MRI ANALYSIS TO DETECT GRAY MATTER TISSUE LOSS
IN MULTIPLE SCLEROSIS

by

KUNIO NAKAMURA

Submitted in partial fulfillment of the requirements
For the degree of Doctor of Philosophy

Dissertation advisor: Elizabeth Fisher, PhD

Department of Biomedical Engineering
CASE WESTERN RESERVE UNIVERSITY

August, 2011
We hereby approve the thesis/dissertation of

Kunio Nakamura

candidate for the Ph.D. degree *.

(signed) Andrew M. Rollins

(chair of the committee)

Elizabeth Fisher

Bruce D. Trapp

David L. Wilson

June 10, 2011

*We also certify that written approval has been obtained for any proprietary material contained therein.
To my mother, my father, and my wife
for their love, patience and support.
# Table of Contents

Table of Contents .............................................................................................................. 1  
List of Tables ..................................................................................................................... 4  
List of Figures .................................................................................................................... 5  
Acknowledgements ........................................................................................................... 8  
List of Abbreviations ........................................................................................................ 9  
Abstract ............................................................................................................................ 11  

## Chapter 1 Introduction ......................................................................................... 13  
1.1 Overview of Multiple Sclerosis ................................................................. 13  
1.2 Clinical Features of MS ............................................................................. 14  
1.2.1 Symptoms and Diagnosis ............................................................... 14  
1.2.2 MS Disease Course .......................................................................... 15  
1.2.3 Clinical Measures of Disability ...................................................... 17  
1.2.4 MS Treatment .................................................................................. 18  
1.3 Pathology in MS ....................................................................................... 19  
1.3.1 Focal Pathology in White Matter ................................................... 19  
1.3.2 Diffuse Pathology in White Matter ............................................... 21  
1.3.3 Pathology in Gray Matter ............................................................... 22  
1.4 Imaging Markers in MS ........................................................................... 25  
1.4.1 Conventional MRI in MS ............................................................... 26  
1.4.2 Advanced MRI to Detect GM Abnormality .................................. 28  
1.5 Brain Atrophy in MS ................................................................................ 35  
1.6 Gray Matter Atrophy in MS ..................................................................... 37  
1.6.1 Volumetric Gray Matter Atrophy .................................................. 38  
1.6.2 Cortical Thickness Measurements in MS ...................................... 42  
1.7 Objectives ................................................................................................. 44  

## Chapter 2 Gray Matter Segmentation Algorithm ............................................. 46  
2.1 Introduction ................................................................................................. 46  
2.1.1 MRI Format and Viewing ............................................................... 46  
2.1.2 Background ..................................................................................... 47  
2.1.3 Overview of Segmentation Algorithm .......................................... 49  
2.2 Segmentation Algorithm .......................................................................... 51  
2.2.1 Image Correction ........................................................................... 51  
2.2.2 Fuzzy C-means Clustering Method .............................................. 54  
2.2.3 Anatomic Probability Map ............................................................ 55  
2.2.4 Morphologic Probability Map ....................................................... 57  
2.2.5 Partial Volume Correction ............................................................. 59  
2.3 Validation .................................................................................................... 60  
2.3.1 Accuracy Test with Simulated MRI .............................................. 60  
2.3.2 Accuracy Test with Manually Traced Real MRIs ....................... 61  
2.3.3 Reproducibility Test ..................................................................... 64  
2.3.4 Comparison to Other Segmentation Methods ......................... 65  
2.3.5 Sensitivity to Lesion Size ............................................................... 67  
2.4 Discussion ................................................................................................. 70
5.2.3 Clinical Correlation of Cortical Thickness............................................................. 144
5.2.4 Disability Progression vs. Thickness and Change.................................................. 145
5.2.5 Predictors of Cortical Thickness and Thinning...................................................... 146
5.2.6 Sample Size Estimate ............................................................................................. 147
5.3 Discussion ............................................................................................................... 147

Chapter 6 Summary and Future Work ................................................................ 153

6.1 Summary ............................................................................................................. 153
6.2 Limitations .......................................................................................................... 155
6.3 Future Work ........................................................................................................ 157
  6.3.1 Higher Field MRI .............................................................................................. 157
  6.3.2 Treatment Effect on Gray Matter Atrophy ......................................................... 157
  6.3.3 Investigation on Pseudoatrophy Effect .............................................................. 158
  6.3.4 Regional Correlation of GM Atrophy and WM Damage .................................. 159
  6.3.5 Regional Correlation of GM Atrophy and Local Damage ............................... 159
6.4 Conclusion .......................................................................................................... 160

Bibliography .................................................................................................................. 161
List of Tables

Table 1-1: MS subtypes ................................................................. 15
Table 1-2: Classification of cortical lesions .................................. 23
Table 1-3: Gray matter atrophy in the literature ......................... 39
Table 2-1: Result of accuracy test using simulated BrainWeb MRI  61
Table 2-2: MRI sequence ............................................................. 62
Table 3-1: MRI sequence for longitudinal study ......................... 77
Table 3-2: Baseline characteristics .............................................. 79
Table 3-3: Characteristics of subjects by EDSS progression .......... 83
Table 3-4: Characteristics of subjects by MSFC progression ........ 83
Table 3-5: Predictors of atrophy from multiple regression analysis 84
Table 4-1: Comparison between longitudinal and cross-sectional methods 128
Table 5-1: Baseline characteristics .............................................. 142
Table 5-2: Clinical and MRI correlations .................................... 145
Table 5-3: Baseline patient characteristics based on MSFC progression 146
Table 5-4: Sample size calculation for cortical thinning ............. 147
### List of Figures

| Figure 1-1: | Common progression pattern in MS | 16 |
| Figure 1-2: | Photographs of fixed normal and MS brain slices | 20 |
| Figure 1-3: | An example of proteolipid protein stained MS cerebrum | 22 |
| Figure 1-4: | MRI of MS patient and definitions of tissue regions | 26 |
| Figure 1-5: | Spectrum of brain atrophy from different MS patients | 35 |
| Figure 1-6: | Serial MRI of RRMS patient over 8.5 years | 36 |
| Figure 2-1: | Overall flowchart of the segmentation method | 50 |
| Figure 2-2: | Segmentation of brain, outer contour, and T2 lesions from FLAIR | 51 |
| Figure 2-3: | Correction of intensity variation | 52 |
| Figure 2-4: | Interleaved motion artifact correction | 53 |
| Figure 2-5: | Co-registration of FLAIR and T1-weighted images | 53 |
| Figure 2-6: | Intensity-based probability map | 55 |
| Figure 2-7: | Anatomic probability map from the Harvard Brain Atlas | 57 |
| Figure 2-8: | Morphologic probability maps for deep gray matter cortical gray matter | 59 |
| Figure 2-9: | Result of gray matter, lesion and brain segmentation | 60 |
| Figure 2-10: | Comparison of automated and manual segmentation | 63 |
| Figure 2-11: | Plots of scan-rescan variability in gray matter fraction | 65 |
| Figure 2-12: | Comparison of three different segmentation techniques | 67 |
| Figure 2-13: | Effect of lesion size on GM volume | 69 |
| Figure 3-1: | Plot of mean annualized rates of atrophy in white matter fraction, gray matter fraction, and brain parenchymal fraction | 81 |
| Figure 3-2: | Plot of gray matter fraction and EDSS | 82 |
| Figure 3-3: | Plot of gray matter fraction and MSFC | 82 |
| Figure 4-1: | Schematic 2D examples of different measures of thickness | 91 |
| Figure 4-2: | Flowchart of cortical longitudinal atrophy detection algorithm | 92 |
| Figure 4-3: | Averaged T1-weighted image in a CLADA-reference space | 94 |
| Figure 4-4: | Input, corrected, registered, and averaged images for CLADA | 96 |
Figure 4-5: Segmentation steps in CLADA

Figure 4-6: Surface deformation for subject-specific surface model

Figure 4-7: Comparison of computational times of self-intersection test with or without bounding box

Figure 4-8: Laplace map in shrink-wrap algorithm

Figure 4-9: Nonlinearly transformed and filtered regional labels

Figure 4-10: Effects of self-intersection test and surface relaxation steps

Figure 4-11: Laplace map in phantom experiment

Figure 4-12: Example of CLADA images

Figure 4-13: Comparison of surfaces with and without the relaxation step

Figure 4-14: Example of longitudinal deformation in atrophied brain

Figure 4-15: Boxplots of percent error in global and regional cortical thickness measurements from scan-rescan tests

Figure 4-16: Surface map of scan-rescan absolute cortical thickness difference

Figure 4-17: Intra-operator variability in CLADA

Figure 4-18: Figurative flowchart of the rapid autopsy program

Figure 4-19: Selection of region of interest from in situ MRI and from corresponding photograph of fixed brain slice

Figure 4-20: Plot of cortical thickness measurements from in situ MRI and photographs of post-mortem brain slices

Figure 4-21: Result of sensitivity test

Figure 4-22: Plot of annualized cortical thickness change measured from T1SE and MPRAGE images

Figure 4-23: Effect of different reference spaces on cortical thickness and cortical thinning

Figure 4-24: Comparison of cortical thickness measured from CLADA and FreeSurfer

Figure 4-25: Scatter plot of scan-rescan variability from full longitudinal and cross-sectional versions of CLADA

Figure 4-26: Effect of MS lesions on the annualized rates of cortical thinning

Figure 5-1: Baseline cortical thickness difference maps
Figure 5-2: Boxplots of cortical thickness at baseline .................................................. 143
Figure 5-3: Annualized change in global cortical thickness and longitudinal evolution of cortical thickness ........................................................................... 144
Acknowledgements

I would like to thank my advisor, Dr. Elizabeth Fisher, for her guidance, support and mentorship, not only during my graduate studies but also during my undergraduate times. This work would not be possible without her patience and kindness. Her expertise in the scientific design of experiments, image analysis, manuscript writing, and presentation has made me a better researcher.

I thank my committee members, Drs. Andrew Rollins (academic advisor), Bruce D. Trapp, and David L. Wilson who provided me with helpful suggestions to improve my research and dissertation.

I am also grateful for the members of the Fisher laboratory, past and present, for many intellectual discussions. I am thankful to the collaborators in the Mellen Center for Multiple Sclerosis Treatment and Research, Department of Neuroscience, and Department of Quantitative Health Sciences as well as the patients and volunteers for their time and cooperation for the studies.

The studies were supported by NIH National Institute of Neurological Disorders and Stroke (P01-NS38667) and National Multiple Sclerosis Society (RG 3099A, RG 3548A2).
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>25FTW</td>
<td>25-Foot Timed Walk</td>
</tr>
<tr>
<td>9HPT</td>
<td>Nine-Hole Peg Test</td>
</tr>
<tr>
<td>BPF</td>
<td>Brain Parenchymal Fraction</td>
</tr>
<tr>
<td>CIS</td>
<td>Clinically Isolated Syndrome</td>
</tr>
<tr>
<td>CLADA</td>
<td>Cortical Longitudinal Atrophy Detection Algorithm</td>
</tr>
<tr>
<td>cMRI</td>
<td>Conventional Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>CNS</td>
<td>Central Nervous System</td>
</tr>
<tr>
<td>CSF</td>
<td>Cerebrospinal Fluid</td>
</tr>
<tr>
<td>CTh</td>
<td>Cortical Thickness</td>
</tr>
<tr>
<td>DMT</td>
<td>Disease Modifying Therapy</td>
</tr>
<tr>
<td>EDSS</td>
<td>Kurtzke Expanded Disability Status Scale</td>
</tr>
<tr>
<td>FCM</td>
<td>Fuzzy C-Means Clustering Method</td>
</tr>
<tr>
<td>FLAIR</td>
<td>Fluid-Attenuated Inversion Recovery</td>
</tr>
<tr>
<td>FOV</td>
<td>Field-Of-View</td>
</tr>
<tr>
<td>GdLV</td>
<td>Gadolinium-Enhancing Lesion Volume</td>
</tr>
<tr>
<td>GM</td>
<td>Gray Matter</td>
</tr>
<tr>
<td>GMF</td>
<td>Gray Matter Fraction</td>
</tr>
<tr>
<td>ICS</td>
<td>Inner Cortical Surface</td>
</tr>
<tr>
<td>MPRAGE</td>
<td>Magnetization Prepared Rapid Gradient Echo</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>MS</td>
<td>Multiple Sclerosis</td>
</tr>
<tr>
<td>MSFC</td>
<td>Multiple Sclerosis Functional Composite</td>
</tr>
<tr>
<td>MTR</td>
<td>Magnetic Transfer Ratio</td>
</tr>
<tr>
<td>NC</td>
<td>Normal Control</td>
</tr>
<tr>
<td>NABT</td>
<td>Normal-Appearing Brain Tissue</td>
</tr>
<tr>
<td>NAWM</td>
<td>Normal-Appearing White Matter</td>
</tr>
<tr>
<td>OASIS</td>
<td>Open Access Series of Imaging Studies</td>
</tr>
<tr>
<td>OCS</td>
<td>Outer Cortical Surface</td>
</tr>
</tbody>
</table>
**OCV:** Outer Contour Volume

**PASAT:** Paced Auditory Serial Addition Test

**PD:** Proton Density

**PPMS:** Primary Progressive Multiple Sclerosis

**ROI:** Region of Interest

**RRMS:** Relapsing-Remitting Multiple Sclerosis

**SPM:** Statistical Parametric Mapping

**SPMS:** Secondary Progressive Multiple Sclerosis

**T1LV:** T1 lesion volume

**T1SE:** T1-weighted Spin Echo

**T2LV:** T2 lesion volume

**WM:** White Matter

**WMF:** White Matter Fraction
MRI Analysis to Detect Gray Matter Tissue Loss in Multiple Sclerosis

Abstract

by

KUNIO NAKAMURA

Multiple sclerosis (MS) has been traditionally characterized by primary demyelination and inflammation in white matter (WM). However, recent histopathologic studies have shown that gray matter (GM) of MS patients is also abnormal. My aim was to develop methods for quantifying GM damage in terms of GM atrophy and cortical atrophy, to investigate the evolution in various MS disease stages, and to assess relevance to clinical status. First, I developed an automated algorithm that segmented GM and WM in magnetic resonance images (MRI) and measured the normalized GM volume. The algorithm was designed to be applicable to MRI of MS patients, which had focal lesions and significant atrophy. The algorithm was validated and applied in a longitudinal study that included patients with clinically isolated syndrome (CIS), relapsing-remitting (RRMS) and secondary progressive (SPMS) MS as well as healthy normal controls. Other conventional MRI markers of focal damage and clinical measures were available to explore the correlations and predictors of GM atrophy. We found that (1) the rate of GM atrophy increased in a stage-dependent manner, which was similar to that of whole brain atrophy, (2) GM atrophy had moderately strong correlations with the disability measures, and (3) predictors of GM atrophy changed from RRMS to SPMS. Next, I developed a
registration and deformable model-based longitudinal method (CLADA, Cortical Longitudinal Atrophy Detection Algorithm) that had high reproducibility and could measure global and regional cortical atrophy in terms of cortical thickness and its change. CLADA was validated and applied to the full longitudinal MRI dataset to explore the evolution of cortical thinning, its clinical correlations, and predictors. The rate of cortical thinning increased with advancing disease, correlated with clinical disability, and distinguished stable and worsening patients with more significance than GM atrophy. In summary, my research showed that GM and cortical atrophy could be measured reliably with the new techniques. Furthermore, application of these methods in MS patients demonstrated that GM and cortical atrophy measurements were (1) relevant both clinically and biologically (2) able to provide insights on MS pathogenesis; and (3) suitable for future clinical trials of potential MS therapies.
CHAPTER 1 INTRODUCTION

In this chapter, an overview of multiple sclerosis (MS) including its clinical, pathologic, and imaging description is provided. Next, various advanced imaging techniques that are sensitive to abnormality in gray matter (GM) are described. The measurement of GM atrophy is presented as a clinically relevant mean to detect overall GM damage, and finally the aims of my project are outlined.

1.1 Overview of Multiple Sclerosis

Multiple sclerosis is a chronic, inflammatory and demyelinating disease of the central nervous system (CNS). Its pathologic hallmarks include inflammation, demyelination, gliosis, and axonal degeneration. Conventional magnetic resonance imaging (MRI) typically shows multiple focal abnormalities in the brain and spinal cord. MS affects approximately 2.5 million worldwide and 300,000 in the United States (Anderson et al., 1992) incurring an approximate annual cost of $10 billion (Whetten-Goldstein et al., 1998). It is the most common cause of non-traumatic neurologic disability among young adults (Weinshenker, 1996). The average age at disease onset is around 30, and women are 2-3 times more susceptible than men. MS symptoms, prognosis, and disease course vary significantly among patients. But, many patients have
a relatively normal life expectancy despite their disability, and only a small portion of patients die prematurely (National Multiple Sclerosis Society, 2011).

1.2 Clinical Features of MS

1.2.1 Symptoms and Diagnosis

Multiple sclerosis is a highly variable disease in terms of symptoms and progression, but it is generally characterized by unpredictable acute attacks of neurological impairment (relapses) followed by chronic progression of disability. There is no surrogate marker that reliably predicts the relapses or progression.

The relapses are highly variable within and among patients. These symptoms include sensory disturbances (numbness, pain, tingling sensation, and coordination problems), motor dysfunctions (spasticity, gait ataxia, and weakness in arm and leg), and visual impairments (Noseworthy et al., 2000). Other MS symptoms include seizure, bladder, bowel, and sexual dysfunctions (Compston and Coles, 2008). Cognitive dysfunction (slow processing speed and inability to concentrate) and memory impairment are also common (Chiaravalloti and DeLuca, 2008). Depression and fatigue are also widely reported (Ziemssen, 2009). Relapses generally persist for days. Patients tend to recover completely or partially at the early disease stage even without any treatment, and a relatively long period of disease inactivity follows (remittance). Relapses become rare during progressive worsening of disability.

Since its symptoms are similar to other conditions, MS can be difficult to diagnose, and the diagnostic criteria for MS have been evolving. The traditional criteria required at least 2 relapses for clinically definite MS (Poser et al., 1983). The current
diagnostic criteria (revised McDonald criteria) integrate MRI and require a dissemination of lesions in both time and space on MRI.\(^1\) Since relapses are rare (0.6-1.1 relapse per year (Scalfari et al., 2010)), the revised McDonald criteria facilitate early diagnosis of MS, allowing early MS treatment.

1.2.2 MS Disease Course

Three MS subtypes are defined according to the pattern of relapse and progression of disability (Lublin and Reingold, 1996) as shown in Table 1-1.

<table>
<thead>
<tr>
<th>MS Subtype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relapsing-remitting MS (RRMS)</td>
<td>Clearly defined relapses with full or partial recovery; periods between disease relapses characterized by a lack of disease progression (remittance)</td>
</tr>
<tr>
<td>Secondary progressive MS (SPMS)</td>
<td>Initial RR disease course followed by progression with or without occasional relapses, minor remissions, and plateaus</td>
</tr>
<tr>
<td>Primary progressive MS (PPMS)</td>
<td>Disease progression from onset with occasional plateaus and temporary minor improvements allowed.</td>
</tr>
</tbody>
</table>

**Relapsing-Remitting and Secondary Progressive MS**

Approximately 85% of MS patients experience relapse onset type of MS where patients initially exhibit an RRMS pattern followed by SPMS (Confavreux and Vukusic, 2006). At the first relapse suggestive of CNS inflammation, patients are diagnosed as clinically isolated syndrome (CIS). At the second relapse or imaging evidence of

\(^1\) The dissemination of lesions in time require either (1) a contrast-enhancing lesion at the different location, at least 3 months after the initial clinical event or (2) a new hyperintense lesion on T2-weighted MRI at least 30 days after the initial clinical event. The dissemination of lesions in space require 3 of the following 4 conditions: (1) at least 1 contrast-enhancing lesion or 9 hyperintense lesions on T2 scans if there is no contrast-enhancing lesion; (2) at least 1 infratentorial lesion; (3) at least 1 juxtacortical lesion; and (4) at least 3 periventricular lesions.
disseminated lesions by time and space, the CIS patients are diagnosed as RRMS (Polman et al., 2005). Most CIS patients who become MS initially present with optic neuritis or sensory symptoms (Weinshenker et al., 1989). RRMS patients generally advance to SPMS where disability worsens with no or occasional relapses (Confavreux and Vukusic, 2008). A typical evolution of disability is illustrated in Figure 1-1 (Trapp et al., 1999a).

**Figure 1-1:** A common progression pattern in MS (dark dashed line). Patients do not experience symptoms until the first relapse (onset). With additional clinical or imaging evidence of CNS inflammation suggestive of MS, patients are diagnosed with relapsing-remitting MS (RRMS) and experience relapses with complete or partial recoveries. During secondary progressive MS (SPMS), patients become more disabled even though the relapses become rare. The gray line shows a hypothesized sub-clinical disease activity in RR-SP MS course. Modified from Trapp et al. (Trapp et al., 1999a).

It is believed that the clinical relapses are only the tip of the iceberg and sub-clinical disease activity occurs even before the diagnosis and continues throughout the disease. Patients may be capable of functionally compensating during RRMS but lose the compensatory ability at a certain threshold, and convert to SPMS (Figure 1-1, gray curve)
(Trapp et al., 1999b). It is uncertain whether relapses predict long-term disability in MS (Confavreux et al., 2000; Ebers et al., 2008; Salfari et al., 2010). A potential surrogate marker of MS disease progression must be sensitive to such sub-clinical disease activity.

**Primary progressive MS**

The rest of MS patients (15%) are primary progressive MS (PPMS) who continuously become disabled without or very rare (in the order of 1 per decade) relapses from the disease onset (Confavreux and Vukusic, 2006; Kremenchtzky et al., 1999; Lublin and Reingold, 1996). Compared to RR-SP MS, the onset age is older (mean=40) and the proportion of men is higher (50%) in PPMS (Miller and Leary, 2007). The pathological and biological difference underlying these clinical phenotypes is not clear.

1.2.3 Clinical Measures of Disability

There are standardized methods to evaluate the level of disability. The most common measure is Kurtzke expanded disability status scale (EDSS) (Kurtzke, 1983), which scores disability on an ordinal rating scale from normal (EDSS=0) to death due to MS (EDSS=10). The pivotal scale is EDSS of 6 where patients require unilateral walking aids. At EDSS of 9.5, which is the last score before death, the patients are “totally helpless bed patient; unable to communicate effectively or eat/swallow.”

A more recent quantitative measure of disability is the multiple sclerosis functional composite (MSFC), which is a combined z-score of 3 tests: (1) 25-foot timed walk (25FTW) for lower extremity function, (2) nine-hole peg test (9HPT) for arm and hand function, and (3) paced auditory serial addition test (PASAT) for cognitive function.
of concentration and processing speed. The benefits with MSFC are that it is objective, quantitative, and less expensive than EDSS because a trained staff can administer the test rather than a neurologist (Rudick et al., 1997). MSFC includes a test of cognition.

Both measures have disadvantages. EDSS is nonlinear, subjective (prone to inter-rater variability), and heavily dependent on ambulation after EDSS of 6. EDSS changes may be reversible over short time periods (Ebers et al., 2008). The drawbacks for MSFC are that PASAT is highly demanding and intolerable by some patients. In longitudinal studies, PASAT tends to improve because of a learning effect (Polman and Rudick, 2010).

Both measures are used in correlational studies of imaging markers. Progression on EDSS is often used in MS clinical trials as the primary outcome measure.

1.2.4 MS Treatment

There is no cure for MS. An ideal cure for MS should halt the disease progression and regain the neurologic deficits. Preventive measures are also not available. Current MS treatments are either symptomatic treatments or disease-modifying therapy (DMT).

For severe acute relapses, high-dose corticosteroids are intravenously injected to reduce inflammation and shorten the relapse severity and duration. However, corticosteroids have not demonstrated superiority over other proven therapies in terms of reducing disease activity and slowing progression (Frohman et al., 2007; Shah et al., 2007).

There are currently 7 DMTs approved in the United States for MS: interferon β-1a subcutaneously or intramuscularly, interferon β-1b, glatiramer acetate, mitoxantrone, natalizumab, and fingolimod. In RRMS, interferon β and glatiramer acetate are most
commonly used. These agents are only moderately effective, as some patients do not respond to treatment (Rudick et al., 2004). Natalizumab and mitoxantrone have potentially serious side effects and are typically used for patients who do not respond to other therapies.

Typically, DMT’s efficacy is shown using the relapse rate or disability progression. But the relapse rate is low (typically 0 or 1 relapse per year), and the changes in EDSS or MSFC are small during the short duration of clinical trials (2 years or less), thus requiring a large number of patients. DMT’s efficacy in preventing long-term disability from these short clinical trials is not well established (Ebers et al., 2008). Therefore, clinical trials for new therapeutic candidates demand sensitive, cost-effective, non-invasive, and objective surrogate markers of disease progression.

1.3 Pathology in MS

MS brains typically contain characteristic multiple foci in WM and these focal areas have been extensively studied due to their appearance on MRI and on post-mortem specimen. However, there is significant damage in non-focal areas including normal-appearing white matter (NAWM) and GM.

1.3.1 Focal Pathology in White Matter

The most obvious pathology in MS is the multiple visible WM abnormalities called lesions or plaques (Figure 1-2, arrows). The lesions are typically located periventricularly or subcortically throughout the brain.
The lesions are areas of demyelination and are extremely heterogeneous as there are various degrees of inflammation, demyelination, remyelination, gliosis, and axonal damage. These lesions are classified according to their inflammatory activity: (1) active lesions are hypercellular demyelinated area with abundant perivascular and parenchymal macrophages; (2) chronic active lesions are demyelinated and have hypercellular border, a hypocellular center, and perivascular cuffs. The hypercellular border often contains macrophages and myelin debris; and (3) chronic inactive lesions are hypocellular and demyelinated throughout the area. These lesions often have gliosis and large extracellular spaces (De Groot et al., 2001; Trapp et al., 1998). Remyelinating lesions are characterized by thinly wrapped and irregularly formed myelin. These lesions are also called shallow plaques (Bruck et al., 2003).
Axonal damage is important pathology in MS and occurs early (Trapp et al., 1998). Possible mechanisms of axonal damage include bystander effects of inflammatory activity (Bo et al., 1994), disturbed energy balance and chronic necrosis (Trapp and Stys, 2009). Unlike myelin content, which can increase with remyelination (Prineas and Connell, 1979), axonal transection is irreversible and believed to be responsible for the progression of irreversible disability (Bjartmar et al., 2000; Trapp et al., 1999b).

