified Huntingtons Disease Rating Scale: Reliability and Consistency. *Movement Disorders*, 11(2):136–142.

[78] Hutton, C., De Vita, E., Ashburner, J., Deichmann, R., and Turner, R. (2008). Voxel-based cortical thickness measurements in MRI. *NeuroImage*, 40(4):1701–1710.

[79] Hutton, C., De Vita, E., and Turner, R. (2002). Sulcal segmentation for cortical thickness measurements. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 2488, pages 443–450. Springer Verlag.

[80] Jack, C. R. J., Petersen, R. C., Xu, Y. C., Waring, S. C., O'Brien, P. C., Tangalos, E. G., Smith, G. E., Ivnik, R. J., and Kokmen, E. (1997). Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology*, 49(3):786–794.

[81] Jaume, S., Rondao, P., and Macq, B. (2005). Open Topology: A Toolkit for Brain Isosurface Correction. *The Insight Journal*.

[82] Jensen, P. S. (1972). Finite difference techniques for variable grids. *Computers & Structures*, 2(1-2):17–29.

[83] Johnson, E. B., Ziegler, G., Penny, W., Rees, G., Tabrizi, S. J., Scahill, R. I., and Gregory, S. (2021). Dynamics of Cortical Degeneration Over a Decade in Huntington's Disease. *Biological Psychiatry*, 89(8):807–816.

[84] Johnson, S. A., Stout, J. C., Solomon, A. C., Langbehn, D. R., Aylward, E. H., Cruce, C. B., Ross, C. A., Nance, M., Kayson, E., Julian-Baros, E., Hayden, M. R., Kieburtz, K., Guttman, M., Oakes, D., Shoulson, I., Beglinger, L., Duff, K., Penziner, E., and Paulsen, J. S. (2007). Beyond disgust: impaired recognition of negative emotions prior to diagnosis in Huntington's disease. *Brain*, 130(7):1732–1744.

[85] Jones, S. E., Buchbinder, B. R., and Aharon, I. (2000). Three-dimensional mapping of cortical thickness using Laplace's equation. *Human Brain Mapping*, 11(1):12–32.

[86] Juan, F. D. D., Julie, C. S., Govinda, P., Andrew, C., Phyllis, C., Gary, F. E., and Nellie, G. K. (2016). Multimodal imaging biomarkers in premanifest and early Huntington's disease: 30-month IMAGE-HD data. *British Journal of Psychiatry*, 208(6):571–578.

[87] Jubault, T., Gagnon, J. F., Karama, S., Ptito, A., Lafontaine, A. L., Evans, A. C., and Monchi, O. (2011). Patterns of cortical thickness and surface area in early Parkinson's disease. *NeuroImage*, 55(2):462–467.

[88] Julien, C. L., Thompson, J. C., Wild, S., Yardumian, P., Snowden, J. S., Turner, G., and Craufurd, D. (2007). Psychiatric disorders in preclinical Huntingtons disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 78(9):939–943.

[89] Karaçali, B. and Davatzikos, C. (2006). Simulation of Tissue Atrophy Using a Topology Preserving Transformation Model. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 25(5).

[90] Kassubek, J., Juengling, F. D., Ecker, D., and Landwehrmeyer, G. B. (2005). Thalamic Atrophy in Huntington's Disease Co-varies with Cognitive Performance: A Morphometric MRI Analysis. *Cerebral Cortex*, 15(846-853).

[91] Kassubek, J., Juengling, F. D., Kioschies, T., Henkel, K., Karitzky, J., Kramer, B., Ecker, D., Andrich, J., Saft, C., Kraus, P., Aschoff, A. J., and Ludolph, A. C. (2004). Topography of cerebral atrophy in early Huntington's disease: a voxel based morphometric MRI study. *Journal of Neurology, Neurosurgery, & Psychiatry*, 75(2):213–220.

[92] Kaye, J. A., Swihart, T., Howieson, D., Dame, A., Moore, M. M., Karnos, T., Camicioli, R., Ball, M., Oken, B., and Sexton, G. (1997a). Volume loss of the hippocampus and temporal lobe in healthy elderly persons destined to develop dementia.

[93] Kaye, J. A., Swihart, T., Howieson, D., Dame, A., Moore, M. M., Karnos, T., Camicioli, R., Ball, M., Oken, B., and Sexton, G. (1997b). Volume loss of the hippocampus and temporal lobe in healthy elderly persons destined to develop dementia. *Neurology*, 48(5):1297–1304.

[94] Khanal, B., Ayache, N., and Pennec, X. (2017). Simulating longitudinal brain MRIs with known volume changes and realistic variations in image intensity. *Frontiers in Neuroscience*, 11(MAR):132.

[95] Kuan, W. L., Kasis, A., Yuan, Y., Mason, S. L., Lazar, A. S., Barker, R. A., and Goncalves, J. (2015). Modelling the natural history of Huntington's disease progression. *Journal of Neurology, Neurosurgery and Psychiatry*, 86(10):1143–1149.

