UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Spinal Cord Imaging in Multiple Sclerosis Patients:
Applications of Machine Learning and Computer Vision Methods

Permalink

https://escholarship.org/uc/item/7bm8m24w

Author

Datta, Esha

Publication Date

2018

Peer reviewed|Thesis/dissertation

Spinal Cord Imaging in Multiple Sclerosis: Applications of Machine Learning
and Computer Vision Methods

by

Esha Datta

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Dedication

This dissertation is dedicated to my mother, Dr. Anindita Datta, who inspired me every day of my childhood with her scientific and artistic pursuits.

Acknowledgements

I would like to thank Roland Henry for all of his guidance as my advisor and for taking a chance to initially hire me. I would like thank all of the members of the Henry lab and my fellow graduate students, Kesshi Jordan, Anisha Keshavan, and Bago Amirbekian. I'd like to thank Nico Papinutto and Regina Schlaeger, for their ground breaking work in spinal cord imaging that made my work possible. Thank you, Sarah Nelson, Sharmila Majumdar, Ari Green, Pratik Mukherjee, and Sri Nagarajan for your mentorship.

The text of chapter 3 is a reprint of the material as it appears in Neuroimage. I would like to thank the co-authors listed in that work for their contributions. Nico Papinutto developed the PSIR sequence to acquire the images, Alyssa Zhu developed a script for zooming and interpolating the images, Julio Carballido-Gamio developed a script to register images by shape, and Roland Henry supervised this work.

I would like to thank my father, Nibir Datta, who encouraged my love of math and science and my sister, Sreoshi Sen Datta, for always believing that I could do anything. I thank my brother-in-law, my nieces (Amiya, Kamala, and Sahana), and my father-in-law for all of support. I would like to thank my mother-in-law, Dr. Haimanti Dorai, whose bravery in life inspires me daily and reminds me to live a life that is bold. I am so grateful to my husband, Amal Dorai, for inciting me to pursue my Ph.D and encouraging me to find the type of work that makes me excited to wake up in the morning, and to my baby, Riju Datta Dorai, who reminds me to live each day with joy and wonder. Finally, I'd like to thank the patients of the EPIC multiple sclerosis study. Every image from this thesis represents a human life that is facing an incredible challenge. Thank you for helping to further our knowledge of MS.

Abstract

Spinal Cord Imaging in Multiple Sclerosis Patients: Applications of Machine

Learning and Computer Vision Methods

Esha Datta

During the last few decades, researchers have struggled to find reliable biomarkers for multiple sclerosis (MS) that could aid in diagnosis, measurement of disease progression, evaluation of treatments in clinical trials, and prediction of treatment effect. Traditional metrics, such as brain and lesion volumes, are poor contenders since they do not reliably reflect clinical metrics. Until recently, spinal cord metrics were also poor contenders, due to the quality limitations of spinal cord imaging. However, with recent technological advances, we are now able to acquire better quality spinal cord images and capture these metrics more accurately. This thesis investigates the potential of using spinal cord images clinically in MS through four different studies. The first study investigates different spinal cord metrics and shows how spinal cord PSIR gradient independently predicts EDSS in RRMS patients. The second study demonstrates two different methods for how spinal cord gray matter can be automatically segmented so that metrics can be easily obtained in a clinical setting. The third study is an investigation of how spinal cord metrics change longitudinally. The fourth study is a voxel-wise analysis of spinal cord metrics that shows local patterns of intensity, gradient, and deformation.

Table of Contents

List of Figures

List of Tables

Chapter 1 - Introduction

1.1 Multiple Sclerosis

MS is a disease that affects around 2.5 million people around the world. In the U.S., there are over 400,000 affected around 200 new cases every week. MS often has a large impact on quality of life, due to its accompanying symptoms, which vary depending on the location of the affected nerves. These symptoms include muscle weakness, vision problems, coordination issues, and sensory problems. The disease course of MS also varies. In relapsing-remitting MS (RRMS), new symptoms occur in isolated attacks, while in progressive MS (PMS), symptoms build up over time. The cause of MS is still not well understood, though we know that environmental factors and genetic susceptibility both play a role. Currently, there is no cure for MS, though there are treatments available to modify disease progression and alleviate symptoms.

MS is a disease in which the immune system mistakenly attacks the insulating covers of nerve cells, called myelin, in the brain and spinal cord. T cells enter into the brain through disruptions in the blood brain barrier and start attacking the myelin, triggering inflammation. The destruction of myelin allows additional water to fill the breach, in addition to the fluid that accumulates during initial swelling. These damaged areas with additional water content are seen as lesions on MRI scans. The resulting myelin damage also affects the ability of nerves to effectively send signals and communicate, which results in a wide variety of symptoms. Neurodegenerative processes then gradually destroy the axons and nerve tissue to cause irreversible atrophy.

Through the use of magnetic resonance imaging (MRI), we are able to visually assess the damage that occurs in the brain and spinal cord from MS. The most notable sign of MS in an MRI scan is the presence of lesions, that develop on the white matter. The effects of inflammation and demyelination are primarily seen in the white matter, while the effects of neuronal and axonal loss are primarily seen in the gray matter.

1.2   The Spinal Cord

The spinal cord is the main pathway for connecting the brain to the peripheral nervous system. From each of its 31 segments, there is a pair of sensory nerve roots and a pair of motor nerve roots. Like the brain, the spinal cord is also made up of white matter and gray matter. The butterfly-shaped center of the cord is made up of the gray matter and different sections correspond to either motor or sensory function. The surrounding white matter is made up of nerve fibers that carry information back and forth from the brain to the rest of the body. These white matter tracts are also organized regionally by function.

Spinal cord abnormalities, such as atrophy, diffuse abnormalities, and focal lesions, play an important role in the diagnosis and prognosis of MS (Kearney, Miller, & Ciccarelli, 2015). Around 90% of patients with clinically definite MS have spinal cord abnormalities (Lycklama et al., 2003) and pathological studies (FOG, 1950) have looked at these abnormalities postmortem in MS patients. According to the McDonald diagnostic criteria for MS, one sign of MS that demonstrates dissemination in time is the presence of focal T2 lesions in the spinal cord (Thompson et al., 2018). Focal spinal cord lesions are often associated with motor symptoms, coordination problems, and bowel and bladder disfunction.

1.3   Motivation: The Search for a Biomarker

Multiple sclerosis is characterized by a complex pathophysiology, which includes inflammation, demyelination, axonal degeneration, and neuronal loss (Kutzelnigg & Lassmann, 2014). For this reason, symptoms vary widely among individuals in type and severity. Because of this broad heterogeneity, a biological marker that reflects the underlying disease process and differentiates patients would prove very useful for tracking disease progression. In the past, studies have relied on metrics such as brain volume and lesion load, but neither of these metrics has a strong correlation with clinical metrics, such as the Expanded Disability Status Scale (EDSS) score. Researchers continue to search for a quantitative and sensitive biological marker that correlates well with clinical disability and can possibly be used to predict disease progression or track progress in clinical trials.

It is well known that the spinal cord plays a large role in MS and may be responsible for much of the motor disability of patients (Bernitsas et al., 2015). However, until recently, MS spinal cord studies were limited due to the poor gray and white matter contrast and low spatial resolution seen in conventional MRI as well as artifacts due to physiological motion of the cord and the adjacent tissues (Stroman et al., 2014; Wheeler-Kingshott et al., 2014). Now, through the enhanced T1 contrast offered by phase-sensitive inversion recovery (PSIR) imaging, we are able to capture higher quality images of the spinal cord gray and white matter within a reasonably short acquisition time (Papinutto et al., 2015). Now, spinal cord imaging has come to play a more vital role (Wheeler-Kingshott et al., 2014) in the investigation of multiple sclerosis (Massimo Filippi & Rocca, 2013).

Recent studies using this imaging modality have shown the promise of spinal cord gray matter area at selected levels as a biomarker for MS (Schlaeger et al., 2014, 2015). Through the use of a regression subset selection, this study showed that spinal cord gray matter area from the C2/C3 axial level was a better predictor of EDSS score when compared to other traditional metrics such as the gray matter and white matter volumes from the brain, the upper cervical spinal cord area (UCCA), the T1 lesion load in the brain, the FLAIR lesion load in the brain, and the number of T2 lesions in the spine. Prompted by this finding, this work aims to investigate the potential of other metrics from the PSIR images of the spinal cord.

Spinal cord imaging has come to play a vital role (Wheeler-Kingshott et al., 2014) in the investigation of many disorders, such as multiple sclerosis (Massimo Filippi & Rocca, 2013) and traumatic spinal cord injury (Freund, Curt, Friston, & Thompson, 2013; Potter & Saifuddin, 2003).*In vivo* gray and white matter volume estimates in the spinal cord may play a critical role in improving our understanding of the disease process in disorders involving both gray and white matter such as multiple sclerosis (Schlaeger et al., 2014, 2015) or in predominantly white matter diseases such as adrenomyeloneuropathy (Israel, Ostendorf, Stiepani, & Ploner, 2005). Spinal cord metrics may serve as biomarkers that reflect clinical outcomes and can be used for the prediction or monitoring of disease progression.

1.4 Presented Studies

This thesis presents four different studies that investigate the potential of using PSIR spinal cord images in a clinical setting for MS patients. The first study investigates intensity and gradient metrics in the spinal cord. In this cohort, spinal cord PSIR gradient independently predicts EDSS in RRMS patients, suggesting that the gradient may be a metric that estimates the microstructural

tissue damage. The second study presents an active countour method for segmenting spinal cord gray matter. This method has competed favorably against 5 other methods in a challenge for automatic spinal cord gray matter segmentation methods. Additionally, a convolutional neural network method is also demonstrated for segmenting spinal cord gray matter. The third study is an investigation of how spinal cord metrics change longitudinally. The fourth study is a voxel-wise analysis of spinal cord metrics that shows local patterns of intensity, gradient, and deformation.

Chapter 2 - Study of Spinal Cord PSIR Gradient and Intensity Metrics

2.1 Introduction

Texture metrics provide a quantitative measurement of the patterns and relationships of voxels in an image. These quantitative texture metrics are useful in studying medical images of MS patients and can illuminate subtle differences that are not immediately apparent to the human eye (Zhang, 2012). In the past, texture analysis has been used in multiple sclerosis to segment lesions, normal-appearing white matter, and white matter and also to examine the relationship between texture severity of lesions and disease progression. Texture features are a measure of the structural regularity in an image, and thus they are a reflection of histopathology and the structural property of a tissue. Several tools are used to study texture, including gray level co-occurrence matrices that compute the probabilities of co-occurring intensity values and spatial frequency analysis of images.

A very simple analysis for comparing image textures can be performed with first order statistics, such as mean and variance of intensity and gradient. The spinal cord gray matter and white matter regions are irregularly shaped areas made up of only a few pixels. For this reason, a simple metric of gradient is more appropriate for capturing texture information than more complicated methods, such as co-occurrence matrices. The gradient is a measure of how much the image intensity changes across space. A low mean image gradient reflects a more uniform image, while a high mean image gradient reflects an image with more variable intensities. Previous studies have investigated global texture metrics in the spinal cord (Mathias, Tofts, & Losseff, 1999) and shown that mean intensity and mean gradient in the spinal cord are correlated with EDSS scores.

This study focuses on regional metrics of gradient and intensity. Since gradient is a relative metric and is less likely to be influenced by intensity differences caused by scanner variations, it may be more useful than measurements of absolute intensity.

This study hypothesizes that intensity and gradient metrics from the upper cervical spinal cord may be sensitive to microstructural damages of tissues and therefore be viable biomarkers that are correlated with clinical disability to a greater degree than more traditional metrics. The regional analysis described in this paper aims to shed further light onto the role of white and gray matter in MS and lead us closer towards a biomarker that reflects clinical disability better than the current standards.

2.2   Methods

The subjects in this study included 20 healthy controls, 92 40 relapsing MS (RMS) patients and 37 18 progressive MS patients (PMS). The age and gender distribution of the control group was similar to the group of RMS patients. The median EDSS score of RMS patients is 2 and the median EDSS score of the PMS patients is 6. The clinical characteristics of the patients are described in Figure 1. A study of gray and white matter area was previously performed in this cohort and correlations with EDSS score were reported (Schlaeger et al., 2014). These correlations are presented again in this study to provide a basis for comparison.

	Controls	Relapsing MS	Progressive MS
Number	20	92	37
Mean Age+/- Standard Deviation	48.6 +/- 12. 2	49.0 +/- 9.3	57.9 +/- 9.3
Percent Female	65.0%	63.0%	51.3%
Mean Disease Duration +/- Standard Deviation	0 +/- 0.0	15.2 +/- 8.5	20.7 +/- 10.9
Median EDSS (Interquartile Range)	0 (0-0)	2 (1.5-2.5)	6 (3.5-6.5)

Figure 1: Subject Demographics

To test for reliability, texture and intensity metrics were also calculated on 8 different healthy subjects who underwent two scans in the same session with repositioning in between scans. The interclass correlations (ICC) were calculated for both the mean intensity and mean gradient over the entire cord.

All subjects in the patient cohort received a standardized clinical neurological exam to determine the EDSS score. In addition, all patients were scanned on a Siemens 3T Skyra scanner equipped with a 20 channel neck-head coil and a 32 channel spine coil within two weeks of their clinical examination. For each patient, axial 2D phase sensitive inversion recovery (PSIR) images were acquired at the intervertebral disc level C2/C3 perpendicular to the cord. The acquisition parameters are listed below.

Imaging Parameters:
3T Skyra scanner
20 channel neck/head coil
32 channel spinal coil
In-plane resolution = .78 X.78 mm^2
Slice thickness = 5 mm
TR = 4000 ms
TE = 3.22 ms
TI = 400 ms
Flip Angle = 10 degrees
3 averages

Figure 2: Imaging Parameters

For each phase sensitive reconstructed image, cord masks were generated on up-sampled interpolated images through semi-automated segmentation using the software JIM6 (Horsfield et al., 2010). In addition, gray matter masks were generated on these same images through manual segmentation by a trained neurologist. To obtain white matter masks, the gray matter masks were subtracted from the cord masks. At the C2/C3 level for each patient, the areas, mean intensities, and the mean intensity gradients were calculated for both the white matter and the gray matter.

Initially, univariate analyses were performed and regression models were used to determine the Spearman rank correlation between each of these metrics and the EDSS score. In addition, a multivariable model was performed to determine how all of these metrics, in addition to covariates such as age, disease duration, and sex, interact to provide a prediction for the EDSS score. A stepwise regression model was used to determine the subset of variables that contribute most to the prediction. Following that, a least squares regression model was used to determine the relationship between the MRI metrics and EDSS score. All statistical analyses were performed using the software JMP (v12).

2.3 Results

2.3.1 Reliability:

For the 8 instances of subjects who underwent two separate scans, the ICC for mean intensity was 0.89 and the ICC for mean gradient was 0.85. A previous study (Schlaeger et al., 2014) looked at the reliability of volumetric metric and calculated intraclass coefficients (ICC). For the assessment of total cord area, the inter-rater ICC was greater than 0.99. For the assessment of gray matter area, the inta-rater reliabity showed an ICC of 0.98 and the inter-rater reliability showed an ICC of 0.91.

2.3.2 Correlation with Clinical Outcomes:

The distribution of the mean intensity and mean gradient metrics in the control and disease type groups are shown in the box and whisker plots below in Figure 3. From the plots it can be observed that the intensity and gradient metrics are fairly consistent in healthy controls, but that the ranges increase with RMS patients and are even larger in PMS patients. In addition, the mean intensities and mean gradients are decreased in RMS patients and even lower for PMS patients when compared with controls. If the disease type is treated as an ordinal variable, Spearman Rho coefficients can be calculated to see how metrics relate to disease type. All of the intensity metrics had p-values of less than 0.0001 when correlated with disease type. However, the correlation of the gray matter area with disease type is the largest.

	Box Plots Separated Out By Disease Type	Spearman's Rho	P Value
WM Area		-.27	.0009
GM Area		-.57	< .0001
Average WM Intensity		-.43	< .0001
Average GM Intensity		-.41	< .0001
Average WM Gradient		.34	< .0001
Average GM Gradient		-.26	.0013
	Controls RMS PMS		
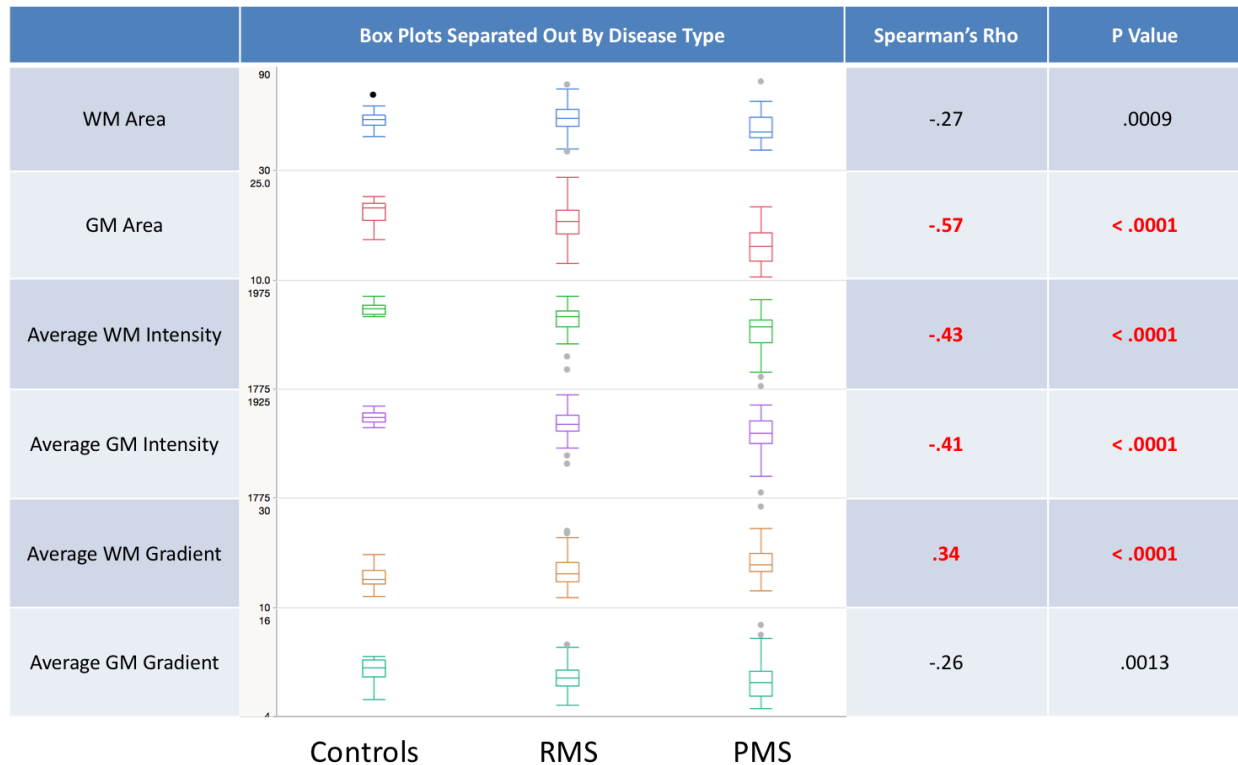
Figure 3: Box and Whisker Plots of Metrics by Disease Type

As seen in Figure 4, in a univariate linear regression analysis, both of the intensity metrics were shown to be significantly correlated with EDSS (p-values less than 0.0001 and Spearman rank correlation magnitudes above 0.3), suggesting that there may be global differences in image intensity that relate to clinical disability. The white matter gradient, but not the gray matter gradient, was shown to be significantly correlated with EDSS. However, the strongest correlation by far was that of the spinal cord gray matter area (p-value < 0.0001, Spearman rank correlation = -0.63),
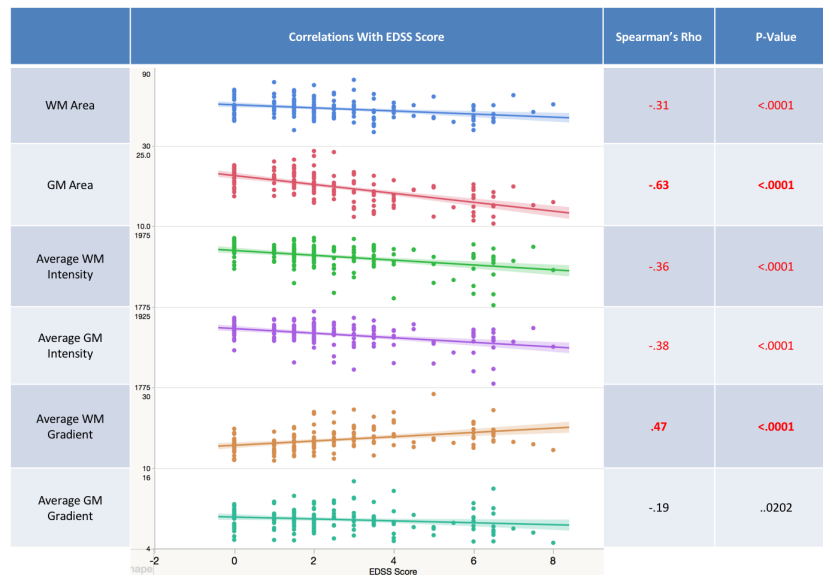
Figure 4: Univariate Linear Regression of Metrics with EDSS Score

Upon further investigation, the gradient data was observed to have quadratic behavior, with an upward trend in RMS patients with lower EDSS scores and a downward trend in PMS patients with higher EDSS scores. Quadratic regression was performed on all of the gradient metrics with EDSS score. The amount of variation in the gradient metrics explained by the EDSS score was higher in the quadratic models for all cases. In particular, the fit of the quadratic models was significantly better for the metrics involving white matter. The fit lines are shown below in Figure 5.