1.3.2 Diffuse Pathology in White Matter

Normal-appearing white matter (i.e., outside of focal lesions, NAWM) is myelinated but not normal as activated microglia, gliosis, and axonal loss can be found (Allen et al., 2001; Trapp et al., 1998). There is significant diffuse inflammation in NAWM of SPMS and PPMS compared to acute or RRMS patients (Kutzelnigg et al., 2005). There is also axonal loss in NAWM of MS compared to controls (Evangelou et al., 2000; Ganter et al., 1999; Peterson et al., 2001), and an increased number of axonal terminal spheroids is found in progressive MS compared to acute or relapsing MS (Kutzelnigg et al., 2005). These studies have shown that NAWM in MS is abnormal and may reflect the progression of disease. Importantly, conventional MRI is not sensitive to pathology in NAWM. The precise mechanism of axonal damage in NAWM is also not known. A possible mechanism is that once axons are damaged by nonspecific inflammatory process, the distal portion of axons undergoes Wallerian degeneration, resulting in axonal degeneration in NAWM. This phenomenon is evidenced by the empty tubes of remaining myelin sheaths (Dutta and Trapp, 2007; Trapp et al., 1998). Axonal
loss in NAWM is also irreversible and considered to contribute to the progression of irreversible disability (Bjartmar et al., 2003).

1.3.3 Pathology in Gray Matter

In addition to pathology in WM, histopathological studies show that GM of MS patients is also demyelinated, and GM demyelination can be equally or more extensive than focal WM demyelination. Vercellino et al. detected cortical demyelination in 15% of the total GM area and WM demyelination in 22% of the total WM area in coronal brain sections (Vercellino et al., 2005). Similarly, Bo et al. found 26.5% demyelinated GM area versus 6.5% in WM (Bo et al., 2003); Gilmore et al found 33% GM and 20% WM demyelinated areas (Gilmore et al., 2009).

*Figure 1-3:* A proteolipid protein (PLP) stained example of MS cerebrum. The figure shows densely myelinated white matter (dark), less myelinated cortex (less dark), white matter demyelination (crosses), and gray matter demyelination (arrows). Courtesy of Dr. Ansi Chang.
Cortical lesions are classified according to their locations (Bo et al., 2003; Kidd et al., 1999). The most common classification is described in Table 1-2 (Bo et al., 2003).

Table 1-2: Widely recognized classification of cortical lesions defined by Bo et al. (2003). Types III and IV may be combined for subpial lesions.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Distribution of cortical lesion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I</td>
<td>GM/WM mixed lesion (leukocortical lesions)</td>
</tr>
<tr>
<td>Type II</td>
<td>Intracortical lesion; completely in the cortex</td>
</tr>
<tr>
<td>Type III</td>
<td>Subpial cortical lesion; extending from pial surface toward WM but not touching it</td>
</tr>
</tbody>
</table>

The percent area of cortical demyelination with respect to the total cortical area increases with the disease stage: from 3% in RRMS to 26% in SPMS in one study (Vercellino et al., 2005), and from 3% in RRMS to 13% in SPMS in another study (Kutzelnigg et al., 2005). Cortical lesions are more frequently found in cingulate gyrus (Bo et al., 2003; Vercellino et al., 2005). In addition to the cerebral cortex (Bo et al., 2003; Kidd et al., 1999; Kutzelnigg et al., 2005; Peterson et al., 2001), GM demyelination can be found in other CNS regions including the spinal cord (Gilmore et al., 2006; Gilmore et al., 2009; Gilmore et al., 2009), cerebellum (Gilmore et al., 2009; Kutzelnigg et al., 2007), hippocampus (Geurts et al., 2007; Papadopoulos et al., 2009; Vercellino et al., 2005), basal ganglia (Vercellino et al., 2005), thalamus (Cifelli et al., 2002), and hypothalamus (Huitinga et al., 2001).

Similar to the focal WM lesions, GM pathology in demyelinated areas includes neuronal, axonal and glial loss. There is also synaptic and dendritic loss in GM lesions (Cifelli et al., 2002; Frischer et al., 2009; Peterson et al., 2001; Vercellino et al., 2005;
Wegner et al., 2006). However, GM lesions are less inflammatory than WM lesions as there is no significant lymphocyte infiltration, complement deposition or BBB breakdown (Brink et al., 2005; Kidd et al., 1999; Peterson et al., 2001; van Horssen et al., 2007).

The pathology in myelinated cortex, i.e., regions outside of cortical lesions, is not well established. Peterson et al. have reported slightly increased transected neurites in the myelinated cortex compared to the control cortex (8±4 par mm$^3$ and 1±1 par mm$^3$, respectively), but this difference is subtle compared to the number of transected neurites in the demyelinated cortical area (4119±540 in active cortical lesions) (Peterson et al., 2001). Wegner et al. also have shown that the neuronal density in the myelinated MS cortex is not statistically different from that in the cortex of control brains (86±13 vs. 89±22 cells/mm$^2$, p>0.20) (Wegner et al., 2006). There is general thinning in the cerebral cortex of MS compared to controls (2.55±0.2mm and 2.84±0.36mm, p=0.027) as measured from 2D post-mortem brain slices, but no significant correlation was found between cortical thickness and the extent of cortical demyelination (Wegner et al., 2006).

Whether GM neurodegeneration occurs as a retrograde degeneration from damaged axons within lesions or independently of focal WM pathology remains highly debatable (Frischer et al., 2009; Geurts et al., 2009; Geurts and Barkhof, 2008; Pirko et al., 2007; Stadelmann et al., 2008). For example, the association between meningeal inflammation and cortical demyelination in the cortex is unclear (Kooi et al., 2009; Magliozzi et al., 2007). Regardless of the dependence on WM pathology, various types of damage occur throughout GM in MS. Further study of GM pathology in living MS patients is required understand its relationship to disability progression.
1.4 Imaging Markers in MS

Histopathologic studies are performed on biopsy and autopsy samples. Biopsy samples may be atypical MS because the purpose of biopsies is to rule out other rare conditions such as sarcoidosis at the diagnosis (National Multiple Sclerosis Society, 2011) and because biopsies are not required for monitoring. Autopsy cases are biased towards the chronic end-stage MS. In vivo imaging techniques such as MRI can non-invasively detect ongoing MS abnormalities. MRI has greater contrast for soft brain tissues and higher sensitivity for MS lesions than computed tomography (Nesbit et al., 1991). MRI provides visual and quantitative measures that can be used to investigate pathologic events, to objectively monitor the disease progression, and to evaluate therapeutic outcomes.

In this section, conventional MRI (cMRI), new MRI techniques, and quantitative image analysis methods are reviewed. An example of MRI in MS and definitions of tissues are shown in Figure 1-4 and described in the following sections.
1.4.1 Conventional MRI in MS

T2-weighted imaging

T2-weighted MR images of MS patients show hyperintense regions in WM (T2 lesions), typically in periventricular and subcortical WM, which correspond to focal MS lesions found at autopsy (Stewart et al., 1984). A problem of distinguishing cerebrospinal
fluid (CSF) and WM lesions on T2-weighted spin echo (T2SE) images (Figure 1-4a) can be resolved by using T2-weighted fluid-attenuated inversion recovery (FLAIR) images because of dark CSF and increased contrast of T2 WM lesions (Figure 1-4b). MRI has a superior sensitivity over clinical relapses as MRI can detect clinically silent disease activity. In fact, MRI-detected disease activity can outnumber clinical activity by 10-fold (Confavreux and Vukusic, 2006). MRI-guided histopathologic comparison reveals that only 55% of T2 lesions are demyelinated, indicating that T2 lesions are nonspecific marker of MS pathology (Fisher et al., 2007). Furthermore, T2 lesion volume, which increases from RRMS to SPMS, eventually levels off (“plateauing”) and does not explain the accumulation of neurologic disability (Li et al., 2006). Indeed, T2 hyperintense lesions are only weakly correlated with clinical disability in most studies (Barkhof, 1999; Filippi et al., 1995; Li et al., 2006; Mammi et al., 1996).

**Proton density-weighted imaging**

Proton density (PD)-weighted images (Figure 1-4c) are sensitive to the density of water protons. PD-weighted images show greater GM-WM contrast, but hyperintense MS lesions may appear similar to GM. PD-weighted MRI can be acquired without an additional scan time when the PD/T2-weighted dual echo sequence is used to acquire T2-weighted MRI. Using both PD/T2-weighted dual echo MRIs is helpful when segmenting T2 lesions.

**T1-weighted imaging**
T1-weighted spin echo images (Figure 1-4d) show that the subsets of T2 lesions are hypointense lesions (T1 lesions). Chronic T1 lesions are called black holes. They significantly correlate with the percent of residual axons stained with Bodian staining and moderately correlate with degree of matrix destruction stained with hematoxylin and eosin, indicating that T1 lesions represent regions of severe tissue destruction (van Walderveen et al., 1998). Contrast enhancement of T1-weighted images using gadopentetate dimeglumine (Gd-DTPA) shows hyperintense lesions (i.e., Gd-lesion or enhancing lesions; Figure 1-4e) at regions where there is blood-brain barrier breakdown. Thus these enhancing lesions represent active inflammation (Kermode et al., 1990). A weekly MRI study has shown that each enhancing lesion lasts about 2-4 weeks (Cotton et al., 2003) with or without clinical decline, emphasizing MRI’s greater sensitivity over clinical symptoms (Lai et al., 1996). These enhancing lesions are also seen at the first appearance of T2 lesions (Kermode et al., 1990). Like T2 lesion volume, enhancing lesions are commonly used in clinical trials due to the high sensitivities (Frank et al., 1994; Miller et al., 1988). However, the correlation between lesion measures and disability is only modest at best, and the number of enhancing lesions decreases in the progressive phase of the disease (Filippi and Rocca, 2007).

1.4.2 Advanced MRI to Detect GM Abnormality

cMRI is highly sensitive to the focal MS pathology in WM, but insensitive to GM pathology. Only 0 to 10% of cortical lesions were detected using cMRI (Geurts et al., 2005b; Kidd et al., 1999; Newcombe et al., 1991). Therefore, non-conventional, advanced MRI techniques have been implemented to improve the detection of GM
damage. GM damage is important because it can be more extensive than WM (Bo et al., 2003; Gilmore et al., 2009; Vercellino et al., 2005) and because conventional imaging markers (i.e., T2 lesion and enhancing lesion volumes) do not fully explain the progression of disability (Filippi and Rocca, 2007; Li et al., 2006). In the following sections, examples of advanced MRI methods are discussed, including magnetic resonance spectroscopy, magnetization transfer ratio, and diffusion tensor imaging as well as new MRI methods to improve the sensitivity of abnormal GM. After these examples, brain atrophy is separately described as a quantitative method to detect overall brain damage.

**Magnetic resonance spectroscopy**

Magnetic resonance spectroscopy (MRS) is an *in vivo* spectroscopic method to quantify regional metabolite compositions using MRI. Among various MRS-detectable metabolites (e.g., choline, creatine, inositol, alanin, and lactate), N-acetylaspartate (NAA) is particularly important in MS because NAA is localized in neurons, neuronal processes and axons; thus, NAA is a specific marker of neuronal damage and neuronal function (Bjartmar et al., 2001).

Acute lesions show a reduction in NAA level which may partially recover after acute phase, suggesting that NAA is decreased in lesions but reversible in the surviving axons (De Stefano et al., 1995b). Alterations in the lipid resonances are shown to precede the appearance of lesions in NAWM (Narayana et al., 1998). Another MRS study has shown that axonal pathology begins early as the NAA level is significantly decreased in NAWM of CIS patients compared to healthy normal controls (Wattjes et al., 2007).
Several studies have correlated the decreased NAA levels to disability levels (Davie et al., 1995; De Stefano et al., 2003). Furthermore, MRS abnormalities are detected in deep GM including thalamus (Cifelli et al., 2002) and hippocampus (Geurts et al., 2006). In the cortex where MRS measurements are difficult (Geurts et al., 2004), the findings are inconsistent (Chard et al., 2002b; Geurts et al., 2006; Kapeller et al., 2001; Sharma et al., 2001).

Several disadvantages of MRS are (1) MRS-based NAA measures are often normalized with respect to creatine concentration, which may vary (De Stefano et al., 1995b), (2) NAA decrease may be reversible in the surviving axons (De Stefano et al., 1995a), (3) large MRS voxels are heavily contaminated by the partial volume effects, which is critical in the cortex, (4) reproducibility is low and varies regionally (Geurts et al., 2004), (5) it is difficult to apply in multi-center clinical trials, and (6) sensitivity to longitudinal change and treatment response are currently uncertain (Barkhof et al., 2009). Nevertheless, MRS is highly specific and provides insights into pathology relevant in MS. Additional metabolites including GABA (gamma-aminobutyric acid), vitamin C and glutamine are also investigated for their potential relevance in MS (Bakshi et al., 2008; Srinivasan et al., 2005).

**Magnetic transfer ratio**

Magnetization transfer (MT) results from interactions of magnetized protons in water and macromolecules (e.g., lipid membrane on myelin). Application of an off-resonance MT pulse can selectively saturate magnetization on macromolecular protons.
A magnetization transfer ratio (MTR) map can be generated from MT images acquired with and without off-resonant MT pulse as follows:

\[
\text{MTR} = \frac{M_0 - M_t}{M_0}
\]

where \(M_0\) is the image without and \(M_t\) is the image with the MT saturation pulse (Wolff and Balaban, 1994). An example of MTR map is shown on Figure 1-4(k).

In post-mortem MRI-histopathologic correlational studies, MTR is strongly correlated to myelin content in WM \((r = -0.84, p < 0.001)\) (Schmierer et al., 2004) and even sensitive to remyelination (Barkhof et al., 2003; Chen et al., 2007; Schmierer et al., 2004). In a longitudinal study, the local inhomogeneity in MTR has predicted subsequent MTR change in the lesions and NAWM because active lesions tend to appear inhomogeneous (inactive center with active edge) while chronic inactive lesions have homogeneous appearance (Chen et al., 2007). MTR alterations occur months before the appearance of contrast enhancing lesions. Therefore, MTR change may be the first MRI-detectable event as MS lesions emerge (Filippi et al., 1998). Other cross-sectional (i.e., single time-point across subjects) studies have shown significant differences between MS patients and healthy normal controls (NC) in whole brain MTR (Loevner et al., 1995) as well as in NAWM and GM MTR, indicating that MTR can detect diffuse pathology (Cercignani et al., 2001; De Stefano et al., 2006; Ge et al., 2001b).

Image acquisition for MTR is not difficult, but MTR values strongly depend on the MRI sequence parameters and saturation pulses (Ropele and Fazekas, 2009). Current studies are investigating MTR’s reliability in multi-center clinical trials (van den
Elskamp et al., 2010) and its sensitivity in detecting subpial cortical lesions (Chen et al., 2010).

**Diffusion tensor imaging**

Diffusion tensor imaging (DTI) is sensitive to random water motion or water diffusion (Basser et al., 1994). The water diffusion is represented by ellipsoidal tensor consisting of three orthogonal eigenvectors ($\lambda_1, \lambda_2, \lambda_3$ where $\lambda_1 > \lambda_2 > \lambda_3$). Various DTI-based parameters can be calculated from the eigenvectors: for example, the trace of apparent diffusion coefficients (sum of 3 eigenvectors = $\lambda_1 + \lambda_2 + \lambda_3$), axial diffusivity ($\lambda_1$), mean diffusivity (mean of 3 eigenvectors = ($\lambda_1 + \lambda_2 + \lambda_3$) / 3), and fractional anisotropy, which is related to the standard deviation of the eigenvectors (anisotropy).

In the CNS, water diffusion is predominantly restricted by the semi-permeable cell membranes of axons and myelin (Pierpaoli et al., 1996), and DTI can indirectly show myelin and axonal abnormalities in NAWM (Ceccarelli et al., 2007). An example of apparent diffusion coefficient map is shown in Figure 1-4(l).

In MS, DTI studies show decreased fractional anisotropy and increased axial and mean diffusivity in lesions and NAWM compared to normal controls (Bammer et al., 2000; Ciccarelli et al., 2001; Filippi et al., 2001). Similar abnormalities have been detected in GM (Ceccarelli et al., 2007; Cercignani et al., 2001; Ciccarelli et al., 2001), but DTI’s pathologic relevance is unclear as there are no directional fibers in GM.

**Imaging of gray matter damage**
Gray matter of MS patients generally appears normal on cMRI, and cMRI’s sensitivity to cortical lesions with respect to histopathology is extremely low (0-10%) (Geurts et al., 2005b; Kidd et al., 1999; Newcombe et al., 1991). The sensitivity is higher for leukocortical lesions (type I, ~30%) due to their involvement in WM, but for intracortical (type II) and subpial lesions (type III), the sensitivity is extremely low (~5%) (Geurts et al., 2005b) despite their abundance as 65-85% of all cortical lesions are the subpial type (Albert et al., 2007; Bo et al., 2003; Peterson et al., 2001). Several factors may explain low sensitivity of cMRI: (1) partial volume effects significantly reduce the contrast due to the complex folding of the thin cortical sheet; (2) there is relatively little myelin in the cortex, and thus the effect of demyelination is small; (3) GM has a higher cellular density, and thus the extracellular space may not be as expandable as in WM and may not significantly affect the density of MRI-detectable protons (Kidd et al., 1999); and (4) cortical lesions are less inflammatory and cause less edema (smaller change in proton density), which also reduces the contrast.

Advanced imaging methods such as double inversion recovery (DIR) (Geurts et al., 2005a), and phase-sensitive inversion recovery (PSIR) (Nelson et al., 2007) have improved the detection of cortical lesions by 150-500% with respect to cMRI, but the majority of cortical lesions are still missed when compared to the histopathologic standard (Seewann et al., 2010). There is no established standard DIR sequence yet, and since DIR is noisy, there is significant inter-rater variability in lesion detection (Geurts et al., 2011). Nevertheless, DIR-based cortical lesions have been detected in all types of MS (Calabrese et al., 2009; Calabrese et al., 2010b; Calabrese et al., 2007a). Preliminary results have shown that the number and volume of DIR cortical lesions increase over 3
years in RRMS and SPMS patients (Calabrese et al., 2010a) and in PPMS (Calabrese et al., 2009), suggesting that cortical lesions may be associated with accumulation of disability.

There is no sensitivity improvement with high-field MRI (4.7T) (Geurts et al., 2008), but ultrahigh-field MRI at 8-9.4T may reveal more cortical lesions (Kangarlu et al., 2007; Schmierer et al., 2010). High-field or multi-sequence MRI has improved classification of cortical lesions (Bagnato et al., 2006; Nelson et al., 2007; Nelson et al., 2008; Tallantyre et al., 2010). However, whole brain coverage is difficult, and high-field MRI scanners are not used for routine clinical imaging studies.

In a post-mortem high-field (4.7T) MRI study of the spinal cord, high-resolution PD-weighted MRI detected 73% of GM lesions. The sensitivity was higher than GM lesions in the cerebral cortex because most of the spinal cord GM lesions were leukocortical and tended to be large along the spine (Gilmore et al., 2009). In vivo spinal cord studies are difficult because (1) spinal cord is small and its MRI is affected by partial volume effects with surrounding bone, fat, and CSF, and (2) spinal cord MRI is also influenced by motion artifact due to respiratory and cardiac movements.

A reliable method to detect GM damage and the translation of advanced technology into clinical practice remain a challenge in MS. Therefore, various quantitative image processing methods that do not require acquisition of sophisticated MRI or high field MRI scanner have been proposed.
1.5 Brain Atrophy in MS

Another imaging marker designed to capture overall disease severity is whole brain atrophy (Rudick et al., 1999). Whole brain atrophy represents the end result of destructive pathology, both focal and diffuse, and can be quantified from cMRI (Fisher and Rudick, 2003). The precise mechanisms of brain atrophy in MS are not known, but demyelination, axonal degeneration, and neurodegeneration as well as synaptic and dendritic loss are believed to lead to brain atrophy.

Measures of whole brain atrophy, such as brain parenchymal fraction (BPF), have been developed to estimate brain volume normalized to head size (Rudick et al., 1999). BPF is decreased in MS (Rudick et al., 1999), and the rate of whole brain atrophy in MS patients is higher than in healthy normal controls (Fisher and Rudick, 2003); the brain volume decreases by 0.2-0.5%/year in normal aging (Fox and Schott, 2004) and by 0.6-1.5%/year in MS (Barkhof et al., 2009; Fisher and Rudick, 2003). Examples BPF measurements in cross-sectional and longitudinal studies are shown in Figure 1-5 and Figure 1-6, respectively.

![Figure 1-5: Examples of spectrum of brain atrophy from different MS patients. Brain parenchymal fraction (BPF) from left to right is 0.87 (a, no atrophy), 0.82 (b, moderate atrophy), and 0.80 (c, severe atrophy).](image-url)
atrophy), and 0.70 (c, severe atrophy). Images are transformed to a standard space using rigid-body registration.

**Figure 1-6:** Registered serial axial FLAIR images from a single RRMS patient over 8.5 years. Brain parenchymal fraction (BPF) is shown with follow-up interval in months (in parenthesis) below each image.

Another commonly used measure of whole brain atrophy is the percent brain volume change (PBVC), calculated by SIENA (Structural Image Evaluation, using Normalisation, of Atrophy) (Smith et al., 2002) in FMRIB Software Library (FSL). SIENA is a longitudinal, registration-based method that calculates shifts in the brain edges. In previous studies, SIENA has been more robust than cross-sectional methods (Altmann et al., 2009) such as SIENAx (Smith et al., 2002) and central cerebral volume (Losseff et al., 1996).
There are several confounding factors in the brain atrophy measurements. Technical factors include inconsistency in the image acquisition (e.g., MRI sequence parameters, scanner hardware, software, image resolution, and subject repositioning). Biological factors such as the level of (de)hydration can change the brain volume by 0.2-0.6% (Duning et al., 2005; Kempton et al., 2009). An important factor relevant in MS is pseudoatrophy effect. This effect is an accelerated brain volume decrease presumably due to resolution of inflammatory edema. Indeed, a significant pseudoatrophy effect has been observed during clinical trials of anti-inflammatory agents (e.g., interferon β-1a and natalizumab) where whole brain atrophy rates are greater in the treated group compared to the placebo group in the first year. The atrophy rates are reversed and show the treatment effect (i.e., reduced atrophy) in the second year (Miller et al., 2007; Rudick et al., 2000). In another study, the changes in WM volume appear to be influenced by inflammatory activity in WM (Tiberio et al., 2005).

Compared to lesion measurements, whole brain atrophy is more strongly correlated to clinical measures such as EDSS and MSFC (Ge et al., 2000; Rudick et al., 2001). Brain atrophy measures such as BPF and PBVC are routinely used as secondary or exploratory outcome measures in large multi-center MS clinical trials (Cohen et al., 2008; Miller et al., 2007; Rudick, 2004).

1.6 Gray Matter Atrophy in MS

Unfortunately, whole brain atrophy measurements may not accurately represent “true” irreversible tissue loss because of the pseudoatrophy effect. Since GM lesions are less inflammatory (Peterson et al., 2001), the effect of pseudoatrophy is believed to be
small in GM (Zivadinov et al., 2007). Therefore, GM atrophy may better reflect true irreversible tissue loss. Furthermore, the neurodegenerative aspect of MS may be more specifically measured by GM atrophy than whole brain atrophy because most of neurons and neuronal cell bodies are in GM (Azevedo et al., 2009; Kandel et al., 2000), and little myelin is found in GM. Therefore, GM atrophy is a potential marker of irreversible tissue loss that is related to MS disease progression and may be useful in clinical trials of anti-inflammatory and neuroprotective therapies.

1.6.1 Volumetric Gray Matter Atrophy

Typical volumetric measurement of GM atrophy involves (1) segmentation of brain GM using high-resolution structural T1-weighted MRI, (2) measurement of normalized GM volume, or GM fraction (GMF), and (3) calculation of annualized change in GMF between two or more time-points.

Previous cross-sectional studies have shown that normalized GM volume is decreased in MS compared to healthy normal controls (Chard et al., 2002a; Inglese et al., 2004; Sastre-Garriga et al., 2004). In longitudinal studies, the rate of GM atrophy is estimated to be approximately 1-2% per year in MS (Horakova et al., 2008; Sastre-Garriga et al., 2005; Tedeschi et al., 2009; Tiberio et al., 2005; Valsasina et al., 2005), which is greater than that due to normal aging (less than 1%) (Evans et al., 2010; Ge et al., 2002; Thompson et al., 2003).

GM atrophy occurs early in MS as evidenced by decreased GMF found in early RRMS compared to NC (Chard et al., 2002a). A longitudinal study reported a higher rate
of atrophy in CIS patients who converted to MS (Dalton et al., 2004), suggesting that GM atrophy is sensitive to sub-clinical damage at the very earliest stage of MS.

The correlations between GM atrophy and disability can vary (Table 1-3). Some small cross-sectional studies and short longitudinal studies have not found correlations between GM atrophy and disability. It is possible that these studies are too small (Tiberio et al., 2005) and that the level of disability is too homogeneous (Chard et al., 2002a). Studies may require a longer duration because the short term change in disability is small and may not be very reliable (Ebers et al., 2008). In fact, a moderately strong correlation of correlation coefficient \( r = -0.58 \) is found between GMF and EDSS in a large cross-sectional study (\( n = 597 \)) (Tedeschi et al., 2005). In a longer longitudinal study, Dalton et al. have compared the various rates of atrophy (whole brain, GM, WM, and ventricle) from a group of CIS patients who have converted to MS within 3 years of first relapse and a group that has remained as CIS, and they have found that GM atrophy has the largest difference (3.3% vs. 1.1%, CIS converting to MS and remaining CIS), indicating that GM is clinically relevant (Dalton et al., 2004). A long-term study of CIS followed over 20 years has shown that GM atrophy reflects the disease subtypes and disability to a greater extent than T2LV or WM atrophy (Fisniku et al., 2008).