[96] Kwon, J. S., McCarley, R. W., Hirayasu, Y., Anderson, J. E., Fischer, I. A., Kikinis, R., Jolesz, F. A., and Shenton, M. E. (1999). Left Planum Temporale Volume Reduction in Schizophrenia. *Archives of General Psychiatry*, 56(2):142–148.

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

[Langbehn et al.] Langbehn, D. R., Hayden, M., and Paulsen, J. S. CAG-Repeat Length and the Age of Onset in Huntington Disease (HD): A Review and Validation Study of Statistical Approaches.

[99] Lange, H. H. (1981). Quantitative changes of telencephalon, diencephalon, and mesencephalon in Huntington's chorea, postencephalitic, and idiopathic parkinsonism. *Verhandlungen der Anatomischen Gesellschaft*, 75:923–925.

[100] Lannois, M. and Pavoit, J.-M. (1897). Deux cas de chorée hereditaire avec autopsies. *Archives de Neurologie*, 4:333–334.

[101] Larson, K. E. and Oguz, I. (2021). Synthetic atrophy for longitudinal surface-based cortical thickness measurement. In Landman, B. A. and Išgum, I., editors, *Medical Imaging 2021: Image Processing*, volume 11596, page 125. SPIE.

[102] Larson, K. E. and Oguz, I. (2022). Synthetic Atrophy for Longitudinal Cortical Surface Analyses. *Frontiers in Neuroimaging*.

[103] Lee, J., Ehlers, C., Crews, F., Niethammer, M., Budin, F., Paniagua, B., Sulik, K., Johns, J., Styner, M., and Oguz, I. (2011). Automatic cortical thickness analysis on rodent brain. In *Proceedings of SPIE–the International Society for Optical Engineering*, page 796248179624811.

[104] Li, B., Niessen, W. J., Klein, S., Ikram, M. A., Vernooij, M. W., and Bron, E. E. (2021a). Learning unbiased group-wise registration (LUGR) and joint segmentation: evaluation on longitudinal diffusion MRI. In Landman, B. A. and Išgum, I., editors, *Medical Imaging 2021: Image Processing*, volume 11596, page 14. SPIE.

[105] Li, H., Zhang, H., Hu, D., Johnson, H., Long, J. D., Paulsen, J. S., and Oguz, I. (2020). Generalizing MRI Subcortical Segmentation to Neurodegeneration. *Machine learning in clinical neuroimaging and radiogenomics in neuro-oncology : Third International Workshop, MLCN 2020, and Second International Workshop, RNO-AI 2020, held in conjunction with MICCAI 2020, Lima, Peru, October 4-8, 2020...*, 12449:139–147.

[106] Li, H., Zhang, H., Johnson, H., Long, J. D., Paulsen, J. S., and Oguz, I. (2021b). Longitudinal subcortical segmentation with deep learning. In Išgum, I. and Landman, B. A., editors, *Medical Imaging 2021: Image Processing*, volume 11596, page 115960D. International Society for Optics and Photonics, SPIE.

[107] Li, Y., Wang, Y., Wu, G., Shi, F., Zhou, L., Lin, W., Shen, D., Alzheimer's Disease, T., and Initiative, N. (2012). Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features.

[108] Liu, D., Long, J. D., Zhang, Y., Raymond, L. A., Marder, K., Rosser, A., Mccusker, E. A., Mills, J. A., and Paulsen, J. S. (2015). Motor onset and diagnosis in Huntington disease using the diagnostic confidence level. *Journal of Neurology*, 262(12):2691–2698.

[109] Long, J. D., Paulsen, J. S., Marder, K., Zhang, Y., Kim, J.-I., and Mills, J. A. (2014). Tracking Motor Impairments in the Progression of Huntington's Disease. *Movement Disorders*, 29(3):311–319.

[110] Lorensen, W. E. and Cline, H. E. (1987). MARCHING CUBES: A HIGH RESOLUTION 3D SURFACE CONSTRUCTION ALGORITHM. *Computer Graphics*, 21(4).

[111] Lynch, D. R. (2005). *Numerical partial differential equations for environmental scientists and engineers; a first practical course*. Springer Science+Business Media, New York, NY.

[112] MacDonald, M. E., Ambrose, C. M., Duyao, M. P., Myers, R. H., Lin, C., Srinidhi, L., Barnes, G., Taylor, S. A., James, M., Groot, N., MacFarlane, H., Jenkins, B., Anderson, M. A., Wexler, N. S., Gusella, J. F., Bates, G. P., Baxendale, S., Hummerich, H., Kirby, S., North, M., Youngman, S., Mott, R., Zehetner, G., Sedlacek, Z., Poustka, A., Frischauf, A. M., Lehrach, H., Buckler, A. J., Church, D., Doucette-Stamm, L., O'Donovan, M. C., Riba-Ramirez, L., Shah, M., Stanton, V. P., Strobel, S. A., Draths, K. M., Wales, J. L., Dervan, P., Housman, D. E., Altherr, M., Shiang, R., Thompson, L., Fielder, T., Wasmuth, J. J., Tagle, D., Valdes, J., Elmer, L., Allard, M., Castilla, L., Swaroop, M., Blanchard, K., Collins, F. S., Snell, R., Holloway, T., Gillespie, K., Datson, N., Shaw, D., and Harper, P. S. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell*, 72(6):971–983.