	Quadratic Regression With EDSS Score	R² of Quadratic Fit	Linear Regression with EDSS Score	R² of Linear Fit
Average Cord Gradient		.114		.056
Average GM Gradient		.034		.030
Average WM Gradient		.124		.075
Average Boundary Gradient		.110		.093
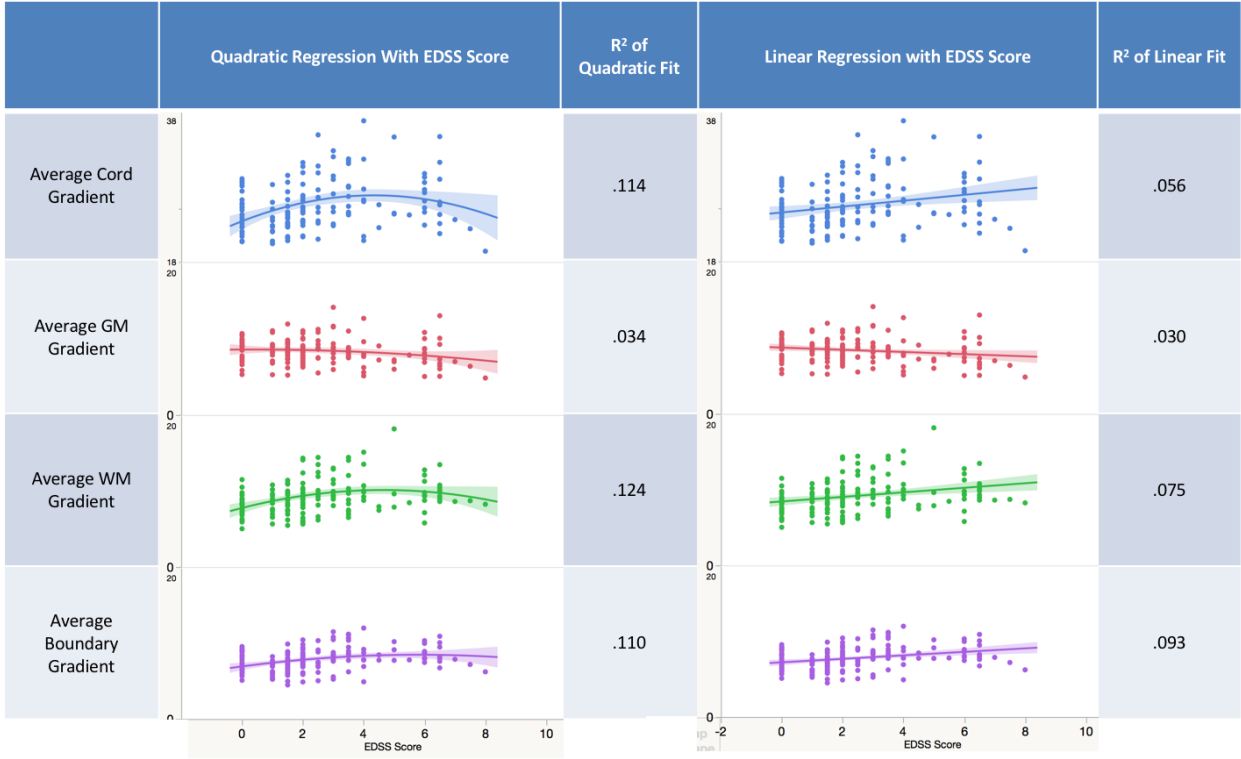
Figure 5: Quadratic and Linear Regression of Gradient Metrics with EDSS Score

To capture the two different trends, the data was separated into two groups, RRMS patients who tend to have lower EDSS scores and PMS patients who tend to have higher EDSS scores. Remarkably, no significant correlations with EDSS score were found in progressive patients, while strong correlations were found in patients with lower EDSS scores for gray matter area and average white matter gradient.

	RMS Patients Correlation With EDSS	Rho	P	PMS Patients Correlation With EDSS	Rho	P
GM Area		-.38	.0002		-.16	.34
WM Area		-.21	.04		.05	.77
Cord Area		-.28	.01		-.08	.65
Avg GM Intensity		-.07	.53		-.15	.36
Avg WM Intensity		.03	.80		-.16	.34
Avg Cord Intensity		.02	.87		-.17	.33
Avg GM Gradient		.10	.35		-.20	.23
Avg WM Gradient		.47	<.0001		-.16	.35
Avg Cord Gradient		.48	<.0001		-.22	.19
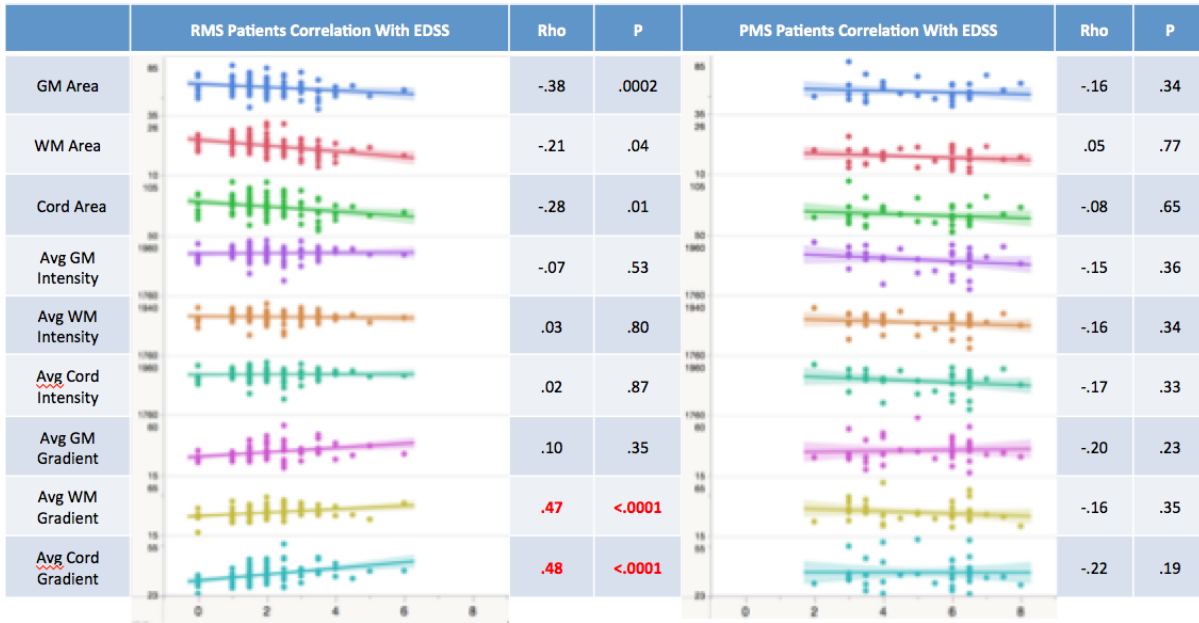
Figure 6: Linear Regression of Metrics with EDSS Score When Separated By Disease Type

To further explore the correlation of average white matter gradient with EDSS score in RRMS patients, several other regions of interest were investigated as shown in Figure 7. A boundary region was created from the single pixel boundary of the gray matter masks. The remainder of the white matter was designated as a region as well. In addition, the total cord was also used as a region of interest.

Figure 7: Gray Matter ROI shown in Blue, Boundary ROI shown in Yellow

Figure 8: Correlations of Gradient Metrics with EDSS Scores in RMS Patients

The results showed that there was a correlation of gradient in the boundary area with EDSS score, but that the correlation was weaker for areas further away from the gray matter. In addition, the correlation is still seen strongly when the average gradient is taken over the entire spinal cord.
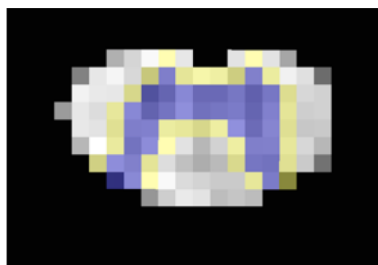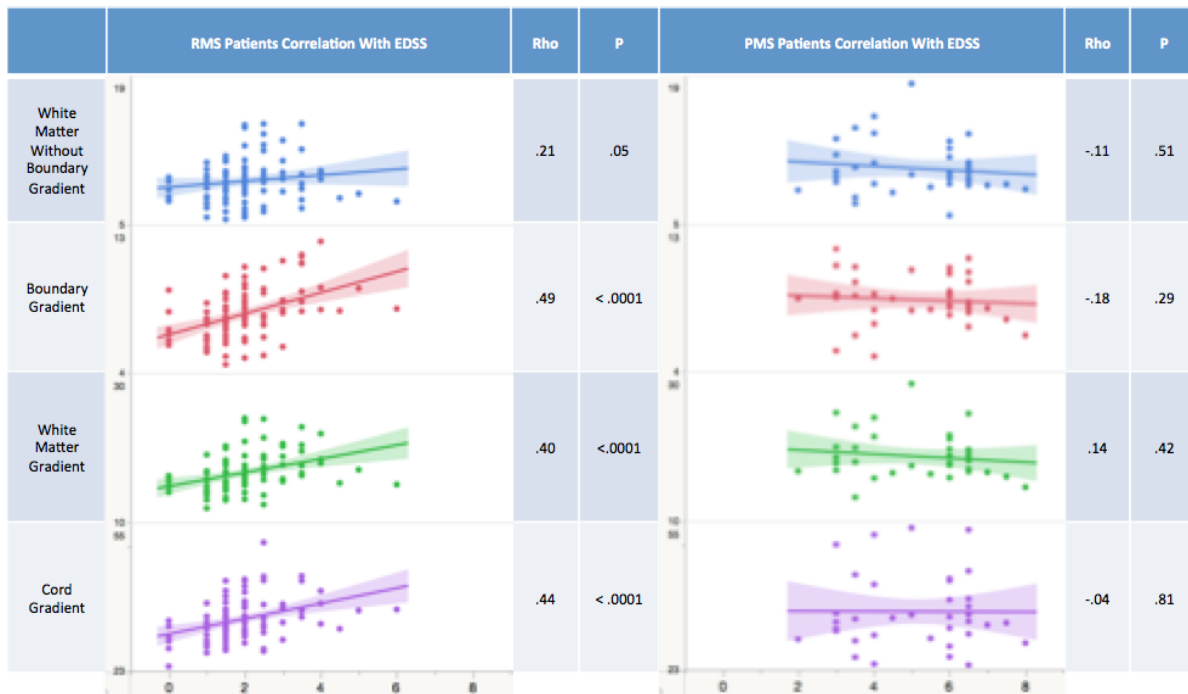
To determine how intensity and gradient metrics interact with more traditional metrics, such as brain and lesion volumes, and other covariates, such as age, gender, and disease duration, in patients with lower EDSS scores, a multivariate linear model was created as well. A stepwise regression model determined the best subset of variables for predicting EDSS. The complete list of metrics, as well as the list of those selected by the stepwise regression is summarized below.

Parameter	Selected for Model	Beta	P-Value
Gray Matter Area	Yes	-.33	.0003
Average Cord Gradient	Yes	.33	.0003
Age	No	-	-
Gender	No	-	-
Disease Duration	No	-	-

Figure 9: Metrics Selected by Stepwise Regression to Predict EDSS for RMS Patients

Using the selected subset of variables, a least squares multivariate model was performed for the RMS patient group. In the final model, the mean cord gradient has a standard beta of .33 with a p-value of .0003 and the gray matter area has a standard beta of -.33 with a p-value of .0003. The fact that both of these variables were selected in the stepwise regression suggests that they both independently contribute to the model of EDSS. In this model, cord gradient and spinal cord area explain 32% of EDSS variance. On its own, spinal cord gray matter area explains 22% of the variance and on its own, cord gradient explains 22% of the variance.

2.4 Discussion:

Based on phase sensitive inversion recovery imaging, this study assesses the potential of cervical spinal cord intensity and texture metrics as biomarkers through a study of the association with clinical disability in a large single center cross sectional study of patients with MS. The associations found between intensity and gradient metrics and disability also help to shed light on the role of white and gray matter in the spinal cord on MS. In particular, white matter gradient metrics were shown to have strong potential as reliable and easy to measure biomarkers with a strong correlation with clinical disability

2.4.1 Reliability:

Of all the spinal cord metrics, total cord area is by far the most reliable, due to the automated method of measurement. Gray matter area has a slightly lower ICC, due to variability from manual operators. Though the mean intensity and mean gradient lag slightly behind, they are still reliable enough to be used as predictors of clinical outcome.

In the future, we hope to investigate how robust these metrics are when obtained with different scanners or different acquisition protocols. However, gradient and intensity metrics are

expected to be quite robust since they are both relative measures. PSIR has been showed to be a quantitative image modality (Papinutto et al., 2015). Since phase reconstruction is performed with a reference image in addition to the acquisition image, the intensity metrics are obtained as relative metrics. Gradient metrics are likely to be more insensitive to scanner or protocol differences than intensity metrics, since they are relative measures of the intensities.

2.4.2 Ease of Measurement:

With PSIR imaging, it is now possible to acquire high quality spinal cord images at selected levels in a very short acquisition time. Cord segmentation can be done automatically with programs such as JIM. Once the cord has been segmented, obtaining the mean cord intensity and the mean cord gradient is an easy task with software. In contrast, obtaining the metric of spinal cord gray matter area requires gray matter segmentation in the cord, a process that is currently manual and must be performed by a trained expert.

2.4.3 Correlation with Clinical Outcomes:

The results suggest that in addition to gray matter area, intensity metrics and the white matter gradient metric from PSIR images are valuable in predicting clinical disability.

While intensity metrics are valuable when predicting EDSS score in the overall patient population, they have lower correlations for patients with less severe EDSS scores. This suggests that the linear regressions for intensity metrics in Figure 3 are driven mainly by the extreme values. This would indicate that intensity metrics are better at predicting large differences in EDSS scores and worse at predicting subtle differences in clinical disability.

White matter gradient is shown to correlate with EDSS score in the overall patient population as well. However, white matter gradient is shown to be most valuable in predicting

EDSS score for patients with less severe EDSS scores. When added together with spinal cord gray matter area, white matter gradient boosts the amount of EDSS variation explained to 32%.

It is notable that unlike white matter gradient, the gray matter gradient did not show any correlation with EDSS score. When investigating the white matter gradient, it was found that the area of white matter directly surrounding the gray matter is the part that contributes most to this correlation. However, this correlation can still be found when the average gradient is taken over the entire cord, which means that meaningful gradient metrics are very easy to obtain.

Our results show that the metric of average spinal cord gradient adds information to the prediction of EDSS and is particularly useful in distinguishing subtle differences in EDSS scores for RMS patients with milder disability. The results of the multivariate analysis suggest that mean cord gradient and gray matter area are independently associated with EDSS in patients with milder disability. By using both of these complementary metrics in conjunction, clinical disability can be predicted to a much higher degree.

2.4.4 Pathological Correlates:

Pathological studies suggest that the metric of gradient is a measure of tissue heterogeneity that reflects the amount of microstructural damage. The metric of intensity is likely to be a reflection of the tissue water content and thus a measure of demyelination.

While controls exhibit rather uniform gradient metrics, results show that the gradient increases in the lower range of EDSS scores, peaking around a value of 4.5 before decreasing in the higher range of EDSS scores. The quadratic behavior of the gradient metrics seen in this study makes sense from a pathological perspective. A healthy spinal cord starts off with more uniform intensity, but as the disease progresses, there is a mix of healthy and diseased tissue and the

17

variation increases. Eventually, there is a return to uniformity as diseased tissue dominates the spinal cord.

It is notable that white matter gradient has a strong correlation with EDSS scores, while gray matter gradient does not. A previous study has shown that spinal cord gray matter area has a strong correlation with EDSS score, while spinal cord white matter area does not. This finding suggests that though there is no sign of atrophy in the white matter, damage is still occurring at a microstructural level. Due to inflammation in the white matter, it is possible that atrophy can not be detected by area changes but can be detected through changes in gradient. Since gray matter is a less ordered tissue than white matter, it is also possible that the gradient metric reflects changes in gray matter to a lesser degree.

A recent post-mortem study suggests that texture metrics, which measure the patterns and relationships of intensity voxels in MRI, are reflections of micro scale structural pathological changes in biological tissue. This study investigated how texture metrics based on the polar Stockwell Transform correlated with tissue pathology in T2 weighted images of post-mortem MS brains (Zhang et al., 2013). This study showed that texture heterogeneity was largest in lesions, followed by diffuse appearing white matter, followed by normal appearing white matter. The results suggest that texture heterogeneity is dominated by demyelination, and is also affected by axonal injury and inflammation.

Texture analysis allows for detection of these more clinically relevant subtle structural tissue alterations. Due to the small size and resolution of the spinal cord regions we are investigating, this study focused on gradient, a very simple metric of texture. Just like the polar Stockwell Transform metrics from the studies previously mentioned, the metric of gradient also measures the coarseness of the image texture. From this, we can surmise that the metric of gradient

similarly measures the homogeneity of the tissue and the degree of partial demyelination. With future higher resolution images, gray-level co-occurrence matrices and spatial frequency methods may prove more useful in uncovering more subtle structural tissue alterations.

2.5 Summary

While recent studies have shown how texture metrics in the overall spinal cord relate to EDSS scores, this is the first study to investigate how these metrics relate to white and gray matter of the cervical spinal cord and also to different disease courses. Our results suggest that structural microdamage as reflected by the gradient in the white matter is much more important than microdamage in the gray matter when predicting clinical disability in RMS patients. In addition, correlations between intensity and disease course can be seen in all regions of the spinal cord, suggesting that this effect is global. In addition, we can see how the boundary area between gray matter and white matter may be the most important location of structural damage when predicting clinical disability.

This study provides further evidence for the potential of spinal cord metrics as biomarkers in multiple sclerosis. This study suggests that cord gradient provides a complementary correlate of EDSS in RMS patients. Cord gradient may be particularly useful in a clinical setting, since it is an easier metric to obtain than gray matter area, which requires segmentation.

Chapter 3 - Gray Matter Segmentation with Active Contours

3.1 Introduction

Certain spinal cord metrics may serve as biomarkers that reflect clinical outcomes and can be used for the prediction or monitoring of disease progression. For example, two recent imaging studies have revealed the potential of cross sectional spinal cord gray matter area as a biomarker correlated with clinical outcome in multiple sclerosis (Schlaeger et al., 2014, 2015). An automatic method for spinal cord gray and white matter segmentation could provide an easy way to routinely obtain these metrics in multiple sclerosis patients for both cross-sectional and longitudinal assessments. This would allow evaluation of these metrics in their ability to capture disease progression, to identify subjects at risk for worsening, as well as to potentially evaluate treatment effects in clinical trials.

For this reason, tissue segmentation methods that were developed for brain MR images have been largely unsuccessful in spinal cord images. Now, through sequences like the phase-sensitive inversion recovery imaging (PSIR) (Papinutto et al., 2015) and T2* weighted MRI (Held, Dorenbeck, Seitz, Fründ, & Albrich, 2003; M.C. Yiannakas et al., 2012) we are able to acquire higher quality images of the spinal cord reliably, and in a reasonably short acquisition time. The added contrast in these images provides the opportunity of delineating the gray matter either manually or automatically.

Studies of spinal gray matter have largely depended on manual segmentation, which is a time consuming process that requires a trained expert and may be susceptible to operator bias. Automatic segmentation methods would allow for spinal cord metrics to be computed routinely for large populations. Furthermore, automatic methods would provide improved reliability given

the natural need to employ multiple operators across cohorts and time for manual methods. Automatic methods would improve the utility of spinal gray matter volumes and areas in multisite studies. This actively developing area is in need of algorithms that can be optimized and tested across sites and pulse sequences.

Few studies have explored algorithms for automatically segmenting spinal cord gray matter (De Leener, Taso, Cohen-Adad, & Callot, 2016). A template-based registration method for segmenting gray matter in the spinal cord (De Leener et al., 2017; Taso et al., 2015) works fairly well for healthy controls but was not tested in patients with gray matter atrophy and on imaging contrasts different from the template's T2* weighting. A semi-automated method that uses a fuzzy connector method has also been proposed, but it requires significant manual editing of the outputs (M.C. Yiannakas et al., 2012). Another study used a multi-atlas method that incorporates non-local statistical fusion (Asman, Smith, Reich, & Landman, 2013). This study required a database of multiple atlases and reported an average Dice coefficient of around 0.75 compared to manual segmentation.

My work focused on the development and evaluation of an automated spinal cord gray matter segmentation method based on Morphological Geodesic Active Contour (MGAC) models. Traditional active contour models are methods used in computer vision where a deformable spline is warped, subject to certain constraints and image forces, until a predefined overall energy is minimized (Kass, Witkin, & Terzopoulos, 1988). The standard solution for contour evolution algorithms involves numerical methods of integration that are computationally costly and may have issues with stability. The original active contours approach also depends on the parameterization of the contour and has trouble handling changes in curve topology. The geodesic active contour method addresses these issues, reduces the need for preprocessing since it utilizes

21

fewer parameters, and is better able to recognize an object with non-ideal edges (Caselles, Kimmel, & Sapiro, 1995). The morphological geodesic active contours method (Marquez-Neila, Baumela, & Alvarez, 2014) is a variation on this approach that allows for fast, stable contour evolution, increases recognition of objects with non-ideal edges, and reduces the need for image preprocessing.

The proposed method is demonstrated on 2D axial images acquired at the intervertebral disc level C2/C3, which has been shown to be clinically relevant in Multiple Sclerosis (Schlaeger et al., 2014). Motion artifacts are also less likely to occur at C2/C3 than at lower levels. The method can potentially be extended for use on other levels and with 3D volumes as well. The MGAC method requires the cord edge delineation as an input. A cross-sectional cord shape template is then registered with this input in an initial step that diminishes the need for pulse sequence dependent templates. This is used in conjunction with an active contour algorithm. While almost all brain and spinal cord automatic algorithms require some editing in a subset of cases, the degree to which they are automated is judged by the frequency and subjective level of the post-hoc editing. The MGAC method we present is shown to require obvious and predictable edits in the segmentations, such as the removal of a small bump or the continuation of a leg, which are needed with low frequency. The MGAC method can also be easily extended to other imaging protocols.

3.2   Methods

3.2.1   Image Acquisitions

45 healthy subjects and 58 multiple sclerosis patients were scanned for this study. Phase sensitive inversion recovery (PSIR) images were acquired in all patients and 42 of the 45 healthy

subjects, and T2* weighted images were acquired in 3 of the healthy subjects. Figure 10 shows the

number of subjects and scans acquired for each investigation.

Subject Type	Scan Type	Number of Subjects	Total Number of Scans	Purpose
Healthy Controls	2D PSIR	20	20	Template Creation
Healthy Controls	2D PSIR	12	12	Accuracy Tests
Healthy Controls	2D PSIR	8	16 (20 including unused scans)	Reproducibility Tests
Healthy Controls	2D PSIR (low and high resolution)	1	2	Validity Tests on High Resolution Images
Healthy Controls	2D T2* weighted	3	3	Versatility Tests on T2* weighted images
Healthy Controls	3D PSIR	1	1	Extension to 3D
MS Patients	2D PSIR	58	58	Patient Data Tests

Figure 10: Number of Subjects and Scans Acquired

The acquisition parameters for each scan type are shown in Figure 11. The 2D scans were

acquired axially at the intervertebral disc level C2/C3 perpendicular to the cord. The T2* weighted

scans were obtained on 2 different scanners with a 2D MEDIC (*Multi-Echo Data Image*

Combination; 6 echo times: 8.1, 13.98, 19.86, 25. 74, 31.62, and 37.5 ms) protocol. One scan was

acquired at a higher resolution. The high resolution 2D PSIR scans and the 2D T2* weighted scans

utilized a 64-channel head/neck coil that was unavailable for the other 2D PSIR scans and the 3D

PSIR scan, previously acquired with a 20-channel head/neck coil. The 2D T2* weighted and 3D

PSIR scans were accelerated with a parallel imaging factor of 2.