### Table 1-3:

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Method</th>
<th>Subject (#)</th>
<th>GM Atrophy Rates</th>
<th>GM-Clinical Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liu (1999)§</td>
<td>Cavalieri</td>
<td>RR (20), SP (20), NC (10)</td>
<td>NA</td>
<td>None</td>
</tr>
<tr>
<td>Ge (2001a)†</td>
<td>3D VIEWNIX</td>
<td>RR (30), NC (25)</td>
<td>NA</td>
<td>No correlation with EDSS</td>
</tr>
<tr>
<td>Chard (2002a)</td>
<td>SPM99</td>
<td>RR (26), NC (27)</td>
<td>NA</td>
<td>No correlation with EDSS</td>
</tr>
<tr>
<td>Quarantelli (2003)†</td>
<td>RMC</td>
<td>RR (50), NC (54)</td>
<td>NA</td>
<td>No correlation with EDSS</td>
</tr>
<tr>
<td>Author</td>
<td>Method</td>
<td>Condition</td>
<td>Change</td>
<td>Parameter</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------</td>
<td>-----------</td>
<td>--------</td>
<td>-----------</td>
</tr>
<tr>
<td>De Stefano (2003)</td>
<td>SIENAx</td>
<td>RR (65), PP (25), NC (18)</td>
<td>NA</td>
<td>GMV-EDSS (RR)§:</td>
</tr>
<tr>
<td>Chard (2004)</td>
<td>SPM99</td>
<td>RR (13), NC (9)</td>
<td>MS: −0.0086/y</td>
<td>None except ΔGMF ARR 18mo prior to study **</td>
</tr>
<tr>
<td>Dalton (2004)</td>
<td>SPM99 (T2WI)</td>
<td>CIS (27), CIS→MS (31)</td>
<td>−0.017/3y</td>
<td>GM atrophy in converted CIS than remaining CIS</td>
</tr>
<tr>
<td>Amato (2004)</td>
<td>SIENAx</td>
<td>RR (41), NC (16)</td>
<td>NA</td>
<td>GMV-Cognitive test score: −0.58**</td>
</tr>
<tr>
<td>Sastre-Garriga (2004)</td>
<td>SPM99</td>
<td>PP (43), NC (45)</td>
<td>NA</td>
<td>GMF-MSFC: 0.34*</td>
</tr>
<tr>
<td>Sanfilippo (2005)</td>
<td>SPM99</td>
<td>MS (41), NC (18)</td>
<td>NA</td>
<td>GMF-EDSS: −0.34*</td>
</tr>
<tr>
<td>Oreja-Guevara (2005)</td>
<td>SIENAx</td>
<td>RR (26)</td>
<td>NA</td>
<td>No ΔEDSS correlation</td>
</tr>
<tr>
<td>Sastre-Garriga (2005)</td>
<td>SPM99 / SIENA</td>
<td>PP (145)</td>
<td>−1.50±1.6%/y</td>
<td>NA</td>
</tr>
<tr>
<td>Tedeschi (2005)† / RMC</td>
<td>RMC</td>
<td>RR (427), SP (140), PP (30), NC (104)</td>
<td>NA</td>
<td>Predictor of GMF atrophy: EDSS + age at disease onset</td>
</tr>
<tr>
<td>Tiberio (2005)</td>
<td>SPM99</td>
<td>RR (21), NC (10)</td>
<td>NC: −1.0%/y</td>
<td>No correlations: GMF-EDSS nor MSFC</td>
</tr>
<tr>
<td>Benedict (2006)</td>
<td>SIENAx</td>
<td>RR (77), SP(42), NC(27)</td>
<td>NA</td>
<td>GMV-PASAT: 0.47**</td>
</tr>
<tr>
<td>Morgen (2006)</td>
<td>VBM + SPM2</td>
<td>RR (19), NC</td>
<td>NA</td>
<td>No correlation with EDSS GMV-PASAT: 0.54 GMV: ci &lt; cp</td>
</tr>
<tr>
<td>Prinster (2006)</td>
<td>VBM + RMC</td>
<td>RR (51), NC</td>
<td>NA</td>
<td>No correlation: regional GMV and EDSS</td>
</tr>
<tr>
<td>Portaccio (2006)</td>
<td>SIENAx</td>
<td>RR (41), NC (16)</td>
<td>NA</td>
<td>No correlation: GMV,WMV-ΔEDSS</td>
</tr>
<tr>
<td>Oreja-Guevara (2006)</td>
<td>SPM99</td>
<td>RR (22)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Horakova (2008)</td>
<td>SIENA</td>
<td>RR (36)</td>
<td>−1.4%/y</td>
<td>NA</td>
</tr>
<tr>
<td>Fisher (2008)</td>
<td>FCM</td>
<td>NC (17), CIS (7), CIS→RR (8), RR (28), RR→SP (7), SP (19)</td>
<td>NC: −0.02%/y</td>
<td>GMF-EDSS: −0.48***</td>
</tr>
<tr>
<td>Fisniku (2008)†</td>
<td>SIENAx</td>
<td>NC (25), CIS (29), RR (33), SP (11)</td>
<td>NA</td>
<td>GMF-MSFC: 0.52***</td>
</tr>
<tr>
<td>Tedeschi (2009)</td>
<td>RMC</td>
<td>RR/SP (267)</td>
<td>−1.97%/2y</td>
<td>GMV-PASAT: 0.32*</td>
</tr>
</tbody>
</table>
Regional GM atrophy can correlate better with specific functional deficits. Using a voxel-based morphometry (VBM) method (Ashburner and Friston, 2000), early regional GM volume loss has been found in deep GM even at the first clinical presentation (Henry et al., 2008), supporting the ability of regional GM atrophy to detect sub-clinical damage. VBM can be used to explore the spatial relationship between local GM tissue loss and focal damage (Ceccarelli et al., 2008; Morgen et al., 2006; Prinster et al., 2006; Sepulcre et al., 2006; Sepulcre et al., 2009) and potentially diffuse NAWM damage. However, VBM requires many statistical voxel-wise comparisons and is believed to be less sensitive to regions where inter-subject variation is naturally high (e.g., cortex). Therefore, methods that measure the volume and its change directly from subject images may be preferred over VBM methods when detecting cortical atrophy (Anderson et al., 2006; Hutton et al., 2009). Using such direct methods, several studies found correlations between cortical GM volume loss and disability (De Stefano et al., 2003) as well as cognitive impairment (Amato et al., 2004; Benedict et al., 2006; Portaccio et al., 2010).
2006), and the cortical GM volume loss predicted most of the variance in cognitive dysfunction (Benedict et al., 2006).

1.6.2 Cortical Thickness Measurements in MS

In addition to GM volumetric studies, recent development of sophisticated image processing tools and high-resolution images have allowed measurement of cortical thickness (CTh) (Aganj et al., 2009; Dale et al., 1999; Das et al., 2009; Goldenberg et al., 2002; Han et al., 2004; Hutton et al., 2008; Jones et al., 2000; Lohmann et al., 2003; MacDonald et al., 2000; Memoli et al., 2004; Miller et al., 2000; Rueda et al., 2010; Scott et al., 2009; Srivastava et al., 2003; Xu et al., 1999; Yezzi and Prince, 2003; Zeng et al., 1999). One study has shown that the brain size is more strongly associated with the cortical surface area than with CTh (Im et al., 2008). Since the association between brain size and CTh is small, CTh does not strictly require head size normalization.

Most current CTh measurement techniques such as FreeSurfer (Dale et al., 1999), registration-based CTh (DiReCT) (Das et al., 2009), cortical reconstruction using implicit surface evolution (CRUISE) (Han et al., 2004), and Anatomic Segmentation using Proximity (ASP) (MacDonald et al., 2000) require high-resolution images, which are not typically acquired for MS patients, and these techniques cannot be applied in low-resolution MRIs. Some are longitudinal methods (e.g., version 4.0 or above of FreeSurfer (Han et al., 2006) and DiReCT (Das et al., 2009), image gradient-based method (Chen et al., 2004), or 4D minimal line integral method (Li et al., 2010)). Han et al. have reported that a longitudinal CTh measurement method possesses smaller variability errors than the independent application of cross-sectional methods using FreeSurfer (Han et al., 2006).
suggesting that the longitudinal methods are preferred for measuring the CTh change. One of the registration-based longitudinal CTh methods can be applied in low-resolution MRIs, but the measurements tend to be restricted to regions where the cortical surface borders CSF (gyral apices and open sulci) and may not cover the entire cortex (Chen et al., 2004). Therefore, there is a need for a longitudinal method that can measure CTh from the entire brain and can be applied to low-resolution T1SE MRIs.

Using current methods, alterations in CTh are reported in normal aging (Resnick et al., 2003; Salat et al., 2004; Thambisetty et al., 2010) as well as in a variety of pathologic conditions such as Alzheimer’s disease (Lerch et al., 2005), frontotemporal dementia (Das et al., 2009), Huntington’s disease (Rosas et al., 2002), schizophrenia (Kuperberg et al., 2003), and MS (Sailer et al., 2003).

In MS, a cross-sectional CTh study has found thinner cortex compared to NC (mean CTh of 2.30±0.14mm and 2.48±0.11mm, respectively, p<0.001), particularly in frontal and temporal lobes (Sailer et al., 2003). Like total GM tissue loss, CTh is reduced early in the disease (Calabrese et al., 2007b), and early thinning is observed in frontal and temporal lobes and spread to other lobes in later advanced stages (Calabrese et al., 2010c; Calabrese et al., 2007b; Sailer et al., 2003). In a large cross-sectional study, CTh is correlated with EDSS in 425 RRMS patients (Charil et al., 2007). Regionally, CTh correlates with a fatigue level (Pellicano et al., 2010) and cognitive impairment (Calabrese et al., 2010c). Thinner cortical areas corresponding to the initial clinical symptom in CIS patients suggestive of MS are found (e.g., patients with optical neuritis onset have thinner occipital lobes, and patients with the pyramidal onset had the thinner motor cortex) (Calabrese et al., 2007b; Jenkins et al., 2011). There has been only one
longitudinal study of CTh in MS patients, and it has found that the rate of cortical thinning is greater in progressively disabling patients compared to stable patients (−3.13±2.88%/yr and 0.06±2.31%/yr, p=0.002, respectively) in 30 MS patients over a year (Chen et al., 2004), further emphasizing that CTh is clinically relevant. However, long-term evolution of cortical thinning has not been reported in MS.

In summary, GM and cortical atrophy are potential surrogate markers of MS disease progression. Since whole brain atrophy may be significantly affected by inflammatory edema, and since most of the neurons are in GM, GM atrophy may be a more specific marker for irreversible tissue destruction than whole brain atrophy is. The temporal evolution of GM and cortical atrophy, as well as their clinical relevance in MS, need to be investigated. A better understanding of the changes in GM would provide more complete information about pathologic processes in MS than one could obtain using only whole brain atrophy measures. Important questions such as the relationship between GM atrophy and focal WM lesions or between GM atrophy and diffuse WM tissue damage could then be addressed.

1.7 Objectives

My research had two aims: (1) to develop and validate an automated cross-sectional method to segment GM, to apply this segmentation method in a longitudinal study, to measure GM atrophy in terms of GMF, and to investigate the clinical relevance of GM atrophy; (2) to develop and validate a longitudinal method for measuring CTh and its change, to apply this method to the same longitudinal study, and to explore its clinical relevance in MS.
In Chapter 2, a new automated *cross-sectional* method to segment GM is described. This segmentation algorithm is designed to be applicable in MRIs of MS patients, which often exhibit significant atrophy and focal abnormal areas. In Chapter 3, the developed segmentation algorithm is applied to baseline and fourth year data of a longitudinal MRI study to characterize the evolution of overall GM tissue loss in a variety of MS patients (CIS, CIS patients who converted to RRMS, RRMS, RRMS patients who advanced to SPMS, and SPMS) with respect to the healthy normal controls. In Chapter 4, a more sophisticated image processing algorithm, Cortical Longitudinal Atrophy Detection Algorithm (CLADA), is described. CLADA is specifically designed to measure the subtle temporal change in CTh. Unlike the segmentation algorithm in Chapter 2, CLADA is a registration and deformable model-based “longitudinal” method and can achieve highly consistent results. Various validation tests are conducted to evaluate its reliability. In Chapter 5, the full longitudinal (semi-)annual MRI data is used to investigate the patterns of cortical atrophy in MS. Finally in Chapter 6, my work is summarized, and possible future directions are provided.
CHAPTER 2  GRAY MATTER SEGMENTATION

ALGORITHM

2.1  Introduction

This chapter describes the gray matter (GM) segmentation algorithm for the measurement of GM atrophy in MS patients. The general overview of the algorithm (Section 2.1.3) is described first, followed by the detailed description of each component (Section 2.2). Various validation tests and their results are reported (Section 2.3) next. The developed algorithm will be applied in a longitudinal study in the next chapter.

The majority of content in this chapter was published in NeuroImage (Nakamura and Fisher, 2009).

2.1.1  MRI Format and Viewing

All MRI images were obtained as either DICOM or vendor specific format and were converted to an in-house format (BIP, Biomedical Image Processing format, Cleveland Clinic). Digital image processing software was written using the BIP library in the C language and compiled on SGI Irix64 and Linux (RedHat and CentOS).

A customized MRI viewer was developed using GTK+ for orthogonal viewing of BIP images on Windows and CentOS Linux. Unlike previous BIP viewers, which were
essentially 2D slice viewers, the new viewer showed 2 orthogonal planes (axial plus sagittal or coronal planes) for better visualization of the cortex. The viewer was also capable of showing color images, which facilitated visual verifications of segmentation results.

2.1.2 Background

Measurement of GM atrophy using normalized GM volume requires segmentation of GM tissue and calculation of GMF over time. Segmentation of brain tissue from MRI is an extensively studied field and many automated software exist. Few algorithms rely solely on image intensity (Schnack et al., 2001) because these approaches are overly sensitive to image artifacts such as radiofrequency (RF) inhomogeneity, $B_0$ inhomogeneity, and aliasing, and cannot adequately account for overlapping intensity distributions across structures. Therefore, to improve segmentation accuracy, most tissue segmentation algorithms combine intensity information with other techniques, such as the use of a priori anatomic information (Chalana et al., 2001; Van Leemput et al., 1999) or edge information through deformable contours (Davatzikos and Prince, 1995; Xu et al., 1999; Zeng et al., 1999). Intensity information is analyzed differently in each approach, including Gaussian mixture models (Andersen et al., 2002; Ashburner and Friston, 2005; Marroquin et al., 2002; Zhang et al., 2001), discriminant analysis (Amato et al., 2003), $k$-nearest neighbor classification (Mohamed et al., 1999), and fuzzy $c$-means clustering (Ahmed et al., 2002; Pham, 1999; Suckling et al., 1999; Zhou and Bai, 2007; Zhu and Jiang, 2003). The use of multiple images has advantages over a single image because the different contrasts can be enhanced between tissues. For example, FLAIR images have
desirable contrast between MS lesions and the normal-appearing brain tissue and can be combined with other images to obtain GM and WM segmentation (Sajja et al., 2006).

There are a few widely available and commonly used brain tissue segmentation methods that use both intensity and a priori anatomic information. These algorithms, such as the segmentation tool in SPM (Ashburner and Friston, 2005) and FAST in FSL (Zhang et al., 2001), have been designed for general use, and therefore, are not necessarily optimized for specific pulse sequences or for application to images from patients with a specific disease. The use of such general programs to segment MR images of MS patients often results in misclassification of MS lesions as GM or CSF due to overlapping intensities, which then requires time-consuming manual editing (Sanfilipo et al., 2005) and introduces operator variability into the measurements. These methods are also prone to classification errors due to partial volume effects between MS lesions and normal tissue. Furthermore, for retrospective image analysis, where image data may not have been acquired using optimal sequences for use with one of the widely available segmentation tools, a customized segmentation method may be required to obtain the most accurate results. SPM also requires high-resolution images of ~1×1×1mm³, and thus not well suited for typical clinical MRIs in MS, which typically have lower resolution (slice thickness = 3-5mm). Therefore, there is a need for new segmentation method that can be applied in routine clinical MRIs of MS for accurate measurements of GM atrophy.

The normalization of GM volume in GMF accounts for the inter-subject variability of head size, and the normalization factor is calculated by the sum of GM, WM and CSF volumes when CSF volume can be reliably measured (Equation 2-1). An alternative method suited for T1-weighted or FLAIR images is to use the outer contour
volume (OCV), which is the volume inside the smoothed brain parenchymal surface (Equation 2-2).

\[
\text{GMF} = \frac{\text{GMV}}{\text{GMV} + \text{WMV} + \text{CSFV}}
\]  
\text{Equation 2-1}

\[
\text{GMF} = \frac{\text{GMV}}{\text{Outer Contour Volume}}
\]  
\text{Equation 2-2}

The annualized rate of GM atrophy is calculated in longitudinal studies by taking the difference of GMF from at least two time-points and dividing by the interval between the two scans.

2.1.3 Overview of Segmentation Algorithm

The overall flow chart, including preprocessing and GM segmentation steps, is shown in Figure 2-1.
Prior to GM segmentation, the brain is segmented from non-brain tissues on FLAIR or proton density- / T2-weighted (PD/T2) dual echo images using a fully automated knowledge-based segmentation method, as previously described (Fisher et al., 1997), and an example of segmentation results are shown in Figure 2-2. The outer contour of the brain (a smoothed surface that includes the ventricles and CSF surrounding the brain) is also segmented. The total brain volume (BV) and volume within the outer contour (outer contour volume, OCV) are calculated. BPF is calculated as the ratio of BV and OCV, BPF = BV / OCV. For MRIs of MS patients, focal T2 hyperintense lesions (T2
lesions) are segmented in the same software using a modified version of the iterated conditional modes (ICM) algorithm (Besag, 1986).

![Image of brain segmentation steps](image)

**Figure 2-2**: Example of input axial FLAIR image (a), segmented brain segmentation (b), outer contour of the brain (c), and focal T2 hyperintense lesions (d).

The GM segmentation step involves T1-weighted images and relies on three probability maps (Figure 2-1 steps 5-7): intensity-based probability, anatomic probability, and morphologic probability. The probability maps are combined to create the tissue masks for GM and WM, which are corrected for partial volume effects. Finally, gray matter fraction (GMF) is calculated as the ratio of GM volume and OCV (Equation 2-2).

### 2.2 Segmentation Algorithm

#### 2.2.1 Image Correction

The image correction steps after segmentation of brain, lesions and outer contour, include inter-slice intensity correction, non-parametric nonuniformity normalization (N3) (Sled et al., 1998), and anisotropic diffusion filtering (Perona and Malik, 1990). The inter-slice intensity correction algorithm estimates and reduces the signal loss due to the $B_0$ inhomogeneity in the superior and inferior slices. A single multiplication factor is estimated for each slice by calculating the mode of the pixel-by-pixel ratios between
contiguous slices within the brain. The N3 algorithm corrects the smooth intensity variation on MR images by deconvoluting the intensity histogram and mapping the estimated ‘true’ intensity back to the image. The anisotropic diffusion filtering reduces the noise while preserving edges. An example of intensity correction using inter-slice intensity correction and N3 algorithms is shown in Figure 2-3.

![Figure 2-3: The intensity variation in original image (a) was corrected with inter-slice intensity correction program and N3 algorithm (b).](image)

In datasets with motion between interleaved sets, the images are corrected by splitting, registering, transforming, and merging the interleaved sets. The registration algorithm has 6 degrees of freedom (6-dof or rigid-body), uses normalized mutual information (NMI) as the measure of image similarity (Pluim et al., 2003), and calculates the optimal parameters using the simplex downhill optimization (Press et al., 2002). The transformation is divided into 2 equivalent midway transformations so that the interpolation artifact is equally distributed among the interleaved sets. Figure 2-4 shows interleaved axial FLAIR image with patient motion (left). The mismatch in interleave sets is visible in the sulci, hyperintense lesions, and brain boundaries (left) and is successfully corrected (right).
Figure 2-4: An example of interleaved motion artifact (left) and corrected image (right). The 2D axial FLAIR image was acquired with two interleaved sets from a MS patient who moved between the interleave sessions. The motion correction algorithm split, rigidly aligned, transformed, and combined the interleaved sets.

For the patient motion between acquisitions of FLAIR and T1-weighted images, the images are co-registered using the same 6-dof NMI registration method. Figure 2-5 shows an example of co-registration of FLAIR and T1-weighted spin echo (T1SE). The brain boundary is calculated from the FLAIR image and overlaid on T1SE. The pre-registration misalignment in (a) is corrected after co-registration using NMI registration (b).

Figure 2-5: FLAIR brain mask boundary is overlaid on T1-weighted image before and after co-registration.
2.2.2 Fuzzy C-means Clustering Method

The first of three probability maps is derived from the grayscale intensities of T1-weighted images. To calculate the intensity-based probability, the grayscale values for all normal-appearing brain tissue (NABT) voxels (i.e., voxels included in the brain mask but not in the lesion mask) are analyzed with a modified fuzzy c-means (FCM) clustering method (Bezdek et al., 1993; Clark et al., 1994). FCM iteratively calculates the fuzzy memberships and the cluster centers (mean tissue intensity) using the following equations:

\[
u_{ik} = \left[ \sum_{j=1}^{c} \left( \frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^2 \right]^{-1} \quad \text{Equation 2-3}
\]

\[v_i = \frac{\sum_{k=1}^{n} (u_{ik})^2 x_k}{\sum_{k=1}^{n} (u_{ik})^2} \quad \text{Equation 2-4}\]

where \(u_{ik}\) is the fuzzy membership of tissue \(i\) at voxel \(k\), calculated from \(x_k\), which is the intensity at voxel \(k\), and \(v_i\), which is the mean intensity for tissue \(i\). There are three tissue classes for GM, WM, and CSF. In addition to the above standard FCM, this algorithm includes additional parameters to factor in non-Gaussian tissue characteristics and local information, as shown:

\[
u_{ik} = \left[ \sum_{j=1}^{c} \left( \frac{w_i x_k - v_j}{w_j (x_k - v_j)} \right)^2 \right]^{-1} \quad \text{Equation 2-5}
\]

\[
u_{ik} = (1 - \beta)u_{ik} + \frac{\beta}{N} \sum_{n=1}^{N} u_{ikn} \quad \text{Equation 2-6}
\]
The additional parameters are $w$ for weighting one tissue over the other to account for various tissue intensity characteristics, and $\beta$, to account for the local fuzzy memberships. The fuzzy membership with local information is $u'_{ik}$, where $u_{ikn}$ is the fuzzy membership for class $i$ of voxel $k$'s local neighbor $n$. Since cortical GM could be very thin, a 2D local region with 4 neighbors is used. The parameters $\beta$ and $w$ are empirically determined based on the best results obtained as compared to manual tracing in terms of the similarity index (Zijdenbos et al., 1994). The similarity index is calculated as below.

$$\text{Similarity Index} = \frac{2TP}{2TP + FP + FN} \quad \text{Equation 2-7}$$

where TP is the number of voxels correctly classified as GM, FN is the number of voxels incorrectly classifying as non-GM, and FP is the number of voxels incorrectly classifying as GM. The similarity index ranges from 0 (no overlap) to 1 (perfect match) with greater than 0.7 is considered good agreement.

![Figure 2-6: T1-weighted MRI (a) and its intensity-based probability map from fuzzy c-means clustering method. The slice is at the same location as the FLAIR image in Figure 2-2.](image)

2.2.3 Anatomic Probability Map
The second GM probability map, the anatomic probability map, is commonly used in brain tissue segmentation (Ashburner and Friston, 2005). The probability maps may come from an average of many brains or a single template. In this GM segmentation method, the Harvard Brain Atlas is used as the template brain (Kikinis et al., 1996). The GM and WM masks from the atlas are morphologically dilated and blurred for the GM and WM probability maps.

Registration between subject image and the Harvard Brain Atlas template image uses an affine 12-dof registration. Unlike typical standard space registration methods, which only uses grayscale MRI from the subject and template, this registration method uses both grayscale image and segmented brain mask. The 12 registration parameters are estimated in five sequential steps as follows: (1) two rotations ($R_Y$ and $R_Z$) based on subject’s left and right brain hemispheric symmetry on grayscale image; (2) global scaling factor ($S_G$) based on the ratio of characteristic brain radii (the peak positions in the brain radius histogram) from the subject brain and atlas brain; (3) x-rotation ($R_X$) and three translations ($T_X, T_Y, T_Z$) with fixed $R_Y, R_Z$, and $S_G$; (4) three orthogonal scaling ($S_X, S_Y, S_Z$) with fixed translation and rotation parameters; and (5) three skewing ($K_X, K_Y, K_Z$) with fixed translation, rotation and scaling parameters. The registration parameters in steps (3)-(5) are determined based on maximizing the overlap between the subject’s brain and the template brain. The anatomic probability map is transformed to the image space using trilinear interpolation. The sequential determination of each registration parameter and the use of both grayscale MRI and brain mask result in good alignment for an affine registration.
2.2.4 Morphologic Probability Map

The third GM probability map is the individualized morphological probability map and is created from morphologic models of cortical GM and deep GM. The anatomic probability map based on the affine template registration alone cannot account for significant atrophy, and the morphologic probability is designed to capture the structural shifts that occur due to marked ventricular enlargement common in MS brains. Furthermore, because of lower contrast in deep GM than cortical GM on typical clinical T1SE MRIs, a separate morphologic deep GM probability map is used to improve the detection of deep GM.

The morphologic deep GM model consists of two 3D ellipsoids Figure 2-8(a). Their shape and location are determined from the directions of the principal axes of the left and right lateral ventricles and the brain’s centroid position. Defining each ellipsoid requires 6 parameters, \( x_0, y_0, z_0, a, b, \) and \( c \) as follows:

\[
\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} + \frac{(z-z_0)^2}{c^2} = 1 \tag{2-8}
\]
where \((x_0, y_0, z_0)\) are the ellipsoid’s centroid location and \((a, b, c)\) are the radii along 3 axes. These 6 parameters are pre-determined during the algorithm calibration using manually placed ellipsoids from 7 MRIs. After the training, the ellipsoid centroid is calculated as \([x_0, y_0, z_0] = C_{brain} + T\) where \(C_{brain}\) is the 3D coordinate of the brain centroid and \(T\) is the offset of \([16.5, 23.5, 6.0]\)mm for the ellipsoid in the right hemisphere and \([-16.5, 23.5, 6.0]\)mm for the left ellipsoid. The ellipsoidal radii are also pre-determined using the length along the principal axes with scaling factors: \((a,b,c) = [0.1031, 0.1355, 0.1362]\) \(L\) where \(L\) is the length of principal axes. Finally, each ellipsoid is transformed along the principal axis of the lateral ventricle on the same side. The ellipsoids cover all the deep GM structures (i.e. caudate, putamen, and globus pallidus) except for the diencephalon, which can be captured by the cortical GM model because it borders the brain surface.

The cortical GM model represents the probability of cortical GM as a function of the distance from the approximated brain surface Figure 2-8(b): the function is 100%-probability for distance \(\leq 4\)mm and linearly decreases to 0% by 10mm. This brain surface is determined by the mid-sagittal plane and by the edges on initial brain segmentation excluding the lateral ventricle edges. The mid-sagittal plane is detected through minimization of left and right hemispheric intensity differences. The lateral ventricles is segmented by a series of morphologic operators (3mm-dilation, seed-fill, 1.5mm-opening, and 1mm-median filter) on the segmented brain mask.
2.2.5 Partial Volume Correction

The three probability maps are multiplied and normalized from 0 to 100% to calculate the posterior probability, and then GM and WM masks are created by setting thresholds at 50%. The GM volume is measured after partial volume correction (Santago and Gage, 1995), and gray matter fraction (GMF) is calculated as the ratio between GM volume and OCV as shown in (Equation 2-2). White matter fraction (WMF) is calculated as BPF – GMF. GM segmentation results are visually verified using a color image that includes partial volume effects (Figure 2-9).
2.3 Validation

Five tests were performed to evaluate the performance of the GM segmentation algorithm: (1) segmentation of simulated MRI datasets and comparison to the gold standard to determine accuracy; (2) segmentation of real MRI datasets and comparison to results from manual tracings as another evaluation of accuracy; (3) segmentation of scan-rescan images to determine the reproducibility; (4) comparison to other widely used segmentation methods; and (5) segmentation of the same image with simulated lesion growth to determine the effects of MS lesions on GM volume.