[113] Majid, D. S., Aron, A. R., Thompson, W., Sheldon, S., Hamza, S., Stoffers, D., Holland, D., Goldstein, J., Corey-Bloom, J., and Dale, A. M. (2011a). Basal ganglia atrophy in prodromal Huntington's disease is detectable over one year using automated segmentation. *Movement Disorders*, 26(14):2544–2551.

[114] Majid, D. S., Stoffers, D., Sheldon, S., Hamza, S., Thompson, W. K., Goldstein, J., Corey-Bloom, J., and Aron, A. R. (2011b). Automated structural imaging analysis detects premanifest Huntington's disease neurodegeneration within 1 year. *Movement Disorders*, 26(8):1481–1488.

[115] Mason, S. L., Zhang, J., Begeti, F., Guzman, N. V., Lezar, A. S., Rowe, J. B., Barker, R. A., and Hampshire, A. (2015). The role of the amygdala during emotional processing in Huntington's disease: From pre-manifest to late stage disease. *Neuropsychologia*, 70:80–89.

[116] Mehrabi, N. F., Waldvogel, H. J., Tippett Phd, L. J., Hogg, V. M., Synek, B. J., and Faull Mbchb, R. L. M. (2016). Symptom heterogeneity in Huntington's disease correlates with neuronal degeneration in the cerebral cortex. *Neurobiology of Disease*, 96:67–74.

[117] Meyer, J. R., Roychowdhury, S., Russell, E. J., Callahan, C., Gitelman, D., and Mesulam, M. M. (1996). Location of the central sulcus via cortical thickness of the precentral and postcentral gyri on MR. *AJNR: American Journal of Neuroradiology*, 17(9):1699.

[118] Minkova, L., Gregory, S., Scahill, R. I., Abdulkadir, A., Kaller, C. P., Peter, J., Long, J. D., Stout, J. C., Reilmann, R., Roos, R. A., Durr, A., Leavitt, B. R., Tabrizi, S. J., and Klöppel, S. (2018). Cross-sectional and longitudinal voxel-based grey matter asymmetries in Huntington's disease. *NeuroImage: Clinical*, 17.

[119] Möncke-Buchner, E., Reich, S., Mücke, M., Reuter, M., Messer, W., Wanker, E. E., and Krüger, D. H. (2002). Counting CAG repeats in the Huntingtons disease gene by restriction endonuclease EcoP15I cleavage. *Nucleic Acids Research*, 30(16):e83.

[120] Morris, L. A., O'Callaghan, C., and Le Heron, C. (2022). Disordered Decision Making: A Cognitive Framework for Apathy and Impulsivity in Huntington's Disease. *Movement Disorders*, 37(6):1149–1163.

[121] Mugler, J. P. and Brookeman, J. R. (1990). Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magnetic Resonance in Medicine*, 15(1):152–157.

[122] Oguz, I. and Sonka, M. (2014a). LOGISMOS-B: Layered Optimal graph image segmentation of multiple objects and surfaces for the brain. *IEEE Transactions on Medical Imaging*.

[123] Oguz, I. and Sonka, M. (2014b). Robust Cortical Thickness Measurement with LOGISMOS-B. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2014*, pages 722–730. Springer International Publishing Switzerland.

[124] Oguz, I., Styner, M., Sanchez, M., Shi, Y., and Sonka, M. (2015). LOGISMOS-B for Primates: Primate Cortical Surface Reconstruction and Thickness Measurement. *Proceedings of SPIE–the International Society for Optical Engineering*, 9413:941313.

[125] Pagan, F., Torres-Yaghi, Y., and Altshuler, M. (2017). The diagnosis and natural history of Huntington disease. *Handbook of Clinical Neurology*, 144:63–67.

[126] Paulsen, J. S. (2011). Cognitive Impairment in Huntington Disease: Diagnosis and Treatment. *Current Neurology and Neuroscience Reports*, 11:474–483.