Scan Type	2D PSIR	2D PSIR (high resolution)	2D T2* weighted	3D PSIR
Intervertebral Levels:	C2/C3	C2/C3	C2/C3	C2-C5
Scanner:	3T Skyra scanner	3T Skyra scanner	3T Skyra scanner	3T Skyra scanner
In plane resolution:	.78 x .78 mm²	.5 x .5 mm²	.5 x .5 mm²	.625 x .625 mm²
Slice thickness:	5 mm	5 mm	5 mm	5 mm
Number of Slices:	1	1	1	12
Bandwidth	250 Hz/Px	250 Hz/Px	250 Hz/Px	250 Hz/Px
TR:	4000 ms	4000 ms	300 ms	4000 ms
TI	400 ms	400 ms		400 ms
TE	3.22 ms	3.22 ms	23 ms	4.23 ms
Flip Angle:	10 degrees	10 degrees	30 degrees	10 degrees
Head/neck coil channels:	20 channels	64 channels	64 channels	20 channels
Number of averages:	3	15	3	2
Acquisition Time:	1 minute 50 seconds	9 minutes 10 seconds	2 minutes 28 seconds	9 minutes 16 seconds

Figure 11: Pulse Sequence Parameters

The patients were selected from a cohort investigated in a previously published study (Schlaeger et al., 2014). The original cohort was screened to include subjects with a diagnosis of MS according to international criteria who were greater than 18 years of age. Any patients with relapses within 4 weeks prior to the visit, use of corticosteroids 4 weeks prior to the MRI exam, a recent history of drug or alcohol abuse, a diagnosis of hepatitis B or C or human immunodeficiency virus, participation in ongoing MS trials with unlicensed drugs, or any concurrent illness or disability were excluded. In this study, we analyzed only those cases without obvious lesions at the C2/C3 intervertebral disc level in the PSIR image, in order to evaluate the algorithm in cases without lesions. Out of 127 patients, 14 were excluded due to image artifacts or lesions that precluded manual gray matter segmentation. 55 additional patients were excluded due to lesions that did not preclude manual segmentation. Figure 12 shows the demographics of the remaining 58 MS patients and the 45 healthy controls included in this study.

	Controls	Relapsing MS Patients	Progressive MS Patients	All Patients
Number	45	46	12	58
Mean Age +/- Standard Deviation	44.2 +/- 14.6	49.6 +/- 8.5	60.7 +/- 11.4	51.9 +/- 10.2
Percent Female	62.2%	63.0%	58.3%	62.1%
Mean Disease Duration +/- Standard Deviation	N/A	14.6 +/- 7.7	17.7 +/- 12.6	15.3 +/- 8.8
Median EDSS (Interquartile Range)	N/A	2 (1.5-2.5)	4.25 (3.5-6.5)	2 (1.5-3.75)
Mean Cross-Sectional Cord Area +/- Standard Deviation	79.67 +/- 7.12 mm^2 (Calculated from 41 subjects with 2D PSIR images)	77.23 +/- 7.73 mm^2	67.66 +/- 6.85 mm^2	75.25 +/- 8.49 mm^2
Mean Cross-Sectional Gray Matter Area +/- Standard Deviation	19.59 +/- 1.58 mm^2 (Calculated from 32 subjects with manual GM segmentations)	18.36 +/- 1.91 mm^2	15.09 +/- 1.49 mm^2	17.68 +/- 2.28 mm^2

Figure 12: Subject Demographics

3.2.2   Creating a Template

In order to obtain initial estimates for the gray matter segmentation for input into the MGAC algorithm, templates were created and registered to each individual patient. A template of the overall cross-sectional shape of the whole cord and a template of the cross-sectional spinal cord gray-matter shape were built using 2D PSIR images at the C2/C3 level in 20 healthy subjects.

To create the templates, whole cord masks and gray matter masks were first created from the images. In order to facilitate manual gray matter segmentations, all images were first up-sampled by a factor of 10 using the python package SciPy and cropped to 300 voxels by 300 voxels. The field of view for each cropped image was thus 30 times the resolution of the original image. For all of the 2D PSIR images analyzed, the cropped subsections used for analysis each

25

measured 23.4 mm by 23.4 mm. The edge of the whole spinal cord and the edge of the spinal cord gray matter from 20 subjects were manually delineated to create masks using the software Jim (v. 6.0, Xinapse Systems, Northants, UK; *www.xinapse.com*). Distance maps were then created from these masks, where the value of each voxel represented the closest distance from the contour.

To create a template for the shape of the whole cord using these distance maps, all subjects were initially registered to a randomly chosen subject to create a mask reflecting the average affine shape and size of the 20 subjects. Registrations were done with multi-resolution affine transformations with 5 degrees of freedom (translation in x and y, scaling in x and y, and rotation). In order to create a template representing the average shape of the whole spinal cord from 20 subjects, an iterative algorithm (Rueckert, Frangi, & Schnabel, 2003) with both multi-resolution affine and nonlinear transformations (Vercauteren, Pennec, Perchant, & Ayache, 2008) was employed. In each iteration, the whole cord masks of all spinal cords were fully registered to the current template. The deformation fields from the nonlinear transformations were then averaged and the inverse was applied to the current template to create a new template. The affine template was used for the first iteration, and in the last iteration of this algorithm, the affine and nonlinear transformations were also applied to the gray matter segmentations to create a probabilistic gray matter template.

To create the gray matter template mask, a threshold of 0.5 was then applied to the probabilistic template. Since only the whole cord mask is used to find the transformations, the resulting probabilistic gray matter template is able to capture the variance in gray matter shape from different individuals. All affine registrations were performed by representing segmentations with distance fields (Carballido-Gamio et al., 2013), while all nonlinear registrations were performed by representing the background with distance fields and the spinal cords with zeros.

The registration software was an internally developed tool (Carballido-Gamio et al., 2013), programmed in MATLAB (The Mathworks, Inc. Natick, MA) to enable distance map based registrations (Reinertsen, Descoteaux, Drouin, Siddiqi, & Collins, 2004; Suh & Wyatt, 2006). Figure 13 shows a diagram of the steps involved in the generation of the templates for the whole cord shape and the spinal cord gray matter shape.
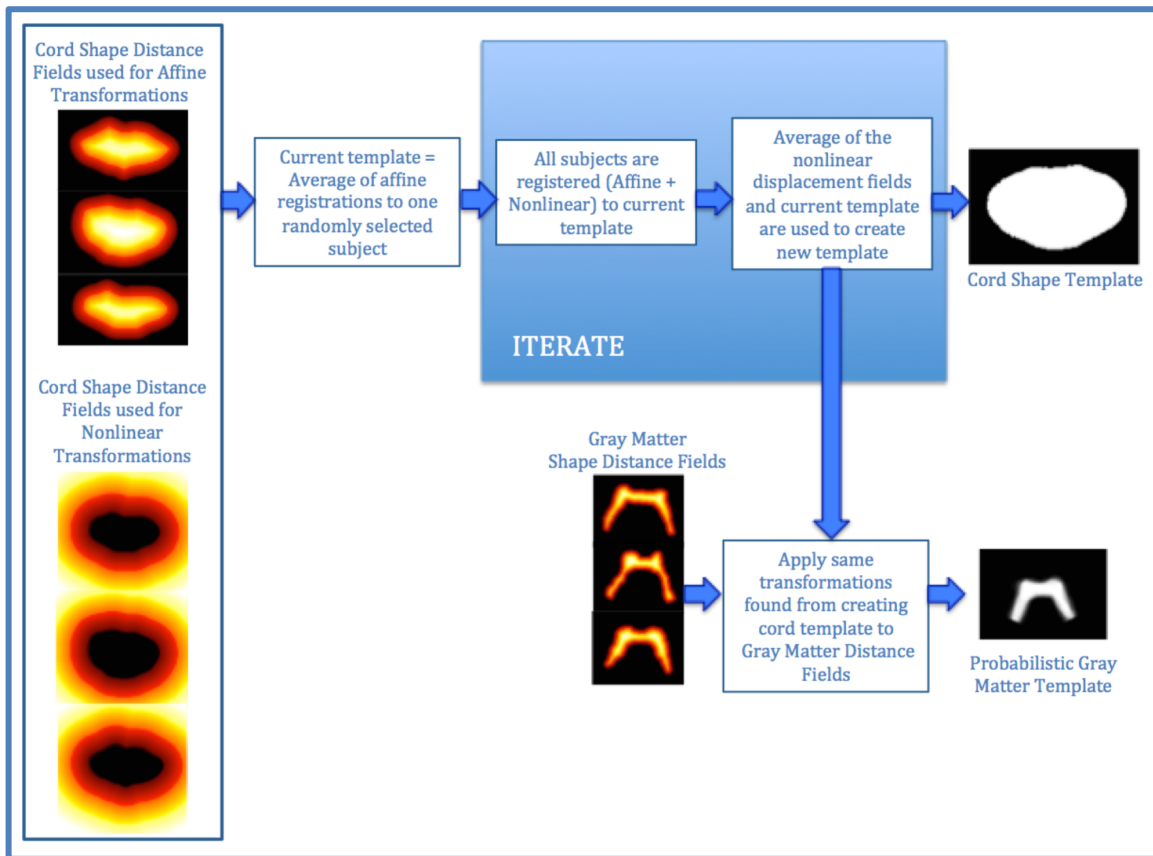

Figure 13: Contour Driven Template Creation

3.2.3   Creating an Initial Guess for Gray Matter Segmentation Based on Registration

To segment the gray matter in an image of the spinal cord, an initial guess of the segmentation must be provided to the active contours algorithm. This initial guess is based on the non-linear transformation of the previously created whole cord template to the delineated whole cord in the image. This process is shown in steps 1-5 in Figure 14. For each image to undergo gray

matter segmentation, the delineation of the edge of the whole cord is required as an input to the algorithm. This delineation is done with the software Jim in a semi-automatic process, which requires the user to provide a landmark for the center of the spinal cord (Horsfield et al., 2010). The previously created whole cord template is then registered with both affine and non-linear transformations to the delineated whole cord shape of each subject. The computed affine and non-linear transformations are then applied to the previously created spinal cord gray matter template. The transformed gray matter template gives a rough idea of the gray matter segmentation in each subject.
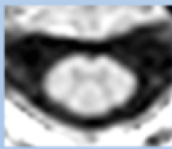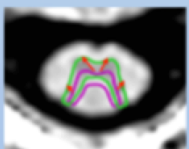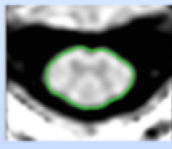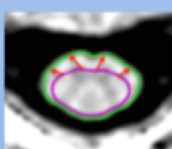
Figure 14: Steps for Segmentation with MGAC

3.2.4 Morphological Geodesic Active Contour Model

The registered gray matter template is used as an initial guess to initialize the geodesic active contour algorithm. The MGAC algorithm uses an open source Python implementation of the morphological geodesic active contour method (https://github.com/pmneila/morphsnakes). To use this implementation, the user provides an initial contour which is then deformed in a method driven by three image forces: a smoothing force that controls the smoothness of the contour, a

28

balloon force that inflates or deflates the contour in areas where information is lacking, and an image attraction force, which drives the contour to the maximum gradient areas in the image. Our parameters were selected according to the methods and guidelines stated in the study that developed this morphological geodesic active contour method (Marquez-Neila et al., 2014) and are listed in Figure 15. While one of the 5 parameters (the sigma stopping criteria) varied with acquisition resolution in this study, a consistent set of parameters can be used in practice provided that the original images are zoomed and interpolated to the same level before the active contours algorithm is applied.

Parameter	Function	Value:
α	Stopping criteria of algorithm	1000
σ	Stopping criteria of algorithm	3 for lower res (2D PSIR) 8 for mid res (3D PSIR) 10 for high res (2D PSIR, T2*W)
μ	Number of successive application of smoothing algorithm	5
θ	Threshold for application of smoothing and ballooning	1
ν	Strength of the balloon force	1

Figure 15: Active Contour Model Parameters

3.3 Results

3.3.1 Accuracy

In the 12 test subjects, the spinal cord gray matter was manually segmented by an experienced neurologist three separate times using the software Jim. These three sets of manual segmentations were then used to validate the accuracy of the MGAC segmentations using Dice similarity coefficients. Dice similarity coefficients were calculated for both the original manual and MGAC segmentations that were created using interpolated images as well as the down sampled manual and MGAC segmentations in the original resolution. In addition, the manual

segmentations were compared against the MGAC segmentations using the Hausdorff distance, which is defined as the greatest of all the distances from a point on one segmentation contour to the closest point on the other segmentation contour.

The results of MGAC and one of the three sets of manual segmentations for the 12 controls are shown in Figure 16. Figure 17 shows how the gray matter areas from MGAC segmentation and manual segmentation compared to each other in the 12 subjects. All of the Hausdorff distances between the MGAC and all three sets of manual segmentations were less than 1.0 mm.
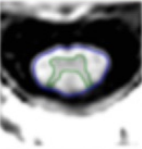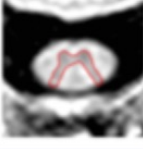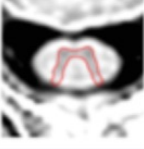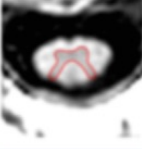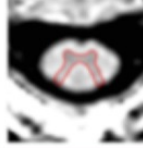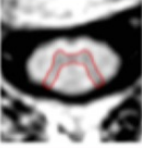
	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6
Manually Segmented						
MGAC Segmentation						
Dice Similarity Coefficient	.88	.88	.89	.89	.89	.87
Low Res Dice Similarity Coefficient	.85	.86	.89	.89	.86	.83
Hausdorff Distance (mm)	.60	.49	.44	.42	.55	.80

	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Avg (Min-Max) for all 3 sets of segmentations
Manually Segmented							
MGAC Segmentation							
Dice Similarity Coefficient	.88	.87	.85	.87	.88	.91	.88 (.82-.93)
Low Res Dice Similarity Coefficient	.86	.83	.84	.82	.91	.93	.88 (.81-.94)
Hausdorff Distance (mm)	.70	.55	.80	.60	.84	.47	.61 (.33-.99)

Figure 16: Manual segmentation (shown in green) vs. MGAC segmentation (shown in red directly below) of spinal cord gray matter for 12 healthy controls. Dice similarity coefficients are reported below the images.

Manual vs MGAC GM Areas at C2-C3

$y = 0.7283x + 6.1222$
$R^2 = 0.74387$

Y-axis: Manual Segmentation GM Area at C2-C3 (Average of 3 areas) in mm^2

X-axis: MGAC Segmentation GM Area at C2-C3 in mm^2
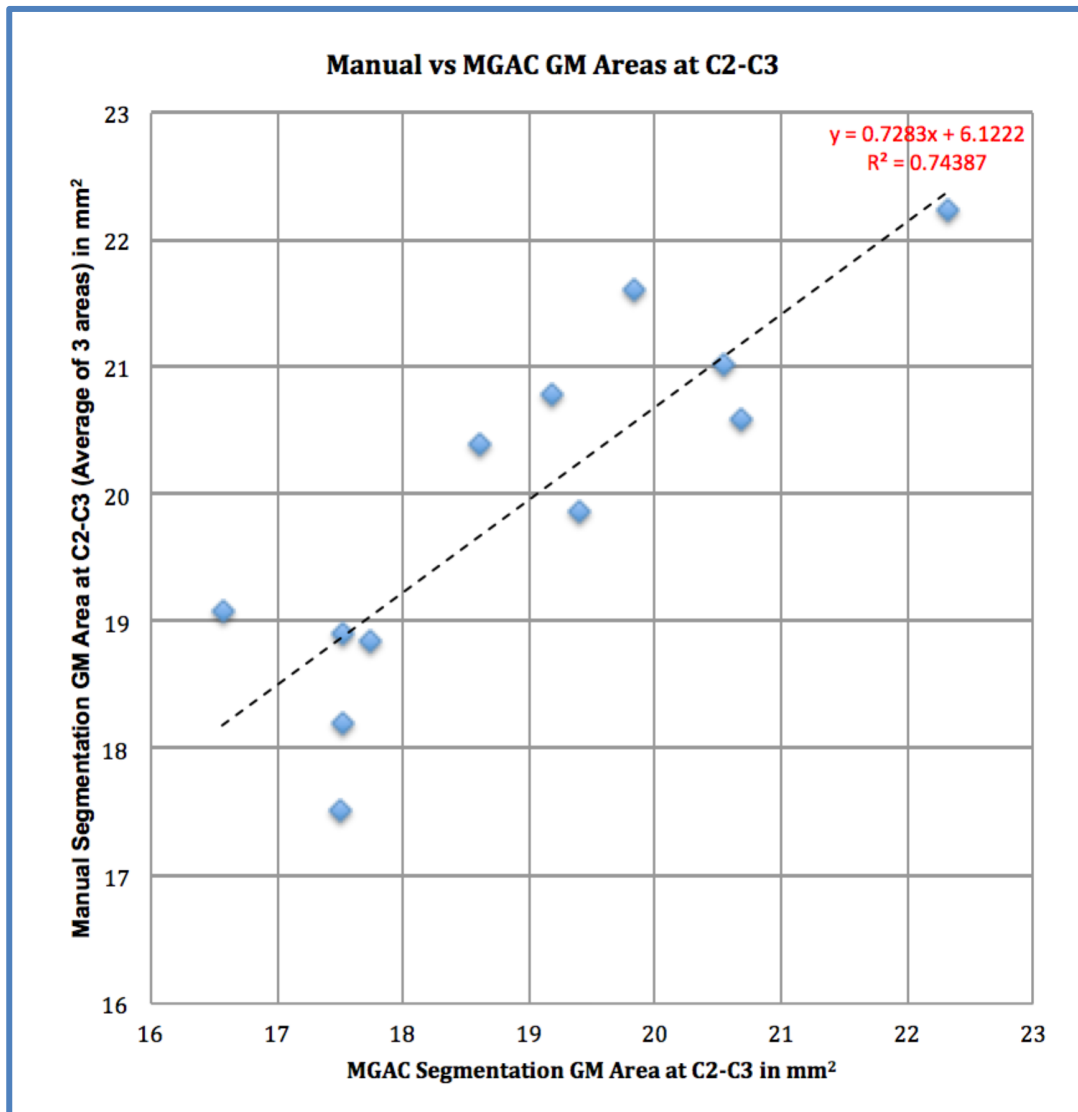
Figure 17: Comparison of spinal cord gray matter areas (in mm2) from the manual and MGAC segmentations in 12 healthy controls. The linear fit line with its equation and the correlation are displayed on the figure.

For comparison between manual and MGAC segmentations, a total of 34 PSIR images acquired from 21 subjects were segmented. 12 subjects had one scan, 8 subjects had both a test and a retest scan (and 2 out of these 8 had an additional test and retest scan that went unused), and 1 subject had both a regular and a high resolution scan. While all of the 34 MGAC segmentations had slight differences from the manual segmentations, 4 had significant differences in the shape topology, as shown in Figure 18. In one case, there was an incomplete continuation of the posterior

horn to the edge of the cord in the MGAC segmentation. This topological difference is understandable in cases with reduced gray matter contrast, where even manual operators find it necessary to extrapolate due to the reduced contrast. In three other MGAC segmentation cases, there was an additional bump on the side that did not appear in the manual segmentations. This topological difference is caused by darker intensities in the white matter that appear on the side of the gray matter. When this occurs, the algorithm detects the border of this hypointense area instead of the border of the gray matter. Currently, these differences are small and can easily be fixed with manual editing. In the future, these differences in topology can be avoided with modifications to the algorithm to constrain the gray matter to hit the edge of the cord mask and also to constrain the algorithm to look for a gradient where the hypointensity is on the correct side of the curve by providing the image contrast as prior information to the algorithm.
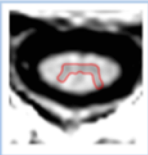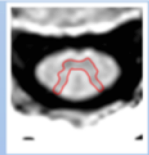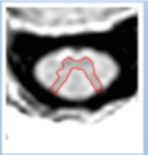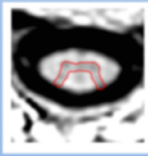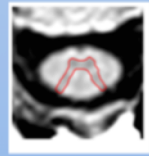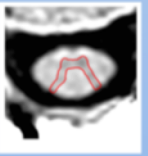
	Subject 1	Subject 2	Subject 3	
MGAC Segmentation Error				
Area (mm²)	18.00	15.05	20.45	21.29
Correct MGAC Segmentation on Subsequent Scan of Same Subject	None Available			
Area (mm²)		18.24	19.81	19.18
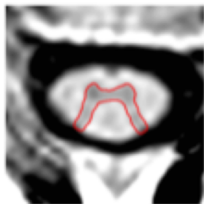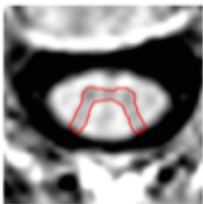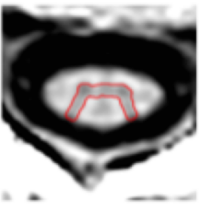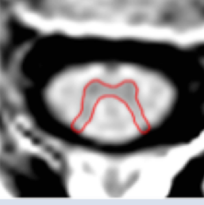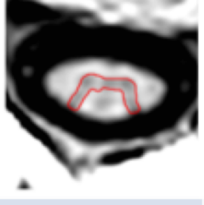
Figure 18: Topological Differences in MGAC Segmentations

3.3.2 Reproducibility

Eight additional subjects were scanned twice in the same session after repositioning. Reproducibility of the MGAC spinal cord gray matter areas was assessed in terms of percent change between test and retest scans. In addition, the intra-class correlation coefficient (ICC) was

calculated between the automatically segmented areas and the average of the manually segmented areas.

The absolute value of the percent change in area between the two scans was calculated for each pair and ranged from 0.54% to 6.61% with a mean of 2.85% and a standard deviation of 1.24%. The eight pairs of scans are shown in Figure 19. This variability corresponds to an interclass correlation coefficient between the test and retest MGAC segmentation of gray matter areas of .88. The twelve subjects with manual segmentations were segmented three times by a neurologist. In comparison, the percent test-retest changes for the twelve subjects with manual segmentations ranged from 0% to 6.51% with a mean of 2.38% and a standard deviation of 2.06%.