2.3.1 Accuracy Test with Simulated MRI

The simulated BrainWeb MRI data (Collins et al., 1998) was used to evaluate the accuracy of segmentation by comparing the segmented tissue masks to the gold standard segmentation. BrainWeb offers various noise levels (0-9%) and RF inhomogeneity levels.
(0-40%). Twelve combinations of noise and inhomogeneity levels were used. The volumetric errors including absolute and percent difference were measured. Since the volumetric difference was a global measure and may average out to small errors, the similarity index (Equation 2-7) was also calculated to evaluate accuracy.

The volumetric error was 1.2% for GM volume at the default 3% noise and 20% RF inhomogeneity level. For the same image, the similarity index for GM was 0.964 where similarity index of 0.7 or above is considered good agreement.

There was a large decrease in the similarity index when the noise level was increased from 7% to 9%. Nevertheless, the segmentation algorithm performed well in the presence of RF inhomogeneity at the various noise levels as evaluated by the similarity index (Table 2-1).

<table>
<thead>
<tr>
<th>Noise levels</th>
<th>RF inhomogeneity levels</th>
<th>0%</th>
<th>20%</th>
<th>40%</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>0.974</td>
<td>0.975</td>
<td>0.967</td>
<td>0.972</td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>0.971</td>
<td>0.974</td>
<td>0.966</td>
<td>0.970</td>
<td></td>
</tr>
<tr>
<td>3%</td>
<td>0.962</td>
<td>0.964</td>
<td>0.956</td>
<td>0.960</td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>0.945</td>
<td>0.949</td>
<td>0.941</td>
<td>0.945</td>
<td></td>
</tr>
<tr>
<td>7%</td>
<td>0.911</td>
<td>0.920</td>
<td>0.913</td>
<td>0.915</td>
<td></td>
</tr>
<tr>
<td>9%</td>
<td>0.866</td>
<td>0.873</td>
<td>0.858</td>
<td>0.866</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>0.938</td>
<td>0.943</td>
<td>0.933</td>
<td>0.938</td>
<td></td>
</tr>
</tbody>
</table>

2.3.2 Accuracy Test with Manually Traced Real MRIs

MRIs from 3 MS patients and 3 normal healthy controls were used to evaluate the segmentation accuracy by using manual tracings in real MRIs as the gold standard. The
images were acquired on a 1.5T Siemens Vision scanner and included a FLAIR and T1SE with the sequence parameters listed in Table 2-2:

**Table 2-2:** MRI sequence.

<table>
<thead>
<tr>
<th>Image type</th>
<th>Sequence detail</th>
<th>Voxel size</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2-weighted fluid-attenuated inversion recovery (FLAIR)</td>
<td>30 contiguous slices&lt;br&gt;FOV = 172×230 mm&lt;br&gt;Matrix size = 192×256&lt;br&gt;TR = 6000 ms&lt;br&gt;TI = 2000 ms&lt;br&gt;TE = 105 ms&lt;br&gt;NSA = 2</td>
<td>0.9×0.9×5mm</td>
</tr>
<tr>
<td>T1-weighted spin echo</td>
<td>30 contiguous slices&lt;br&gt;FOV = 172×230 mm&lt;br&gt;Matrix size = 192×256&lt;br&gt;TR = 800 ms&lt;br&gt;TE = 15 ms&lt;br&gt;NSA = 1</td>
<td>0.9×0.9×5mm</td>
</tr>
</tbody>
</table>

Abbreviations: FOV = field of view, TR = repetition time, TI = inversion time, TE = echo time, NSA = number of signal average

The images were processed through the automated GM segmentation algorithm and, separately, GM was manually traced in each image.

The leave-one-out cross-validation approach was used to avoid testing and training on the same dataset. Because the optimization of the segmentation parameters (\(w\) and \(\beta\)) used the same manually traced images, these optimized parameters were not used for this part of validation. Instead, for each of the six cases, the parameters were optimized separately using only other five cases and excluding the test data.

An example of accuracy test with manually traced real MRI is shown in Figure 2-10.
Figure 2-10: An example of input image (a), manual segmentation (b), and automated segmentation (c) from MS patient. Visually, the automated segmentation method showed good correspondence compared to the manual segmentation. (d) There was an excellent correlation between the automated and manual volume measurements from 3 MS patients and 3 controls. The mean absolute error for GM volume was 3.1±2.6%, and the Pearson correlation coefficient was 0.993. (e) The Bland-Altman plot showed that there is a bias between manual and automated segmentations.

The mean absolute error (standard deviation) for GM volume was 3.1% (2.6), and the automated and manually derived tissue volumes for each image set were highly correlated (Pearson correlation coefficient, $r = 0.993$). The similarity index of automated segmentation compared to manually tracing was also good; the mean similarity index was 0.841 and 0.836 for controls and MS patients, respectively. Bland-Altman plot showed a bias (Figure 2-10e) where brains with larger GM volume had larger absolute difference in GM volume. Since there is no intrinsic bias in the automated segmentation method, the
observed bias may be related to the manual tracing. A post hoc analysis showed that BPF and GM volume were correlated in these 6 subjects ($r = 0.87$). It was possible that with the 5mm slice thickness, manual tracings tended to be too conservative along the GM-CSF border in deep sulci, and this conservative manual tracing might have led to an underestimation of true GM volume. While the accuracy of manual tracing is a common pitfall in validation, the correlation coefficient of 0.993 and similarity indices of 0.841 and 0.836 assured that our segmentation is accurate.

2.3.3 Reproducibility Test

For the reproducibility test, 9 MS patients were imaged at 3 different time-points within 2 weeks. Each image set was processed separately through the automated GM segmentation algorithm. The reproducibility of the algorithm was evaluated by the coefficient of variation (COV = standard deviation / mean) for GM volumes and GMF.

The COV for GM volumes and GMF was 1.0% and 1.1%, respectively (Figure 2-11). This reproducibility of approximately 1% appeared to be sufficient to detect GM atrophy in MS, which was approximately 1-2%/year.
Figure 2-11: (a) Gray matter fractions and (a) difference from the individual mean gray matter fraction from scan-rescan MRI of 9 MS patients. Each subject was scanned 3 times over a 2 week-period.

2.3.4 Comparison to Other Segmentation Methods

The results of the segmentation algorithm were quantitatively and qualitatively compared to the results of other common freely available segmentation algorithms: FMRIB's Automated Segmentation Tool (FAST in FSL) (Zhang et al., 2001) and
Statistical Parametric Mapping (SPM) (Ashburner and Friston, 2005). For the BrainWeb images, which had a high-resolution 1mm isotropic voxel size, the similarity indices were compared to the published values (Ashburner and Friston, 2005; Ferreira da Silva, 2007). For low-resolution T1SE images, there were no published values. Therefore, 3 methods (the proposed method, FAST and SPM) were applied on T1SE from the 6 MRIs that were manually traced. The similarity index was calculated using the manual segmentation as the gold standard. As a qualitative assurance, a blinded neuroradiologist reviewed the segmentation results and scored them as 1 = excellent, 2 = good, 3 = acceptable, and 4 = unacceptable.

The similarity index measured from applying our GM segmentation method to BrainWeb image (0.94) was higher than the published values: 0.93 and 0.90 for SPM and FAST, respectively (Ashburner and Friston, 2005; Ferreira da Silva, 2007). Therefore, our algorithm performed slightly better than the commonly available segmentation methods using the BrainWeb image.

On T1SE images from 6 subjects, the average similarity index (s.d.) was 0.8381 (0.01), 0.7887 (0.04), and 0.7084 (0.08) for our method, SPM, and FSL, respectively. For T1SE MRIs, our method out-performed other commonly used methods when compared to the manual tracings. An example of segmentation with input images is shown in Figure 2-12.
An experienced neuroradiologist was asked to perform a blinded review of the GM segmentation results for each of the 3 methods for 4 brains. The results showed that the new method resulted in acceptable GM segmentation in all cases while some results from other methods were unacceptable due to the misclassifications of lesions as GM and non-GM classified as GM. The mean subjective scores demonstrated consistency with the quantitative evaluation (new method: 2.0, FSL: 2.5, and SPM: 2.5; 1 = excellent, 2 = good, 3 = acceptable, and 4 = unacceptable), showing at least in these images, the new method is superior to other commonly used segmentation methods.

2.3.5 Sensitivity to Lesion Size

Finally, the segmentation algorithm was tested for the effects of MS lesions in WM on FLAIR images (T2LV). Masks of segmented MS lesions were morphologically dilated with 3D spherical kernels of various sizes (1.2, 1.5, 1.9, 2.1, 2.6, 2.8, 2.9, 3.5, and 3.9 mm) in order to simulate different sized MS lesions within the same MRIs. The dilated lesion masks were masked by the brain tissue mask to ensure that ‘lesion’ voxels did not extend into non-parenchymal space and masked out from the GM mask. The
automated segmentation was then performed on the same MRI repeatedly using different lesion masks and GM volume was measured. This testing was repeated in 18 different MS patients to ensure that lesions of various starting sizes and locations were included. The linear regression was calculated between GM volume and T2LV to measure the effect of enlarging MS lesions.

The linear regression analysis showed there was a clear and fairly consistent effect of total T2LV on the measured GM volume for each case; as the lesion size was systematically increased, the GM volume decreased (Figure 2-13). The lesion volume and the GM volume were inversely proportional with a mean (s.d.) slope of $-0.26 (0.07)$. Since T2LV are highly variable even in the same patient, this systematic effect suggested that the GM volumes need to be corrected.
Figure 2-13: Effect of lesion size on GM volume. From T2-weighted image (a), lesions were segmented and overlaid on T1-weighted image (b), which were subsequently dilated by (1.9mm in c and 2.9mm in d) spherical kernel. Using T1-weighted image (e), gray matter was segmented within the normal-appearing brain tissue as shown in (f) original segmentation, (g and h) segmentation after simulated lesion growth by 1.9- and 2.9mm-dilation, respectively.

(i) The plot showed the effect of lesion size on automated gray matter volume measurements in 18 different MS patients. The average slope was $-0.26$. 

$69$
2.4 Discussion

2.4.1 Segmentation Method

This chapter described a new, fully automated method for GM segmentation in brain MRIs of MS patients. The method combined probability maps derived from intensity information, anatomic information, and morphology and segmented GM from low-resolution clinical brain MRIs.

Effect of MS lesions

The use of both anatomic and morphologic probability maps was a unique aspect of this algorithm. These maps provided patient-specific information about the locations of cortical GM and deep GM structures that was more precise than the use of either the anatomic or morphologic map alone. This was a key step for correct tissue classification in MS brains since T1 hypointense lesions and their surrounding partial volume voxels would be misclassified as GM. The misclassification of lesion voxels was a common problem with the application of general-use brain segmentation software, such as SPM (Ashburner and Friston, 2005) and FAST (Zhang et al., 2001), to segment GM in MS patients. Correcting the misclassification required time-consuming manual editing to obtain an acceptable segmentation (Sanfilipo et al., 2005).

The misclassification of lesions as GM was a potential problem for both cross-sectional and longitudinal studies. In cross-sectional studies, comparison of SPM-derived GM volumes between different subjects without correcting misclassified GM voxels might lead to a result that is actually a composite of both GM and lesion volumes, which, in turn, might lead to misinterpretation of the results. As shown in Figure 2-13(i), initial
GM volumes were not strongly correlated to lesion volume, but uncorrected GM volumes might artificially appear to be correlated to lesion volume if the misclassification of T1 hypointense lesions was left uncorrected.

In longitudinal studies, this dependence of GM volume on lesion volume presented an even bigger problem because MS lesions are highly dynamic, and the misclassified lesion volume change might be even greater than the true GM volume change. By performing the lesion segmentation step separately, masking out the lesions, and combining intensity, anatomic, and morphologic probability maps to segment GM, the misclassification of MS lesions could be avoided.

**Effect of atrophy**

The use of two different probability maps, anatomic and morphologic, was also important for segmentation of MS brains because many patients had a significant degree of atrophy, leading to the enlarged lateral ventricles. For this reason, a simple affine transformation of an anatomic reference might not provide a useful estimate of GM location, particularly for deep GM structures. The morphologic probability map improved the estimate of GM location because it was derived directly from the patients’ image data and, therefore, the degree of brain atrophy for that individual was taken into account.

2.4.2 Validation Tests

**Accuracy tests**

In a comparison to published results using BrainWeb images, the new segmentation method performed comparably well (Ashburner and Friston, 2005;
Another accuracy test was the comparison to manual segmentation of real MRIs. This was done because the simulated BrainWeb images were very different from typical MRIs acquired for MS patients. The typical clinical images acquired for MS studies are low-resolution MRIs (slice thickness = 3-5mm) as in the images used here. While the developed algorithm clearly agreed with manually segmented images, the numerical accuracy results were only moderately good. This was not surprising given that the “gold standard” for comparison was manual segmentation, which was known to be error-prone and highly subjective.

In fact, there was an observed bias in the comparison between manual and automated segmentation (Figure 2-10e) where the volume difference was greater for larger brains. This bias appeared to be caused by the differences in the ability to distinguish the GM-CSF border in deep sulci. Brains with low GM volume typically had a significant degree of whole brain atrophy and vice versa. These brains with significant atrophy had very large sulci with clear GM-CSF edges, whereas the brains without significant atrophy had ambiguous GM-CSF edges due to partial volume effects in the tight sulcal spaces. The brains that were manually traced for this validation study had a very wide range in brain parenchymal fraction (BPF = 0.74 – 0.87), and therefore a wide range in GM-CSF separation in the deep sulci, which might have made this bias evident.

**Scan-rescan reproducibility test**

The scan-rescan test was performed in order to evaluate the applicability of the new segmentation method for longitudinal studies of GM atrophy, wherein the extent of GM tissue loss could be estimated by the difference in GM fractions obtained at 2
different time-points. The COV of approximately 1% from repeated scans obtained within 2 weeks demonstrated that the algorithm was highly reproducible and similar to other methods (0.4% (Benedict et al., 2006), 0.4% (Sastre-Garriga et al., 2005), 0.7% (Chard et al., 2002a), 0.8% (Sanfilipo et al., 2005), 1.0% (Furby et al., 2010), 1.9% (Ge et al., 2001a)). This variability was greater than that of BPF, which was about 0.2% (Fisher and Rudick, 2003). The typical rate of GM atrophy in MS patients was estimated to be about 1-2% per year (Horakova et al., 2008; Sastre-Garriga et al., 2005; Tedeschi et al., 2009; Tiberio et al., 2005; Valsasina et al., 2005) as shown on Table 1-3. Therefore, this method would be most appropriate for application to longitudinal studies of duration 2 years, or longer. For short-term studies, a more precise method for measurement of GM tissue loss would be needed (e.g., longitudinal registration-based methods (Chapter 4)).

**Comparison to other methods**

The developed method was directly compared to other typical segmentation methods (SPM and FAST). The results showed that while the new method performed equally well when applied to high-resolution images (i.e., BrainWeb), it out-performed the other methods when applied to low-resolution images. This was critical because typical MRIs acquired for MS are lower-resolution. Therefore, the new method would be applicable for retrospective analysis of images from MS patients such as existing clinical trial datasets, and would be capable of determining the rates of GM atrophy without acquiring new images.

**Effect of T2 lesions on GM volume**
The last validation test investigated if the GM volume was affected by changes in T2LV. Unlike the issue of misclassified lesion voxels discussed above (Chard et al., 2002a; Sanfilipo et al., 2005), this effect stemmed from the fact that unsupervised clustering algorithms were sensitive to the voxel intensities that were fed to it as input. Thus, even a relatively minor change in the input voxels could have an effect on the final clustering result. In the developed algorithm, lesion voxels were masked out before the FCM algorithm. Therefore, as the lesions grew or shrank, the actual voxels that were used as input in the intensity-based classification step would change. This was shown by the simulations to have a clear effect on the resulting GM volumes. If left uncorrected in a longitudinal study of MS patients, some portion of the change in GM volume would appear to be correlated to T2LV and its change, simply from this technical issue. A systematic test was performed to calculate the extent and significance of this association, so that the effects of T2LV could be estimated using the mean slope of the GM volume versus T2LV regression lines.

2.4.3 Conclusion

In summary, the results indicated that the new segmentation algorithm could be used for reliable measurement of GM atrophy in MS patients if the follow-up duration was 2 years or longer. It also suggested that the individual GM volume should be corrected for T2LV before statistical group comparisons. The issues inherent to the analysis of MRIs from MS patients had been addressed directly in the design of the algorithm.
CHAPTER 3  MEASUREMENT OF GRAY MATTER ATROPHY IN MS

3.1 Introduction

In this chapter, the newly developed GM segmentation algorithm was applied to a longitudinal MRI study to characterize GM atrophy in MS patients with the most common phenotypes who had entered into a prospective longitudinal study. Specifically, the rates of GM atrophy were measured from patients with clinically isolated syndrome (CIS), relapsing-remitting MS (RRMS), and secondary progressive MS (SPMS), and compared to the GM atrophy rate of age- and gender-matched healthy normal control (NC) subjects. This chapter addressed the pattern of GM tissue loss in these patients over the course of 4 years. The correlations between GM tissue loss and other MRI measures of tissue damage, and between GM tissue loss and clinical worsening were also investigated.

The majority of content in this chapter was published in Journal of the Neurological Sciences (Rudick et al., 2009) and Annals of Neurology (Fisher et al., 2008).

3.2 Materials and Methods

3.2.1 Subjects
MS patients and healthy NC were recruited from the Mellen Center for Multiple Sclerosis Treatment and Research at Cleveland Clinic. MS patients were diagnosed according to the standard criteria (Polman et al., 2005). In detail, RRMS patients had two or more discrete relapses with significant neurological recovery in the prior 3 years, and SPMS patients needed to experience continued deterioration for at least 6 months, with or without superimposed relapses with a prior history of at least two relapses. Patients were allowed to be on disease modifying treatment. CIS patients had an episode of neurologic dysfunction typical of MS (e.g., optic neuritis, transverse myelitis, and brainstem syndrome). Conversion of CIS to MS was based on a clinical relapse. Age-matched healthy subjects with a normal neurological examination were also recruited.

Clinical evaluations for patients included Kurtzke Extended Disability Status Scale (EDSS) (Kurtzke, 1983), 25-foot timed ambulation (25FTW), 9-hole peg test (9HPT), 3-second Paced Auditory Serial Addition Test (PASAT), relapse history, and medications. MS Functional Composite (MSFC) was determined from the 25FTW, 9HPT, and PASAT by normalization to a published MS reference group (Fischer et al., 1999).

Disease worsening was defined on MSFC as sustained change of 20% or more from baseline for any of the 3 components of the MSFC sustained for 2 consecutive visits and on EDSS as worsening from baseline by 1.0 EDSS point (or 0.5 point for those with EDSS >=5.5 at baseline) also sustained at 2 consecutive visits. Thus, for both MSFC and EDSS progression, worsening was required to persist for ≥6 months.

For group comparisons, patients were categorized into 5 groups using the baseline and the end-of-study classifications: (1) CIS patients, (2) CIS patients who converted to
MS, (3) RRMS patients, (4) RRMS patients who converted to SPMS, and (5) SPMS patients.

3.2.2 MRI

MRI was acquired axially on a 1.5-Tesla Siemens scanner and included a T2-weighted, fluid-attenuated inversion recovery (FLAIR), proton density (PD)-weighted images acquired with and without a magnetization transfer (MT) saturation pulse for calculation of magnetization transfer ratio (MTR) map, and T1-weighted spin echo (T1SE) images acquired before and after the injection of standard-dose gadolinium contrast agent (Gd-DTPA, T1Gd). MRIs were acquired annually for NC and semi-annually for the patients. Subjects were excluded if they had corticosteroid therapy within 2 months. The details of the image acquisitions are listed on Table 3-1.

<table>
<thead>
<tr>
<th>Image type</th>
<th>Sequence detail</th>
<th>Voxel size</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2-weighted fluid-attenuated inversion recovery (FLAIR)</td>
<td>TR = 6000s, TI = 2000ms, TE = 105 ms</td>
<td>0.9×0.9×5mm</td>
</tr>
<tr>
<td>Proton density-weighted 3D gradient echo image without MT saturation pulse</td>
<td>TR = 35 ms, TE = 6 ms, Flip angle = 3 degrees</td>
<td>0.9×0.9×5mm</td>
</tr>
<tr>
<td>Proton density-weighted 3D gradient echo image with MT saturation pulse</td>
<td>TR = 35 ms, TE = 6 ms, Flip angle = 3 degrees, MT pulse: 7.68 ms 250 Hz Gaussian applied 1.5 kHz off resonance; H1 = 8.8 μT; once per TR interval</td>
<td>0.9×0.9×5mm</td>
</tr>
<tr>
<td>T1-weighted spin echo image without contrast</td>
<td>TR = 800 ms, TE = 15 ms</td>
<td>0.9×0.9×5mm</td>
</tr>
<tr>
<td>T1-weighted spin echo image with contrast</td>
<td>TR = 800 ms, TE = 15 ms, Acquired 5 min after injection of gadolinium-DTPA (0.1 mmol/kg body weight)</td>
<td>0.9×0.9×5mm</td>
</tr>
</tbody>
</table>

3.2.3 Image Analysis
Image analysis was used to estimate brain volumes including BPF, GMF, and WMF. BPF was calculated using the previously described method as the brain volume (BV) divided by the outer contour volume (OCV) (Fisher et al., 1997). The gray matter volume (GMV) was calculated using the new segmentation algorithm (Nakamura and Fisher, 2009) and was adjusted to account for the lesion volume (T2-hyperintense lesion volume, T2LV) as derived in the previous chapter (Section 2.3.5): 

\[ \text{GMV Adjusted} = \text{GMV Measured} + 0.26 \times \text{T2LV} \]

Thus, GMF was calculated as the final \[ \text{GMV Adjusted} \] divided by the OCV. WMF was calculated as BPF minus GMF.

Other conventional measures were T2LV, T1-hypointense lesion volume (T1LV), and gadolinium-enhancing lesion volume (GdLV). Additionally MTR maps were calculated from the PD-weighted image pair acquired with and without an MT saturation pulse using Equation 1-1 (Dousset et al., 1992). The mean MTR of NABT (MTR_{NABT}) and mean lesion MTR (MTR_{Lesion}) were calculated from the voxels included within the brain and T2 lesion masks, respectively. The mean MTRs were divided by the mean MTR of normal-appearing WM (MTR_{NAWM}) to calculate MTR contrast ratios (CR).

3.2.4 Statistical Analysis

The MRI measures were calculated for the baseline and 4-year data. Changes were calculated and compared among MS subgroups and healthy control subjects using analysis of covariance (ANCOVA). For categorical variables, a \( \chi^2 \) test was performed. Spearman’s rank correlation coefficient (SRCC) was used for correlations between GM atrophy and clinical disability. Pearson’s correlation coefficient was used to assess correlations between GMF and WMF, and between GMF and age. To compare groups in
MSFC and EDSS progression status, t-tests were used. For this comparison, longer clinical data was used (mean = 6.6 years). To determine the predictors of GM atrophy, a multiple regression model was developed using a set of baseline MRI measures and their 4-year changes, and adjusted $R^2$ was reported.

### 3.3 Results

#### 3.3.1 Baseline Characteristics

Table 3-2 summarizes the baseline characteristics for each subject group as classified at the 4-year follow up.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>NC</th>
<th>CIS</th>
<th>CIS→MS</th>
<th>RRMS</th>
<th>RRMS→SPMS</th>
<th>SPMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>17</td>
<td>7</td>
<td>8</td>
<td>28</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>Mean age (SD), yr</td>
<td>41.6 (8.1)</td>
<td>44.9 (10.1)</td>
<td>36.1 (7.1)</td>
<td>39.7 (8.4)</td>
<td>42.2 (7.0)</td>
<td>49.7 (7.4)</td>
</tr>
<tr>
<td>Female, (%)</td>
<td>10 (59)</td>
<td>6 (86)</td>
<td>5 (63)</td>
<td>23 (82)</td>
<td>4 (57)</td>
<td>14 (74)</td>
</tr>
<tr>
<td>Mean disease duration (SD), yr</td>
<td>NA</td>
<td>0.3 (0.1)</td>
<td>0.5 (0.7)</td>
<td>6.7 (5.1)</td>
<td>18.5 (11.1)</td>
<td>17.4 (5.0)</td>
</tr>
<tr>
<td>Mean EDSS (SD)</td>
<td>NA</td>
<td>0.9 (0.9)</td>
<td>1.2 (0.4)</td>
<td>2.0 (1.5)</td>
<td>4.8 (1.6)</td>
<td>5.4 (1.3)</td>
</tr>
<tr>
<td>Mean MSFC (SD)</td>
<td>NA</td>
<td>0.36 (0.37)</td>
<td>0.51 (0.46)</td>
<td>0.39 (0.62)</td>
<td>−0.34 (0.78)</td>
<td>−1.04 (1.49)</td>
</tr>
<tr>
<td>Mean T2LV (SD), ml</td>
<td>NA</td>
<td>2.5 (2.3)</td>
<td>7.6 (8.3)</td>
<td>20.7 (17.8)</td>
<td>42.4 (24.1)</td>
<td>43.8 (26.0)</td>
</tr>
<tr>
<td>Mean T1LV (SD), ml</td>
<td>NA</td>
<td>0.15 (0.22)</td>
<td>0.29 (0.48)</td>
<td>1.74 (2.49)</td>
<td>8.17 (6.89)</td>
<td>8.73 (9.00)</td>
</tr>
<tr>
<td>Gd+ (%)</td>
<td>0</td>
<td>14</td>
<td>38</td>
<td>18</td>
<td>29</td>
<td>16</td>
</tr>
<tr>
<td>Mean MTR$_{ Nabt}$ (SD)</td>
<td>35.6 (0.75)</td>
<td>35.5 (0.98)</td>
<td>35.9 (0.81)</td>
<td>35.3 (1.01)</td>
<td>34.6 (0.35)</td>
<td>34.35 (1.19)</td>
</tr>
<tr>
<td>Mean MTR$_{LesionCR}$ (SD)</td>
<td>NA</td>
<td>0.89 (0.12)</td>
<td>0.92 (0.05)</td>
<td>0.92 (0.04)</td>
<td>0.88 (0.03)</td>
<td>0.90 (0.05)</td>
</tr>
<tr>
<td>Mean BPF (SD)</td>
<td>0.862 (0.012)</td>
<td>0.861 (0.008)</td>
<td>0.85 (0.023)</td>
<td>0.84 (0.027)</td>
<td>0.81 (0.02)</td>
<td>0.801 (0.04)</td>
</tr>
<tr>
<td>Mean GMF (SD)</td>
<td>0.554 (0.015)</td>
<td>0.551 (0.010)</td>
<td>0.55 (0.016)</td>
<td>0.537 (0.018)</td>
<td>0.519 (0.017)</td>
<td>0.528 (0.032)</td>
</tr>
<tr>
<td>Mean WMF (SD)</td>
<td>0.308 (0.011)</td>
<td>0.309 (0.009)</td>
<td>0.300 (0.017)</td>
<td>0.304 (0.016)</td>
<td>0.291 (0.015)</td>
<td>0.280 (0.016)</td>
</tr>
</tbody>
</table>

Abbreviations: BPF = brain parenchymal fraction; CIS = patients who had a clinically isolated syndrome and did not meet the criteria for a diagnosis of clinically definite MS over 47 years.
3.3.2 Group Comparison of Gray Matter Atrophy

The annualized rates of WMF, GMF, and BPF are plotted in Figure 3-1 for each group. The figure shows that the rates of GM atrophy increased with MS stage in comparison to the NC: 3.4-fold greater in patients converting from CIS to RRMS, 8.1-fold greater in RRMS patients, 12.4-fold greater in patients converting from RRMS to SPMS, and 14-fold greater in SPMS patients. The GM atrophy rates in the combined set of RRMS patients (CIS→RRMS and RRMS stable) and combined set of SPMS patients (RRMS→SPMS and SPMS) were significantly greater than GM atrophy rate in NC (p = 0.05 and p = 0.005, respectively). In contrast, WM atrophy rates were similar in all MS subtypes, at approximately threefold greater than in healthy control subjects.
3.3.3 Clinical Correlation of Gray Matter Atrophy

GMF was correlated with clinical disability scores. At baseline, GMF had the strongest correlation with MSFC score (SRCC = 0.49, p<0.0001) while the strongest WMF correlation was with 9HPT (SRCC = 0.63, p<0.0001). At the last visit, GMF correlation was greatest with MSFC (SRCC = 0.56, p<0.0001) and was similar between RRMS and SPMS subgroups. GMF was also moderately correlated with EDSS for the group as a whole (SRCC = −0.48, p<0.0001). Correlations differed between the RRMS and SPMS groups for EDSS and for the individual components of the MSFC.