[127] Paulsen, J. S., Long, J. D., Johnson, H. J., Aylward, E. H., Ross, C. A., Williams, J. K., Nance, M. A., Erwin, C. J., Westervelt, H. J., Harrington, D. L., Bockholt, H. J., Zhang, Y., McCusker, E. A., Chiu, E. M., Panegyres, P. K., Cross, S., Ryan, P., Epping, E. A., Preston, J., Goh, A., Antonopoulos, S., Loi, S., Raymond, L., Decolongon, J., Fan, M., Coleman, A., Mallone, W. M., Suter, G., Varvaris, M., Yoritomo, N., Griffith, J., Loy, C., Gunn, D., Guttman, M., Sheinberg, A., Law, A., Quaid, K., Wesson, M., Wojcieszek, J., Perlmutter, J., Barton, S., Smith, S., Barker, R. A., Mason, S., Guzman, N. V., Perlman, S., Clemente, B., Jones, R., Wood-Siverio, C., Factor, S. A., Samii, A., Macaraeg, A., Lee, J., Tedesco, M., Maxwell, B., Kumar, R., Erickson, D., Nickels, B., Marshall, F., Chesire, A., Wodarski, M., Hickey, C., Geschwind, M. D., Sha, S., Satris, G., Ahmed, A., Reece, C., Bura, A., Mourany, L., Pillai, J., Mazzoni, P., Marder, K., Wasserman, P., Craufurd, D., Bek, J., Howard, E., Warner, T., Burrows, M., Orth, M., Süßmuth, S., Barth, K., Trautmann, S., Schwenk, D., Eschenbach, C., Wheelock, V., Kjer, L., Martin, A., Farias, S., Miedzybrodzka, Z., Rae, D., D'Alessandro, M., Suchowersky, O., Chua, P., Komiti, A., Rosas, H. D., Rosser, A., Price, K., Hunt, S., Jankovic, J., Ondo, W., Martin, W., King, P., Wieler, M., Sran, S., de Yébenes, J. G., and Dubinsky, R. (2014a). Clinical and biomarker changes in premanifest Huntington disease show trial feasibility: A decade of the PREDICT-HD study. *Frontiers in Aging Neuroscience*, 6(APR).

[128] Paulsen, J. S., Long, J. D., Ross, C. A., Harrington, D. L., Erwin, C. J., Williams, J. K., Westervelt, H. J., Johnson, H. J., Aylward, E. H., Zhang, Y., Bockholt, H. J., and Barker, R. A. (2014b). Prediction of manifest huntington's disease with clinical and imaging measures: A prospective observational study. *The Lancet Neurology*, 13(12):1193–1201.

[129] Paulsen, J. S., Magnotta, V. A., Mikos, A. E., Paulson, H. L., Penziner, E., Andreasen, N. C., and Nopoulos, P. C. (2006). Brain Structure in Preclinical Huntingtons Disease. *Biological Psychiatry*, 59(1):57–63.

[130] Peavy, G. M., Jacobson, M. W., Goldstein, J. L., Hamilton, J. M., Kane, A., Gamst, A. C., Lessig, S. L., Lee, J. C., and Corey-Bloom, J. (2010). Cognitive and Functional Decline in Huntington's Disease: Dementia Criteria Revisited. *Movement disorders : official journal of the Movement Disorder Society*, 25(9):1163.

[131] Penney, J. B., Vonsattel, J. P., MacDonald, M. E., Gusella, J. F., and Myers, R. H. (1997). CAG repeat number governs the development rate of pathology in huntington's disease. *Annals of Neurology*, 41(5):689–692.

[132] Pereira, J. B., Ibarretxe-Bilbao, N., Marti, M. J., Compta, Y., Junqué, C., Bargallo, N., and Tolosa, E. (2012). Assessment of cortical degeneration in patients with Parkinson's disease by voxel-based morphometry, cortical folding, and cortical thickness. *Human Brain Mapping*, 33(11):2521–2534.

[133] Pham, D. L. and Prince, J. L. (1999). Adaptive fuzzy segmentation of magnetic resonance images. *IEEE transactions on medical imaging*, 18(9):737–752.

[134] Pla, P., Orvoen, S., Saudou, F., David, D. J., Humbert, S., Adam Samuels, B., Belzung, C., and Francois Rabelais, U. (2014). Mood disorders in Huntington's disease: from behavior to cellular and molecular mechanisms.

[135] Pruessmann, K. P., Weiger, M., Scheidegger, M. B., and Boesiger, P. (1999). SENSE: sensitivity encoding for fast MRI. *Magnetic resonance in medicine*, 42(5):952–962.

[136] Rangone, H., Humbert, S., and Saudou, F. (2004). Huntington's disease: How does huntingtin, an anti-apoptotic protein, become toxic? *Pathologie Biologie*, 52(6):338–342.

[137] Reilmann, R., Leavitt, B. R., and Ross, C. A. (2014). Diagnostic criteria for Huntington's disease based on natural history. *Movement Disorders*, 29(11):1335–1341.

[138] Reiner, A., Dragatsis, I., and Dietrich, P. (2011). Genetics and Neuropathology of Huntington's Disease. *International review of Neurobiology*, 98:325.

[139] Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual and Motor Skills*, 8(7):271.

[140] Reuter, M., Rosas, H. D., and Fischl, B. (2010). Highly accurate inverse consistent registration: A robust approach. *NeuroImage*, 53(4):1181–1196.

[141] Reuter, M., Schmansky, N. J., Rosas, H. D., and Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*, 61(4):1402–1418.

[142] Rodda, R. A. (1981). Cerebellar atrophy in Huntington's disease. *Journal of the Neurological Sciences*, 50:147–157.

[143] Rosas, H., Hevelone, N., Zaleta, A., Greve, D., Salat, D., Fischl, B., and Hev-elone, N. (2005). Regional cortical thinning in preclinical Huntington disease and its relationship to cognition for Neuroimaging of Aging and Neurodegenerative Diseases. *Neurology*, 65(5):745–747.