	Test-Retest Subject 1	Test-Retest Subject 2	Test-Retest Subject 3	Test-Retest Subject 4
Test Image MGAC Segmentation				
Retest Image MGAC Segmentation				
Percent Change	-2.26%	0.66%	0.54%	-6.61%
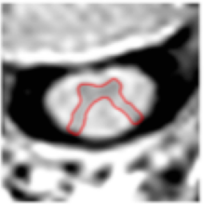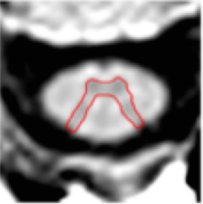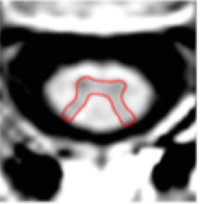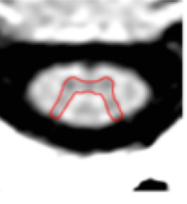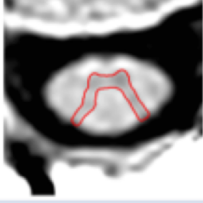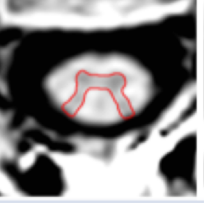
	Test-Retest Subject 5	Test-Retest Subject 6	Test-Retest Subject 7	Test-Retest Subject 8
Test Image MGAC Segmentation				
Retest Image MGAC Segmentation				
Percent Change	-5.43%	-3.23%	-0.95%	3.11%

Figure 19: Area percent changes for areas segmented with MGAC of 8 subjects who were each scanned twice with repositioning
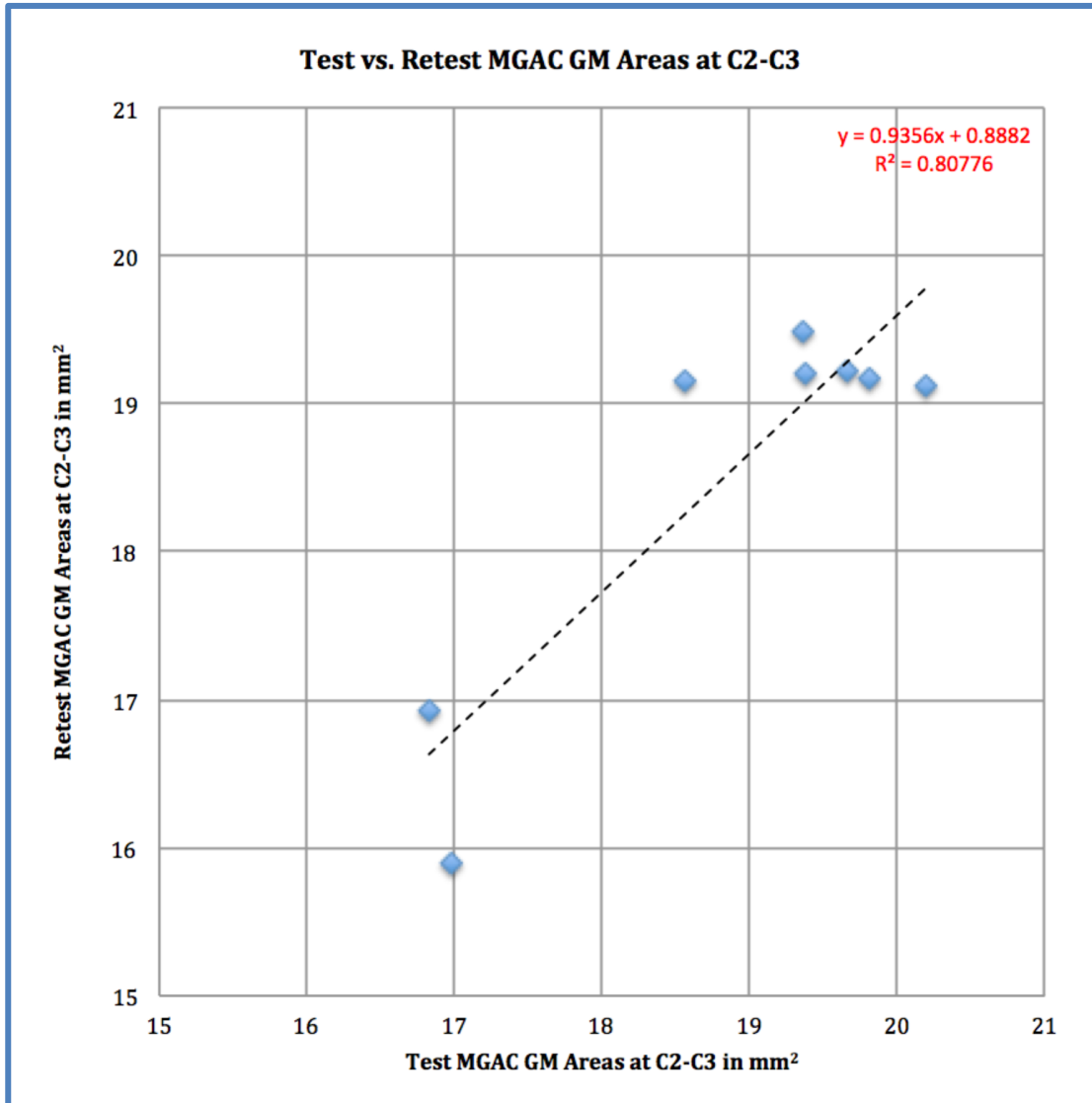
Figure 20: Comparison between the areas (in mm2) from the test and retest MGAC segmentations in 8 healthy controls. The linear fit line with its equation and correlation are displayed on the figure.

3.3.3 Validity

While no gold standard exists, we hypothesize that partial volume effects may overinflate the estimation of gray matter area and that this over-inflation is larger with the manual segmentation than with the MGAC segmentation. This hypothesis was formed after detecting a bias between the MGAC and manual approaches with the MGAC method yielding systematically lower gray matter area values for all subjects. We evaluated evidence for this hypothesis based on

one subject scanned at both low and high resolution as described in Figure 11. As shown in Figure 21, each of these scans was both manually and automatically segmented for spinal cord gray matter. Validity of the spinal cord gray matter areas computed from MGAC segmentation was assessed in terms of the percent change in cross sectional area from scans of different resolution.

We assume that the higher resolution scans will provide a more realistic estimate of gray matter segmentation than the low resolution scans, due to the higher level of detail. The percent change was 25% between the manual segmentations from high and low resolution images, while the percent change was 13% between the MGAC segmentations from high and low resolution images. These results suggest that both manual and MGAC methods may overestimate the gray matter area, with larger bias for the manual approach.
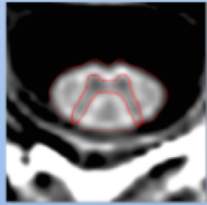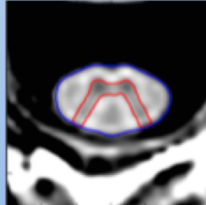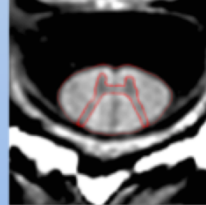
	Manual Segmentation: Low Resolution	MGAC Segmentation: Low Resolution	Manual Segmentation: High Resolution	MGAC Segmentation: High Resolution
Image				
Area (mm²)	18.09	17.02	14.12	14.88

Figure 21: Manual and MGAC segmentations of low and high resolution images

For the higher resolution images, it was necessary to tune the parameters of the MGAC algorithm in a different manner than for the lower resolution images. When keeping all other parameters constant at $\alpha = 1000$, $\mu = 5$, $\theta = 1$, $\nu = 1$ while increasing sigma, the errors in the contour decrease, but the detail in the horn contours also decreases.

3.3.4 Versatility

As a proof of concept, 2D T2* weighted images from four subjects were also segmented using the MGAC algorithm. Versatility of the MGAC method was assessed through visual inspection of the segmented spinal cord gray matter area from the T2* weighted scans.

In Figure 22, successful gray matter segmentations from 3 T2* weighted images acquired with the same protocol on 2 different Skyra scanners are shown. In two of the T2* subjects, there were slight errors in the cord segmentation due to hypointense blood vessels, seen in the blue contours in Figure 10. However, these minor errors did not have a large impact on the algorithm and gray matter segmentation was still successful in these cases. In the future, we plan to assess the performance on T2* weighted images with further testing on higher quality 3D T2* weighted images.
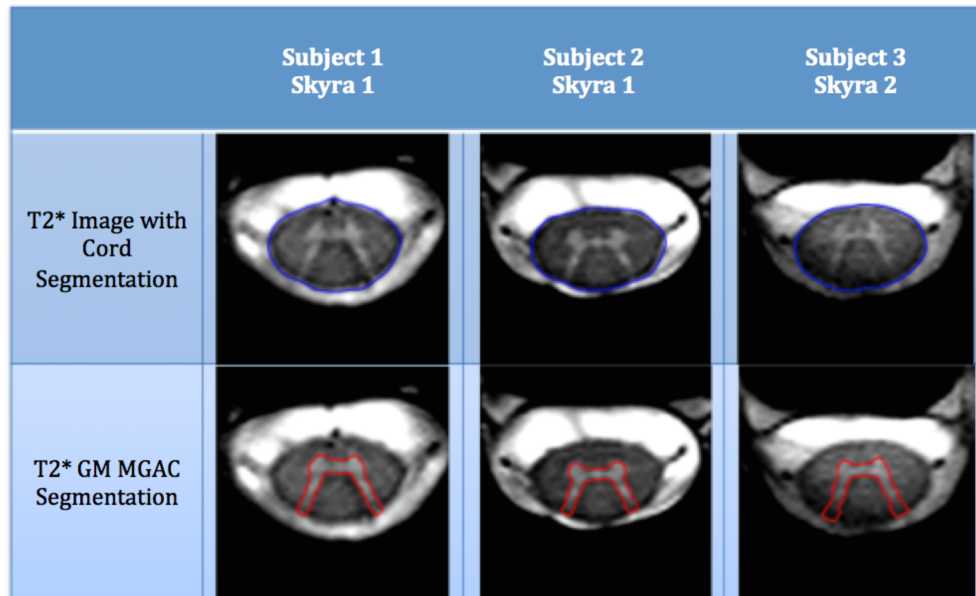
Figure 22: MGAC segmentation of T2 weighted images*

3.3.5   Potential for Clinical Applications

To demonstrate the potential for this method to be used in clinical applications, PSIR images from 58 multiple sclerosis patients who did not have obvious lesions at the C2C3 level were also segmented using the MGAC algorithm. The segmentations were assessed based on Dice similarity coefficients and Hausdorff distances between the manual and MGAC segmentations. Figure 23 shows the MGAC segmentations of these 58 patients in red overlaid with the manual segmentations in green. Each patient is labeled with the Hausdorff distance between the MGAC and manual segmentations. Of the 58 patients, 12 had Hausdorff distances greater than 1.0 mm, signifying the presence of significant errors. These instances were not found to be associated with the level of spinal cord atrophy or gray matter atrophy. Figure 24 shows how the gray matter areas from MGAC segmentation and manual segmentation compared to each other in the 46 patients with Hausdorff distances less than 1.0 mm between the segmentations. Figure 25 shows the MGAC segmentation failures in several MS patients with lesions at the C2/C3 level.
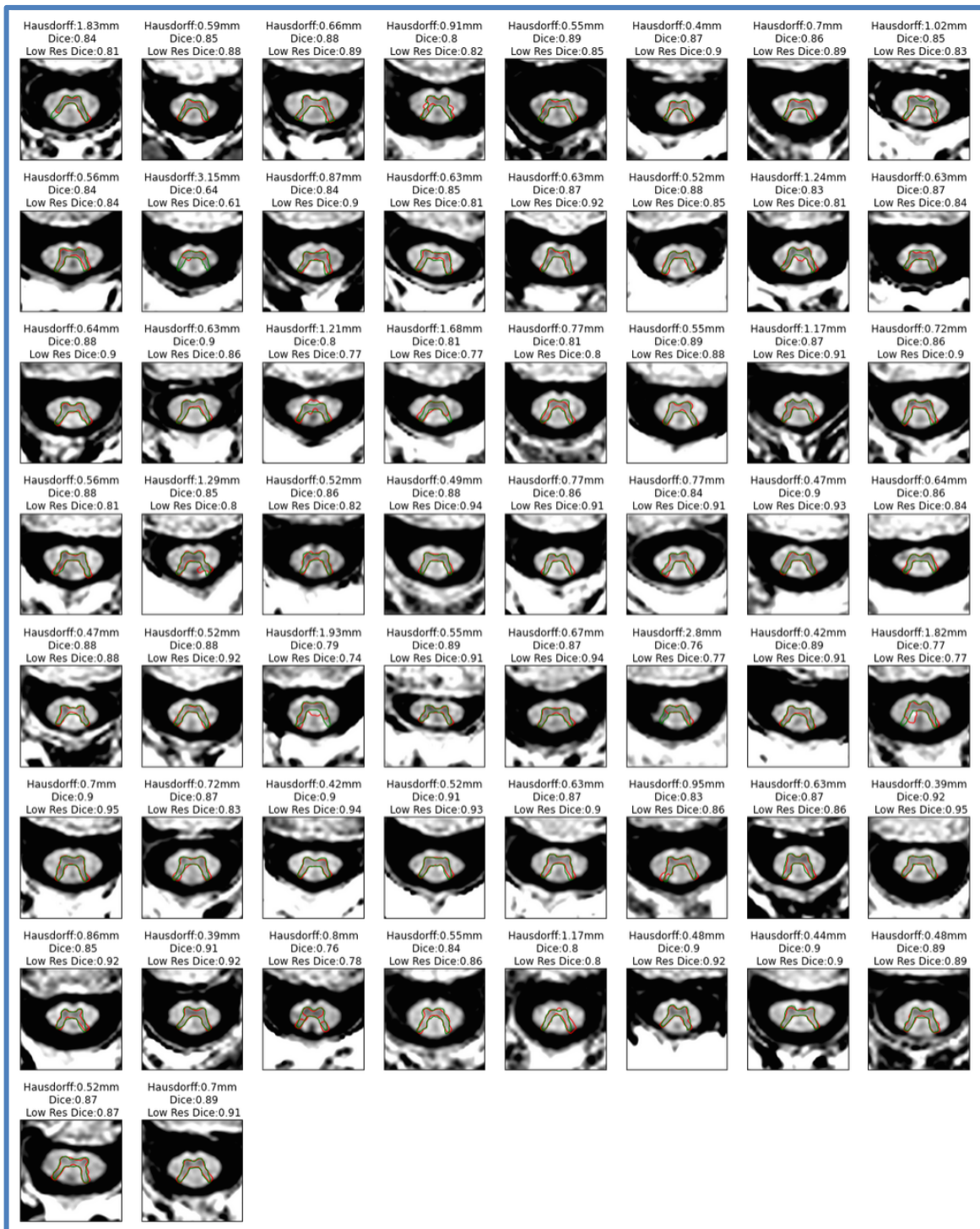
Figure 23: MGAC segmentation (red) and manual segmentation (green) of PSIR images from MS patients. Hausdorff distances between the MGAC and manual segmentations, Dice Similarity Coefficients, and Dice Similarity Coefficients in the original resolution are reported above each image.
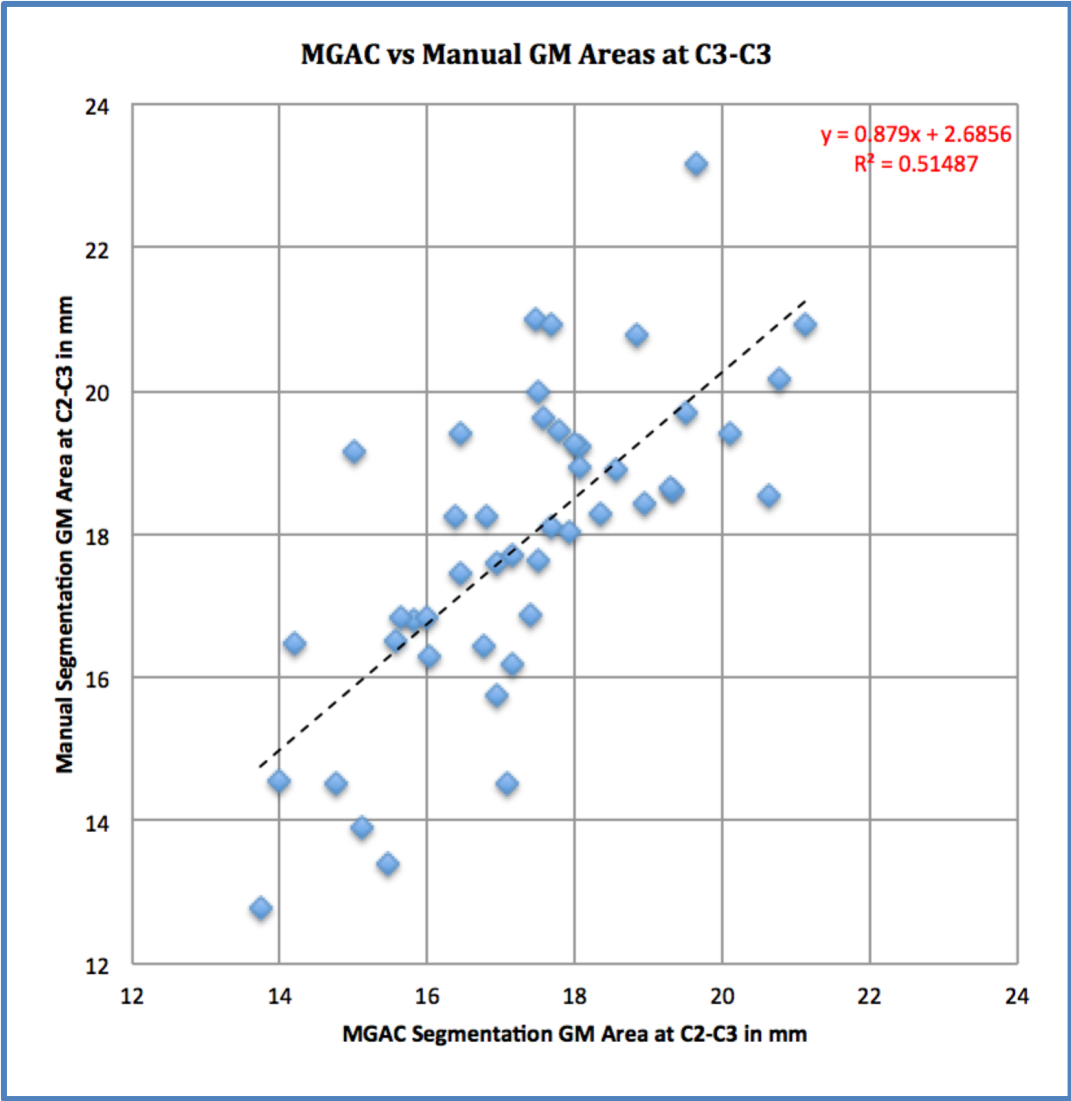
Figure 24: Comparison between the areas (in mm2) from the manual and MGAC segmentations in 46 MS patients. The equation of the linear fit model and the correlation is displayed on the graph.
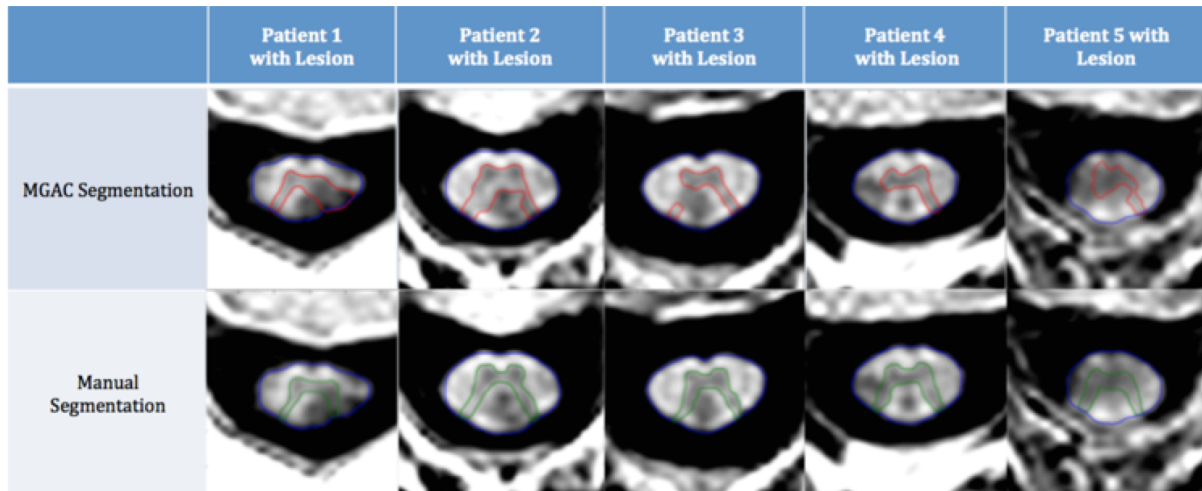
Figure 25: Examples of MGAC Segmentation Failures in MS Patients with Lesions

3.3.6 Extension to 3D Images and Other Cord Levels

One 3D PSIR image was used to demonstrate how this method could be extended for use with three-dimensional scans or 2D multi-slice scans including different cord levels. The 2D C2/C3 template was registered to every slice and the MGAC algorithm was applied in the usual manner to every slice. Segmentations were assessed through visual inspection. While there are errors in a few slices, the segmentation is still successful in several different levels of the spinal cord with substantial differences in gray matter shape. Though the 2D template was made with C2/C3 images, results show that it can still yield successful segmentations at surrounding levels with roughly similar gray matter structure. This example demonstrates the potential for the more general use of this algorithm in the future.
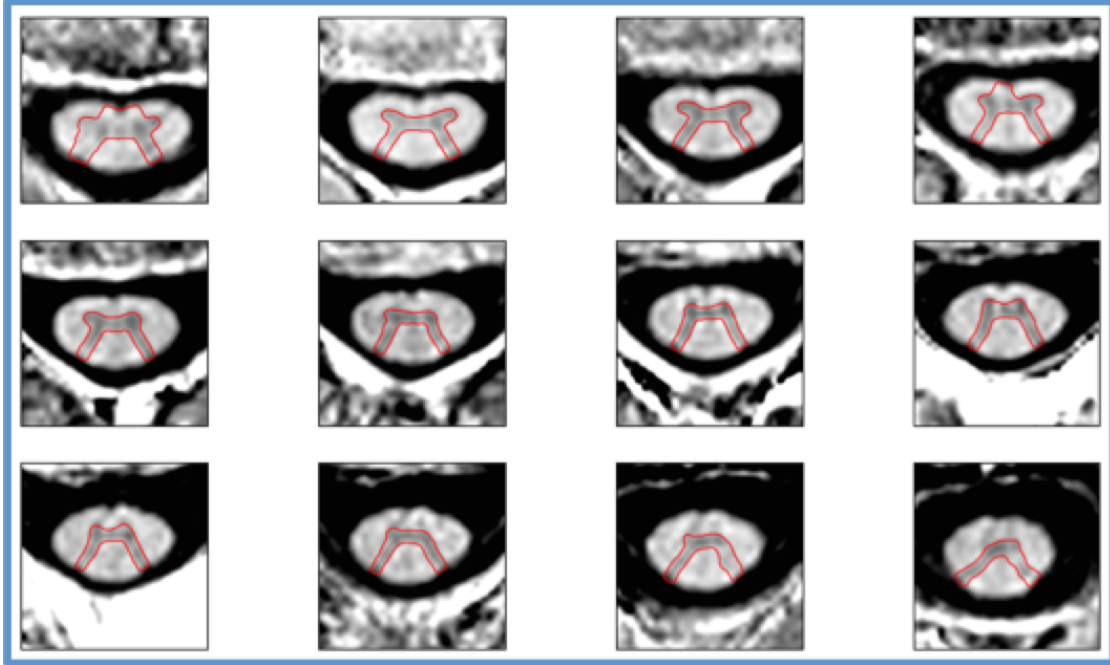
Figure 26: MGAC segmentation of 12 slices from a 3D PSIR T1 weighted image

3.4 Discussion

The MGAC segmentation method presented in this study provides an accurate and reproducible method of segmenting gray and white matter in the non-diseased spinal cord, thus enabling the computation of valuable spinal cord metrics such as gray matter area. The value of grey matter areas and volumes as future biomarkers of MS progression is indicated by previous studies that suggest that manually segmented spinal cord gray matter area provides a strong correlate with disability and progressive phenotype in multiple sclerosis. Additionally, these regions of interest can be used to determine quantitative imaging metrics within the spinal cord gray and white matter. Routine, robust estimates of spinal cord gray and white matter areas may provide outcome markers in clinical trials and potentially be of prognostic value.