Figure 3-1: Plot of mean annualized rates of atrophy in white matter fraction, gray matter fraction, and brain parenchymal fraction. For abbreviations, see Table 3-2. (Fisher et al., 2008)
Figure 3-2: Plot of gray matter fraction and EDSS. There was a moderate correlation (Spearman’s rank correlation coefficient = 0.56, p < 0.0001). Red dots represent RRMS and blue dots represent SPMS. (Fisher et al., 2008)

Figure 3-3: Plot of gray matter fraction and MSFC. There was a moderate correlation (Spearman’s rank correlation coefficient = −0.48, p < 0.0001). Red dots represent RRMS and blue dots represent SPMS. (Fisher et al., 2008)

When MS patients were classified into those who worsened and remained stable according to the EDSS and MSFC scores over the duration of 6 years, the group with worsening disease had larger mean annualized rate of atrophy in GMF and WMF (Table 3-3 for EDSS and Table 3-4 for MSFC). This difference between worsening and stable patients was significant when the status was defined by MSFC whereas EDSS-defined
groups did not reach statistical significance. Atrophy in WM was similar with either of the worsening definitions (Rudick et al., 2009).

Table 3-3: Characteristics of subjects by EDSS progression.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Worse</th>
<th>Stable</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>21</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Mean age (SD), yr</td>
<td>44.4 (7.24)</td>
<td>42.1 (9.83)</td>
<td>0.3</td>
</tr>
<tr>
<td>Mean disease duration at baseline (SD), yr</td>
<td>10.8 (8.89)</td>
<td>10.5 (8.38)</td>
<td>0.88</td>
</tr>
<tr>
<td>MS classification, (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIS</td>
<td>3 (14)</td>
<td>5 (12)</td>
<td>0.86</td>
</tr>
<tr>
<td>RRMS</td>
<td>11 (53)</td>
<td>25 (59)</td>
<td></td>
</tr>
<tr>
<td>SPMS</td>
<td>7 (33)</td>
<td>12 (29)</td>
<td></td>
</tr>
<tr>
<td>EDSS at baseline (SD)</td>
<td>3.10 (2.56)</td>
<td>3.36 (2.02)</td>
<td>0.69</td>
</tr>
<tr>
<td>MSFC at baseline (SD)</td>
<td>−0.33 (1.36)</td>
<td>0.004 (1.04)</td>
<td>0.33</td>
</tr>
<tr>
<td>BPF at baseline (SD)</td>
<td>0.83 (0.04)</td>
<td>0.83 (0.04)</td>
<td>0.92</td>
</tr>
<tr>
<td>GMF at baseline (SD)</td>
<td>0.53 (0.03)</td>
<td>0.53 (0.02)</td>
<td>0.91</td>
</tr>
<tr>
<td>WMF at baseline (SD)</td>
<td>0.29 (0.02)</td>
<td>0.29 (0.02)</td>
<td>0.84</td>
</tr>
<tr>
<td>BPF % Δ/year (SD)</td>
<td>−0.30 (0.26)</td>
<td>−0.26 (0.24)</td>
<td>0.65</td>
</tr>
<tr>
<td>GMF % Δ/year (SD)</td>
<td>−0.34 (0.40)</td>
<td>−0.25 (0.39)</td>
<td>0.38</td>
</tr>
<tr>
<td>WMF % Δ/year (SD)</td>
<td>−0.28 (0.47)</td>
<td>−0.25 (0.63)</td>
<td>0.86</td>
</tr>
</tbody>
</table>

Abbreviation: see Table 3-2.

Table 3-4: Characteristics of subjects by MSFC progression.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Worse</th>
<th>Stable</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>29</td>
<td>34</td>
<td>NA</td>
</tr>
<tr>
<td>Mean age (SD), yr</td>
<td>44.6 (9.7)</td>
<td>41.3 (8.3)</td>
<td>0.15</td>
</tr>
<tr>
<td>Mean disease duration at baseline (SD), yr</td>
<td>15.0 (8.9)</td>
<td>6.7 (6.0)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>MS classification, (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIS</td>
<td>2 (7)</td>
<td>6 (18)</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>RRMS</td>
<td>12 (41)</td>
<td>24 (71)</td>
<td></td>
</tr>
<tr>
<td>SPMS</td>
<td>15 (52)</td>
<td>4 (11)</td>
<td></td>
</tr>
<tr>
<td>EDSS at baseline (SD)</td>
<td>4.35 (2.05)</td>
<td>2.35 (1.91)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>MSFC at baseline (SD)</td>
<td>−0.51 (1.34)</td>
<td>0.24 (0.84)</td>
<td>0.012</td>
</tr>
<tr>
<td>BPF at baseline (SD)</td>
<td>0.81 (0.04)</td>
<td>0.84 (0.03)</td>
<td>0.01</td>
</tr>
<tr>
<td>GMF at baseline (SD)</td>
<td>0.53 (0.03)</td>
<td>0.54 (0.02)</td>
<td>0.08</td>
</tr>
<tr>
<td>WMF at baseline (SD)</td>
<td>0.29 (0.02)</td>
<td>0.30 (0.02)</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>BPF % Δ/year (SD)</td>
<td>−0.38 (0.24)</td>
<td>−0.20 (0.22)</td>
<td>0.01</td>
</tr>
<tr>
<td>GMF % Δ/year (SD)</td>
<td>−0.40 (0.47)</td>
<td>−0.18 (0.28)</td>
<td>0.03</td>
</tr>
<tr>
<td>WMF % Δ/year (SD)</td>
<td>−0.22 (0.44)</td>
<td>−0.30 (0.68)</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Abbreviation: see Table 3-2.
3.3.4 Predictors of Gray Matter Atrophy

MRI predictors of atrophy were different between the RRMS and SPMS groups, and for GM atrophy and WM atrophy, as shown in Table 3-5. Changes in MRI measures over the course of the study were not predictive of BPF change in the SPMS group. However, in the RRMS group, on-study changes were strongly predictive of concurrent change in BPF, accounting for 72% of the variance in atrophy rate in the RRMS patients. The predictors of GM atrophy included both focal ($\Delta$T2LV and baseline MTR$_{Lesion}$), and diffuse damage (baseline MTR$_{NABT}$) in the RRMS group. Interestingly, however, no predictor was found for GM atrophy in the SPMS group.

<table>
<thead>
<tr>
<th>Table 3-5: Predictors of atrophy from multiple regression analysis.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent variable</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>BPF, % $\Delta$/year</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>GMF, % $\Delta$/year</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>WMF, % $\Delta$/year</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: $\Delta$ = change; subscript “0” = baseline data.

3.4 Discussion

In this study, patients with CIS, RRMS, and SPMS were studied concurrently with age- and sex-matched NC subjects over 4 years in an observational protocol. The results showed that GM atrophy accelerated in patients with more advanced stages of MS. The annualized rates of whole brain atrophy increased with disease stage, from less than 0.2% in CIS patients converting to RRMS to almost 0.4% in patients with SPMS.
Expressed as fold increase from NC, whole brain atrophy increased progressively from 2-fold normal in CIS patients converting to RRMS to 5.8-fold normal in SPMS patients. WM atrophy was relatively constant at 3-fold normal across all disease categories. GM atrophy increased from 3.4-fold normal in CIS patients converting to RRMS to 14-fold normal in SPMS patients.

Although several short-term studies reported greater rates of GM atrophy than WM atrophy in MS patients at various stages of disease (Chard et al., 2004; Dalton et al., 2004; Oreja-Guevara et al., 2005; Sastre-Garriga et al., 2005; Tiberio et al., 2005; Valsasina et al., 2005), an increasing contribution of GM tissue loss to whole brain atrophy was never seen. This finding suggested that the pathological process changed as the disease advanced. The results suggested that MS initially affected both WM and GM equally but became more prominent in GM over time. This was an important finding since MS was conventionally thought as a WM disease.

A second important finding was that GM atrophy correlated with disability. Numerous prior reports have documented correlations between whole brain atrophy and neurological disability, and varying degrees of correlations between GM atrophy and disability, as shown in Table 1-3. The results here showed moderately strong correlations between GMF and MSFC, 25FTW, 9HPT, PASAT and EDSS. In addition to correlations between atrophy and disability scores, there was a trend toward a greater rate of GM atrophy, but not WM atrophy, in patients with sustained EDSS worsening. The findings of greater GM atrophy rates with increasing disease duration and advancing disease stage, together with the disability correlations, would underscore the clinical relevance of GM atrophy in MS and support the use of GM atrophy in clinical trials, particularly in SPMS.
A third important finding was that GM atrophy was more closely related to the disease progression status on MSFC than EDSS. This is not surprising since EDSS heavily depends on ambulation while MSFC includes a test of cognitive function where cognitive dysfunction is generally attributed to GM damage. This finding is consistent with previous reports, which found moderate to fairly strong correlation between GM tissue loss and cognitive tests (Amato et al., 2004; Benedict et al., 2006).

Lastly, the predictors of atrophy were different between RRMS and SPMS as well as between GM and WM. T2LV changes, MTR_{Lesion}, and MTR_{NABT} accounted for a large proportion of the variance in whole brain atrophy and WM atrophy in both RRMS and SPMS. These same markers explained some of the variance in GM atrophy in RRMS. In contrast, there were no observed correlates of GM atrophy in SPMS patients. It is currently unclear if neurodegenerative GM tissue loss in MS occurs as retrograde degeneration from damaged axons or as a process that is independent of focal WM pathology (e.g., cortical/meningeal inflammation or primary neurodegeneration) (Frischer et al., 2009; Geurts et al., 2009; Geurts and Barkhof, 2008; Pirko et al., 2007; Stadelmann et al., 2008; Vercellino et al., 2005). These results might shed light on the role of primary or secondary neurodegeneration in MS. The results here suggested that the processes responsible for GM atrophy changed over the course of disease. In the early stage of disease, GM atrophy appeared to be partly related to WM tissue damage. However, as the disease progressed, pathological processes resulting in GM tissue loss evolved but were no longer related to WM tissue damage because none of the conventional lesion or MTR measures reflecting WM damage predicted GM tissue loss in SPMS. It was possible that
GM lesions, not visible by conventional MRI, were the major contributor to GM atrophy in SPMS (Kutzelnigg et al., 2005).

One potential limitation in investigating GM atrophy in clinical trial is that the duration of typical MS clinical trials is shorter (1-2 years). In this 4-year longitudinal study, the average GM atrophy was approximately 0.3% per year, or 1.2% over 4 years, and the scan-rescan reproducibility was 1.0% (Section 2.3.3), indicating that this cross-sectional GM segmentation method was sensitive enough for detecting GM atrophy. However, in shorter studies, such cross-sectional methods might not be sensitive enough in detecting GM atrophy or might require a large sample size. Therefore, a more robust longitudinal registration-based method might be needed.
CHAPTER 4  CORTICAL LONGITUDINAL ATROPHY DETECTION ALGORITHM

4.1  Introduction

This chapter describes a new algorithm (Cortical Longitudinal Atrophy Detection Algorithm, or CLADA), to measure cortical atrophy by measuring global and regional cortical thickness (CTh) and change in CTh using serially acquired MRIs. CLADA is designed to be sensitive to small changes in MRI. The chapter is organized as follows: description of CLADA (Sections 4.1-4.2), validation tests and results (Section 4.3), and discussion (Section 4.4).

The majority of this chapter was published in NeuroImage (Nakamura et al., 2011).

4.1.1  Background

The measurement of CTh from brain MRI is a fairly new technique, and few algorithms have been published. Generally, these methods can be classified into voxel-based and surface-based methods. The voxel-based methods include the gradient-based approach (Chen et al., 2004), Eulerian partial differential equation (PDE) approach (Acosta et al., 2009; Das et al., 2009; Hutton et al., 2008; Yezzi and Prince, 2003),
Laplace equation approach (Jones et al., 2000), level set approach (Han et al., 2004; Zeng et al., 1999), or minimal line integral method (Aganj et al., 2009). The gradient-based method uses intensity gradient perpendicular to the brain surface and can be robust against smoothly varying intensity inhomogeneity. The Eulerian PDE and Laplace equation-based methods define CTh as the trajectory between two surfaces (as in an electric field). The level set methods use an implicit explicit model where the voxel values represent the distance from the surface, and CTh can be calculated from the distance map. The minimal line integral methods define CTh as the minimal straight distance from one surface to another. Typically, the voxel-based methods are computationally fast but heavily dependant on the accuracy of segmentation, are difficult to impose constraints (e.g., surface smoothness and thickness smoothness), and require separate topological processing (Das et al., 2009; Han et al., 2003; Rueda et al., 2010).

On the other hand, methods that use explicit surface models (collections of points and polygons) (Dale et al., 1999; MacDonald et al., 2000; Xu et al., 1999) are time-consuming but can apply constraints to find the surfaces, are inherently topologically correct (MacDonald et al., 2000), and sub-voxel in their nature.

The definition of CTh differs among various methods and includes the distance along the surface normal (normal distance) (Chen et al., 2004), minimal distance (Aganj et al., 2009; Li et al., 2010; Miller et al., 2000), minimal distance averaged from both surfaces (Dale et al., 1999), distance from solving the Laplace equation (Laplace distance) (Hutton et al., 2008; Jones et al., 2000; Yezzi and Prince, 2003), and straight distance between corresponding points (link distance) (MacDonald et al., 2000)
These definitions can result in the same thickness values when two surfaces are parallel (Figure 4-1a where the red and yellow curves represent the surfaces; e.g., the yellow curve for GM-WM surface and the red curve for GM-CSF surface) or when their curvature is the same (Figure 4-1b). In other situations, thickness varies among the four definitions. The normal distance is intuitive, but the lines along the surface normal can cross each other when the curvature of the two surfaces is different (Figure 4-1c; white lines = normal distance from the yellow surface and black = link distance). In this case, the normal distance measured from the red surface will be similar to the link distance (black). The normal distance does not guarantee the correspondence as shown in Figure 4-1f where white arrows do not return to the same point. Thus, the thickness measurements can be different in each direction (i.e., the distance from GM-WM surface to the GM-CSF surface versus the distance from GM-CSF surface to the GM-WM surface). The minimal distance is simple but also does not guarantee the correspondence. The direction of the distance may not be perpendicular to the surface (Figure 4-1d, white lines = minimal distance and black lines = link distance). A more mathematical definition is the Laplace distance, which guarantees that the direction is perpendicular to the surface at the surface and that the end points correspond to each other (Figure 4-1e; white curves = Laplace distance and black lines = link distance). The curve connecting the surface points can bend according to this definition. On the other hand, the link distance does not curve and may not be perpendicular to the surface (Figure 4-1a-e, black lines).
Figure 4-1: Schematic 2D examples of different measures of thickness (distance along the surface normal, minimal distance from one surface, distance between corresponding points, and distance from solving the Laplace equation) in various situations where two surfaces are represented by yellow and red curves.

(a) When the surfaces are parallel, thickness is the same among the 4 methods. (b) When the curvature on the 2 surfaces is the same, thickness still remains the same among the 4 methods. (c) When the surfaces are not parallel, the distance along the surface normal from yellow surface (white) and the distance between corresponding points (black) may differ. Note that the distance along the surface normal from the red surface will be similar to the distance between corresponding points (black). (d) When the curvatures on 2 surfaces are different, thickness using the minimal distance (white) from yellow surface may be different from the distance between corresponding points (black). Note that the minimal distance from the red surface is not necessarily the same as the minimal distance from the yellow surface. (e) When both surfaces are convex, the distance from Laplace equation (white) guarantees that the curve is normal to both surfaces at the surface whereas the straight distance between corresponding points is not normal to the surface. Note that the distance from solving Laplace equation can curve. (f) The non-correspondence is shown here as the arrows do not return to the same point.

One study has shown low sensitivity with the minimal distance measure and high variability with the normal distance measure while the link distance is the most reliable (Lerch and Evans, 2005).

CLADA is an explicit surface method and its CTh is measured using the definition of the link distance. Therefore, CTh is measured at each time-point on each corresponding pair of vertices. Finally, the rate of change in CTh is calculated for the measurement of cortical atrophy.

4.1.2 Overview of CLADA
A simplified flowchart of CLADA is shown in Figure 4-2. CLADA combines images from all time-points, creates averaged template image, generates a model of the cerebral cortex consisting of 2 explicit surfaces (for inner GM-WM interface and GM-CSF interface, Figure 4-12(d) and (e) respectively) based on the template image, and then deforms the model for each individual time-point for the measurements of CTh and CTh change.

Figure 4-2: Simplified overall flowchart of cortical longitudinal atrophy detection algorithm (CLADA), starting from serial acquisition of T1-weighted images, registration, averaging, segmentation, subject-specific deformable model, longitudinal deformation, and finally cortical thickness measurements.

4.1.3 Deformable Model Format and Viewing

Explicit surfaces are modeled with Object Oriented Graphics Library (OOGL) in OFF file format and implemented in the C language. The OFF file format essentially is a collection of 3D points (vertices) and polygons represented by the vertex indices. In CLADA, all surface polygons are triangular to ensure a unique normal vector for each polygon. An edge file, which saves the list of neighboring vertex indices, is additionally
created to reduce the computation time. The OFF files are rendered and viewed in geomview\textsuperscript{2}.

4.2 Algorithm Description

4.2.1 Image Acquisition

CLADA requires serially acquired T1-weighted images as input. CLADA can process both high-resolution MRI (e.g., magnetization prepared rapid gradient echo (MPRAGE), fast low angle shot (FLASH), or spoiled gradient recalled (SPGR)) and low-resolution 2D images (e.g. T1 spin echo (T1SE)), but the consistency in serial images is critical (i.e., scanner, scanner hardware, software, and imaging sequence). Additional imaging modalities such as FLAIR or T2-weighted images are helpful, but not required, for differentiating the brain, non-brain and abnormal MS lesions.

4.2.2 Pre-processing

Input images are first visually inspected and re-oriented to the default BIP image direction (image orientations of $x =$ left-to-right, $y =$ anterior-to-posterior, and $z =$ superior-to-inferior) as shown in Figure 4-3. Image artifacts (e.g., motion, aliasing, ghosting, and ringing) are individually inspected and recorded. For 2D interleaved sequences, the interleaved patient motion is corrected as in the pre-processing for GM segmentation (Section 2.2.1).

\begin{footnotesize}
\footnote{geomview: http://www.geomview.org/}
\end{footnotesize}
The images are then corrected for $B_0$ and $B_1$ field inhomogeneity by 3 different methods. The first intensity correction recovers the signal drop-off by determining a multiplication factor for each slice, as previously described in Section 2.2.1. The second pre-processing is the N3 correction and also previously described (Sled et al., 1998). Thirdly, a parametric bias field correction (PABIC) is applied (Styner et al., 2000). This algorithm estimates the true intensities by a third-order parametric Legendre polynomial model. The image intensities are initially normalized using a simple scaling during preprocessing and more accurately using a differential bias correction method at the longitudinal deformation (Lewis and Fox, 2004). This differential bias correction do not necessarily eliminate or reduce the intensity nonuniformity in each image, but the resulting bias fields in serial images become more consistent, thus making the measurement of longitudinal atrophy more reliable.

The baseline image is rigidly transformed to a reference space as in Figure 4-3. Since CLADA is a cerebral cortical analysis algorithm, the partial coverage for the cerebellum is not a problem. The same standard space registration method is used as in Section 2.2.3.
4.2.3 Registration and Averaging

During this step, serially acquired images are registered and averaged to create a template image and to improve the signal-to-ratio. For intra-subject registration, non-brain constrained symmetric registration algorithm (NBCSR) is used to register the follow-up images to the baseline image (Chen et al., 2008). Briefly, NBCSR involves 6 steps: (1) whole head rigid-body registration, (2) brain-only affine registration, (3) affine registration in cropped nonbrain region where the top of skull and neck were cropped to eliminate variable appearance, (4) brain-only 6-degrees-of-freedom registration with fixed scaling and skewing from the previous step, (5) repeating steps 1-4 in the backward direction and (6) combination of the resulting forward and inverse transformations to achieve symmetric registration. The correlation coefficient is used as the cost function. This registration matrix is multiplied to the baseline-to-reference registration matrix, and all images are transformed with spline interpolation in a resampled reference space of $0.75 \times 0.75 \times 0.75 \text{mm}^3$ (Figure 4-4).

The field-of-view (FOV) masks are also created and transformed to the same reference space to remove an artifact of averaging missing voxels in case of incomplete brain coverage. Therefore, T1-weighted images from different time-points are added voxel-by-voxel and divided by the sum of the FOV masks for an average image. If FLAIR or T2-weighted images are acquired, they are also registered and transformed to the reference image using NBCSR with normalized mutual information as the cost function.
Figure 4-4: Example of the input (months 0, 6, 12, 18 and 24), corrected, registered, and averaged images.
4.2.4 Segmentation

The goal of the segmentation here is to create a cortical GM-WM boundary mask (Figure 4-5i-j for axial and coronal views), which will be used to initialize the explicit GM-WM surface in the next step. The overall scheme of segmentation is similar to FreeSurfer (Dale et al., 1999) in that the image was segmented into GM, WM, and CSF and a cortical GM–WM boundary is generated through a combination of segmentation and atlas-based classification (Dale et al., 1999). For consistency and efficiency in CLADA, this segmentation step is carried out only once for each subject rather than for each of the images separately.

The subject-specific average image is segmented using the method from Chapter 2. The segmented brain mask (Figure 4-5b) is further processed using morphologic operations to fill in the lateral ventricles as described in Section 2.2.4 (Figure 4-5c). Next, the GM-WM boundary mask, which contains the cerebral white matter, lateral ventricles, and deep gray matter, is created for initial surface deformation. The deep gray matter structures (caudate, putamen, and globus pallidus), cerebellum and brainstem are isolated once in the Harvard Brain Atlas (Kikinis et al., 1996). ART (Automatic Registration Toolbox, version 2009-03-12) nonlinear registration is used to register the subject image and the atlas (Ardekani et al., 2005). Using the resulting transformation, the masks for the cerebellum and brainstem are non-linearly transformed to the subject space, disconnected, and eliminated from subsequent analysis steps (Figure 4-5 g and h, respectively). The deep gray structures are added to the GM-WM boundary mask; the transformed masks of caudate (red) and putamen (turquoise) are shown in Figure 4-5f and the segmented WM

3 http://www.nitrc.org/projects/art/
mask in Figure 4-5e. If necessary, MS lesions are segmented on FLAIR or PD/T2 dual echo images using iterative conditional modes (Besag, 1986) and added back into the GM-WM boundary mask (Figure 4-5d). On MRI of MS patients, it is critical that juxtacortical lesions are properly segmented so that the surface deformation into lesions is avoided (Figure 4-5 k and l; the inlets show zoomed areas of the same juxtacortical lesion with axial and coronal views, respectively). Typical subcortical and periventricular lesions do not interface with the cortex and do not affect the initial deformation. Manual editing is performed to correct any remaining segmentation errors due to lesion holes (if no FLAIR image is available) and misclassification of blood vessels as white matter, which may occur in low-resolution images due to partial volume effects.

![Figure 4-5: Examples of detailed segmentation steps in CLADA. (a) Averaged MPRAGE, (b) brain mask overlay, (c) lateral ventricle overlay, (d) MS lesion overlay, (e) white matter overlay, (f) coronal view of deep gray matter structures (red = caudate and turquoise = putamen), (g) cerebellum overlay, (h) brainstem overlay, (i,j) axial and coronal views of the final GM-WM boundary mask, and (k,l) axial and coronal views of the same juxtacortical lesion, which require proper segmentation as shown by the inlets.](image)

4.2.5 Deformable Model

As in other explicit surface models (Dale et al., 1999; MacDonald et al., 2000), two explicit surfaces, one for the inner cortical surface (ICS, between GM and WM) and
one for the outer cortical surface (OCS, between GM and CSF) are constructed next. The OCS is difficult to detect due to small separations between opposing surfaces within tight sulci and partial volume effects, and thus the ICS is created first with the GM-WM boundary mask (Dale et al., 1999; Teo et al., 1997).

The approach used in CLADA combines some aspects of both FreeSurfer (Dale et al., 1999) and anatomic segmentation using proximities (ASP, (MacDonald et al., 2000)). Instead of simultaneously deforming the two surfaces from the start, as in ASP, a single spherical surface shrink-wraps to the GM-WM boundary mask first, to define the ICS, then it deforms outward to define the OCS. Only after the two surfaces are generated, simultaneous deformation is applied to both surfaces.

An example of deformation is shown in Figure 4-6 starting from (a) sphere, (b) intermediate shrink-wrap with long 1.5mm edges, (c) ICS after shrink-wrap, (d) initial OCS, and (e) final subject-specific cortical surface model after simultaneous surface deformation. The shrink-wrap, OCS initialization and simultaneous deformation are described next.
**Figure 4-6:** Example of surface deformation for subject-specific dual cortical surface model. The top row shows the rendered deformed surface, and the bottom row shows image with the surfaces shown in yellow and red for inner cortical surface and outer cortical surface, respectively. From left to right, (a) initial spherical surface (yellow), (b) after the shrink-wrap at large edge length (2mm), (c) further shrink-wrap with 1mm-edge length, (d) initial estimate of outer cortical surface, and (e) final subject specific cortical deformable model after simultaneous deformation.