[144] Rosas, H. D., Feigin, A. S., and Hersch, S. M. (2004). Using Advances in Neuroimaging to Detect, Understand, and Monitor Disease Progression in Huntington's Disease. *Neurotherapeutics*, 1(2):263–272.

[145] Rosas, H. D., Goodman, J., Chen, Y. I., Jenkins, B. G., Kennedy, D. N., Makris, N., Patti, M., Seidman, L. J., Beal, M. F., and Koroshetz, W. J. (2001). Striatal volume loss in HD as measured by MRI and the influence of CAG repeat. *Neurology*, 57(6):1025–1028.

[146] Rosas, H. D., Koroshetz, W. J., Chen, Y. I., Skeuse, C., Vangel, M., Cudkowicz, M. E., Caplan, K., Marek, K., Seidman, L. J., Makris, N., Jenkins, B. G., and Goldstein, J. M. (2003). Evidence for more widespread cerebral pathology in early HD: An MRI-based morphometric analysis. *Neurology*, 60(10):1615–1620.

[147] Rosas, H. D., Liu, A. K., Hersch, S., Glessner, M., Ferrante, R. J., Salat, D. H., Van Der Kouwe, A., Jenkins, B. G., Dale, A. M., and Fischl, B. (2002). Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology*, 58(5):695–701.

[148] Rosas, H. D., Salat, D. H., Lee, S. Y., Zaleta, A. K., Pappu, V., Fischl, B., Greve, D., Hevelone, N., and Hersch, S. M. (2008). Cerebral cortex and the clinical expression of Huntington's disease: Complexity and heterogeneity. *Brain*, 131(4):1057–1068.

[149] Ross, C. A., Aylward, E. H., Wild, E. J., Langbehn, D. R., Long, J. D., Warner, J. H., Scahill, R. I., Leavitt, B. R., Stout, J. C., Paulsen, J. S., Reilmann, R., Unschuld, P. G., Wexler, A., Margolis, R. L., and Tabrizi, S. J. (2014). Huntington disease: natural history, biomarkers and prospects for therapeutics. *Nature Publishing Group*, 10:204–216.

[150] Rüb, U., Seidel, K., Heinsen, H., Vonsattel, J. P., den Dunnen, W. F., and Korf, H. W. (2016). Huntington's disease (HD): the neuropathology of a multisystem neurodegenerative disorder of the human brain. *Brain Pathology*, 26(6):726–740.

[151] Ruocco, H. H., Bonilha, L., Li, L. M., Lopes-Cendes, I., Cendes, F., and Cendes, F. (2008). Longitudinal analysis of regional grey matter loss in Huntington disease: effects of the length of the expanded CAG repeat. *Journal of Neurology, Neurosurgery, & Psychiatry*, 79:130–135.

[152] Rycroft, C. H. (2009). VORO++: A three-dimensional Voronoi cell library in C++. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 19(4):041111.

[153] Salat, D. H., Buckner, R. L., Snyder, A. Z., Greve, D. N., Desikan, R. S., Busa, E., Morris, J. C., Dale, A. M., and Fischl, B. (2004). Thinning of the cerebral cortex in aging. *Cerebral Cortex*, 14(7):721–730.

[154] Sanderson, C. and Curtin, R. (2016). Armadillo: a template-based C++ library for linear algebra. *Journal of Open Source Software*, 1:26.

[155] Scahill, R. I., Andre, R., Tabrizi, S. J., and Aylward, E. H. (2017). Structural imaging in premanifest and manifest Huntington disease. In *Handbook of Clinical Neurology*, volume 144, pages 247–261. Elsevier B.V.

[156] Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G. P., Davies, S. W., Lehrach, H., and Wanker, E. E. (1997). Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. *Cell*, 90(3):549–558.

[157] Schroeder, W., Martin, K., and Lorensen, B. (2006). *The Visualization Toolkit*. Kitware, 4th edition.

[158] Schwartz, E. L. (1993). Computer-Aided Neuroanatomy of Macaque Visual Cortex. In *Computational Neuroscience*, page 295315. MIT Press, Cambridge, MA, USA.

[159] Schwartz, E. L. and Merker, B. (1986). Computer-Aided Neuroanatomy: Computer-Aided Geometry of Cortical Surfaces and an Optimal Flattening Algorithm. *IEEE Computer Graphics and Applications*, 6(3):36–44.

[160] Schwartz, E. L., Shaw, A., and Wolfson, E. (1989). A Numerical Solution to the Generalized Mapmaker's Problem: Flattening Nonconvex Polyhedral Surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(9):1005–1008.

[161] Shiee, N., Bazin, P.-L., Cuzzocreo, J. L., Ye, C., Kishore, B., Carass, A., Calabresi, P. A., Reich, D. S., Prince, J. L., and Pham, D. L. (2014). Reconstruction of the Human Cerebral Cortex Robust to White Matter Lesions: Method and Validation. *Human Brain Mapping2*, 35:3385–3401.