To prove useful in clinical practice, an algorithm should be sufficiently robust so that it can be used on a variety of images acquired with a variety of scanners. The MGAC algorithm does not

depend on the absolute intensity values of the acquired image. Since the MGAC algorithm is based on finding the maximum gradient, it depends only on the relative intensity differences of the image. In other words, the algorithm simply requires an image with sufficient contrast between gray and white matter. In principle, the MGAC algorithm could be applied to other MR images that are used for the spinal cord 10,11, provided they offer a sufficient level of gray to white contrast. Our results showed that the MGAC algorithm was able to successfully segment gray matter areas in the T2* weighted images that were tested. Furthermore, in contrast to methods using intensity-based template registration 13,14, the dependence on pulse sequence specific generation of templates is reduced since only the shapes of the cord and gray matter templates are used, and these form only the initial input to the algorithm. The use of the same template was tested on another PSIR pulse sequence with higher resolution with positive results. Future work is needed nonetheless to explore and determine the sensitivity to shape templates for other different acquisition schemes.

Segmenting spinal cord gray matter with the MGAC algorithm is currently a semi-automated process, since the algorithm requires the delineation of the cord as an input. This cord delineation may be obtained with the software JIM in a semi-automatic process that requires a landmark. Currently, there are several tools (Amann et al., 2016) in development to automate the delineation of the edge of the spinal cord, such as PropSeg (Marios C. Yiannakas et al., 2016) from the Spinal Cord Toolbox. In the future, as these tools develop, this initial step can be automated so that MGAC will be fully automated.

The template used in this study was created using the manual gray matter segmentations of 20 2D PSIR images from healthy subjects. The number of subjects used to create this template is similar to that of the publicly available template (Fonov et al., 2014) from the Spinal Cord Toolbox, which was created with 16 subjects. Publicly available templates can be used with the MGAC

algorithm, though cohort specific templates that better capture the contrast, signal to noise, and other characteristics of a data set are likely to work best.

The MGAC gray matter areas tend to be slightly smaller than the gray matter areas from manual segmentation. From our study of higher resolution images, we found that both manual and MGAC approaches may be overestimat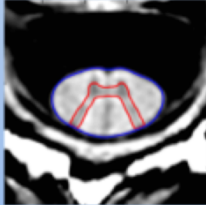ing the gray matter areas, with larger bias for the manual segmentations. This suggests that the MGAC segmentations may be more accurate than the manual segmentations. From visual inspection, it appears that the MGAC segmentations may leave less of a boundary around the gray matter in areas with partial volume averaging issues.

For the higher resolution images, it was necessary to tune the parameters of the MGAC algorithm to different values from those used for the lower resolution images. While tuning parameters, there is a tradeoff between segmentation errors that occur with complex contours and the loss of detail that occurs with more simple contours. A more rigorous exploration of parameters may help to alleviate this problem. It may also be beneficial to use a higher resolution template.

Automatic segmentation will enable routine measurements for tracking, but it is important to ensure that these measurements are as consistent as those done by manual segmentation. The mean percent change for the manual segmentations (2.38% +/- 2.06%) was slightly lower than mean percent change for the MGAC segmentations (2.85% +/- 1.24%). However, the standard deviation of the percent change for the manual segmentations is larger than that of the MGAC segmentations. For this reason, the ICC may be a more meaningful metric. The ICC for the reliability of the MGAC spinal cord gray matter segmentation was 0.88 when the algorithm was tested on two different subsequent scans of the same person. One study (Papinutto et al., 2015) reported an average coefficient of variation of 2.75% when the same operator segmented subsequent C2/C3 PSIR scans of the same person and an ICC value of 0.916 when different

45

operators manually segment the same C2/C3 PSIR scan from a healthy control. However, the ICC for MGAC on test-retest data (0.88) is very similar to the ICC from the manual segmentation results on the same data (0.91), signifying the reliability of this method. This level of accuracy may still not be enough to detect annual changes in gray matter in the clinic. The expected annual change to gray matter area in certain pathologies may be less than 2%, and thus would not be reliably inferred by either manual segmentations or MGAC segmentations at a single subject level.

While several automated and semi-automated methods for the segmentation of spinal cord gray matter area have been proposed, some of these require significant manual intervention, some are lacking in accuracy or robustness, and none have been demonstrated with images of patients with spinal cord atrophy. One proposed method based on a fuzzy connector method (M.C. Yiannakas et al., 2012) requires a significant amount of input from a trained expert. This method requires a seed region of interest to capture a partial segmentation and substantial manual editing is required to capture the rest of the segmentation.

Another study presented an automatic multi-atlas non-local statistical fusion method (Asman et al., 2013). The study exploring this method reported Dice similarity coefficients from 3D volumes ranging around 0.55 to 0.85 with a mean around 0.75. Our current method reports Dice similarity coefficients ranging from 0.82 to 0.93 in the 2D slice at C2/C3, with a mean of 0.89 when validated against manual segmentations. In the future, we hope to conduct a direct comparison of these two methods with the same data set.

Template-based tools (De Leener et al., 2017; Taso et al., 2015) have been shown to work fairly well for healthy controls. One study used a method that registered a spinal cord template to each individual subject to determine the segmented spinal cord gray matter area. This study reported intra-observer Dice coefficients of 0.90+/- 0.01 and an ICC value of 0.80. In comparison,

the MGAC method Dice coefficients are similar at 0.88 +/- 0.03 but the ICC is higher at 0.88. This template-based segmentation method is based on registration with a template of normal spinal cord structure and has not yet been tested on patients who experience gray matter atrophy that is independent from overall cord atrophy or on subjects acquired with a different imaging protocol from the template.

This method competed against five other methods in the Spinal Cord Gray Matter Segmentation Challenge in 2015. The MGAC method scored high amongst the methods in True Positive Rate, indicating the highest level of specificity. In addition, the MGAC method scored amongst the highest in both Skelotonized Hausdorff Distance and Skelotonized Median Distance, demonstrating the methods ability to determine the underlying shape of the GM. However, MGAC did not score as highly in True Negative Rate, representing a lower level of sensitivity. This lower sensitivity is also seen in the lower Mean Absolute Surface Difference and Positive Predictive Value scores. These results suggest that the MGAC method is excellent at determining the underlying shape of the GM, but may overestimate the GM volumes compared to human raters. One strong advantage of the MGAC algorithm over the other methods is its ability to work on images with different contrasts. This algorithm was developed for use on PSIR images, but has also been shown to work well on T2*weighted images.

In the future, we plan to address the current limitations of the MGAC method. Of the 34 automatic segmentations performed on healthy controls, there was one case with a major topological difference from the manual segmentation and three cases with minor topological differences from the manual segmentation due to insufficient contrast in the image. Both of these types of errors are due to issues with the balance of prior information and image information. By

modifying the algorithm so that these common mistakes are recognized, these errors may be avoided.

Our results show that in 46 out of 58 patient cases, the MGAC segmentation had a Hausdorff distance less than 1.0 mm from the manual segmentation. However, there are still more challenges to address for the clinical use of this algorithm. The patient cohort did not include patients with obvious lesions that may confound the MGAC algorithm. Since the segmentation is expected to be further from the prior contour in patients than in the controls, the balance of prior information and image information may need to be adjusted for better recognition of atrophy. In addition, the sample of 58 multiple sclerosis patients may not fully capture the variability of morphological shapes that may be seen in other pathologies with gray matter atrophy.

This paper demonstrated the use of MGAC in 2D scans at the C2/C3 level, and used a single case to show one method for extending the algorithm to other similar cord levels or to 3D images. However, there are many other strategies that could be employed as well. Instead of using registration of a 2D C2/C3 slice template to each slice, slice level cross-sectional templates can be created for the other cervical and thoracic levels and applied appropriately. Another option is to use a 3D template that can be registered to the 3D cord shape. Subsequently, a 3D morphological geodesic active surface model can then be applied in a similar way. Alternatively, MGAC can be applied to the C2/C3 slice of the volume that matches the template. The result can than be used as the prior for the subsequent slice and using an iterative approach, the process can be propagated throughout the 3D volume.

This study has shown that the MGAC segmentation method is successful in estimating gray matter volume in healthy subjects and in MS patients without evident lesions at the imaged level. However, the true power of this method will come from accurate segmentation in other types of

patients, who may have confounding features. From our initial investigation, we have shown that the presence of significant lesions often leads to errors and provides inaccurate results. In Figure 13, the first four examples shown could easily be corrected with simple edits, such as the removal of a small bump or continuation of an existing curve. However, the last example shows a case where the segmentation could not be salvaged with simple manual editing. In these cases, the lack of contrast between lesions and gray matter may undermine the accuracy of any method, including manual segmentation. With further tuning of these parameters, we can change the balance for how much the algorithm depends on the prior shape information versus the intensity information in the image. In the future, a lesion mask could also be input to the MGAC algorithm to aid in segmentation. This should help the algorithm provide accurate assessments for the gray matter area in patients with spinal cord lesions.

In conclusion, this study presents a semi-automated method for segmentation of spinal cord gray and white matter that can be used with a variety of image protocols. Compared to other proposed methods, the MGAC method shows a similar level of accuracy against manual segmentations and a higher level of precision between test and retest scans. This method will allow for easy and routine longitudinal measurements of spinal cord gray and white matter areas, which can be used to further improve our understanding of neurodegenerative processes in disorders affecting the spinal cord such as MS.

3.5 Additional Implementation of Convolutional Neural Networks for Spinal Cord Gray Matter Segmentation

Convolutional neural networks have proven to be very effective in the use of classification and segmentation tasks and this method has recently gained popularity in medical imaging

49

(Greenspan, van Ginneken, & Summers, 2016). A neural network is an artificial intelligence computing method that is inspired by biological neural processes. In a computational neural network, a collection of nodes sends signals to each other through connections, just as biological neurons send signals to each other via synapses. These nodes are organized in layers and signals travel from the input layer to the output layer. The connections each have a weight that controls the signal strength. Each node has a threshold and when the aggregate of the received signals passes the threshold, the node sends its own signal. Using a technique called backpropagation, the error from each neuron is calculated after processing a batch of data, and the network can learn the correct signal weights with gradient descent. Neural networks gained popularity as Graphical Processing Units (GPUs) made the computation of back-propagation possible for many layered networks. Methods using neural networks won many international competitions in pattern recognition, including

Convolutional neural networks (CNNs) are a specialized type of neural networks that have an architecture explicitly designed for image inputs and constrained in a sensible way so that the number of parameters decreases. In a convolutional neural network, there are convolutional layers that are made up of a set of learnable filters. Each of these filters is slid across the width and height of the input image to produce an output that gives a filter response at every pixel in the image. As the network learns the correct weights, it will learn the correct filters that activate when they encounter certain features, such as a horizontal edge or a circular pattern. In 2012, a convolutional neural network method called AlexNet (Krizhevsky, Sutskever, & Hinton, 2017) won the ImageNet Large-Scale Visual Recognition Challenge, and since that time, convolutional neural networks have been used for many different applications.

Several different architectures have shown to be useful in the field of semantic segmentation at the pixel level. Originally, semantic segmentation was performed using a traditional architecture of a convolutional neural network (Farabet, Couprie, Najman, & LeCun, 2013) to label each pixel with a label. A later study (Long, Shelhamer, & Darrell, 2015) proposed using fully convolutional networks where convolutions with kernels covering the entire input region were used instead of fully connected layers. This made semantic segmentation much more efficient since it used fewer parameters. The most typical CNN architectures used in medical imaging segmentation are fully convolutionalneural networks and U-Nets (Ronneberger, Fischer, & Brox, 2015). U-Nets use data augmentation to use training data more effectively and were specifically designed for the application of biomedical image segmentation with smaller training data sets.A recent study investigated the application of spinal cord gray matter and used deep dilated convolutions to segment in 3D MPRAGE images (Perone, Calabrese, & Cohen-Adad, 2018). This study was tested on the data set from the 2015 ISMRM Spinal Cord Gray Matter Segmentation Challenge and achieved an overall dice similarity coefficient of .85.

3.6 Methods

In this work, I have implemented a similar architecture of deep dilated convolutions and trained it solely on zoomed PSIR images from the C2/C3 area of the spinal cord taken from our Siemens Skyra scanner. In this way, the weights learned by the network are more specific to our application and the classification should be more accurate. I have also similarly used a Dice loss function rather than a traditional loss function of cross-entropy, since only a small portion of the original image is made up of the gray matter to be segmented.

The model was implemented in Python using Keras with TensorFlow backend and the code can be found on github at *https://github.com/edatta/SpinalCordGMSegmentCNN*. A dataset of 130 patients and 20 healthy controls was used to train, validate, and test the model. The training set consisted of 130 images, while the validation and test set each contained 15 images that were randomly selected. The model was subsequently run on 883 images from 370 patients and the results were assessed visually.

3.7 Results

After training and validating the model using 145 images, the resulting accuracy on the 15 image test set was .987 with a Dice coefficient of .789. The results of the test set are shown below in Figure 27. The segmentations delineated by an expert neurologist are shown in blue, while the results of the convolutional neural network are shown in blue.
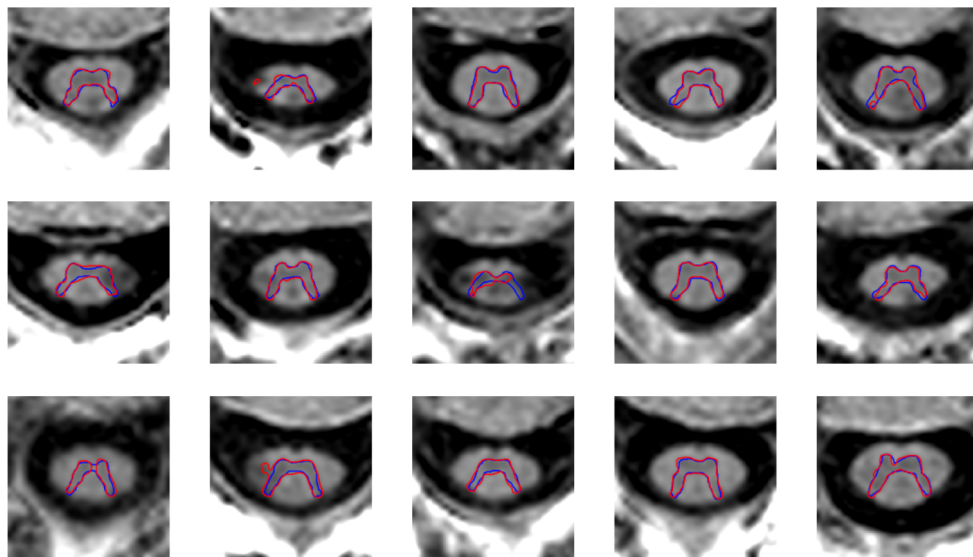
Figure 27: Test Set Results

3.8 Discussion

From the results, we can see some of the most common mistakes made by the convolutional network. While the active contour method creates a single contour, the convolutional neural network judges each voxel independently. For this reason, the resulting mask is often disconnected or of the wrong topography. However, with a larger training set, we would expect that the frequency of these mistakes would greatly decrease.

Many other common methods seem to rely too much on prior information and often find a gray matter segmentation when there is little information in the image. For this reason, these methods may be overestimating the accuracy rate. It is important to judge methods critically and look at the accuracy rate only on images that can be judged by human readers as well. Our results show that when segmentation fails on a certain time point for a certain subject, the segmentation is likely to fail similarly at a later time point. This fact is reassuring since it indicates that these failures are driven by biological differences rather than by image quality or by the algorithm. algorithm. In the future, a more accurate model could be achieved by creating a larger training set.

Chapter 4 - Piloting a Longitudinal Spinal Cord Study

4.1 Introduction

As new developments unfold, spinal cord imaging has grown increasingly important in the investigation of multiple sclerosis (MS). MRI measures of cord size have been shown to be sensitive to the axonal loss and pathological processes seen in MS (Kearney et al., 2015). Recent studies have revealed the potential of cross-sectional spinal cord area from phase-sensitive inversion recovery (PSIR) images as a biomarker correlated with clinical disability in MS (Schlaeger et al., 2014, 2015). This finding suggests that cross-sectional spinal cord area may also be useful as a metric in clinical trials to evaluate the effectiveness of treatments in improving clinical disability.

In order to be used effectively in a clinical study, a metric must have a sufficiently large effect size in the patient population that the improvement due to treatment can be observed with a reasonable number of subjects. This means that if the longitudinal changes are small, the metric must also be sensitive and highly reproducible. In addition, the metric should have a strong correlation with clinical metrics since treatments are designed to improve disability and quality of life and not just underlying biological effects of the disease. The aim of a clinical trial is to measure the difference of the metric in a treatment group relative to a comparator group. If the treatment is successful, the metric in the treated group should approach the value of the metric that is seen in healthy subjects or MS patients with a milder disease course.

Traditionally, MS clinical trials have used metrics such as whole brain atrophy, cortical gray matter atrophy, and subcortical gray matter atrophy to investigate the effectiveness of treatments and many studies have attempted to quantify the sample sizes required to use these

metrics in a clinical trial (Altmann et al., 2009; Anderson, Bartlett, Fox, Fisniku, & Miller, 2007; Kim et al., 2017; Nakamura et al., 2014). However, none of these studies have taken into account differences in atrophy that may be due to other covariates that may affect cohorts differently. Many studies (Bakshi, Dandamudi, Neema, De, & Bermel, 2005; Bonati et al., 2011; Brex et al., 2001; M Filippi et al., 1997; Furby et al., 2010; Lin, Tench, Evangelou, Jaspan, & Constantinescu, 2004; Losseff et al., 1996; Lukas et al., 2015; Nijeholt et al., 1998; Rashid et al., 2006; M. A. Rocca et al., 2011) have investigated the role of upper cervical cord area in MS. In addition, many studies have been conducted to determine the relationship of spinal cord metrics with clinical disability (Jacobsen et al., 2014; Sormani, Arnold, & De Stefano, 2014; Steenwijk et al., 2016). The majority of these studies investigate the correlation of metrics with EDSS score. However, EDSS score is an imperfect measure in many ways since it reflects a variety of motor, sensory, and cognitive symptoms and none of these studies have compared the specific relationships of these metrics with motor disability or cognitive disability.

Inherently, spinal cord atrophy is much more difficult to measure precisely than whole brain or brain gray matter atrophy. Due to the small size of the spinal cord, segmentation mistakes have a much larger effect when measuring the annual percent change between scans. Perhaps for this reason, very few clinical trials have used spinal cord atrophy as an outcome measure (Kalkers, Barkhof, Bergers, van Schijndel, & Polman, 2002; Leary et al., 2003; Montalban et al., 2009). As spinal cord imaging has improved, recent studies (Cawley et al., 2018; Kearney et al., 2014) have used small cohorts to conduct preliminary investigations into determining the potential of using spinal cord atrophy in clinical trials and quantifying the effect sizes. One of these studies determined that segmentations done with an active surface model on a PSIR image are much more reproducible than those done with T1-weighted images or those done with an edge detection

method (Kearney et al., 2014). In this study, we provide estimates based on active contour models of 2D PSIR images (Papinutto et al., 2015) at the C2/C3 level, an efficient approach to quantify spinal cord atrophy, using a large cohort that includes both progressive and relapse-onset MS patients.

The study presented in this paper calculates the sample sizes required to detect a treatment effect on atrophy rates in upper cervical cord area, whole brain volume, and brain gray matter volume. The sample sizes are calculated by comparing patient cohorts to three different reference groups that represent the ideal target for treated patients. These three groups represent non-progressive patients with stable EDSS scores, without cognitive decline, and without motor decline. The calculations were performed after the atrophy rates were adjusted for covariates, such as age, gender, disease duration, and number of relapses in the previous two years.