**Shrink-Wrap Method for Inner Cortical Surface**

CLADA's shrink-wrap algorithm is initialized from a sphere (Figure 4-6a) and employs (1) multi-resolution remesh function, (2) method for fast surface intersection detection, and (3) method for surface relaxation on proximal surfaces to improve efficiency and accuracy.

For remeshing, the algorithm described by Botsch and Kobbelt is implemented to manipulate the mesh configuration and distribution (Botsch and Kobbelt, 2004). Briefly, this remesh algorithm iteratively (1) splits edges longer than 4/3 of the target length, (2) collapses edges shorter than 4/5 of the target, (3) flips edges to reduce the variations in the number edges among vertices, and (4) tangentially smooths the surfaces. For efficiency, a multi-resolution approach is implemented by decreasing the target length from 5 mm to 1 mm. The algorithm is computationally efficient, maintains consistent topology, has small geometric distortions due to its use of an iterative tangential relaxation, and results in consistent distribution of vertices (i.e., 6 or 7 edges per vertex and approximately 60°-angles in many triangles). This remesh function reduces the variability in the number of vertex neighbors, as seen by typical marching cube algorithms (Lorensen and Cline, 1987).

A fast triangle–triangle intersection test is used to detect self-intersections (Moller, 1997) and to prevent physically impossible overlaps. The intersection tests are performed
with fast recursive bounding boxes (Zomorodian and Edelsbrunner, 2000) to decrease the computation time. Figure 4-7 shows the log-plot of the computation time versus the input object’s number of vertices (approximately ranging from 3,200 to 320,000). The bounding boxes reduces the computation time from $O(n^2)$ to $O(n)$.

Figure 4-7: A log-scaled plot comparing the computational times from the basic self-intersection test method and improved method with bounding boxes (BBox).

The surface relaxation on proximal surfaces minimizes the local surface area by smoothing when an object has a tunnel, a common artifact in low-resolution images with a significant partial volume effect. The algorithm detects proximal surface regions by extending the surface deformation and testing for surface intersections but not actually applying the deformation. If a potential surface intersection is detected, a local smoothing operation is performed on this surface region by averaging the neighboring vertex positions.

A Laplace map is created from the brain and GM-WM boundary masks to guide the direction of the deformation (Figure 4-8). It is particularly helpful in deformation into deep sulci. This map is similar to the gradient vector flow (Xu and Prince, 1998) and is a
non-increasing contour map with its maximum at the edge of initial brain mask and the minimum at the edge of the target GM-WM boundary mask. The Laplace equation is a second-order PDE of a scalar field “u” and solved using a finite difference method by approximation (Haidar et al., 2005).

\[ \Delta u = 0 \]
\[ \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \sum_{n=1}^{N} u_n \]  

Equation 4-1

where \( \Delta \) is the Laplace operator, \( u(x,y,z) \) is the 3-dimensional scalar field, and \( N \) is the number of neighbors, which is set to 6 in CLADA. In a matrix form, this equation becomes

\[ \bar{A}\bar{u} = \bar{b} \quad \text{where} \]
\[ A_{ij} = \begin{cases} 
1 & \text{if on boundary} \\
6 & \text{if } i = j \\
-1 & \text{if } i = j_{\text{neighbor}} \\
0 & \text{otherwise}
\end{cases} \]  

Equation 4-2

and where the matrix \( A \) is an \( NV \times NV \) sparse matrix, \( NV \) is the number of voxels within the brain mask, \( b \) is the value of \( u(x,y,z) \) for boundary condition or 0 if inside the object, and matrix elements can be either 1, 6, −1, or 0. The scalar field, \( u(x,y,z) \), is calculated for the voxels inside the brain but outside the GM-WM border mask. At the brain boundary, \( u(x,y,z) \) is set to a large number (e.g., 65535) and the GM-WM boundary \( u(x,y,z) \) is set to 0. The masks are resampled to even higher resolution at 0.4mm isotropic voxel, so the exact solution of this matrix at higher resolution requires an inversion of a 2million-by-2million matrix, which is not practical. Therefore, a conjugate gradient method is used to approximate the solution of this matrix. An example of the Laplace map is shown in
Figure 4-8. After the calculation of the Laplace map, sobel directional filters are applied to calculate the direction of deformation (shown by small black vectors on Figure 4-8 b).

![Figure 4-8](image)

**Figure 4-8:** (a) The log10-transformed Laplace map, log10\(u(x,y,z)\), is shown and color-coded from dark blue \((u=0 \text{ around GM-WM border})\), green \((u=32768)\), yellow \((u=45875)\) to dark red \((u=65536 \text{ near the brain surface})\). (b) The sobel directional filters were applied on the Laplace map to calculate direction of surface deformation (small black vectors). The region came from the medial frontal lobe (white rectangle in (a)).

At the end of the shrink-wrap, the ICS typically contains approximately 200,000 vertices and matches the GM-WM boundary mask (Figure 4-6c).

**Outer Cortical Surface**

After ICS detection, the initial OCS is estimated by iteratively expanding ICS outward (Figure 4-6d). A target OCS vertex is established for each ICS vertex using the surface model, averaged image, and sobel-filtered gradient image according to the following rules: (1) the intensity at the OCS vertex must be above the estimated mean OCS intensity, \(\mu_{OCS}\), which is the mean intensity of GM-CSF edge voxels. The edge voxels must have high gradient, above the 75th percentile for all gradients in the brain.
mask, and intensity between mean GM and CSF intensities (obtained from segmentation).

(2) The first derivative of the intensity profile along the OCS surface normal must be negative (i.e., decreasing intensity). (3) The OCS must not intersect another surface. (4) The OCS vertex must be contained within the brain mask, but not in the cerebellum or brainstem masks. (5) The distance between the ICS vertex and the target OCS vertex must be less than predefined distance of 3 mm. Although the cortex is thicker than 3 mm in some regions, the initial estimate is confined to 3mm to provide a stable and realistic starting point that is close to the final surface. As in ASP (MacDonald et al., 2000), the correspondence between each inner surface vertex and outer surface vertex is maintained for calculation of CTh.

**Simultaneous Surface Deformation**

Finally, the two cortical surfaces are deformed simultaneously with an additional internal energy term to ensure CTh smoothness:

\[
\frac{\partial v}{\partial t} = \alpha F_{\text{surf}} + \beta F_{\text{thck}} + \gamma F_{\text{imag}}
\]

Equation 4-3

\[
F_{\text{surf}} = v_{\text{local}} - v
\]

Equation 4-4

\[
F_{\text{thck}} = CTh_{\text{local}} n - v
\]

Equation 4-5

\[
F_{\text{imag}} = \frac{\lambda}{\nabla I(v) + \lambda} \left\{ 2H(I(v) - \mu_{CS}) - 1 \right\}
\]

Equation 4-6

where \( \frac{\partial v}{\partial t} \) is the iterative deformation of 3D vertex position, \( v \), determined by the internal surface smoothness force \( F_{\text{surf}} \), internal thickness smoothness force \( F_{\text{thck}} \), external image force \( F_{\text{imag}} \) and the weighting constants \( (\alpha, \beta, \gamma) \). The surface
smoothness term ($F_{surf}$) pulls the vertex towards the local average point ($v_{local}$); similarly, the thickness smoothness pulls the vertex towards the point that corresponds to the averaged local thickness ($CTh_{local}$) along the surface normal ($\hat{n}$). The image force is determined by the image gradient at the vertex ($\nabla I(v)$), a constant lambda, $\lambda$, which regulates the gradient, and a Heaviside function ($H$) of image intensity ($I(v)$) and estimated edge intensity ($\mu_{CS}$). Lambda ($\lambda$) is the 95% value of the image gradient within the brain. The weighting constants, $\alpha$, $\beta$ and $\gamma$, are set to 0.5, 1.5, 2.0, respectively, as determined through trial-and-error to produce good results. The integral solution is calculated by the fourth order Runge-Kutta method, and the deformation iterates until the deformation size ($\Delta C$) is smaller than predefined threshold ($\Sigma \Delta C_n / d_n / N < 1\%$) where $d_n$ is the distance between $n$’th vertex on ICS and OCS and $N$ is the total number of vertices.

The resulting model is the subject-specific cortical model based on the averaged image and is used as initialization for longitudinal deformation in individual images. The remesh function is not applied so as to maintain consistent ICS-OCS correspondence.

4.2.6 Longitudinal Deformation

The subject-specific cortical model is subsequently deformed to fit the images from individual time-points. The cortical model rather than the image is transformed to prevent interpolation artifacts (Han et al., 2006). In this step, the remesh function is not applied in order to ensure that consistent ICS–OCS correspondence is maintained. The same deformation process as described above is applied iteratively for each time-point until the change is below a predefined 1% threshold as before.
4.2.7 Measurement Cortical Thickness and Change

Cortical thickness is determined for each surface vertex as the distance between corresponding vertices on the ICS and OCS surfaces (link distance). The global CTh is estimated as the surface-area-weighted average: \( CTh_{\text{global}} = \frac{\sum A_n d_n}{\sum A_n} \) (where \( A_n \) is the triangular surface areas surrounding the \( n^{\text{th}} \) ICS and OCS vertices). Although the superior aspect of the brainstem as well as the diencephalon, hippocampus and amygdala are a part of the ICS and OCS surfaces, these non-cortical regions are excluded from the global CTh measurement by masking with the registered Harvard Brain Atlas (Kikinis et al., 1996) regions. For regional analysis, the atlas is transformed and filtered (in surface space, Figure 4-9). Surface-area-weighted regional CTh is measured for each anatomic region in the same manner as for global thickness. Regional labels remain constant across time-points for consistent longitudinal analysis.

![Figure 4-9: Nonlinearly transformed and filtered regional labels for the frontal lobe (blue), parietal lobe (yellow), temporal lobe (orange), occipital lobe (red), and insular area (cyan). The regional cortical thickness and thinning are measured by surface-area-weighted average within each label. The color is mapped on the inner cortical (GM-WM) surface (d).](image-url)
4.3 Validation

Individual elements of the CLADA’s deformable model were tested using simulated data prior to testing on real brain MRIs. These tests were done to demonstrate efficacy of the self-intersection test, surface relaxation step used for tracking the convoluted inner surface, and the Laplace-based shrink-wrap algorithm. Further tests were performed to evaluate CLADA in terms of its (1) scan-rescan reproducibility (2) intra-rater reproducibility, (3) accuracy, (4) sensitivity to change in CTh, (5) sensitivity to input image type, (6) sensitivity to reference time-point, (7) comparison with FreeSurfer, (8) bias in estimating atrophy, (9) comparison to a cross-sectional method, and finally (10) effect of MS lesions on CTh measurements.

4.3.1 Phantom Experiments

Tests using simulated data and brain MRIs demonstrated the ability of CLADA to track convoluted surfaces (Figure 4-10). While deformation from a sphere to a smaller sphere did not cause any problems, even without the self-intersection test, a sphere with a tunnel caused problems when the surface was deformed without testing for self-intersections because the surface continued to grow on opposite sides. Although physically impossible in actual brain WM, such tunnels and disconnections can result from segmentation errors. This could be particularly problematic for low-resolution MRIs with significant partial volume effects. For spheres with straight tunnels, deformations that included the self-intersection test resulted in essentially the same final surface with or without the inclusion of the relaxation step. However, if there was a narrowing in the center of the tunnel, similar to the shape of the cortex when WM voxels are misclassified.
due to partial volume effects, the result of the self-intersection method alone will depend on the position of the initial object. The relaxation step moved the self-intersection point by relaxing and minimizing the surface area, thus reducing the dependence on the initial condition.

**Figure 4-10:** Effects of including or excluding the self-intersection test and surface relaxation steps shown as a collection of cross-sections of 3D objects (black) with the initial starting point of the deformable surface (blue) and the resulting surface after deformation (red). The images show cross-sectional views of a deformation starting from the blue object into the red object with the binary image in black. Top row: With a sphere and well-located initial object, all methods perform equally well. Second row: When there is a tunnel in the middle of the sphere, the self-intersection test is needed to stop the deformation. Otherwise, the surfaces continue to move toward the other side. Third and fourth rows: When the tunnel narrowing is located at the center, the same results are obtained with or without relaxation. However, when the center is shifted to the right, as in the fourth row, the relaxation method accurately shifts to the right with the narrowing.
Next, the use of the Laplace map to determine the direction of deformation was evaluated by another phantom experiment (Figure 4-11). Given an initial surface (blue) and a target object (gray “L”-shaped object), the surface iteratively shrank inward as shown by colored contours (b). “t” is the number of iterations and colored coded from blue (t=0), light blue (t=9), pink (t=17), orange (t=21), to red (t=29). The surface eventually self-intersected and could no longer move at t=29. Using the gradient of Laplace map (c, short vectors) in addition to surface normal, the deformable surface shrank fairly smoothly towards the target object (d).

Figure 4-11: The use of Laplace map was evaluated in a phantom experiment. Given an initial blue surface and L-shaped target object (a), the surface iteratively shrank inward as shown by colored contours (b). t is the number of iterations and colored coded from blue (t=0), light blue (t=9), pink (t=17), orange (t=21), to red (t=29). The surface eventually self-intersected and could no longer move at t=29. Using the gradient of Laplace map (c, short vectors) in addition to surface normal, the deformable surface shrank fairly smoothly towards the target object.

4.3.2 MRI Experiments
The typical CLADA result on actual MRI is shown in Figure 4-12.

![Figure 4-12: Example of CLADA images showing (a) averaged T1-weighted MPRAGE images, (b) inner cortical surface (yellow) and outer cortical surface (red), (c) color-coded regional patterns of atrophy and growth in this subject, and (d and e) rendering of inner and outer cortical surfaces, respectively.](image)

CLADA was also tested on actual low-resolution MRIs. Axial images with 5 mm thick slices were reformatted in the coronal plane and resampled to have isotropic voxels with significant partial volume averaging (Figure 4-13a). The effects of including both the self-intersection and the relaxation step (Figure 4-13c) can be appreciated over using the self-intersection test alone (Figure 4-13b, blue arrow).

![Figure 4-13: Comparison of the resulting surface detected with and without the relaxation step included in the deformation. (a) resampled low-resolution MRI (b) deformed cortical surface detected with the inclusion of the self-intersection test but without surface relaxation (c) deformed cortical surface detected with the inclusion of the self-intersection test and surface relaxation step. The blue arrow indicates the error where the surfaces moved incorrectly due to a tunnel artifact.](image)

Figure 4-14 shows CLADA results from a patient with Alzheimer’s disease taken from Open Access Series of Imaging Studies database (OASIS) (Marcus et al., 2007).
Significant atrophy is visible in the ventricles over the course of 1223 days (3.4 years). Despite the significant blurring in the averaged image (upper left), CLADA was able to find ICS and OCS properly and successfully deformed both surfaces even in the presence of severe atrophy. The figure also shows that the intra-subject registration (NBCSR) performed well even in the presence of significant atrophy.

**Figure 4-14:** An example of longitudinal deformation in the subcortical region including the posterior lateral ventricle (v) of a demented subject. Despite blurring, CLADA was able to detect the cortical surfaces in the averaged image and to longitudinally follow both inner (yellow) and outer (red) edges even in the presence of severe atrophy.

### 4.3.3 Scan-Rescan Reproducibility

The reproducibility of CLADA’s global and regional CTh measurements was evaluated using scan-rescan MRI datasets. Reproducibility was determined for both high-resolution 3D MPRAGE and standard-resolution 2D T1SE images. Ten scan-rescan datasets of young controls from the OASIS database (Marcus et al., 2007) were used for the high-quality image analysis. Subjects were scanned twice within 10 days on a 1.5T scanner with the same MPRAGE sequence [repetition time (TR) = 9.7msec; echo time
(TE) = 4.0msec; flip angle (FA) = 10°; inversion time (TI) = 20msec; slice thickness (THK) = 1.25mm; gap = 0mm; number of slices = 128; matrix size = 256×256; in-plane resolution (IPR) = 1mm×1mm; number of signal averages (NSA) = 4]. The lower resolution 2D T1SE images were acquired from 9 MS patients on a 1.5 T Siemens Magnetom Vision with the following sequence parameters: TR=800 ms, TE=20 ms, THK=5.0 mm; gap=0 mm; number of slices=30; matrix size=192×256, IPR=0.9 mm×0.9 mm. The MS patients were scanned twice, with one week between acquisitions.

CLADA was applied to analyze the images from both datasets and to measured global and regional CTh, and calculated absolute difference and percent error ( = absolute difference / average) for each subject. The percent errors in global and regional CTh were plotted as boxplots.

CLADA performed better on MPRAGE than T1SE images. The mean (standard deviation) scan-rescan absolute error in global CTh was 0.015 mm (0.012) and 0.029 mm (0.019) for MPRAGE image pairs and T1SE image pairs, respectively. The mean (s.d.) percent error in global CTh was 0.45% (0.38) and 0.77% (0.49) for MPRAGE and T1SE, respectively. Regional CTh measures had lower percent error for MPRAGE images in the temporal, parietal and occipital regions (Figure 4-15).
Figure 4-15: Boxplots of percent error in global and regional cortical thickness measurements from scan-rescan tests. Reproducibility was measured from datasets consisting of T1-weighted spin echo images of 9 MS patients (blue) and MPRAGE of 10 OASIS subjects (green).

For visualization of regional variability, surface error maps similar to Han and colleagues (Han et al., 2006) were generated for each subject. The intensity of each surface voxel was set to the difference in CTh from the nearest vertex and the resulting error image was smoothed (FWHM=0.9 mm). To create mean surface error maps, the images were registered to the MNI152 template using ANTS SyN nonlinear registration (Avants et al., 2008). The error images were averaged across subjects in the MNI space, inverse transformed back to native space of one subject, and mapped to the cortical surface.

Scan–rescan CTh error maps are shown in Figure 4-16. Overall, MPRAGE images showed more homogeneous errors less than 0.3 mm, and lower-resolution T1SE images resulted in larger errors up to 0.7 mm. There were slightly larger errors in the inferior region of the brains, but in general the errors were homogeneous over the entire cortex, especially for MPRAGE.
Figure 4-16: The absolute cortical thickness difference was averaged in MNI space and mapped back to the cortical surface. The red surface indicates large errors and dark blue indicates no error (0mm difference in CTh). MPRAGE images (top) showed very low variability (less than 0.3 mm) whereas spin echo images (bottom) had differences of up to 0.7 mm.

4.3.4 Intra-operator Reproducibility Test

Intra-operator variability was assessed using a subset of these cases (five OASIS MPRAGE and five MS T1SE). A single operator analyzed these scan-rescan cases twice using CLADA, approximately 4 days apart. Variability was calculated as the absolute difference in global and regional CTh and Pearson correlation coefficient of CTh change. An intra-operator variability surface map was generated in the same manner as for scan–rescan variability.

The mean (s.d.) intra-operator absolute difference in global mean CTh from the MPRAGE and T1SE image sets were 0.01 (0.02) mm and 0.01 (0.01) mm, respectively.
The mean (s.d.) absolute differences in CTh change were 0.0009 (0.0012) mm/yr and 0.001 (0.001) mm/yr for MPRAGE and T1SE, respectively. The repeated CTh change measurements were highly correlated in both MPRAGE (slope=1.07, r=0.997, p<0.0001) and T1SE (slope=0.999, r=0.988, p<0.0001). The intra-operator error was much smaller than the scan-rescan error, indicating that the effect of manual editing is small.
Figure 4-17: The intra-operator variability was measured from repeated analysis of longitudinal T1SE images from five MS patients (left column, a, c, and e) and longitudinal OASIS MPRAGE images from 5 demented or non-demented subjects (right b, d, and f). (a,b) Differences in absolute cortical thickness; (c,d) regional error map; (e,f) global and regional cortical thickness changes from repeated analyses. For both lower-resolution T1SE and higher-resolution MPRAGE images, the repeated measurements were highly correlated near the unity line (slope = 0.999, r = 0.988, p < 0.0001 for T1SE; slope = 1.07, r = 0.997, p <0.0001 for MPRAGE).

4.3.5 Accuracy using Post-mortem Images

To evaluate the accuracy of the CTh measurements, CLADA was used to measure CTh from post-mortem MRIs acquired in situ and then compared to a “gold standard” thickness measured in 2D digital photographs of the same brain slices.

Our rapid autopsy protocol (Fisher et al., 2007) involved the following as described in Figure 4-18: (1) clinicians at the Mellen Center for Multiple Sclerosis Treatment and Research recruited prospective patients, obtained consent, and entered the demographic information into the database. (2) Upon the death of an MS patient, appropriate personnel were called, the cadaver was transported to the Cleveland Clinic, and in situ MRI was obtained. The relevant sequence here was MPRAGE [TR=1900 ms, TE=1.71 ms, THK=1.2 mm; number of slices=120; matrix size=256×256, IPR=1 mm]. (3) The brain was removed from the cadaver, and the left cerebral hemisphere was fixed in 4% paraformaldehyde for at least 4 weeks. (4) After the fixation was complete, the fixed brain was placed in a customized slicing box with MRI-sensitive markers and imaged again with the same MPRAGE sequence except with 1mm isotropic voxel size and 4 signal averages to accurately identify the marker locations. (5) Post-fixation MPRAGE (Figure 4-18(5) left) with MRI-sensitive markers were acquired (yellow arrow on the contrast enhanced image (Figure 4-18 (5) right) –this post-fixation MRI was used
solely to locate the corresponding slicing plane in the post-mortem MRI, and thus no CTh measurements were made from post-fixation MRI. (6) Sliced brain tissue was photographed. (7) The left brain hemisphere was segmented in *in situ* MRI. (8) The *in situ* MPRAGE was registered to post-fixation MPRAGE using affine and nonlinear registration tools (FLIRT and FNIRT in FSL (Smith et al., 2004)) to account for brain shrinkage due to fixation. (9) Guided by the MRI-sensitive markers, the registered *in situ* MPRAGE was re-oriented to the slicing plane. Each brain slice was visually inspected and corrected for misalignment.

![Figure 4-18: A figurative flowchart of the rapid autopsy program at the Cleveland Clinic.](image)

To compare CTh measurements obtained from CLADA using 3D MRI to CTh measurements from 2D photographs of the brain slices, four constraints were set to identify valid regions of interest (ROI): (1) Corresponding vertices from the inner and outer cortical surfaces had to be within 0.2 mm from the co-registered image/photograph
plane; (2) The absolute value of the dot product of unit normals from the brain slice plane and the cortical surfaces had to be less than 0.3 to ensure they were approximately perpendicular; (3) Large ROIs that spanned gyri and sulci had to be subdivided to minimize spatial variability as gyrus thickness tended to be larger than sulcal thickness (2.9 mm and 2.2 mm respectively, (Fischl and Dale, 2000)); and (4) As determined by careful visual inspection, areas that appeared to have any misregistration were excluded. ROIs were manually selected from all major (frontal, temporal, parietal, and occipital) lobes throughout the hemisphere. For each valid ROI determined in the MRIs based on the criteria above, the same ROI was manually selected in the photograph based on landmarks and the mean CTh was measured within the ROI. To measure CTh from the photographs, the WM–GM inner edge and GM–CSF outer edge were manually determined, and CTh was estimated using the Laplace method (Haidar and Soul, 2006). Finally, the Pearson correlation and absolute differences between the photograph CTh and MRI CTh measurements were calculated.

Twenty-seven slices from four brains imaged post-mortem were used in the accuracy test. Out of a total of 113 ROIs initially selected for analysis, 10 ROIs were omitted from three slices due to misregistration. Thus, accuracy analysis was completed on the remaining 103 ROIs. An example of matching photograph and MRI is shown in Figure 4-19.
CTh measurements from CLADA analysis of MRIs were linearly correlated to CTh measurements from photographs of the same tissue slices \((r=0.68, \ p<0.001)\). The slope of the regression line \((1.06)\) was slightly greater than unity (Figure 4-20). The mean (s.d.) difference between MRI and photograph CTh measurements was \(0.17 \ (0.54) \ \text{mm}\) and the mean (s.d.) absolute difference was \(0.43 \ (0.36) \ \text{mm}\). Detailed inspections on the large errors revealed that CLADA included dura mater and resulted in increased CTh from CLADA compared to CTh from sliced brain. The mean error was fairly small, considering the voxel size of 1 mm in these MRIs.
The cortical thickness measurements from MRI and photographs correlated with slope = 1.06, R = 0.68, and p < 0.001. The best fit line is shown in black, and dotted red line is the ideal case.

4.3.6 Sensitivity using Simulation of Atrophy

The sensitivity of CLADA for detection of small changes was tested using simulations of regional brain atrophy or growth. ROIs were manually selected from MRIs of five OASIS datasets. The ROIs on the brain surface were very slightly deformed by a known amount at sub-voxel level, either inward for atrophy or outward for growth. The direction of deformation was normal to the brain surface mask and determined by 1.6 mm-FWHM Gaussian and sobel directional filters. The magnitude of the deformation was normalized to maximum of 0.3 mm at the brain edge after the 1.6 mm-FWHM Gaussian filter. CLADA was applied to the original and deformed images to measure changes in CTh. The error was calculated as the mean absolute difference between the measured results and the actual applied deformations. CLADA's sensitivity was calculated by TP/(TP+FN) where TP is the number of surface points CLADA correctly
deformed more than 0.1 mm and FN is the number of surface points CLADA did not deform even though the image was deformed more than 0.1 mm. A plot of the actual versus measured deformations was plotted, and the slope of best-fit line was calculated.

For the sensitivity test, a total of 8172 surface points were collected from 20 ROIs in five OASIS datasets (Figure 4-21). The maximum applied deformation was 0.3 mm. The sensitivity or the true positive rate with a cut-off of 0.1 mm detection was 86%, and the slope of the best-fit line was 0.75. Measured changes were strongly correlated to applied simulated changes ($r=0.958$, $p<0.001$, Figure 4-21). The mean absolute error (s.d.) between measured and applied changes in CTh was 0.02 (0.02) mm. This result showed that CLADA was sensitive to very small, sub-voxel change.

**Figure 4-21:** An example of the original (a), deformed (b), and difference (c) images, and (d) The results of the sensitivity test. The black solid line is the best fit line of the measured thickness change and the applied deformation (slope = 0.75, Pearson correlation coefficient = 0.958, $p<0.001$). The red dotted line is the ideal line.
4.3.7 Comparison of Image Sequences

Next, CLADA results from conventional 2D T1SE images and from high-contrast 3D MPRAGE images were compared. The slice thickness for 2D T1SE was 3mm and MPRAGE was 1.2mm. Longitudinal MRI from 15 secondary progressive multiple sclerosis (SPMS) patients included axial T1SE images (TR=675 ms; TE=17 ms; THK=3 mm; number of slices=48; matrix size=256×192; in-plane resolution=0.9mm×0.9 mm; NSA=1) acquired on a 1.5 T Siemens Symphony scanner and axial MPRAGE images [TR=1900 ms; TE=1.71 ms; FA=8°; TI=900 ms; THK=1.2 mm; number of slices=128; matrix size=256×256; IPR=1×1 mm; NSA=1] acquired on the same day on a 3 T Siemens Magnetom Trio scanner. Images were acquired at baseline and weeks 2, 4, 26, 52, and 78. The T1SE and MPRAGE images were analyzed separately. The annualized global CTh change was calculated as the difference between CTh measurements at baseline and last follow-up divided by the time interval (78 weeks or 1.5 years). The Pearson correlation was determined between annualized global CTh change measurements obtained from T1SE and MPRAGE images.