[162] Si, H. (2015). TetGen, a Delaunay-Based Quality Tetrahedral Mesh Generator. *ACM Trans. Math. Softw.*, 41(2).

[163] Singh-Bains, M. K., Waldvogel, H. J., and Faull, R. L. M. (2016). The role of the human globus pallidus in Huntington's disease. *Brain Pathology*, 26(6):741–751.

[164] Smith, A. (1982). *Symbol digit modalities test (SDMT) manual*. Western Psychological Services, Los Angeles, CA.

[165] Smith, S. M., De Stefano, N., Jenkinson, M., and Matthews, P. M. (2001). Normalized accurate measurement of longitudinal brain change. *Journal of computer assisted tomography*, 25(3):466–475.

[166] Smith, S. M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P. M., Federico, A., and De Stefano, N. (2002). Accurate, Robust, and Automated Longitudinal and Cross-Sectional Brain Change Analysis. *NeuroImage*, 17(1):479–489.

[167] Snowden, J. S., Craufurd, D., Thompson, J., and Neary, D. (2010). Psychomotor, Executive, and Memory Function in Preclinical Huntington's Disease. *http://dx.doi.org/10.1076/jcen.24.2.133.998*, 24(2):133–145.

[168] Snowden, J. S., Gibbons, Z. C., Blackshaw, A., Doubleday, E., Thompson, J., Craufurd, D., Foster, J., Happé, F., and Neary, D. (2003). Social cognition in frontotemporal dementia and Huntington's disease. *Neuropsychologia*, 41(6):688–701.

[169] Sonne, J., Reddy, V., and Beato, M. R. (2022). Neuroanatomy, Substantia Nigra. In *StatPearls*. StatPearls Publishing, Treasure Island, FL.

[170] Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., and Toga, A. W. (2003). Mapping cortical change across the human life span. *Nature neuroscience*, 6(3):309–315.

[171] Sowell, E. R., Thompson, P. M., Leonard, C. M., Welcome, S. E., Kan, E., and Toga, A. W. (2004). Longitudinal Mapping of Cortical Thickness and Brain Growth in Normal Children. *Journal of Neuroscience*, 24(38):8223–8231.

[172] Squitieri F, Berardelli A, Nargi E, Castellotti B, Mariotti C, and Cannella M (2000). Atypical movement disorders in the early stages of Huntington's disease: clinical and genetic analysis. *Clin Genet*, 58:50–56.

[173] Stoebner, Z. A., Hett, K., Lyu, I., Johnson, H., Paulsen, J. S., Long, J. D., and Oguz, I. (2022). Comprehensive shape analysis of the cortex in Huntington's disease. *Human Brain Mapping*.

[174] Stout, J. C., Paulsen, J. S., Queller, S., Solomon, A. C., Whitlock, K. B., Campbell, J. C., Carlozzi, N., Duff, K., Beglinger, L. J., Langbehn, D. R., Johnson, S. A., Biglan, K. M., and Aylward, E. H. (2011). Neurocognitive Signs in Prodromal Huntington Disease. *Neuropsychology*, 25(1):1.

[175] Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6):643–662.

[176] Sukumar, N. and Bolander, J. E. (2003). Numerical Computation of Discrete Differential Operators on Non-Uniform Grids. *CMES*, x(x):1–15.

[177] Sullivan, J. M., Charron, G., and Paulsen, K. D. (1997). A three-dimensional mesh generator for arbitrary multiple material domains. *Finite Elements in Analysis and Design*, 25(3-4):219–241.

[178] Tabrizi, S. J., Langbehn, D. R., Leavitt, B. R., Roos, R. A., Durr, A., Craufurd, D., Kennard, C., Hicks, S. L., Fox, N. C., Scahill, R. I., Borowsky, B., Tobin, A. J., Rosas, H. D., Johnson, H., Reilmann, R., Landwehrmeyer, B., and Stout, J. C. (2009). Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *The Lancet Neurology*, 8(9):791–801.

[179] Tabrizi, S. J., Reilmann, R., Roos, R. A., Durr, A., Leavitt, B., Owen, G., Jones, R., Johnson, H., Craufurd, D., Hicks, S. L., Kennard, C., Landwehrmeyer, B., Stout, J. C., Borowsky, B., Scahill, R. I., Frost, C., and Langbehn, D. R. (2012). Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: Analysis of 24 month observational data. *The Lancet Neurology*, 11(1):42–53.

[180] Tabrizi, S. J., Scahill, R. I., Durr, A., Roos, R. A., Leavitt, B. R., Jones, R., Landwehrmeyer, G. B., Fox, N. C., Johnson, H., Hicks, S. L., Kennard, C., Craufurd, D., Frost, C., Langbehn, D. R., Reilmann, R., and Stout, J. C. (2011). Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: The 12-month longitudinal analysis. *The Lancet Neurology*, 10(1):31–42.