4.2   Methods

4.2.1   Patient Cohort

The patients in this study were selected from a larger observational cohort that was screened to include subjects who were older than 18 years of age with a diagnosis of MS or clinically isolated syndrome (CIS) according to international criteria. The cohort included patients with relapsing MS (RMS) and progressive MS (PMS). The PMS group included patients with both primary progressive MS (PPMS) and secondary progressive MS (SPMS). Any patients who had ongoing symptoms from a recent relpase, a recent history of drug or alcohol abuse, a diagnosis of hepatitis B or C or human immunodeficiency virus, participation in ongoing MS trials with unlicensed drugs, or any concurrent illness or disability were excluded. In this study, we analyzed only those subjects who had undergone scans that were at least 180 days apart. In addition, any

patients with active flare of MS symptoms were excluded as well. 196 patients were identified who met these criteria. Many of these patients were treated with various medications as described in Table 6 in the eAppendix. Our study did not include an analysis of the PPMS patient group, since it only included 10 patients.

4.2.2 Reference Groups

In this study, we defined three different reference groups to represent the ideal clinically responsive patients in a hypothetical clinical trial. All three of these groups were made up of RMS patients. The first group represents patients without general disease progression and includes RMS patients whose Expanded Disability Status Scale (EDSS) scores had not increased over the study period. The second group represents patients without motor decline and includes RMS patients whose dominant 9-hole peg test (9HPT) scores had not increased over time by more than 2 seconds. The third group represented patients without cognitive decline and included RMS patients whose Symbol Digit Modalities Test (SDMT) scores had not decreased over time. We chose conservative definitions of clinical stability by including only those patients showing no increase in EDSS or SDMT and using only a 2 second allowance for increases in 9HPT. The remaining patients were used as hypothetical comparator groups.

4.2.3 Image Acquisitions

For this study, 2D phase sensitive inversion recovery (PSIR) images were acquired axially at the intervertebral disc level C2/C3 perpendicular to the cord in 196 MS patients with a 3T Skyra scanner and a 20 channel head-neck coil. In addition, T1-weighted brain sagittal MPRAGE images were acquired in the same patients. The acquisition parameters are shown in Table 1.

	T1-Weighted Image Acquisition Parameters	C2/C3 2D PSIR Acquisition Parameters
Voxel Size:	1 x 1 x 1 mm^3	.78 x .78 x 5 mm^3
FOV:	240 x 256 x 176 mm	200 x 200 x 5 mm
TR:	2300 ms	4000 ms
TI:	900 ms	400 ms
TE:	2.98 ms	3.22 ms
Flip Angle:	9 degrees	10 degrees
Acquisition Time:	5 minute 12 seconds	1 minute 50 seconds

Table 1 Pulse sequence parameters for T1 weighted brain images

4.2.4   Measuring Atrophy Rates

At each time point, the cross sectional cord area at C2/C3 was measured using the software JIM (Horsfield et al., 2010) to apply active contour models on 2D PSIR images (Papinutto et al., 2015) and the brain gray matter volume was measured using SIENAX (Smith et al., 2002, 2004). The annual percent change in spinal cord area and the annual percent change in brain gray matter volume were both calculated. In addition, the whole brain percent volume change was measured using SIENA (Smith et al., 2002, 2004). For each reference group and the corresponding patient groups used for comparison, the average and standard deviation of the atrophy rates were calculated. In this analysis, the clinically stable comparator cohorts are assumed to be the "treated" group in a hypothetical study while the remaining patients are assumed to be in the "comparator" group.

4.2.5   Adjusting for Covariates

To account for the effect of covariates such as the age, gender, disease duration, number of relapses in the previous 2 years, and the baseline metric at the first timepoint, adjusted values were computed using an ANCOVA analysis with LASSO regression used to aid in covariate selection. For each analysis, a LASSO regression is performed with atrophy rate as the outcome variable and reference group membership and the covariates as the input variables. For a hypothetical clinical

study, the reference group membership variable represents whether the subject has received the treatment or not. The adjusted mean and adjusted standard deviation are estimated using the least squares mean and the root mean square error from this model. If the coefficient of reference group membership is set to zero in the LASSO regression, this indicates that the atrophy rates cannot be adequately distinguished based on reference group membership or that reference group membership is highly correlated with one of the other covariates.

4.2.6 Power Calculation

The least square means and root mean square errors from the clinically stable reference group and the hypothetically untreated group are then used to determine the required sample size to distinguish a treatment effect. Since the clinically stable reference group is made up of patients, power calculations were done for a 100% treatment effect which assumed that the metrics in the hypothetically treated group would fully reach the levels of the clinically stable patients. Our calculations were done for 80% power, which is equivalent to a 4 to 1 weighting of false negative risk and false positive risk. We also assumed a Type I error rate or false positive risk of 5%. The calculations assume a two-sample, one-sided test to determine whether the mean annual atrophy rate of a treated patient group is detectably less than the mean annual atrophy rate of a patient comparator group.

The sample sizes are calculated with the following formula:

$$n = (\sigma_A^2 + \sigma_B^2)(\frac{z_{1-\alpha} - z_{1-\beta}}{\mu_A - \mu_B})^2$$

In this formula, we specify the required sample size per arm, the least squares mean annual atrophy rate from the clinically stable reference group, the least squares mean atrophy rate from

the patient comparator group, the root mean square error from the model, the desired power, and the desired Type I error rate.

4.3 Results

4.3.1 Demographics

Table 2 shows the demographics of the patient cohorts as well as the clinically stable reference groups. In addition, the table shows the medications included in each treatment tier.

	RMS	SPMS	All PMS	All Patients	RMS wo Increased EDSS	RMS wo Motor Decline	RMS wo Cognitive Decline
Number	159	27	37	196	90	144	86
Mean Age +/- SD	47.09 +/- 10.25	53.7 +/- 9.8	53.49 +/- 10.71	48.3 +/- 10.61	46.8 +/- 10.46	46.99 +/- 10.19	47.02 +/- 10.27
% Female	71.7%	55.56%	56.76%	68.88%	68.89%	70.83%	70.93%
Mean Disease Duration +/- SD	13.3 +/- 8.96	20.67 +/- 8.29	18.3 +/- 9.52	14.24 +/- 9.26	12.94 +/- 9.22	13.19 +/- 9.02	13.2 +/- 8.0
Mean Days Between Timepts +/- SD	753.66 +/- 267.07	738.85 +/- 254.01	783.57 +/- 236.06	759.31 +/- 261.18	751.54 +/- 263.23	775.58 +/- 262.19	747.21 +/- 274.02
Treatment Tiers (1/2/3/NR)	53/42/16/48	5/11/6/5	7/13/6/11	60/55/22/59	32/22/14/22	47/39/15/43	31/19/9/27
Mean Number of Relapses in Previous 2 Years	0.45 +/- 0.75	0.19 +/- 0.4	0.14 +/- 0.35	0.39 +/- 0.7	0.49 +/- 0.77	0.45 +/- 0.72	0.4 +/- 0.72
Median EDSS (Interquartile Range)	2.0 (1.5-2.5)	5.5 (4.0-6.0)	5.5 (4.0-6.0)	2.0 (1.5-3.5)	2.0 (1.62-3.0)	2.0 (1.5-2.62)	2.0 (1.5-2.5)
Mean Cross Sectional Cord Area at Baseline +/- SD	76.68 +/- 8.68	66.29 +/- 8.75	67.28 +/- 8.01	74.91 +/- 9.3	76.98 +/- 8.76	77.08 +/- 8.89	76.04 +/- 8.34
Mean Brain Volume at Baseline +/- SD	1504.7 +/- 83.62	1454.66 +/- 61.71	1461.06 +/- 63.57	1496.46 +/- 81.88	1509.64 +/- 82.41	1506.27 +/- 83.78	1499.44 +/- 85.1
Mean Brain GM at Baseline +/- SD	779.08 +/- 49.1	754.04 +/- 41.68	753.66 +/- 40.61	774.28 +/- 48.55	779.01 +/- 47.83	779.44 +/- 50.06	776.52 +/- 50.25
Mean Cortical GM at Baseline +/- SD	633.52 +/- 40.46	615.02 +/- 35.72	614.84 +/- 34.59	629.99 +/- 40.02	634.23 +/- 39.33	634.25 +/- 41.11	631.77 +/- 42.62

Treatment Tier	Treatments
1	Copaxone, Avonex, Rebif, Betaseron, Aubagio, IVSM, Extavia, Ampyra, and Low Dose Naltrexone
2	Gilenya and Tecfidera
3	Rituximab and Tysabri

Table 2 Subject and reference group demographics

4.3.2 Spinal Cord, Whole Brain, and Brain Gray Matter Atrophy Rates

Table 3 lists the atrophy rates observed in the clinically stable "treated" reference groups in comparison to the patient "comparator" groups.

		Reference Group	RRMS and CIS	SPMS	All PMS	All Patients
RMS wo Increased EDSS	Spinal Cord Annual Change	-0.23% +/- 1.53%	-0.56% +/- 1.56%	-0.96% +/- 2.04%	-1.03% +/- 1.95%	-0.72% +/- 1.72%
	Brain Annual Change	-0.42% +/- 0.39%	-0.41% +/- 0.49%	-0.42% +/- 0.3%	-0.54% +/- 0.43%	-0.46% +/- 0.48%
	Brain GM Annual Change	-0.52% +/- 1.06%	-0.74% +/- 1.41%	-0.43% +/- 0.92%	-0.78% +/- 1.24%	-0.75% +/- 1.35%
	Cortical GM Annual Change	-0.78% +/- 1.87%	-1.37% +/- 2.56%	-0.44% +/- 1.81%	-1.3% +/- 2.78%	-1.34% +/- 2.64%
RMS wo Motor Decline	Spinal Cord Annual Change	-0.32% +/- 1.45%	-0.88% +/- 2.26%	-0.96% +/- 2.04%	-1.03% +/- 1.95%	-0.99% +/- 2.04%
	Brain Annual Change	-0.41% +/- 0.39%	-0.55% +/- 0.75%	-0.42% +/- 0.3%	-0.54% +/- 0.43%	-0.54% +/- 0.54%
	Brain GM Annual Change	-0.59% +/- 1.23%	-0.88% +/- 1.13%	-0.43% +/- 0.92%	-0.78% +/- 1.24%	-0.81% +/- 1.21%
	Cortical GM Annual Change	-1.03% +/- 2.28%	-1.01% +/- 1.51%	-0.44% +/- 1.81%	-1.3% +/- 2.78%	-1.22% +/- 2.48%
RMS wo Cognitive Decline	Spinal Cord Annual Change	-0.33% +/- 1.41%	-0.43% +/- 1.71%	-0.96% +/- 2.04%	-1.03% +/- 1.95%	-0.63% +/- 1.81%
	Brain Annual Change	-0.36% +/- 0.43%	-0.49% +/- 0.44%	-0.42% +/- 0.3%	-0.54% +/- 0.43%	-0.51% +/- 0.43%
	Brain GM Annual Change	-0.45% +/- 1.34%	-0.81% +/- 1.05%	-0.43% +/- 0.92%	-0.78% +/- 1.24%	-0.8% +/- 1.12%
	Cortical GM Annual Change	-0.82% +/- 2.44%	-1.28% +/- 1.88%	-0.44% +/- 1.81%	-1.3% +/- 2.78%	-1.29% +/- 2.23%

Table 3 Atrophy rates

4.3.3 Required Sample Size for a Clinical Study

For each analysis performed, Table 4 lists the sample size required to adequately detect a 100% treatment effect from a clinically stable "treated" group after atrophy rates are adjusted for covariates. Table 5 lists the sample sizes calculated to adequately detect a treatment effect from an assumed reference value of zero atrophy using unadjusted atrophy rates. Additional calculations of sample size based on the original unadjusted atrophy rates are shown in Table 6. Note: NS indicates that in this analysis, there was no significant difference between the adjusted values from the "treated" group and the "comparator" group since treatment did not show up as a significant covariate in the regression model. This could possibly indicate that our cohort was too small to determine any actual difference.

61

		RRMS and CIS	SPMS	All PMS	All Patients
RMS wo Increased EDSS	Spinal Cord Annual Change	NS	26	24	88
	Brain Annual Change	NS	NS	NS	NS
	Brain GM Annual Change	NS	NS	NS	255
	Cortical GM Annual Change	145	NS	116	139
RMSwo Motor Decline	Spinal Cord Annual Change	NS	NS	53	41
	Brain Annual Change	NS	NS	NS	NS
	Brain GM Annual Change	NS	NS	NS	NS
	Cortical GM Annual Change	NS	NS	NS	631
RMS wo Cognitive Decline	Spinal Cord Annual Change	NS	NS	183	NS
	Brain Annual Change	NS	NS	NS	117
	Brain GM Annual Change	NS	NS	NS	158
	Cortical GM Annual Change	568	NS	NS	326

Table 4 Sample sizes calculated using adjusted atrophy rates and reference groups of clinically stable patients

	Treatment Effect	RMS	SPMS	All PMS	All Patients
Spinal Cord Annual Change	30%	1514	620	495	781
Spinal Cord Annual Change	50%	545	224	179	282
Brain Annual Change	30%	137	72	87	150
Brain Annual Change	50%	50	26	32	54
Brain GM Annual Change	30%	497	631	350	445
Brain GM Annual Change	50%	179	228	126	160
Cortical GM Annual Change	30%	638	2352	624	529
Cortical GM Annual Change	50%	230	847	225	191

Table 5 Sample Sizes Calculated Using Reference Groups with Unadjusted Atrophy Rates

		RRMS and CIS	SPMS	All PMS	All Patients
RMS wo Increased EDSS	Spinal Cord Annual Change	279	77	61	137
	Brain Annual Change	NS	NS	160	2406
	Brain GM Annual Change	426	NS	257	355
	Cortical GM Annual Change	180	NS	251	202
RMS wo Motor Decline	Spinal Cord Annual Change	143	96	74	89
	Brain Annual Change	204	5766	116	145
	Brain GM Annual Change	201	NS	531	385
	Cortical GM Annual Change	NS	NS	1096	2038
RMSwo Cognitive Decline	Spinal Cord Annual Change	2613	95	73	345
	Brain Annual Change	128	416	69	102
	Brain GM Annual Change	137	NS	193	154
	Cortical GM Annual Change	287	NS	368	317

Table 6 Sample sizes calculated using reference groups with unadjusted atrophy rates

4.4 Discussion

Sample size calculation is a key aspect in the planning of any trial. This study estimated sample sizes for clinical trials in MS for spinal cord metrics and conventional brain metrics based on a large cohort of MS patients.

In this study, we chose to use groups of patients who had not progressed in their clinical symptoms across MS disabilty, sensorimotor, and cognitive domains as references. Traditionally, studies that perform sample size calculations often choose to either assume a reference group of healthy controls or use a comparison value of zero atrophy. However, a reference group with zero percent change may be unrealistic because atrophy also occurs due to aging and other factors. The choice of a patient rather than a healthy control reference group is also more suitable since a successfully treated MS patient may not have the atrophy characteristics of a non-diseased subject in terms of abatement of disability progression. For example, in relapsing patients with and without EDSS increases over 2 years, we found no significant decrease for brain atrophy though there was a significant difference in cord area atrophy. In addition, we chose to use atrophy estimates that

were adjusted for covariates such as age, gender, disease duration, and number of relapses in the previous two years when calculating required sample sizes. Most of the cited studies make no allowances for these differences when calculating sample sizes. Adjusting for these factors provides a more conservative estimate since it ensures that the difference in the "treated" and "comparator" groups is truly due to disease progression, rather than differences in cohort demographics. These adjustments are important since we know that atrophy is dependent on both age and disease duration. Also, since males and females may be differentially represented in progressive MS cohorts, gender adjustment can also be important. By adjusting for number of relapses, we better replicate the conditions of a clinical trial which typically recruits patients with a history of relapses. Furthermore, we chose conservative definitions for our reference groups by including only those patients showing no increase in EDSS or SDMT scores and at least 2 second increases in 9HPT scores.

In contrast to many previous studies of spinal cord atrophy, the cohort used in our study was relatively large and included 196 subjects to better capture the variation of metrics in the patient population. In addition, the cohort was stratified so that the effect in different patient groups could be quantified and patient reference groups were used to better simulate the conditions of a clinical trial.

By obtaining measurements of spinal cord atrophy, brain atrophy, and gray matter atrophy in the same set of patients, we were able to directly compare and better understand their suitability for clinical trials. By using three different reference groups, we were also able to see whether certain metrics are better suited for reflecting different aspects of disability. Previous studies have shown that cross sectional spinal cord areas are strongly associated with EDSS scores (Schlaeger et al., 2014, 2015). However, since EDSS scores represent an amalgam of different types of

disability, this association is difficult to interpret. To get a better understanding of clinical disability, we used SDMT scores to assess cognitive decline and the nine-hole peg test assess sensorimotor decline.

Whole brain atrophy and gray matter atrophy are both outcome measures that are already commonly used in clinical trials. Several studies have already attempted to quantify the required sample sizes needed for clinical trials when using whole brain atrophy as an outcome marker. In RRMS patients, it was shown that 123 patients would be required to show a 30% treatment effect in a two year study (Anderson et al., 2007) using estimates of brain atrophy from SIENA. This study calculated sample size using a reference group of healthy controls and calculations were based on an analysis of change with estimated means and variances from a linear mixed model. When using the unadjusted values with similar assumptions of zero atrophy and 30% treatment effect in RMS patients, our study found a similar required sample size of 137 as seen in Table 5. In SPMS patients, 32 patients were required to show a 50% treatment effect in a two year study (Altmann et al., 2009) of brain atrophy from SIENA. This study calculated sample size using an estimated reference of zero atrophy and a comparison of mean rates of change estimated from longitudinal linear mixed models with several time points. When using the unadjusted values with the same assumptions for a reference, we calculated a similar required sample size of 26 as seen in Table 5. However, our study was unable to find any detectable change after we accounted for covariates.

In addition, many methods have been used to quantify both cortical and deep gray matter atrophy in the brain due to MS. A recent study (Nakamura et al., 2014) calculated the required sample size per arm to detect cortical gray matter atrophy in the brain using different methods, such as FreeSurfer, Jacobian integration, SIENAX, and SPM, using a cohort of 287 relapsing MS

patients. This study showed that for a 50% treatment effect at 80% power, a sample of 248 was required when using SIENAX estimates of cortical gray matter volume. This calculation did not account for normal aging and used data from a multi-center placebo controlled study of add-on oral steroids. Our study found that under similar assumptions of a 50% treatment effect with a reference of zero atrophy, a sample size of 230 was required when using cortical brain gray matter volume measured by SIENAX as seen in Table 5. Another study (Kim et al. 2017), which used FIRST to measure deep gray matter atrophy in PMS patients, showed that a sample size of 242 patients was required to detect a 50% treatment change. In this study, sample sizes were calculated using estimated means and residual standard deviation from a random-effect linear regression model and a reference of zero atrophy was assumed. Our study did not make estimates for subcortical gray matter so we are unable to compare in this case. In our study, once the atrophy rates were adjusted for covariates, significant differences between the groups could no longer be seen.

As spinal cord imaging has improved, several studies have been published to investigate the potential of using spinal cord atrophy as a measure in clinical trials. A recent study (Cawley et al., 2018) calculated the effect size of spinal cord atrophy in 26 primary progressive MS patients compared to 18 controls over a one year period to show that spinal cord atrophy is a feasible outcome measure in a clinical trial. This study used the parameters from these samples to estimate that the required sample size to detect a 50% treatment effect at 80% power was 55 for primary progressive patients. Our study was unable to provide any insights about required sample size for PPMS patients since our cohort only contain 10 PPMS patients. In another study (Kearney et al. 2014), the required sample size per arm for 6 month and 12 month placebo-controlled treatment trials was calculated from a study of 15 MS patients (7 RRMS and 8 SPMS patients) and 15

controls using two different acquisitions and two different methods of measuring total cord area. This study found that the best results came from measuring spinal cord area with an active surface model on PSIR images, which is the same method used in our study. The mean changes over six months were measured for a group of patients and a reference group of controls to estimate the treatment effect. In this study, for a treatment effect of 50% and power of 80%, the calculated sample size per arm was 89 for a 12 month trial. When we used similar assumptions of a 50% treatment effect with a reference of zero atrophy, we calculated a required sample size of 282 in our larger cohort. The difference in calculated sample size from this study may be due to the difference in cohort size. However, with a patient reference group and adjustments for covariates, the calculated required sample sizes for spinal cord atrophy rate was 97 for a reference group of patients without increased EDSS and 41 for a reference group of patients without motor decline as seen in Table 5.

Our results show that spinal cord atrophy has the potential to be useful in many clinical trial scenarios, since changes can be detected in a variety of different patient groups when compared to many different reference groups. In contrast, neither whole brain atrophy nor cortical gray matter atrophy were found to be useful metrics in clinical trials using patient reference groups once differences due to gender, age, or disease duration were corrected. As expected, this study shows that spinal cord atrophy is most useful when using a reference group of non-progressive patients without increased EDSS scores. The sample sizes are slightly larger for a reference group of non-progressive patients without motor decline and larger still for a reference group of non-progressive patients without cognitive decline. This suggests that spinal cord atrophy is much more indicative of motor disability than of cognitive disability.

While it is certainly important that an outcome measure's treatment effect can be detected, it is also very important that this treatment effect is reflective of an improvement in clinical disability. To prove useful to a patient, a treatment for MS must affect disability progression and not just the underlying biological effects of the disease. Several studies have shown connections between patients with whole brain or brain gray matter atrophy and patients with an increase in EDSS score. A meta study of 13 clinical trials showed that the treatment effects on whole brain atrophy were correlated with the treatment effects on EDSS scores (Sormani et al., 2014). Another study (Jacobsen et al., 2014) showed that patients whose EDSS scores had increased by at least one point had significantly increased whole brain, cortical, and putamen volume changes over five years and had significantly increased whole brain volume changes over 10 years. Other studies have shown connections between patients with whole brain or brain gray matter atrophy and a decline in either motor or cognitive function. A study of cortical atrophy patterns showed that cortical atrophy patterns in MS are non-random and clinically relevant and that certain patterns were more associated either motor or cognitive dysfunction (Steenwijk et al., 2016). EDSS scores, which are sensitive to motor dysfunction, were correlated with atrophy in the sensorimotor cortex while cognitive dysfunction metrics were correlated with atrophy in the bilateral posterior cingulate cortex and bilateral temporal pole. However, the connection between biological atrophy and the worsening of clinical symptoms still needs to be investigated further. Another study showed that the measurements of cortical and deep gray matter atrophy vary significantly with the method of analysis (Popescu et al., 2016) and that this variation affects the clinical interpretation of these metrics.