CTh measurements differed for T1SE and MPRAGE input images (Figure 4-22). The mean global CTh at baseline was 2.71±0.21 for MPRAGE and 3.57±0.23 mm for T1SE (p<0.0001). The mean annualized changes in CTh measured from MPRAGE and T1SE images were −0.031±0.02 mm and −0.024±0.02 mm, respectively. The correlation was 0.569 (p=0.027). The percent change in CTh for MPRAGE and T1SE was −1.14% for MPRAGE and −0.69% for T1SE. Although the measured rates of cortical atrophy was smaller with T1SE datasets compared to MPRAGE datasets, CLADA was able to
detected atrophy in all patients from both sequences. The results also showed that CTh significantly differed when different sequences were used.

![Comparison of CLADA Results from T1SE & MPRAGE](image)

**Figure 4-22**: Scatter plot of annualized cortical thickness change measured from T1SE and MPRAGE images in 15 MS patients. The slope was 0.423 (r=0.569, p=0.027).

### 4.3.8 Effect of Reference Image

Registration-related variability was investigated by changing the reference time-point from baseline to follow-up. Eight longitudinal OASIS datasets (each with 2 time-points) were analyzed (Marcus et al., 2007). CLADA was run twice for each set: once with the baseline image used as the reference space, and then again with the follow-up image used as the reference. The absolute difference was calculated for the global and regional CTh measurements, and a surface error map was generated to compare the annualized CTh changes measured each way.
The results from repeated CLADA analysis with either baseline or follow-up image as the reference space showed a mean absolute difference in mean global CTh of 0.013±0.005 mm. The regional CTh absolute differences (s.d.) were 0.015 (0.016) in frontal, 0.023 (0.021) in temporal, 0.014 (0.013) in parietal, and 0.011 (0.012) mm in occipital lobes. The boxplot and the surface error map are shown in Figure 4-23.

The mean error due to the choice of reference was 0.013mm, which was comparable to the scan-rescan mean absolute difference of 0.015mm. Therefore, the effect of reference image was very small.
Figure 4-23: Effect of different reference spaces: (a) boxplot of absolute global and regional
cortical thickness differences; (b) plot of change in cortical thickness using baseline as the
reference space and using the follow-up image as the reference space (slope = 1.00, r = 0.997,
p < 0.0001); (c) surface error map indicating regions where cortical thickness differed.

4.3.9 Comparison with FreeSurfer

CLADA was compared to FreeSurfer since it was widely used, freely available
and extensively validated. CLADA and FreeSurfer were applied on a subset of dataset
(seven of the SPMS patient) described in the “Sequence Comparison” section. FreeSurfer
version 4.5.0 was used on the baseline MPRAGE images. FreeSurfer CTh was measured
within the same regions as CLADA to enable direct comparison of the global and
regional CTh results. Regional labels (global, frontal, temporal, parietal and occipital
ROIs) created from CLADA were used in FreeSurfer as described in FreeSurfer ROI
analysis workflow 4. For FreeSurfer analysis, semi-automatically segmented MS lesions
were filled with a mean normal-appearing WM intensity estimated during segmentation.
FreeSurfer was applied to the lesion-filled MRIs without manual intervention. To reduce
analysis time, this step was performed in place of correcting the lesions after cortical
reconstruction. The results were visually inspected for errors related to the lesions. A
detailed CTh difference map was generated in MNI152 space by nonlinear registration to
visualize regional differences in CTh measured by CLADA and FreeSurfer.

CLADA CTh measurements were correlated with FreeSurfer CTh measurements,
both globally (r=0.73, p=0.04) and regionally (Figure 4-24). The global mean absolute
difference between CLADA and FreeSurfer CTh was 0.37 mm. Regional CTh
differences were slightly higher in temporal lobes (0.42 mm) than in other lobes (frontal=

4 http://surfer.nmr.mgh.harvard.edu/fswiki/VolumeRoiCorticalThickness
0.37 mm, parietal=0.35 mm, and occipital=0.31 mm). The detailed surface error map showed that differences were generally less than 0.5 mm, but the temporal and frontal lobes showed larger differences. Overall, CLADA CTh measurements were higher than FreeSurfer CTh measurements. In this cross-validation test, accuracy could not be determined, but the high correlation and relatively small difference showed that CLADA was in agreement with a previously validated method.

**Figure 4-24**: Cortical thickness measured using CLADA and FreeSurfer in seven SPMS patients (a). The cortical thickness measurements were correlated ($r=0.70, p<0.04$) and
4.3.10 Bias in Estimating Atrophy

To further study the bias in estimating atrophy, the method described by (Yushkevich et al., 2010) was applied to the data from the Reproducibility Test (Section 4.3.3). This approach analyzed scan-rescan data as if the data were from an actual longitudinal study where for each subject, the repeated scans are randomly assigned to be either “baseline” or “follow-up.” Changes in global and regional CTh were statistically compared to truth (no change) using t-tests.

The estimation of bias using the scan-rescan datasets was not significantly different from zero. For the high-resolution OASIS images, the p-values were 0.657, 0.673, 0.354, 0.684, and 0.672 for global, frontal, temporal, parietal, and occipital CTh, and for the low-resolution scan-rescan images the p-values were 0.383, 0.112, 0.983, 0.758, and 0.633 for global, frontal, temporal, parietal, and occipital CTh, respectively. The results showed that the bias in estimating atrophy by CLADA was not statistically significant.

4.3.11 Comparison to a Cross-sectional Method

Next, the full version of CLADA was compared to a cross-sectional version of CLADA where images from individual time-points were analyzed independent of each other (i.e., no intra-subject registration and no averaging) in terms of variability errors, similar to the study by Han and colleagues (Han et al., 2006). In our study, global and
regional (frontal, parietal, temporal, and occipital) CTh were measured from 3 scan-rescan studies: (1) high-resolution, high-contrast (4 signal average) MPRAGE from OASIS database, as in “Scan-rescan Reproducibility (4.3.3),” (2) high-resolution MPRAGE (NSA = 1) from “Comparison of Image Sequence (4.3.7),” acquired within a month (baseline, week 2, and week 4), (3) low-resolution T1SE from “Scan-rescan Reproducibility (4.3.3),” also acquired within a month (baseline, week 2, and week 4). The variability error was defined as mean absolute deviation (\( \Sigma |CTh_n - \mu_{CTh}|/N \) where \( n=1,2,\cdots,N \), \( N \) was the number of images, and \( \mu_{CTh} \) was the average CTh from different sessions) and the percent error was calculated as [mean absolute deviation] / \( \mu_{CTh} \times 100\% \). The t-tests were used to compare the percent error between CLADA and cross-sectional method.

When the variability error in global CTh from dataset 1 (high-resolution, high-contrast MRI) was compared between cross-sectional version of CLADA and full CLADA, there was no significant improvement (0.25% vs. 0.27%; \( p = 0.84 \); difference of 0.02% in Table 4-1: Column = Global and Row = Dataset 1; and Figure 4-25: green circles). However, CLADA out-performed when lower quality MRIs are used (Table 4-1). Regionally, variability errors from full CLADA were consistently smaller than those from the cross-sectional method.

These results indicated that the longitudinal CLADA benefited from its longitudinal processing and that this benefit would be greater for MRIs with lower signal-to-noise ratio or lower resolution.

Table 4-1: Results of comparison between longitudinal and cross-sectional methods. The difference in the variability error is tabulated here; thus larger numbers indicate more
improvements for the longitudinal method. Note that there is no negative number indicating that longitudinal method consistently out-performed the cross-sectional method.

<table>
<thead>
<tr>
<th></th>
<th>Global</th>
<th>Frontal</th>
<th>Parietal</th>
<th>Temporal</th>
<th>Occipital</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data 1</td>
<td>0.02%</td>
<td>0.13%</td>
<td>0.39%</td>
<td>0.15%</td>
<td>0.49%</td>
</tr>
<tr>
<td>Data 2</td>
<td>0.80%</td>
<td>0.88%</td>
<td>1.02%</td>
<td>0.80%</td>
<td>1.49%</td>
</tr>
<tr>
<td>Date 3</td>
<td>1.85%</td>
<td>2.18%</td>
<td>2.13%</td>
<td>1.80%</td>
<td>3.18%</td>
</tr>
</tbody>
</table>

**Figure 4-25:** The scan-rescan variability errors from the full longitudinal CLADA and cross-sectional version of CLADA. Points below the black solid line represent data where full longitudinal CLADA was better. Green circle = high-resolution, high-contrast MPRAGE (dataset 1). Blue triangle = high-resolution MPRAGE (dataset 2). Red square: low-resolution T1SE (dataset 3).

4.3.12 Effects of MS Lesions

Like the systematic effect of MS lesions (T2LV) on GMF (Section 2.3.5), the effect of MS lesions on CTh was investigated. Nine MS patients who had varying levels of T2LV were selected. The enlarging lesions were simulated by dilating the T2 lesion
masks by 1mm-, 2mm-, and 3mm-spherical kernels. CTh was measured using CLADA with original or dilated lesion masks.

The variability in baseline CTh was measured by the coefficient of variations. The effect of MS lesions on cortical thinning was determined by plotting the mean absolute difference in annualized rate of cortical thinning. The coefficient of variation was not measured for cortical thinning because the atrophy rate could be zero where the coefficient of variation became undefined.

The effect of MS lesions on the CTh measurements showed the average coefficient of variation (s.d.) of 0.29% (0.21%). The linear regression of CTh and T2LV showed a slope of 0.0007 mm/ml. The effect on the annualized rate of cortical thinning showed that the variation was very small (mean s.d. = 0.003; Figure 4-26). Therefore, there was no significant effect of lesions on the measurement of CTh and change in CTh.
Figure 4-26: The effect of MS lesions on the annualized rates of cortical thinning. This plot shows that the variation due to the enlarging MS lesions was very small: averaged variance was 7.50e-7. An increase in T2LV is indicated by lighter plot colors.

4.4 Discussion

This chapter described an efficient new method for measurement of cortical atrophy in MRIs acquired for longitudinal studies. Existing software packages for CTh measurements had been intended for analysis of high-contrast images with approximately 1 mm³ voxels. A unique feature of CLADA was that it was designed specifically to measure CTh change from lower-resolution 2D images with 3-5 mm slice thickness. The deformable model incorporated relaxation on self-intersection regions to account for WM segmentation errors, which were common in lower-resolution images with significant partial volume effects. Our tests showed that the algorithm was still successful in detecting cortical surfaces and estimating CTh change in 2D images with 3-5 mm slice thickness. However, as expected, the errors from lower-resolution images were higher than those obtained with high-resolution 3D MPRAGE images. These results suggested that although studies using older, lower-resolution MRIs might require larger sample sizes or longer duration, CLADA was applicable for analysis of retrospective datasets, such as those typically acquired for MS clinical trials.

CLADA used explicit rather than implicit surfaces for the cortical model because explicit representations seemed better suited to model highly convoluted surfaces (Zeng et al., 1999) with smoothness constraints (Acosta et al., 2009). An important feature of our deformable cortical surface detection algorithm was the incorporation of remeshing and self-intersection tests. These components led to an efficient multi-resolution shrink-
wrap algorithm that maintained surface topology. Topological inconsistency is a common problem in cortical surface detection algorithms initialized from a WM mask (Fischl et al., 2001; Han et al., 2006). The multi-resolution remeshing approach achieved highly complex shapes while still maintaining topology. These functions were important because deformable models require good initialization to avoid local minima. The savings in computation time were substantial compared to equivalent methods using explicit surfaces. Measurement of CTh with CLADA takes approximately 15 hours for a single subject with six time-points, whereas similar methods, such as FreeSurfer or ASP, require approximately 20-30 hours for a single image at a single time-point (Dale et al., 1999; MacDonald et al., 2000).

The strength of CLADA was the longitudinal nature of the algorithm, as demonstrated by the high reproducibility. CLADA used images from all time-points to create an average image with high SNR and to estimate the cortical surface by creating a subject-specific cortical surface model. This model was then deformed to fit the patient’s images from individual time-points. User interaction was kept to a minimum – only the mask for the subject’s average image was manually edited, if necessary. Thus, CLADA took less time to analyze follow-up images once the initial model has been created.

4.4.1 Validation Tests

Various validation tests were performed to quantitatively evaluate our new algorithm in terms of reproducibility, accuracy, sensitivity, performance with different types of MRIs, effect of reference space, comparison with FreeSurfer, bias in estimating atrophy, comparison with a cross-sectional method, and effect of enlarging MS lesions.
Reproducibility

The reproducibility test results demonstrated a relatively low scan-rescan error (<1%) in all regional CTh measurements. This was expected given that CLADA is a longitudinal algorithm, which incorporates images over all time-points. The intra-operator variability error was in the order of scan-rescan reproducibility, suggesting that the effect of manual editing is very small. The intra-operator variability in longitudinal change was consistent and small (up to 0.007 mm/yr in temporal lobe).

Accuracy test

Accuracy validation through comparison with post-mortem brain tissue is a very difficult task (Das et al., 2009). The majority of earlier reports on CTh measurements from autopsied brains did not account for three dimensions (Mann, 1991; Wegner et al., 2006; Wiley et al., 1991). Rosas et al. (2002) used autopsied brain tissue to validate the accuracy of MRI CTh measurements, but restricted measurements to the crowns of gyri where the curvature is minimal. To compare 3D MRI-based CTh measurements to 2D tissue-based CTh, careful co-registration of image slices to tissue slices and selection of valid ROIs from MRI were used to ensure that the 3D thickness vectors were in the plane of the corresponding tissue slices. Otherwise the measurements would not be expected to be comparable. These steps allowed the comparison of measurements even in curved regions and deep sulci.

The results demonstrated that CTh as measured by CLADA was correlated to CTh measured in fixed tissue at the same locations. The slope indicated that MRI-based
measurements were slightly higher than those in the fixed tissue, likely due to fixation effects. The absolute error of 0.43 mm was fairly low, considering the voxel size of 1.2 mm$^3$ and complex effects of fixation. The CLADA measurements were higher than photographs only by $0.17 \pm 0.54$ mm on average (median = 0.09 mm, mode = 0.0 mm with bin size = 0.1 mm). While the positive mean CTh offset appeared to show that the CLADA CTh might be consistently higher, the fact that the median offset was 0.09 mm with a 0.0 mm-mode relative to the photograph-derived CTh showed that a few data points with large errors might have skewed the means.

Indeed several points had large positive errors due to the inclusion of dura mater: four ROIs with the largest measurements errors included dura mater in CTh measured by CLADA. The dura mater is difficult to remove with a single MPRAGE image (van der Kouwe et al., 2008), and additional imaging modalities may be helpful in future applications of CLADA.

Other complex factors might have affected the results of this validation test. First, there was an attempt to minimize fixation effects by using affine registration but it might not be entirely removed because CLADA-based CTh was larger. In addition to tissue shrinkage, the fixation effect becomes nonlinear in the focal abnormal areas, as more damaged tissues tended to deform more during the fixation. Furthermore, the fixed brain hemispheres tended to be flattened as the midsaggital plane was placed at the bottom of the container of fixative. Second, two different types of CTh measurements were compared: the distance between corresponding points with CLADA in MRI, and the Laplace thickness in photographs. CLADA thickness was defined by straight lines while the Laplace thickness might curve, which may have artificially decreased the error at the
curved areas (Figure 4-1) (Jones et al., 2000). There are various ways of defining the corresponding points between the inner and outer cortical surfaces for calculation of thickness as discussed in Section 4.1.1. The Laplace method was utilized here because it had unique surface correspondence for CTh measurements even in regions of high curvature. Another important limitation is the accuracy of the gold standard. While CTh was carefully and accurately measured using high-resolution photographs (pixel size < 0.1 mm), factors like fixation and unclear tissue edges can affect the accuracy of the gold standard.

**Sensitivity Test**

The sensitivity test showed that CLADA was able to detect sub-voxel level change. The detected change tended to be slightly less than the applied deformation partly due to the surface and CTh smoothness terms incorporated into CLADA. There was also a trend for larger errors with greater deformation. This observation may be related to an artifact of the applied deformation, which may be misaligned with the cortical surface normal, and the absolute errors increase proportionally with the deformation field. As demonstrated by the comparison of measured to applied deformations, CLADA appeared to have lower sensitivity for positive change (increases in CTh) than for negative change (cortical thinning), with slopes of 0.67 and 0.78, respectively. This difference was primarily due to the limited range of motion for applied deformations in the positive direction in cortical regions with small sulci, which are common in these young normal controls. Measurements of positive deformations selectively applied to gyri surrounded by larger sulci confirmed this explanation. In these
restricted post hoc analyses, the positive deformations had similar sensitivity as negative
deformations (slope = 0.82). Finally, part of the deviation from ideal slope (1.0) may be
attributed to bias introduced by interpolation. In this test, the warped image was
interpolated but the original image was not, and biased interpolation has been shown to
lower sensitivity (Yushkevich et al., 2010). Our simulation method avoided deep sulci
and used crowns of gyri to properly simulate atrophy or growth with simple image
warping. Rather than to estimate accuracy, the main goal of these simulations was to
determine if CLADA was sufficiently sensitive to detect very small changes in CTh. If
necessary, more sophisticated methods for simulating global or regional atrophy using
finite element or topology preserving models (Camara et al., 2006; Karacali and
Davatzikos, 2006) could be implemented for future studies.

**Comparison of Image Sequences**

The image sequence comparison showed that CTh was (on average) 0.8 mm
larger when measured in MRIs with $1 \times 1 \times 3$ mm$^3$ voxels as compared to MRIs with ~1
mm$^3$ voxels. Most likely, blurred tissue boundaries due to partial volume effects led to
apparently thicker cortex. This comparison confirmed the importance of scanning
subjects with the same sequence in longitudinal studies.

There was surprisingly good correlation between CLADA measurements from
MPRAGE images and the lower-resolution T1SE images acquired in the same session,
with the exception of one outlier in the patient group. The outlier in Figure 4-22 appeared
to be related to significant motion artifacts evident in the T1SE images at both time-
points. The difference in the scanner field strength may have affected these results to
some extent. However, the objective of this test was to compare CLADA results from lower-resolution MRIs to that of higher-resolution, higher-contrast images in a longitudinal setting. These data support the use of CLADA to analyze retrospective MRI datasets.

**Comparison with FreeSurfer**

CLADA also showed good agreement with FreeSurfer, the most widely used software package for estimation of CTh. These comparisons showed that CLADA measurements are consistently higher than FreeSurfer by 0.36±0.14mm, on average (median = 0.34 mm, mode = 0.3 mm with bin size = 0.1 mm). This test is a cross-validation analysis to demonstrate comparability with a widely used method rather than a comparison to a gold standard. One possible interpretation is that the FreeSurfer measurements are accurate and CLADA has a consistent sub-voxel offset of approximately 0.3 mm. However, it may also be the case that neither method is completely accurate. Correlational agreement is informative in cross-validation comparisons where the absolute truth is not known (Smith et al., 2007).

**Bias in Estimating Atrophy**

The bias in estimating atrophy was evaluated in several ways. First the reference image was changed to investigate the registration-related bias. The choice of reference image did have a small effect on CTh measurements, despite our efforts such as image averaging with symmetric linear registration, cortical model construction from the averaged template, and transformation of cortical model rather than the image. The bias
estimation using scan-rescan in contrast to the clinical datasets was small, indicating that CLADA can be reliably applied to longitudinal CTh studies. Thus, construction of an unbiased template remains a difficult task (Guimond et al., 2000; Joshi et al., 2004; Reuter and Fischl, 2011). Comparison of CTh measurements using either baseline or the follow-up image as the reference space resulted in small differences, which may be the source of scan-rescan error. The longitudinal change analysis was less affected by the choice of reference space. The differences in CTh change were on the order of 0.001 mm/yr, which was smaller than the mean atrophy rates of −0.05 mm/yr in controls (after converting from percent change) (Chen et al., 2004), −0.0036 mm/yr in MS, −0.0021 mm/yr in controls (Sampat et al., 2010) in longitudinal studies; in cross-sectional studies, the estimated rates of cortical atrophy were −0.0016 mm/yr in controls (Salat et al., 2004), −0.0086 mm/yr in controls (Hutton et al., 2009), and −0.01 mm/yr in RRMS (Charil et al., 2007).

4.4.2 Conclusion

Overall, the results show that CLADA can be used for reliable quantitative measurement of CTh change in longitudinal MRI studies.

In MS, histopathological studies show that the cortex is demyelinated (Kidd et al., 1999) without evidence of inflammation (Peterson et al., 2001). Therefore, cortical atrophy most likely reflects severe cortical damage and neuronal loss rather than effects related to the resolution of edema or pseudoatrophy (Nakamura et al., 2010; Zivadinov et al., 2007).
There is also increasing MRI evidence of cortical damage. Cross-sectional and short-term longitudinal studies have found that MRI-detected CTh change is related to patient disability (Calabrese et al., 2007b; Charil et al., 2007; Chen et al., 2004) and cognitive impairment (Amato et al., 2004; De Stefano et al., 2003). The overall clinical relevance and predictive value of CTh changes for determination of disability progression over the longer term (>1 year) are yet to be determined, and the next chapter describes 4-year longitudinal study to address the clinical relevance of cortical thinning in MS.
CHAPTER 5  CORTICAL ATROPHY IN MS

This chapter describes a longitudinal clinical study of cortical atrophy in MS. The newly developed CTh measurement algorithm was applied to the same longitudinal data as in Chapter 3 to study the evolution of cortical thinning in CIS, RRMS, SPMS, and NC. Clinical correlations and predictors of cortical thinning were also investigated.

This chapter has been prepared for a manuscript submission to Annals of Neurology.

5.1 Materials and Methods

The subjects, MRI, and image analysis are described in Chapter 3. An important difference here is that CLADA was applied to the full longitudinal data. Thus, for each subject, CTh was measured at nine time-points for MS subjects and five time-points for NC. The annualized change in global cortical thickness ($\Delta$CTh) was estimated by the slope of the regression line between CTh and the interval time from baseline. Measurements for baseline CTh and $\Delta$CTh were compared between MS and NC as well as between subgroups (CIS, RRMS, and SPMS) using analysis of covariance (ANCOVA). Baseline CTh and $\Delta$CTh were also compared between the stable versus worsening patients using ANCOVA with age as a covariate.
We calculated Spearman’s rank correlation coefficients for the correlational analysis for CTh and \( \Delta \text{CTh} \) with imaging variables (BPF, GMF, WMF, T2LV, T1LV, GdLV, and MTR\text{NABT}) and clinical measures (EDSS, MSFC and three components of MSFC). We then used multiple regression analysis to determine the independent estimators of baseline CTh and \( \Delta \text{CTh} \) among the imaging variables (T2LV, T1LV, GdLV, MTR\text{NABT}, MTR\text{NAWM}, and MTR\text{Lesion}).

The sample size was estimated for 1-year proof-of-concept study to detect a 60% effect size with 80% power at a 5% significance level. UnifyPow, an SAS module for sample-size analysis was used. The results of measurements from RRMS, SPMS, and combined CIS+RRMS+SPMS were used to calculate the sample size.

## 5.2 Results

### 5.2.1 Baseline Characteristics

The baseline characteristics are shown in Table 5-1, and CTh difference map is shown in Figure 5-1: (a) RRMS minus NC, (b) SPMS minus NC, and (c) SPMS minus RRMS. CTh was significantly lower in the MS group compared to NC (4.19 ± 0.38 mm versus 4.52 ± 0.28 mm, \( p < 0.001 \)), and there was a trend for lower baseline CTh in patients with more advanced disease, where CTh was 4.44 ± 0.22 mm in CIS, 4.21 ± 0.35 mm in RRMS, and 3.96 ± 0.42 mm in SPMS subjects (Figure 5-2).
Figure 5-1: Baseline cortical thickness difference maps (a) RRMS minus controls (NC), (b) SPMS minus NC, and (c) SPMS minus RRMS. Blue indicates smaller thickness, and red show growth.

Table 5-1: Baseline characteristics.

<table>
<thead>
<tr>
<th></th>
<th>NC (n=17)</th>
<th>CIS (n=7)</th>
<th>RRMS (n=36)</th>
<th>SPSM (n=27)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD) Age, years</td>
<td>41.6 (8.1)</td>
<td>44.2 (10.9)</td>
<td>39.8 (8.4)</td>
<td>45.7 (7.8)</td>
</tr>
<tr>
<td>Mean (SD) disease duration, years</td>
<td>NA</td>
<td>0.25 (0.1)</td>
<td>5.7 (5.4)</td>
<td>17.0 (7.6)</td>
</tr>
<tr>
<td>Mean (SD) EDSS</td>
<td>NA</td>
<td>1.0 (0.76)</td>
<td>1.9 (1.5)</td>
<td>5.1 (1.5)</td>
</tr>
<tr>
<td>Mean (SD) MSFC</td>
<td>0.55 (0.27)</td>
<td>0.36 (0.35)</td>
<td>0.43 (0.58)</td>
<td>-0.82 (1.34)</td>
</tr>
<tr>
<td>Mean (SD) T2 lesion volume, ml</td>
<td>NA</td>
<td>2.5 (2.3)</td>
<td>18.1 (17.2)</td>
<td>42.7 (24.9)</td>
</tr>
<tr>
<td>Mean (SD) T1 lesion volume, ml</td>
<td>NA</td>
<td>0.15 (0.22)</td>
<td>1.72 (2.89)</td>
<td>8.3 (8.31)</td>
</tr>
<tr>
<td>Mean (SD) BPF</td>
<td>0.862 (0.012)</td>
<td>0.862 (0.008)</td>
<td>0.843 (0.026)</td>
<td>0.804 (0.04)</td>
</tr>
<tr>
<td>Mean (SD) GMF</td>
<td>0.554 (0.015)</td>
<td>0.553 (0.010)</td>
<td>0.541 (0.019)</td>
<td>0.520 (0.028)</td>
</tr>
<tr>
<td>Mean (SD) WMF</td>
<td>0.308 (0.011)</td>
<td>0.309 (0.009)</td>
<td>0.302 (0.017)</td>
<td>0.283 (0.016)</td>
</tr>
</tbody>
</table>

Abbreviations: NC = healthy normal control, CIS = clinically isolated syndrome, RRMS = relapsing-remitting MS, SPMS = secondary progressive MS, SD = standard deviation, EDSS = extended disability status scale, MSFC = MS functional composite, CTh = cortical thickness.
Figure 5-2: Boxplots showing the results of group comparisons of baseline cortical thickness. Boxes indicate the quartiles, vertical lines are the extreme values, and the thick bars are the medians. There was a trend for lower cortical thickness for more severe disease subtype: healthy normal controls (NC, 4.52 ± 0.28mm), clinically isolated syndrome (CIS, 4.44 ± 0.22 mm), relapsing-remitting (RR, 4.21 ± 0.35mm), and secondary progressive (SP, 3.96 ± 0.42mm) MS.