[181] Tabrizi, S. J., Scahill, R. I., Owen, G., Durr, A., Leavitt, B. R., Roos, R. A., Borowsky, B., Landwehrmeyer, B., Frost, C., Johnson, H., Craufurd, D., Reilmann, R., Stout, J. C., and Langbehn, D. R. (2013). Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: Analysis of 36-month observational data. *The Lancet Neurology*, 12(7):637–649.

[182] Tabrizi, S. J., Schobel, S., Gantman, E. C., Mansbach, A., Borowsky, B., Konstantinova, P., Mestre, T. A., Panagoulias, J., Ross, C. A., Zauderer, M., Mullin, A. P., Romero, K., Sivakumaran, S., Turner, E. C., Long, J. D., and Sampaio, C. (2022). A biological classification of Huntington's disease: the Integrated Staging System. *The Lancet Neurology*, 21(7):632–644.

[183] The Petilla Interneuron Nomenclature Group (2008). Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature Reviews Neuroscience*, 9:557–568.

[184] Thompson, J. C., Harris, J., Sollom, A. C., Stopford, C. L., Howard, E., Snowden, J. S., and Craufurd, D. (2012). Longitudinal evaluation of neuropsychiatric symptoms in Huntington's disease. *Journal of Neuropsychiatry and Clinical Neurosciences*, 24(1):53–60.

[185] Thompson, P. M., Hayashi, K. M., de Zubicaray, G., Janke, A. L., Rose, S. E., Semple, J., Herman, D., Hong, M. S., Dittmer, S. S., Doddrell, D. M., and Toga, A. W. (2003). Dynamics of Gray Matter Loss in Alzheimer's Disease. *The Journal of Neuroscience*, 23(3):994–1005.

[186] Thompson, W. K. and Holland, D. (2011). Bias in tensor based morphometry Stat-ROI measures may result in unrealistic power estimates. *NeuroImage*, 57(1):1–4.

[187] Torrico, T. J. and Munakomi, S. (2022). Neuroanatomy, Thalamus. In *StatPearls*. StatPearls Publishing, Treasure Island, FL.

[188] Tustison, N. J., Avants, B. B., Cook, P. A., Song, G., Das, S., van Strien, N., Stone, J. R., and Gee, J. C. (2013). The ANTs cortical thickness processing pipeline. In Weaver, J. B. and Molthen, R. C., editors, *SPIE Medical Imaging 2013: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 8672, page 86720K. SPIE.

[189] Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., Kandel, B. M., van Strien, N., Stone, J. R., Gee, J. C., and Avants, B. B. (2014). Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *NeuroImage*, 99:166–179.

[190] Tustison, N. J., Holbrook, A. J., Avants, B. B., Roberts, J. M., Cook, P. A., Reagh, Z. M., Duda, J. T., Stone, J. R., Gillen, D. L., Yassa, M. A., Initiative, f. t. A. D. N., Philip, A., Reagh, Z. M., Duda, J. T., Stone, J. R., Gillen, D. L., and Michael, A. (2017). The ANTs Longitudinal Cortical Thickness Pipeline. *bioRxiv*, page 170209.

[191] Van Essen, D. C. and Drury, H. A. (1997). Structural and Functional Analyses of Human Cerebral Cortex Using a Surface-Based Atlas.

[192] Van Essen, D. C. and Maunsell, J. H. (1980). Twodimensional maps of the cerebral cortex. *Journal of Comparative Neurology*, 191(2):255–281.

[193] Van Essen, D. C., Zekit, S. M., Van, D. C., and Zeki, S. M. (1978). The topographic organization of rhesus monkey prestriate cortex. *The Journal of Physiology*, 277(1):193–226.

[194] Van Leemput, K., Bakkour, A., Benner, T., Wiggins, G., Wald, L. L., Augustinack, J., Dickerson, B. C., Golland, P., and Fischl, B. (2009). Automated Segmentation of Hippocampal Subfields From Ultra-High Resolution In Vivo MRI. *Hippocampus*, 19(6):549.

[195] Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. (1999). Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(March):897–908.

[196] von Economo, C. and Koskinas, G. N. (1925). *Die Cytoarchitektonik der Hirnrinde des erwachsenen Menschen*. Springer, Berlin, Germany.

[197] Vonsattel, J. P., Myers, R. H., Stevens, T. J., Ferrante, R. J., Bird, E. D., and Richardson, E. P. (1985). Neuropathological classification of Huntington's disease. *Journal of neuropathology and experimental neurology*, 44(6):559–577.

[198] Vonsattel, J. P. G. and DiFiglia, M. (1998). Huntington Disease. *Journal of Neuropathology & Experimental Neurology*, 57(5):369–384.

[199] Vonsattel, J. P. G., Keller, C., and Cortes Ramirez, E. P. (2011). Huntington's disease - neuropathology. *Handbook of Clinical Neurology*, 100:83–100.

[200] Walker, F. (2007). Huntington's disease. *The Lancet Neurology*, 369(9557):218–228.