There are several limitations to this study design. This study makes the assumption that a reference group of clinically stable non-progressive patients is a better proxy for a successfully

treated group that results in no increase in disability in a clinical trial than a reference group of non-diseased subjects or an assumed comparative atrophy rate of zero. While novel, we believe this to be a reasonable assumption that likely results in more accurate power calculations. In addition, the average time between the longitudinal scans is around 2 years, so we lack data for long-term atrophy of the brain or spinal cord. For example, in cases where we did not detect significant differences in brain atrophy between the "treated" and "comparator" groups, a longer follow up would likely result in detectable differences. Alternatively, a larger cohort may have produced a measurable estimate of the number of subjects required to detect a difference. Many of the patients in this study are already being treated with existing medications. The treatment profiles of the groups differ since progressive patients tend to be treated with more potent medications. While the cohort of 196 patients is relatively large, there are only 10 PPMS patients in the cohort which prevents us from making any meaningful insights about this patient subgroup. In the future, we hope to repeat this study with a cohort that has a larger number of PPMS patients and a better matched treatment profile for each patient group and that has been followed for a period of five to ten years.

Our results suggest that spinal cord atrophy may be the most versatile metric for use in clinical trials since a treatment effect on clinical disability as measured by EDSS score can be detected with a feasible sample size, even after cohort differences in age, gender, and disease duration have been accounted for. This metric seems to primarily reflect motor disability, while whole brain and gray matter atrophy rates perform better in indicating differences in cognitive disability. Given the relatively early developments of spinal cord atrophy measurements, it is also possible that we will see improvements in the precision of these measurements that will provide greater statistical power.

Chapter 5 - Voxelwise Analysis of the Spinal Cord

5.1 Introduction

Studying how different regions of the spinal cord relate to clinical outcomes in multiple sclerosis could help to uncover the pathological mechanisms that underlie the disease. While, studies have investigated regional relationships with multiple sclerosis clinical metrics in the brain (Di Perri et al., 2008), few regional studies have been done on the spinal cord, due to previous constraints in imaging such a small structure.

One previous study (Valsasina et al., 2013) used voxel-based morphometry to look at metrics based on four quadrants of the spinal cord over six levels and found that atrophy primarily occurred in the anterior sections of the cord due to age. Another previous study (Maria A Rocca et al., 2013) performed a voxel-wise analysis on T2 and T1 weighted cervical cord images in MS patients to show the distribution of atrophy and lesions in different phenotypes. This study found that significant atrophy was found in BMS, SPMS, and PPMS patients and was more likely to be seen in the posterior and lateral cord portions. No significant differences were found between left, right, anterior or posterior cord sections for lesion load.

One study found that tensor based morphometry can be used to demonstrate gray matter atrophy in the aging spinal cord (Taso et al., 2015). In this study, regions of atrophy in elderly patients at the C2/C3 level were shown to occur in the anterior horn area. This is similar to the results found from post-mortem studies that showed that neuronal loss due to aging seems to occur in the anterior horns of the gray matter in the spinal cord (Cruz-Sánchez, Moral, Tolosa, de Belleroche, & Rossi, 1998; Terao et al., 1996).
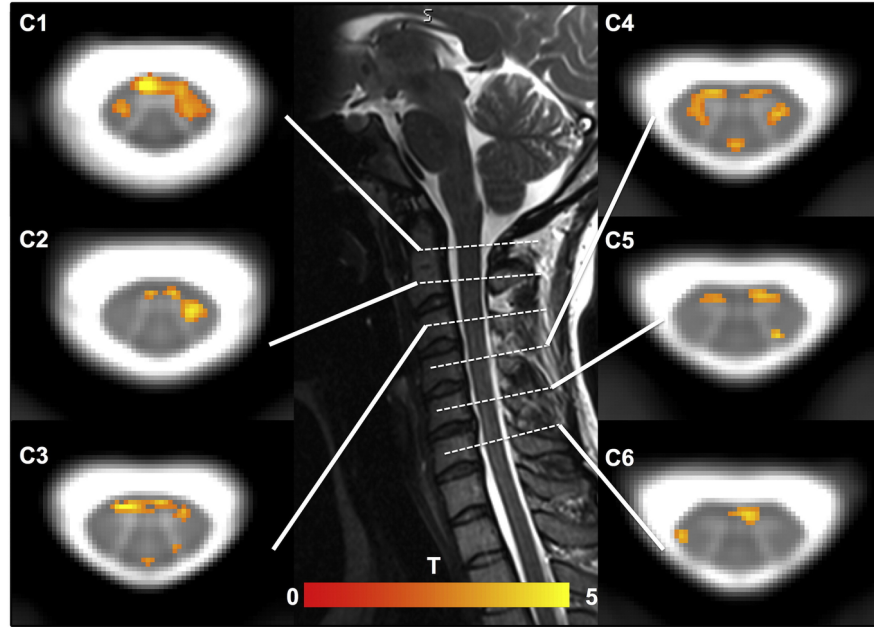
Figure 28: Results of the TBM analysis between young and elderly volunteers (2 samplet-test, p<0.05 FDR corrected, k>1, SPM8) overlaid on the whole population (n=65) T2*-w template. The red/yellow clusters indicate regions of atrophy in the elderly subjects (Taso et al., 2015)

A recent study has used quantitative metrics from diffusion tensor imaging to assess microstructural tissue properties in spinal cord gray matter in MS patients (Kearney et al., 2014). This study found abnormalities in the posterior and lateral portions of the spinal cord in MS patients. Another recent study (Valsasina et al., 2018) concluded that hypointense lesions were more prevalent in the posterior cord and that this difference was reflected in differences in clinical phenotype and disability.

Figure 29: T1-weighted hypointense lesion probability maps in MS patients (Valsasina et al., 2018)

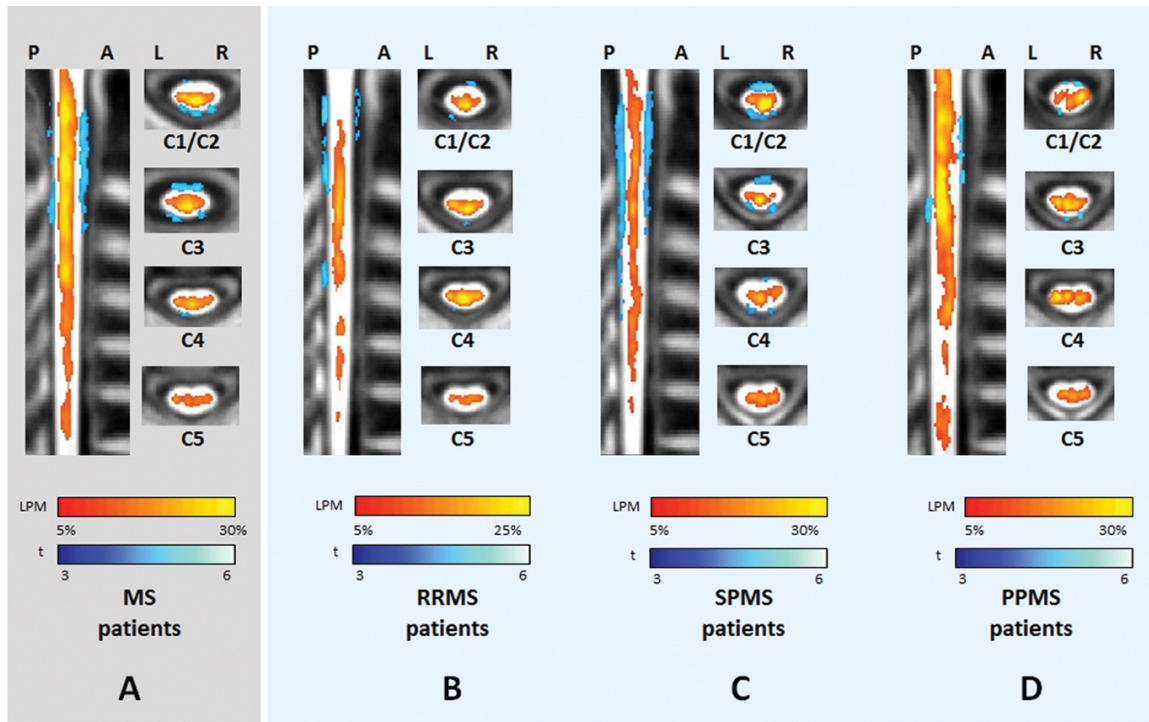In the study presented here, a voxel-wise analysis was done to investigate three different attributes: intensity, texture, and shape. Using the quantitative intensity from PSIR images, we can get a local metric of the water content of the tissue. By using the magnitude of the gradient at each voxel, we can get a texture metric that represents the local heterogeneity of the tissue. Similarly, with tensor based morpohometry, the Jacobian determinant gives a metric that represents the local anatomical shape differences from a control template. The increased quality from the PSIR images allows us to study the regional differences at a much finer level than previous studies.

5.2 Methods

The cohort used for this study was the same as the cohort referred to in Figure 1. The subjects in this study included 20 healthy controls, 92 40 relapsing MS (RMS) patients and 37 progressive MS patients (PMS). The age and gender distribution of the control group was similar

to the group of RMS patients. The median EDSS score of RMS patients is 2 and the median EDSS score of the PMS patients is 6. The clinical characteristics of the patients are described in Figure 1.

All subjects in the patient cohort received a standardized clinical neurological exam to determine the EDSS score. In addition, all patients were scanned on a Siemens 3T Skyra scanner equipped with a 20 channel neck-head coil and a 32 channel spine coil within two weeks of their clinical examination. For each patient, axial 2D phase sensitive inversion recovery (PSIR) images were acquired at the intervertebral disc level C2/C3 perpendicular to the cord. The acquisition parameters are listed in Figure 2.

For each phase sensitive reconstructed image, cord masks were generated on up-sampled interpolated images through semi-automated segmentation using the software JIM6 (Horsfield et al., 2010). In addition, gray matter masks were generated on these same images through manual segmentation by a trained neurologist.

Using the images from the 20 healthy control subjects, a spinal cord template was created using ANTS. The white matter tract atlas from the Spinal Cord Toolbox was registered to this template and gray matter ROIs were hand drawn as well as well to create regional ROIs in the template space.
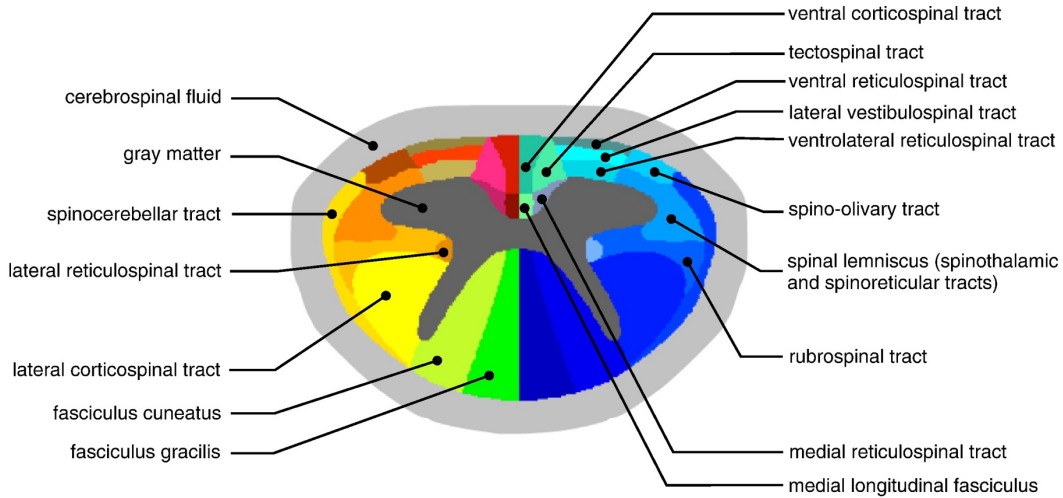
Figure 30: Spinal Cord Regions From Spinal Cord Toolbox (De Leener et al., 2017)

For each image, a gradient map was also created that represents how much the image intensity changes across space. In addition, each image was registered to a template and a Jacobian determinant map was created that represented each voxel's volume change in the transformation.

The original PSIR images from all subjects were then non-linearly transformed to the common template space. The gradient maps and the Jacobian determinant maps underwent the same transformation. This study focused on the relationship of three different local metrics (intensity, gradient, and Jacobian determinants) with three different clinical metrics (EDSS score, 9 hole peg test, and 25 foot walk) and the relationship with two demographic metrics (age and disease duration).

Three different sets of regional analyses were performed using this data:

1. At each voxel, a regression model was created and a corresponding Spearman coefficient was calculated for the relationships of each of the three local metrics with each of the clinical metrics and each of the demographic metrics.

2. For each regional ROI, a regression model was created and a corresponding Spearman coefficient was calculated for the relationship of the average local metric in the region with each of the clinical metrics.

3. At each voxel, a multivariate ridge regression model was created for the relationship of each of the three local metrics with each of the clinical metrics, using the demographic metrics as covariates. At each voxel, a corresponding r squared value was calculated from the multivariate model.

5.3   Results

5.3.1   Voxel-wise Univariate Models

From the univariate voxel-wise models, we can see that a large clusters appear in the relationships of intensity with clinical metrics such as EDSS score, 9 hole peg test, and 25 foot walk between the horns of the gray matter area. In addition, there is a large cluster in the relationship of intensity with age. There are no large clusters in the maps for the relationships with Jacobian determinant or gradient.
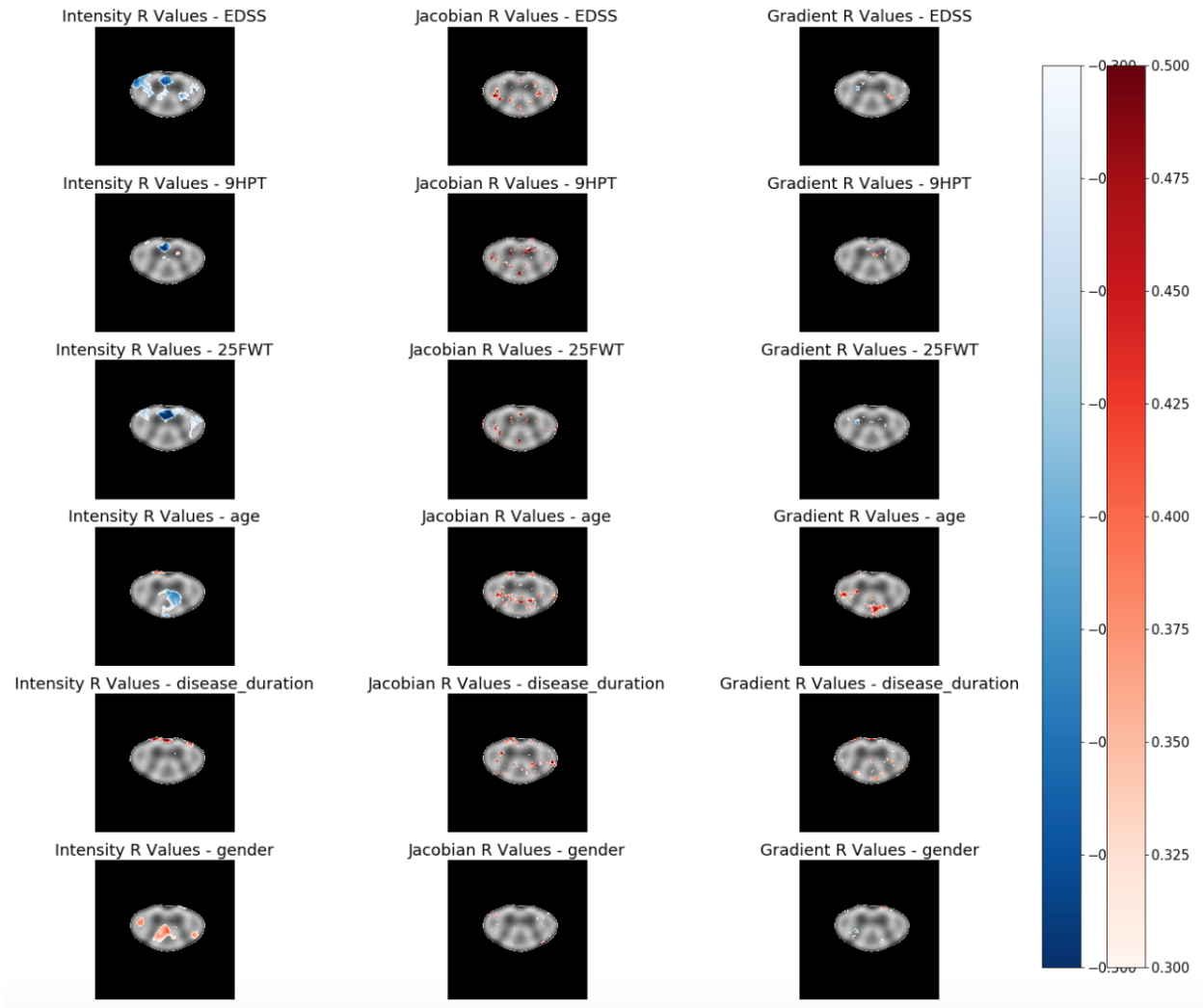
Figure 31: R Values from Voxel Wise Univariate Regression Models

5.3.2 Regional Univariate Models

From the regional models, we see that that two different tract areas in the spinal cord appear significant in the relationship of intensity and clinical metrics. The tracts with the highest correlations of average intensity with EDSS include the lateral corticospinal tract and the medial longitudinal fasciculus with the medial reticulospinal tract. In addition, the regional analysis shows the relationships of Jacobian determinants with EDSS and age. This regional analysis suggests that atrophy correlated to EDSS score is located more in the lateral corticospinal tracts and that atrophy due to age is in the posterior region of the spinal cord around the gray matter.
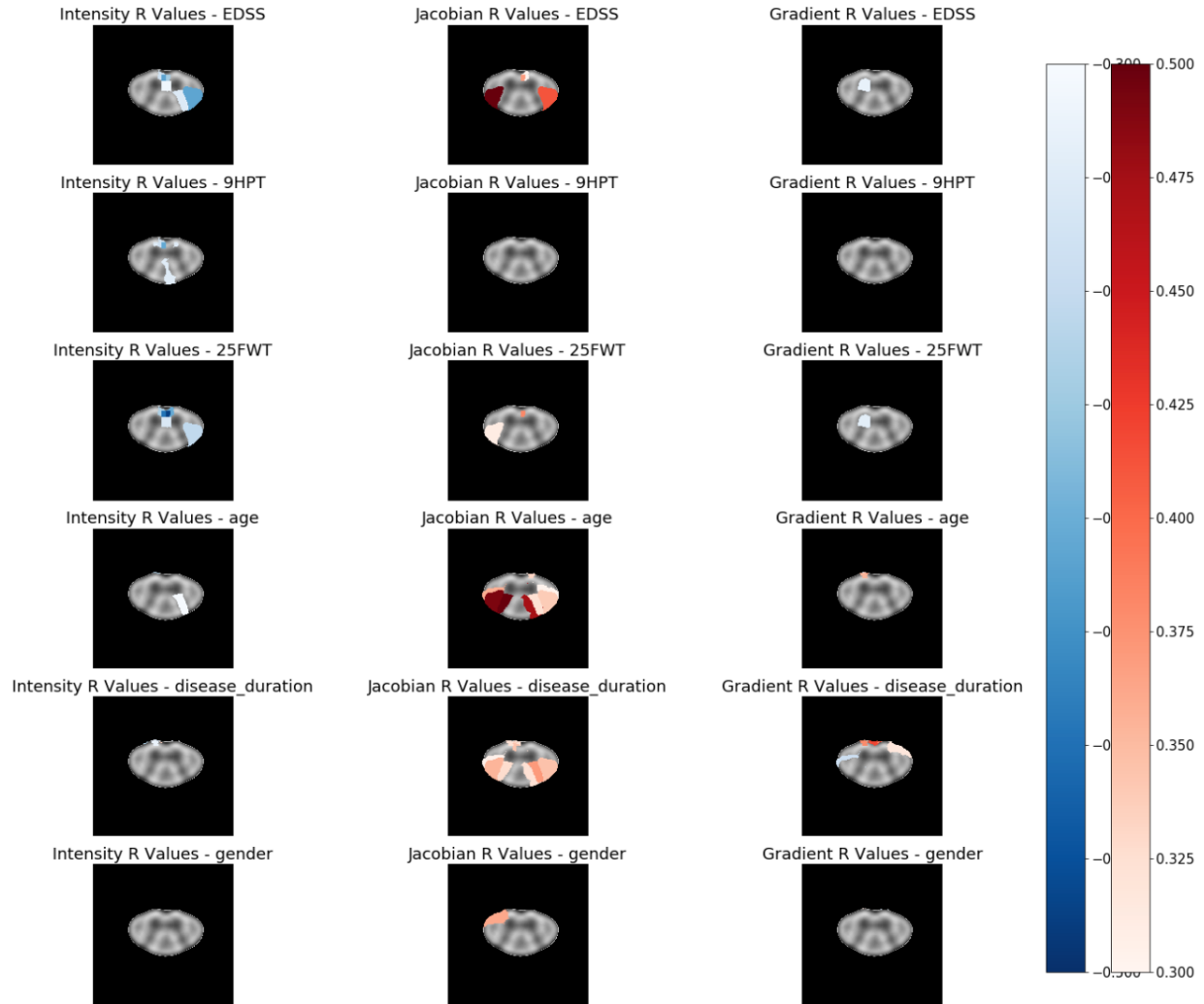
Figure 32: R Values from Regional Univariate Analysis Models

5.3.3 Multivariate Voxel-wise Models with Age and Gender as Covariates

By using a multivariate model, we can separate out effects due to age or gender. These models show again that the correlations with intensity are centered around the lateral corticospinal tract and the medial longitudinal fasciculus and medial reticulospinal tract. The areas where Jacobian determinants correlate most with EDSS seem to be located in the lateral corticospinal tracts and the anterior horns of the gray matter. The areas where gradient correlates most with EDSS seem to be located near the boundaries of the gray matter and the white matter.