5.2.2 Cortical Thinning

The annualized changes (ΔCTh) were $-0.010 \pm 0.018$, $-0.013 \pm 0.012$, $-0.018 \pm 0.016$, and $-0.025 \pm 0.022$ mm/yr for NC, CIS, RRMS, and SPMS, respectively (Figure 5-3a). ΔCTh in SPMS subjects was significantly different from NC (p = 0.008) but other differences did not reach statistical significance. The longitudinal evolution of global cortical thinning for each subgroup is shown in Figure 5-3b.
Figure 5-3: (a) Boxplots showing the annualized change in global cortical thickness ($\Delta$CTh) for each compared group (NC: 0.010 ± 0.018mm/year, CIS: 0.013 ± 0.012mm/year, RRMS: 0.018 ± 0.016mm/year, SPMS: 0.025 ± 0.022mm/year). (b) The same data is plotted as a function of time (error bars are standard errors). Missing data is estimated by linear regression.

5.2.3 Clinical Correlation of Cortical Thickness

At baseline, there were moderate correlations between CTh and EDSS as well as between CTh and MSFC (Table 5-2). All components of MSFC were correlated with CTh and the strongest correlations were observed with 9HPT.
### Table 5-2: Spearman rank correlation coefficients

<table>
<thead>
<tr>
<th></th>
<th>Baseline Cortical Thickness</th>
<th>Cortical Thickness Change</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correlations with Demographic and Disease Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>−0.46 (p&lt;0.0001)</td>
<td>−0.10 (p=0.43)</td>
</tr>
<tr>
<td>MS Duration</td>
<td>−0.60 (p&lt;0.0001)</td>
<td>−0.21 (p=0.08)</td>
</tr>
<tr>
<td>EDSS</td>
<td>−0.58 (p&lt;0.0001)</td>
<td>−0.20 (p=0.1)</td>
</tr>
<tr>
<td>MSFC</td>
<td>0.54 (p&lt;0.0001)</td>
<td>0.20 (p=0.1)</td>
</tr>
<tr>
<td>25’ walk</td>
<td>−0.55 (p&lt;0.0001)</td>
<td>−0.11 (p=0.37)</td>
</tr>
<tr>
<td>9-hole peg test</td>
<td>0.62 (p&lt;0.0001)</td>
<td>0.17 (p=0.16)</td>
</tr>
<tr>
<td>PASAT</td>
<td>0.30 (p=0.01)</td>
<td>0.12 (p=0.33)</td>
</tr>
<tr>
<td><strong>Correlations with MRI Measures:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPF</td>
<td>0.68 (p&lt;0.0001)</td>
<td>0.22 (p=0.07)</td>
</tr>
<tr>
<td>GMF</td>
<td>0.55 (p&lt;0.0001)</td>
<td>0.12 (p=0.33)</td>
</tr>
<tr>
<td>WMF</td>
<td>0.59 (p&lt;0.0001)</td>
<td>0.22 (p=0.07)</td>
</tr>
<tr>
<td>T2LV</td>
<td>−0.62 (p&lt;0.0001)</td>
<td>−0.38 (p=0.001)</td>
</tr>
<tr>
<td>T1LV</td>
<td>−0.60 (p&lt;0.0001)</td>
<td>−0.28 (p=0.02)</td>
</tr>
<tr>
<td>GdLV</td>
<td>0.15 (p=0.2)</td>
<td>−0.34 (p=0.004)</td>
</tr>
<tr>
<td>NABT MTR</td>
<td>0.46 (p&lt;0.0001)</td>
<td>0.29 (p=0.02)</td>
</tr>
</tbody>
</table>

Abbreviations: EDSS = Kurtzke Expanded Disability Status Scale; MSFC = MS Functional Composite; PASAT = Paced Auditory Serial Addition Test; BPF = Brain Parenchymal Fraction; GMF = Gray Matter Fraction; WMF = White Matter Fraction; T2LV = T2 Lesion Volume; T1LV = T1 Lesion Volume; GdLV = Gadolinium Lesion Volume; NABT = Normal-Appearing Brain Tissue; MTR = Magnetization Transfer Ratio

The baseline CTh correlated with other concurrent atrophy measures (BPF, GMF and WMF). Baseline CTh also correlated with baseline T1LV and T2LV but not with GdLV. The correlations between baseline MTR and baseline CTh were also moderate. ∆CTh was related to whole brain and GM atrophy rates (r = 0.48, p < 0.0001 and r = 0.38, p = 0.001, respectively) and cumulative GdLV (r = 0.36, p = 0.003).

#### 5.2.4 Disability Progression vs. Thickness and Change

When patients were classified into EDSS-progressed and stable groups, there were trends for lower CTh and more atrophy in patients who progressed. For baseline CTh, the means were 4.17 ± 0.35 mm vs. 4.20 ± 0.41 mm (p=0.78), and ∆CTh means were −0.023 ± 0.021 mm/yr vs. −0.017 ± 0.014 mm/yr (p=0.18) for progressed and stable
groups, respectively. However, when progression was defined by MSFC, compared to stable patients, the patients who progressed had significantly lower baseline CTh (4.07 ± 0.41 mm vs. 4.29 ± 0.33, p=0.02) and more ΔCTh (−0.027 ± 0.019 mm/year vs. −0.013 ± 0.013 mm/year, p=0.001). The baseline characteristics for these subgroups are shown in Table 5-3.

Table 5-3: Baseline patient characteristics based on MSFC progression.

<table>
<thead>
<tr>
<th></th>
<th>MSFC Progressed (n=31)</th>
<th>MSFC Stable (n=39)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD) age, years</td>
<td>44.5 (9.7)</td>
<td>41.8 (8.7)</td>
<td>0.23</td>
</tr>
<tr>
<td>Mean (SD) disease duration, years</td>
<td>14.1 (9.4)</td>
<td>5.9 (6.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mean (SD) EDSS</td>
<td>4.2 (2.1)</td>
<td>2.2 (1.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mean (SD) MSFC</td>
<td>−0.47 (1.3)</td>
<td>0.26 (0.8)</td>
<td>0.009</td>
</tr>
<tr>
<td>Mean (SD) T2 lesion volume, ml</td>
<td>34.0 (25.0)</td>
<td>19.7 (21.6)</td>
<td>0.014</td>
</tr>
<tr>
<td>Mean (SD) T1 lesion volume, ml</td>
<td>5.7 (7.9)</td>
<td>2.9 (5.0)</td>
<td>0.10</td>
</tr>
<tr>
<td>Mean (SD) BPF</td>
<td>0.816 (0.042)</td>
<td>0.841 (0.029)</td>
<td>0.008</td>
</tr>
<tr>
<td>Mean (SD) GMF</td>
<td>0.528 (0.030)</td>
<td>0.539 (0.019)</td>
<td>0.083</td>
</tr>
<tr>
<td>Mean (SD) WMF</td>
<td>0.288 (0.017)</td>
<td>0.302 (0.018)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

For abbreviations, see Table 5-1. * p < 0.05, ** p <0.01

5.2.5 Predictors of Cortical Thickness and Thinning

Multivariate analysis demonstrated that age (standard regression coefficient, $\beta = 0.31$, $p = 0.001$), baseline T2LV ($\beta = 0.49$, $p < 0.001$), and baseline MTR$_{NABT}$ ($\beta = 0.23$, $p = 0.03$) were independent correlates of baseline CTh. The overall adjusted $R^2$ for this model was 0.57. Predictors for ΔCTh were baseline T2LV ($\beta = 0.36$, $p = 0.002$) and cumulative GdLV ($\beta = 0.32$, $p = 0.006$) with an overall model $R^2$ was 0.24.
5.2.6 Sample Size Estimate

Sample size estimates for a one-year study are shown in Table 5-4. For a trial in SPMS, 92 patients per arm would be required to detect a 60% effect size with 80% power at a 5% significance level. For a trial in RRMS, 233 patients per arm would be required.

<table>
<thead>
<tr>
<th></th>
<th>80% Power</th>
<th>90% Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>RRMS+SPMS+CIS</td>
<td>166</td>
<td>221</td>
</tr>
<tr>
<td>RRMS only</td>
<td>233</td>
<td>312</td>
</tr>
<tr>
<td>SPMS only</td>
<td>92</td>
<td>100</td>
</tr>
</tbody>
</table>

5.3 Discussion

This chapter characterized cortical thinning in patients with CIS, RRMS, and SPMS. There are several important findings from this study. First, our baseline results confirmed previous cross-sectional studies that the cortex was thinner at baseline in MS patients compared to NC (Sailer et al., 2003) and thinner in more advanced patients (Calabrese et al., 2007b). The baseline CTh correlated with lesion load as well as with all measures of disability (EDSS, MSFC, 25FTW, 9HPT, and PASAT), further emphasizing that CTh is a clinically relevant marker of MS disease progression. Surprisingly, the correlation between MSFC and CTh was driven more strongly by 25FTW and 9HPT and least by PASAT. In previous studies, correlations between PASAT and measures of global GM atrophy were higher than for cortical atrophy (Fisher et al., 2008)(Amato et al., 2004; Benedict et al., 2006; Morgen et al., 2006). Benedict et al. reported the variance in PASAT was best described by normalized GM volume and to a lesser degree by
normalized cortical GM volume (Benedict et al., 2006), thus, the PASAT may be more closely associated with tissue loss in deep GM structures. Second, the longitudinal analysis suggested there was greater cortical thinning in MS patients as compared to age-matched NC and greater cortical atrophy as MS stage advanced. This finding was not observed in an earlier report on longitudinal CTh in MS (Chen et al., 2004), possibly due to the smaller sample size and shorter duration in that study. The result is in concordance with our previous report on longitudinal GM atrophy, where GMF change was found to be 8-fold higher in RRMS and 14-fold higher in SPMS compared to NC (Chapter 3) (Fisher et al., 2008). Third, when we examined the MRI predictors of CTh and cortical thinning (ΔCTh), we found that both focal and diffuse damage (T2LV and MTRNABT) were independently associated with baseline CTh, accounting for 57% of the variance. However, only focal lesion measures (T2LV and cumulative GdLV) predicted subsequent cortical thinning, accounting for only 24% of the variance. Finally, we determined that measurement of cortical thinning would be feasible as an outcome measure in MS clinical trials, particularly for proof-of-concept studies in SPMS.

The predictors of CTh and cortical thinning suggested that cortical atrophy might be the end result of a variety of brain insults: physiologic (age), focal (T2LV) and diffuse (MTRNABT) damage. However, a large amount of the variance was not accounted. Additional pathologic processes that conceivably contribute to cortical thinning include cortical demyelination, synaptic loss, and glial loss (Wegner et al., 2006). However, cMRI cannot detect such damage. The relationship between focal WM damage and cortical pathology and its dependence on the phase of MS have been studied within both the histopathological (Frischer et al., 2009; Kutzelnigg et al., 2005) and imaging
communities (Bendfeldt et al., 2009; Calabrese et al., 2007a; Ceccarelli et al., 2008; Charil et al., 2007; Sailer et al., 2003; Sepulcre et al., 2006). Our evidence showed that both focal (T2LV, \( \beta = 0.50 \)) and diffuse damage (MTR\textsubscript{NABT}, \( \beta = 0.23 \)) predicted the baseline CTh, but only focal damage (T2LV, \( \beta = 0.33 \), and cumulative GdLV, \( \beta = 0.31 \)) remained in the model of \( \Delta \text{CTh} \), suggesting, at least in our MS patients, that focal WM lesions played a significant role in the development of cortical atrophy.

Another question concerned the relationship between volumetric GM atrophy measures and CTh measures (Winkler et al., 2009), and which approach might be more sensitive and robust for detection of irreversible neuronal loss and ultimately MS clinical progression. There have been many previous studies investigating the clinical and biological relevance of GMF. GMF has been shown to be correlated with lesion volumes, 

\citep{Chard2002a, Dalton2004, Fisher2008, Fisniku2008, Ge2001a, Quarantelli2003, Sanfilipo2005}, disability scores \citep{Chard2002a, Chard2004, Ge2001a, Sanfilipo2005, Tiberio2005, DeStefano2003}, and cognitive impairment \citep{Amato2004, Amato2007, Benedict2006, DeStefano2003}. The strength of the correlations varied with study population and length of follow-up, but generally GMF correlated more strongly with clinical measures than other conventional MRI measurements. CTh was also correlated to disability \citep{Charil2007, Sailer2003} and cognitive impairment \citep{Calabrese2010c}. Interestingly CTh appeared to have regional specificity: thinner primary motor cortex was found in patients with pyramidal onset, and thinner visual cortex was found in patients with optical neuritis onset when compared to patients with other types of symptoms at first presentation \citep{Calabrese2007}. Fatigue correlated to CTh in
parietal lobes (Pellicano et al., 2010). These results suggested that cortical atrophy was significantly involved in MS and was clinically relevant.

In this study, a moderate correlation was found between baseline CTh and concurrent GMF ($r = 0.55$, $p < 0.001$). This weaker than expected relationship might be attributable to deep GM and cerebellar atrophy, which were known to be significant in MS (Audoin et al., 2006; Cifelli et al., 2002; Ramasamy et al., 2009) but not included in CTh measurements. The correlation between annualized GMF percent change during the study and annualized change in global CTh was low ($r = 0.38$, $p = 0.001$), and might indicate variability in regional GM atrophy (Audoin et al., 2006). Thus, cortical and deep GM atrophy might evolve differently and further longitudinal studies for each GM component as well as regional cortical atrophy are needed to determine the specific relationship and clinical deficit.

Some important technical issues related to the study are worth mentioning. First, we measured CTh from T1-weighted spin echo images (T1SE) with 5mm slice thickness, which are typical images acquired in prior MS clinical trials, but not ideal for estimating CTh. These are the images available from the standardized MRI protocol used for our long-term observational study of atrophy in MS, which began in 1999. Most of the previous CTh studies in MS used FreeSurfer, but it requires high-resolution, high-contrast MRI and is not applicable to T1SE image. The analysis software, CLADA, was designed to reliably measure $\Delta$CTh from these images, but higher-resolution and higher-contrast 3D images can potentially improve the sensitivity. Despite this limitation, CTh differences were detected between MS and control subjects in the present study. Several factors contributed to the detection of cortical thinning: (1) the longitudinal nature of
CLADA which uses registration and deformable models and was sensitive to subtle changes and (2) the use of multiple time-points (9 time-points during 4-year semi-annual MRI in MS and 5 time-points in controls) to calculate the linear regressions instead of a simple percent difference from 2 time-points. Second, there was a potential influence of MS lesions on CTh measurements. Most MS lesions were periventricular and did not significantly influence the cortical surface detection except during segmentation. In CLADA, a lesion mask was created on the subject-specific average image to eliminate systematic segmentation errors. However, juxtacortical lesions might still affect CTh measurements, and the use of T1SE was advantageous over MPRAGE because juxtacortical lesions are rarely visible on T1SE whereas MPRAGE is sensitive to these lesions. In other surface-based cortical surface detection algorithms, these hypointense juxtacortical lesions might require manual correction and result in higher variability because of inconsistent mesh constructions. Intracortical lesions are not detectable in T1SE images and rarely visible even with MPRAGE (Nelson et al., 2008). Some intracortical lesions may be visible (Bagnato et al., 2009) but most of the cortical lesions cannot be detected by cMRI, even by advanced MRI methods such as double inversion recovery (Geurts et al., 2005a) and phase-sensitive inversion recovery (Nelson et al., 2007). Thus, cortical lesions were unlikely to affect CTh measurements in our study.

Our results indicate that cortical atrophy is clinically relevant in MS, and therefore, cortical atrophy is a potential marker of MS progression. However, the source of cortical GM tissue loss remains unknown, and such tissue loss may be caused by a combination of both WM and GM damage. Further studies are needed to investigate the effect of cortical and WM lesions on cortical thinning. The lower degree of inflammation
in the cortex and the likely association between CTh and neuronal density make CTh a potentially valuable outcome measure for MS clinical trials testing neuroprotection.
CHAPTER 6  SUMMARY AND FUTURE WORK

6.1 Summary

In this work, new methods to measure GMF and CTh from brain MRI were
developed and validated. Using these techniques and longitudinally acquired MRI of MS
patients and healthy normal controls, we determined the rates of GM and cortical atrophy.
The rates of GM and cortical atrophy in MS patients were greater than those in healthy
controls, correlated with clinical disability, and were different among the patients who
worsened and remained stable. The main results and conclusion are summarized here,
followed by possible future direction.

In Chapter 2, a new automated GM segmentation method to measure normalized
GM volume (GMF) was developed and validated. The method was designed to be
applicable for brain MRI of MS patients where typical MS features such as focal lesions
and significant brain atrophy were taken into account. We validated this method in terms
of reproducibility, accuracy, and sensitivity to lesion size. The average scan-rescan
reproducibility using T1SE images from 9 MS patients, acquired 3 times over 2 weeks,
was 1.1% for GMF. The accuracy, as measured by the similarity index, was 0.938 and
0.836 in comparison to gold standard in simulated MRIs and in comparison to manual
tracing in real MRIs, respectively. The accuracy was better than commonly used software,
particularly when analyzing low-resolution clinical MRIs. We determined that the
method could be reliably used for clinical MRIs of MS patients and for longitudinal study of duration 2 years or longer. We also discovered that the size of lesions systematically affected GM volume, and this bias could be statistically corrected.

In Chapter 3, we applied the segmentation algorithm to a 4-year longitudinal study of MS patients and age-matched healthy normal controls and measured GM atrophy. GM atrophy was calculated by measuring GMF at baseline and at year 4, subtracting the two measurements, and dividing by the interval time (4 years). There was a stage-dependent acceleration in the annualized rates of GM atrophy in MS patients while WM atrophy remained constant among MS patients. There were moderately strong correlations between GM atrophy and clinical measures of disability. The rates of GM atrophy differed when the MS group was subdivided into those who worsened and those remained stable. These findings demonstrated the clinical relevance of GM atrophy in MS. We also found that predictors of GM atrophy differed between RRMS and SPMS groups, which suggested different mechanisms of neurodegeneration in RRMS and SPMS.

In Chapter 4, we developed and validated CLADA, which measured cortical atrophy in terms of change in CTh. This algorithm was a registration and deformable model-based longitudinal method. CLADA used images from all time-points to create a subject-specific template image, generate a deformable cortical surface model, and deform the model to fit each image from different time-points. The validation tests for CLADA included various reproducibility measures, accuracy, sensitivity to change, comparison of image sequence, bias in estimating atrophy, comparison of full and cross-sectional versions of CLADA, and systematic effect of WM lesions. The validation tests
showed that CLADA was reliable, more reproducible than cross-sectional method, could measure CTh change from low-resolution MRI, and had small bias in estimating atrophy.

In Chapter 5, CLADA was applied to the full longitudinal data from Chapter 3 (nine MRIs acquired semi-annually for patients and five MRIs acquired annually for controls) to study cortical atrophy in MS. The results showed that CLADA detected cortical atrophy in MS patients. Similar to GM atrophy measured by GMF, cortical atrophy was greater in patients with more advanced disease and correlated with clinical disability. Cortical atrophy distinguished stable and worsening patients with more significance than GM atrophy and was predicted by baseline T2 lesion volume and cumulative Gd-enhancing lesion volume in the MS group.

Overall this work demonstrated the clinical importance of GM and cortical pathology and showed that GM and cortical atrophy could be used as a reliable imaging marker of clinical progression, in particular for assessing potential neuroprotective therapies in MS. It also suggested that mechanisms of RRMS and SPMS differed which might imply different therapeutic strategies are necessary for neuroprotection in different MS stages.

6.2 Limitations

The limitations of each study were described in the individual chapters, and here we summarize the limitations that affected all studies.

- The GM lesions were not studied in this project even though these lesions were believed to be relevant (Calabrese et al., 2007a; Kutzelnigg and Lassmann, 2005). Current conventional MRI techniques are insensitive to cortical lesions (Geurts et al.,
2005b; Kidd et al., 1999; Newcombe et al., 1991), and the use of new MRI techniques such as double inversion recovery (DIR) or phase sensitive inversion recovery (PSIR) might be advantageous in investigating the relationship between GM damage and clinical correlation. However, DIR could only detect about 30% of cortical lesions compared to the histopathological standard despite its improved sensitivity (Seewann et al., 2010). Furthermore, there were no standard pulse sequences to acquire DIR, and the inter-rater variability is high (Geurts et al., 2011).

- The studied patients were allowed to be on DMTs over the course of the longitudinal study, which might have reduced the rate of atrophy. Patients could also be switching drugs, which further complicated the interpretation of our results. However, the medication is recorded and considered in the analysis. Furthermore, the use of untreated patients would have ethical concerns and potential bias as untreated patients might have the benign course due to their genetic disposition. The use of corticosteroids also affected MRI measurements, and thus the patient visits were delayed to minimize the possible artifact.

- During the course of the study, the scanner was changed from Siemens Vision to Siemens Symphony. We acquired images from 9 MS patients with varying degrees of atrophy and lesions from both scanners on the same day and used them for calibrating the measurements. The calibration involved a linear regression of GMF and CTh measurements and CLADA weighting parameter ($\alpha = 1.2$). The consistency of longitudinal data was also monitored by the measurements from healthy normal controls as seen in Figure 3-1 and Figure 5-3(b).
6.3 Future Work

There are numerous potential studies that can build on the methods developed here. We provide five examples of such potential studies. Two of them (6.3.2 and 6.3.3) have already started with promising results (Fisher et al., 2010; Nakamura et al., 2010).

6.3.1 Higher Field MRI

Current clinical study used images from 1.5T MRI scanners. Higher field MRIs including 3T and 7T have become increasingly accessible, and the images from high field scanners can provide smaller voxel size, shorter scanning time, and/or higher signal-to-noise ratio (SNR). Segmentation algorithms benefit from high field MRI because (1) high resolution MRI reduces partial volume effects, (2) shorter scanning time reduces the risk of patient motion, and (3) high SNR clarifies the borders between different tissues. Some disadvantages also exist: (1) there are greater susceptibility artifacts, (2) geometric distortions also increase due to both object-induced susceptibility and non-uniform main magnetic field, (3) the specific absorption rate (SAR) is increased and may prevent using long sequences with numerous RF pulses, and (4) variations in phase cause signal decrease (signal drop-off) and increase (hot spot). Therefore, the developed methods will likely benefit from higher field MRI but may need to be updated with advancing MRI technologies.

6.3.2 Treatment Effect on Gray Matter Atrophy

This future project investigates the treatment effect on GM and cortical atrophy. Although it is well established that whole brain atrophy responds to therapies (Barkhof et
al., 2009), the treatment effect on GM atrophy is not well established (Bendfeldt et al., 2010; Zivadinov et al., 2007). The response to treatment is essential for GM atrophy to be a surrogate marker and to be applied in clinical trials. The image analysis software developed in Chapter 2 and Chapter 4 can be directly used in retrospective analysis of previous longitudinal clinical trials. If the treatment effect is confirmed, GM atrophy measures can be used to evaluate future therapeutic candidates.

6.3.3 Investigation on Pseudoatrophy Effect

Another potential application of GM atrophy is to study the pseudoatrophy effect. It is believed that the resolution of inflammatory edema causes apparent whole brain atrophy in the treated patients during the first year of anti-inflammatory therapy and that reduced whole brain atrophy in the second year reflects the treatment effect (Section 1.5). Since inflammation is absent or significantly reduced in GM (Kidd et al., 1999; Peterson et al., 2001), the observed acceleration of whole brain atrophy is likely caused by significant WM atrophy in the first year, and the observed treatment effect on whole brain atrophy is likely due to reduced GM atrophy in the treated patients in the second year. The tools we developed here can be used to test this hypothesis. This longitudinal study will be similar to the first future study but require at least one intermediate time-point to study the temporal evolution of whole brain, GM, and WM atrophy. If no pseudoatrophy effect is found in GM after anti-inflammatory therapies, GM atrophy becomes even more attractive marker of disease progression in MS because it implies that GM or cortical atrophy are more specific to irreversible damage than whole brain atrophy.
6.3.4 Regional Correlation of GM Atrophy and WM Damage

Another potential future work is to study the relationship between GM atrophy and WM damage. The relationship between GM and WM damage is uncertain and warrants further studies. In the case of secondary neurodegeneration, initial WM damage causes axonal transection, which leads to Wallerian, retrograde, and trans-synaptic degeneration, and ultimately GM atrophy. This hypothesis can be supported by a high correlation between GM atrophy and WM damage. Previous studies have been inconclusive using voxel-based morphometry methods (Battaglini et al., 2009; Bendfeldt et al., 2009; Bendfeldt et al., 2009; Bendfeldt et al., 2010; Sepulcre et al., 2009), which may not be sensitive enough for small change (Anderson et al., 2006; Hutton et al., 2009).

In this research, global measures of GM tissue loss were used (GMF and global CTh), and we found that the relationship between GM atrophy and focal (T2LV) WM and diffuse whole brain (MTR_{NABT}) damage changed from RRMS to SPMS. In the future study, regional measures can be used to confirm this relationship. Since global measures do not account for spatial correspondence between GM and WM, regional comparisons can provide more specific relationship. For this study, frequent MRI is advantageous because delay effects can be explored.

6.3.5 Regional Correlation of GM Atrophy and Local Damage

If the cause of GM atrophy is not WM pathology, either diffuse or focal, the presumed cause is local GM damage. Since there is no reliable method to visualize entire GM damage in MS, there is currently no *in vivo* MRI study that has examined this
relationship. However, in previous post-mortem and animal studies, there has been no
correlation between GM atrophy and local GM demyelination (Pomeroy et al., 2008;
Pomeroy et al., 2010; Wegner et al., 2006). However, these studies were restricted to 2D
histological slides when studying cortical atrophy, which is known to be inaccurate
(Fischl and Dale, 2000). Therefore, if applied to the autopsy cases, CLADA may provide
more accurate and definitive insight into the relationship between GM atrophy and GM
demyelination.

6.4 Conclusion

In conclusion, I have characterized the evolution of GM and cortical atrophy,
demonstrated that these atrophy measures were clinically relevant in MS, and provided in
vivo insights on possible MS mechanisms. The developed algorithms have been validated
and shown to be reliable. The use of GM atrophy and cortical atrophy appear to be highly
promising as a potential imaging marker in clinical trials, and further elucidation of GM
and cortical atrophy will provide even better understanding of MS mechanisms and aid in
the development of effective therapies in MS.


Increased yield of active demyelinating and (p)reactive lesions. Brain 124(Pt 8):1635-1645.