[201] Wang, G., Zhang, X., Su, Q., Shi, J., Caselli, R. J., and Wang, Y. (2015). A Novel Cortical Thickness Estimation Method based on Volumetric Laplace-Beltrami Operator and Heat Kernel. *Medical image analysis*, 22(1):1.

[202] Wang, Q., Chen, Y., Zhang, N., and Gu, Y. (2021). Medical image inpainting with edge and structure priors. *Measurement: Journal of the International Measurement Confederation*, 185.

[203] Warner, J. H., Long, J. D., Mills, J. A., Langbehn, D. R., Ware, J., Mohan, A., and Sampaio, C. (2022). Standardizing the CAP Score in Huntingtons Disease by Predicting Age-at-Onset. *Journal of Huntington's Disease*, 11(2):153–171.

[204] Widmaier, E. P., Raff, H., Strang, K. T., and Sheope, T. C. (2019). *Vanders Human Physiology: the Mechanisms of Body Function*. McGraw-Hill Education, New York, NY, 15th edition.

[205] Wijeratne, P. A., Garbarino, S., Gregory, S., Johnson, E. B., Scahill, R. I., Paulsen, J. S., Tabrizi, S. J., Lorenzi, M., and Alexander, D. C. (2021a). Revealing the Timeline of Structural MRI Changes in Premanifest to Manifest Huntington Disease. *Neurol Genet*, 7:617.

[206] Wijeratne, P. A., Johnson, E. B., Gregory, S., Georgiou-Karistianis, N., Paulsen, J. S., Scahill, R. I., Tabrizi, S. J., and Alexander, D. C. (2021b). A Multi-Study Model-Based Evaluation of the Sequence of Imaging and Clinical Biomarker Changes in Huntingtons Disease. *Frontiers in Big Data*, 4.

[207] Willingham, D. B., Nissen, M. J., and Bullemer, P. (1989). On the development of procedural knowledge.

[208] Wolz, R., Heckemann, R. A., Aljabar, P., Hajnal, J. V., Hammers, A., Lötjönen, J., Rueckert, D., Alzheimer's Disease, T., and Initiative, N. (2010). Measurement of hippocampal atrophy using 4D graph-cut segmentation: Application to ADNI .

[209] Xia, T., Chartsias, A., Wang, C., and Tsaftaris, S. A. (2019). Learning to synthesise the ageing brain without longitudinal data. *Medical Image Analysis*, 73.

[210] Xu, C., Pham, D. L., Rettmann, M. E., Yu, D. N., and Prince, J. L. (1999). Reconstruction of the Human Cerebral Cortex from Magnetic Resonance Images. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 18(6):467.

[211] Xu, C. and Prince, J. L. (1998). Generalized gradient vector flow external forces for active contours. *Signal Processing*, 71(2):131–139.

[212] Yendiki, A., Stevens, A., Augustinack, J., Salat, D., Zollei, L., and Fischl, B. (2008). Anatomical Priors for Global Probabilistic Diffusion Tractography. *MICCAI Workshop on Computational Diffusion MRI*, page 5461.

[213] Yezzi, A. J. and Prince, J. L. (2003). An Eulerian PDE Approach for Computing Tissue Thickness. *IEEE Transactions on Medical Imaging*, 22(10):1332–1339.

[214] Yuille, A. L. (1991). Deformable templates for face recognition. *Journal of cognitive neuroscience*, 3(1):59–70.

[215] Yushkevich, P. A., Avants, B. B., Das, S. R., Pluta, J., Altinay, M., and Craige, C. (2010). Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: an illustration in ADNI 3 T MRI data. *NeuroImage*, 50(2):434–445.

[216] Yushkevich, P. A., Pluta, J., Wang, H., Wisse, L. E., Das, S. R., and Wolk, D. (2016). FAST AUTOMATIC SEGMENTATION OF HIPPOCAMPAL SUBFIELDS AND MEDIAL TEMPORAL LOBE SUBREGIONS IN 3 TESLA AND 7 TESLA T2-WEIGHTED MRI. *Alzheimer's & Dementia*, 12(7):P126–P127.

[217] Zhang, H., Bakshi, R., Bagnato, F., and Oguz, I. (2020). Robust Multiple Sclerosis Lesion Inpainting with Edge Prior. *Machine Learning in Medical Imaging*, 12436:120–129.

[218] Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57.

[219] Zhang, Y., Long, J. D., Mills, J. A., Warner, J. H., Lu, W., Paulsen, J. S., and the PREDICT-HD Investigators and Coordinators of the Huntington Study Group (2011). Indexing disease progression at study entry with individuals at-risk for Huntington disease. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 156(7):751–763.

[220] Zilles, K. and Amunts, K. (2012). Architecture of the Cerebral Cortex. In Mai, J. K. and Paxinos, G., editors, *The Human Nervous System*, chapter 23, pages 836–895. Academic Press, San Diego, CA, United States, 3rd edition.

[221] Zipursky, R. B., Lim, K. O., Sullivan, E. V., Brown, B. W., and Pfefferbaum, A. (1992). Widespread cerebral gray matter volume deficits in schizophrenia. *Archives of general psychiatry*, 49(3):195–205.