77
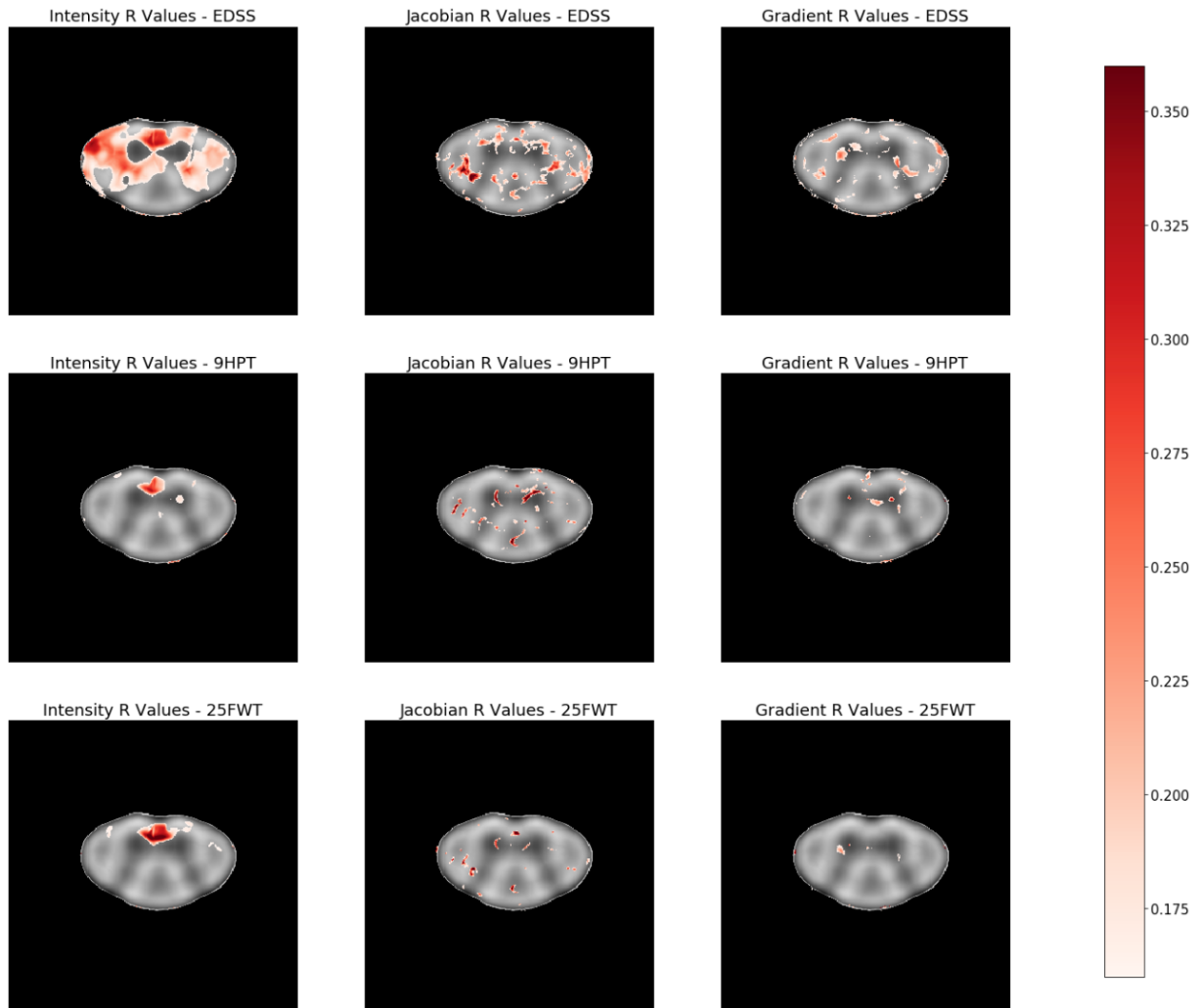
Figure 33: R Squared Values from Multivariate Voxelwise Regression with Age and Disease Duration as Covariates

5.4 Discussion

From our analysis, we find that there are two main areas of strong correlations between EDSS and intensity. While the size and resolution limits of spinal cord imaging make it difficult to accurately locate specific tracts, we can identify general functional areas that seem to be affected.

The first area appears in the relationships with EDSS score, but not in the relationships with nine hole peg test and the twenty five foot walk. This area affected includes the lateral

corticospinal tracts, lateral reticulospinal tracts, and rubrospinal tract. All of these tracts are extrapyramidal descending motor tracts that are involved in limb control, locomotion, or posture control. The lateral corticospinal tract controls fine movement of ipsilateral limbs, the lateral reticulospinal tract inhibits excitatory axial extensor muscles in the trunk and proximal limb muscles, and the rubrospinal tract primarily controls flexion in the upper extremities.

The second area appears in the relationships with EDSS score, nine hole peg test, and twenty five foot walk, and encompasses the medial longitudinal fasciculus and the medial reticulospinal tract. Both of these tracts have a motor function. The primary function of the medical longitudinal fasciculus is to carry information about the direction that the eyes should move and the medial reticulospinal tract is a excited anti-gravity extensor muscles in the trunk and proximal limb muscles.

In the figure below, we can see clearly that the two primary areas affected correspond to the two primary areas where descending motor tracts are located. Because EDSS score primarily reflects motor disability, it is not surprising that this is the case.In the future, it may be advantageous to additionally study the areas that most differ between multiple sclerosis patients and in healthy patients.
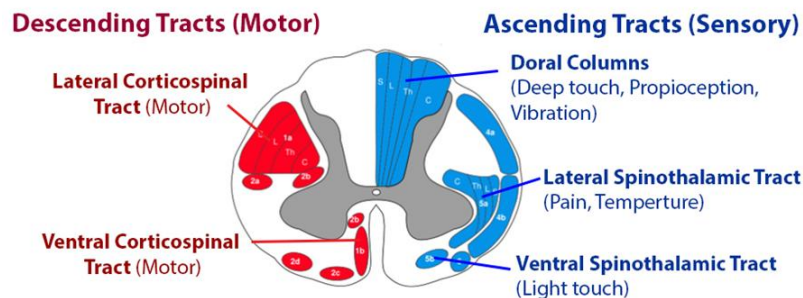

Figure 34: Motor and Sensory Function in the Spinal Cord ("Incomplete Spinal Cord Injuries - Spine - Orthobullets," n.d.)

Bibliography

Altmann, D. R., Jasperse, B., Barkhof, F., Beckmann, K., Filippi, M., Kappos, L. D., … Miller, D. H. (2009). Sample sizes for brain atrophy outcomes in trials for secondary progressive multiple sclerosis. *Neurology*, *72*(7), 595–601. https://doi.org/10.1212/01.wnl.0000335765.55346.fc

Amann, M., Pezold, S., Naegelin, Y., Fundana, K., Andělová, M., Weier, K., … Sprenger, T. (2016). Reliable volumetry of the cervical spinal cord in MS patient follow-up data with cord image analyzer (Cordial). *Journal of Neurology*, *263*(7), 1364–1374. https://doi.org/10.1007/s00415-016-8133-0

Anderson, V. M., Bartlett, J. W., Fox, N. C., Fisniku, L., & Miller, D. H. (2007). Detecting treatment effects on brain atrophy in relapsing remitting multiple sclerosis: Sample size estimates. *Journal of Neurology*, *254*(11), 1588–1594. https://doi.org/10.1007/s00415-007-0599-3

Asman, A. J., Smith, S. A., Reich, D. S., & Landman, B. A. (2013). Robust GM/WM segmentation of the spinal cord with iterative non-local statistical fusion. *Medical Image Computing and Computer-Assisted Intervention : MICCAI … International Conference on Medical Image Computing and Computer-Assisted Intervention*, *16*(Pt 1), 759–67. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24505736

Bakshi, R., Dandamudi, V. S. R., Neema, M., De, C., & Bermel, R. A. (2005). Measurement of Brain and Spinal Cord Atrophy by Magnetic Resonance Imaging as a Tool to Monitor Multiple Sclerosis. *Journal of Neuroimaging*, *15*(4 Suppl), 30S–45S.

https://doi.org/10.1177/1051228405283901

Bernitsas, E., Bao, F., Seraji-Bozorgzad, N., Chorostecki, J., Santiago, C., Tselis, A., … Khan, O. (2015). Spinal cord atrophy in multiple sclerosis and relationship with disability across clinical phenotypes. *Multiple Sclerosis and Related Disorders*, *4*(1), 47–51. https://doi.org/10.1016/j.msard.2014.11.002

Bonati, U., Fisniku, L. K., Altmann, D. R., Yiannakas, M. C., Furby, J., Thompson, A. J., … Chard, D. T. (2011). Cervical cord and brain grey matter atrophy independently associate with long-term MS disability. *Journal of Neurology, Neurosurgery & Psychiatry*, *82*(4), 471–472. https://doi.org/10.1136/jnnp.2010.205021

Brex, P. A., Leary, S. M., O'Riordan, J. I., Miszkiel, K. A., Plant, G. T., Thompson, A. J., & Miller, D. H. (2001). Measurement of spinal cord area in clinically isolated syndromes suggestive of multiple sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, *70*(4), 544–7. https://doi.org/10.1136/JNNP.70.4.544

Carballido-Gamio, J., Harnish, R., Saeed, I., Streeper, T., Sigurdsson, S., Amin, S., … Lang, T. F. (2013). Proximal femoral density distribution and structure in relation to age and hip fracture risk in women. *Journal of Bone and Mineral Research*, *28*(3), 537–546. https://doi.org/10.1002/jbmr.1802

Caselles, V., Kimmel, R., & Sapiro, G. (1995). Geodesic active contours. *IEEE International Conference on Computer Vision*, *22*(1), 694–699.

Cawley, N., Tur, C., Prados, F., Plantone, D., Kearney, H., Abdel-Aziz, K., … Ciccarelli, O. (2018). Spinal cord atrophy as a primary outcome measure in phase II trials of progressive

multiple sclerosis. *Multiple Sclerosis Journal*, *24*(7), 932–941.

https://doi.org/10.1177/1352458517709954

Cruz-Sánchez, F. F., Moral, A., Tolosa, E., de Belleroche, J., & Rossi, M. L. (1998). Evaluation

of neuronal loss, astrocytosis and abnormalities of cytoskeletal components of large motor

neurons in the human anterior horn in aging. *Journal of Neural Transmission*, *105*(6–7),

689–701. https://doi.org/10.1007/s007020050088

De Leener, B., Lévy, S., Dupont, S. M., Fonov, V. S., Stikov, N., Louis Collins, D., … Cohen-

Adad, J. (2017). SCT: Spinal Cord Toolbox, an open-source software for processing spinal

cord MRI data. *NeuroImage*, *145*(Pt A), 24–43.

https://doi.org/10.1016/j.neuroimage.2016.10.009

De Leener, B., Taso, M., Cohen-Adad, J., & Callot, V. (2016). Segmentation of the human spinal

cord. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *29*(2), 125–153.

https://doi.org/10.1007/s10334-015-0507-2

Di Perri, C., Battaglini, M., Stromillo, M. L., Bartolozzi, M. L., Guidi, L., Federico, A., & De

Stefano, N. (2008). Voxel-Based Assessment of Differences in Damage and Distribution of

White Matter Lesions Between Patients With Primary Progressive and Relapsing-Remitting

Multiple Sclerosis. *Archives of Neurology*, *65*(2), 236–243.

https://doi.org/10.1001/archneurol.2007.51

Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2013). Learning Hierarchical Features for

Scene Labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8),

1915–1929. https://doi.org/10.1109/TPAMI.2012.231

Filippi, M., Colombo, B., Rovaris, M., Pereira, C., Martinelli, V., & Comi, G. (1997). A

longitudinal magnetic resonance imaging study of the cervical cord in multiple sclerosis.

Journal of Neuroimaging : Official Journal of the American Society of Neuroimaging, *7*(2),

78–80. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9128443

Filippi, M., & Rocca, M. A. (2013). Linking disability and spinal cord imaging outcomes in MS.

Nature Reviews Neurology, *9*(4), 189–190. https://doi.org/10.1038/nrneurol.2013.40

FOG, T. (1950). TOPOGRAPHIC DISTRIBUTION OF PLAQUES IN THE SPINAL CORD IN

MULTIPLE SCLEROSIS. *Archives of Neurology And Psychiatry*, *63*(3), 382.

https://doi.org/10.1001/archneurpsyc.1950.02310210028003

Fonov, V. S., Le Troter, A., Taso, M., De Leener, B., Lévêque, G., Benhamou, M., … Cohen-

Adad, J. (2014). Framework for integrated MRI average of the spinal cord white and gray

matter: The MNI–Poly–AMU template. *NeuroImage*, *102*, 817–827.

https://doi.org/10.1016/j.neuroimage.2014.08.057

Freund, P., Curt, A., Friston, K., & Thompson, A. (2013). Tracking Changes following Spinal

Cord Injury. *The Neuroscientist*, *19*(2), 116–128.

https://doi.org/10.1177/1073858412449192

Furby, J., Hayton, T., Altmann, D., Brenner, R., Chataway, J., Smith, K. J., … Kapoor, R.

(2010). A longitudinal study of MRI-detected atrophy in secondary progressive multiple

sclerosis. *Journal of Neurology*, *257*(9), 1508–1516. https://doi.org/10.1007/s00415-010-

5563-y

Greenspan, H., van Ginneken, B., & Summers, R. M. (2016). Guest Editorial Deep Learning in

Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging*, *35*(5), 1153–1159. https://doi.org/10.1109/TMI.2016.2553401

Held, P., Dorenbeck, U., Seitz, J., Fründ, R., & Albrich, H. (2003). MRI of the abnormal cervical spinal cord using 2D spoiled gradient echo multiecho sequence (MEDIC) with magnetization transfer saturation pulse. A T2* weighted feasibility study. *Journal of Neuroradiology. Journal de Neuroradiologie*, *30*(2), 83–90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12717293

Horsfield, M. A., Sala, S., Neema, M., Absinta, M., Bakshi, A., Sormani, M. P., … Filippi, M. (2010). Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: application in multiple sclerosis. *NeuroImage*, *50*(2), 446–55. https://doi.org/10.1016/j.neuroimage.2009.12.121

Incomplete Spinal Cord Injuries - Spine - Orthobullets. (n.d.). Retrieved June 22, 2018, from https://www.orthobullets.com/spine/2008/incomplete-spinal-cord-injuries

Israel, H., Ostendorf, F., Stiepani, H., & Ploner, C. J. (2005). Spinal Cord Atrophy in Adrenomyeloneuropathy. *Archives of Neurology*, *62*(7), 1157. https://doi.org/10.1001/archneur.62.7.1157

Jacobsen, C., Hagemeier, J., Myhr, K.-M., Nyland, H., Lode, K., Bergsland, N., … Zivadinov, R. (2014). Brain atrophy and disability progression in multiple sclerosis patients: a 10-year follow-up study. *Journal of Neurology, Neurosurgery & Psychiatry*, *85*(10), 1109–1115. https://doi.org/10.1136/jnnp-2013-306906

Kalkers, N. F., Barkhof, F., Bergers, E., van Schijndel, R., & Polman, C. H. (2002). The effect of

    the neuroprotective agent riluzole on MRI parameters in primary progressive multiple

    sclerosis: a pilot study. *Multiple Sclerosis Journal*, *8*(6), 532–533.

    https://doi.org/10.1191/1352458502ms849xx

Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International*

    Journal of Computer Vision, *1*(4), 321–331. https://doi.org/10.1007/BF00133570

Kearney, H., Miller, D. H., & Ciccarelli, O. (2015). Spinal cord MRI in multiple sclerosis—

    diagnostic, prognostic and clinical value. *Nature Reviews Neurology*, *11*(6), 327–338.

    https://doi.org/10.1038/nrneurol.2015.80

Kearney, H., Yiannakas, M. C., Abdel-Aziz, K., Wheeler-Kingshott, C. A. M., Altmann, D. R.,

    Ciccarelli, O., & Miller, D. H. (2014). Improved MRI quantification of spinal cord atrophy

    in multiple sclerosis. *Journal of Magnetic Resonance Imaging*, *39*(3), 617–623.

    https://doi.org/10.1002/jmri.24194

Kim, G., Chu, R., Yousuf, F., Tauhid, S., Stazzone, L., Houtchens, M. K., … Bakshi, R. (2017).

    Sample size requirements for one-year treatment effects using deep gray matter volume

    from 3T MRI in progressive forms of multiple sclerosis. *International Journal of*

    Neuroscience, *127*(11), 971–980. https://doi.org/10.1080/00207454.2017.1283313

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep

    convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90.

    https://doi.org/10.1145/3065386

Kutzelnigg, A., & Lassmann, H. (2014). Pathology of multiple sclerosis and related

inflammatory demyelinating diseases. In *Handbook of clinical neurology* (Vol. 122, pp. 15–58). https://doi.org/10.1016/B978-0-444-52001-2.00002-9

Leary, S. M., Miller, D. H., Stevenson, V. L., Brex, P. A., Chard, D. T., & Thompson, A. J. (2003). Interferon beta-1a in primary progressive MS: an exploratory, randomized, controlled trial. *Neurology*, *60*(1), 44–51. https://doi.org/10.1212/WNL.60.1.44

Lin, X., Tench, C. R., Evangelou, N., Jaspan, T., & Constantinescu, C. S. (2004). Measurement of Spinal Cord Atrophy in Multiple Sclerosis. *Journal of Neuroimaging*, *14*(3), 20–26. https://doi.org/10.1177/1051228404266265

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3431–3440). IEEE. https://doi.org/10.1109/CVPR.2015.7298965

Losseff, N. A., Webb, S. L., O'Riordan, J. I., Page, R., Wang, L., Barker, G. J., … Thompson, A. J. (1996). Spinal cord atrophy and disability in multiple sclerosis. A new reproducible and sensitive MRI method with potential to monitor disease progression. *Brain : A Journal of Neurology*, *119 (Pt 3)*, 701–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8673483

Lukas, C., Knol, D. L., Sombekke, M. H., Bellenberg, B., Hahn, H. K., Popescu, V., … Vrenken, H. (2015). Cervical spinal cord volume loss is related to clinical disability progression in multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, *86*(4), 410–418. https://doi.org/10.1136/jnnp-2014-308021

Lycklama, G., Thompson, A., Filippi, M., Miller, D., Polman, C., Fazekas, F., & Barkhof, F.

(2003). Spinal-cord MRI in multiple sclerosis. *The Lancet. Neurology*, *2*(9), 555–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12941578

Marquez-Neila, P., Baumela, L., & Alvarez, L. (2014). A Morphological Approach to Curvature-Based Evolution of Curves and Surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(1), 2–17. https://doi.org/10.1109/TPAMI.2013.106

Mathias, J. M., Tofts, P. S., & Losseff, N. A. (1999). Texture analysis of spinal cord pathology in multiple sclerosis. *Magnetic Resonance in Medicine*, *42*(5), 929–35. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10542352

Montalban, X., Sastre-Garriga, J., Tintoré, M., Brieva, L., Aymerich, F., Río, J., … Rovira, À. (2009). A single-center, randomized, double-blind, placebo-controlled study of interferon beta-1b on primary progressive and transitional multiple sclerosis. *Multiple Sclerosis Journal*, *15*(10), 1195–1205. https://doi.org/10.1177/1352458509106937

Nakamura, K., Guizard, N., Fonov, V. S., Narayanan, S., Collins, D. L., & Arnold, D. L. (2014). Jacobian integration method increases the statistical power to measure gray matter atrophy in multiple sclerosis. *NeuroImage. Clinical*, *4*, 10–7. https://doi.org/10.1016/j.nicl.2013.10.015

Nijeholt, G. J., van Walderveen, M. A., Castelijns, J. A., van Waesberghe, J. H., Polman, C., Scheltens, P., … Barkhof, F. (1998). Brain and spinal cord abnormalities in multiple sclerosis. Correlation between MRI parameters, clinical subtypes and symptoms. *Brain : A Journal of Neurology*, *121 (Pt 4)*, 687–97. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9577394

Papinutto, N., Schlaeger, R., Panara, V., Caverzasi, E., Ahn, S., Johnson, K. J., … Henry, R. G.
(2015). 2D phase-sensitive inversion recovery imaging to measure in vivo spinal cord gray
and white matter areas in clinically feasible acquisition times. *Journal of Magnetic
Resonance Imaging*, *42*(3), 698–708. https://doi.org/10.1002/jmri.24819

Perone, C. S., Calabrese, E., & Cohen-Adad, J. (2018). Spinal cord gray matter segmentation
using deep dilated convolutions. *Scientific Reports*, *8*(1), 5966.
https://doi.org/10.1038/s41598-018-24304-3

Popescu, V., Schoonheim, M. M., Versteeg, A., Chaturvedi, N., Jonker, M., Xavier de Menezes,
R., … Vrenken, H. (2016). Grey Matter Atrophy in Multiple Sclerosis: Clinical
Interpretation Depends on Choice of Analysis Method. *PLOS ONE*, *11*(1), e0143942.
https://doi.org/10.1371/journal.pone.0143942

Potter, K., & Saifuddin, A. (2003). MRI of chronic spinal cord injury. *The British Journal of
Radiology*, *76*(905), 347–352. https://doi.org/10.1259/bjr/11881183

Rashid, W., Davies, G. R., Chard, D. T., Griffin, C. M., Altmann, D. R., Gordon, R., … Miller,
D. H. (2006). Increasing cord atrophy in early relapsing-remitting multiple sclerosis: a 3
year study. *Journal of Neurology, Neurosurgery & Psychiatry*, *77*(1), 51–55.
https://doi.org/10.1136/jnnp.2005.068338

Reinertsen, I., Descoteaux, M., Drouin, S., Siddiqi, K., & Collins, D. L. (2004). Vessel Driven
Correction of Brain Shift (pp. 208–216). Springer, Berlin, Heidelberg.
https://doi.org/10.1007/978-3-540-30136-3_27

Rocca, M. A., Horsfield, M. A., Sala, S., Copetti, M., Valsasina, P., Mesaros, S., … Filippi, M.

(2011). A multicenter assessment of cervical cord atrophy among MS clinical phenotypes. *Neurology*, *76*(24), 2096–2102. https://doi.org/10.1212/WNL.0b013e31821f46b8

Rocca, M. A., Valsasina, P., Damjanovic, D., Horsfield, M. A., Mesaros, S., Stosic-Opincal, T., … Filippi, M. (2013). Voxel-wise mapping of cervical cord damage in multiple sclerosis patients with different clinical phenotypes. *Journal of Neurology, Neurosurgery, and Psychiatry*, *84*(1), 35–41. https://doi.org/10.1136/jnnp-2012-303821

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation (pp. 234–241). Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

Rueckert, D., Frangi, A. F., & Schnabel, J. A. (2003). Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, *22*(8), 1014–1025. https://doi.org/10.1109/TMI.2003.815865

Schlaeger, R., Papinutto, N., Panara, V., Bevan, C., Lobach, I. V., Bucci, M., … Henry, R. G. (2014). Spinal cord gray matter atrophy correlates with multiple sclerosis disability. *Annals of Neurology*, *76*(4), 568–580. https://doi.org/10.1002/ana.24241

Schlaeger, R., Papinutto, N., Zhu, A. H., Lobach, I. V, Bevan, C. J., Bucci, M., … Henry, R. G. (2015). Association Between Thoracic Spinal Cord Gray Matter Atrophy and Disability in Multiple Sclerosis. *JAMA Neurology*, *72*(8), 897–904. https://doi.org/10.1001/jamaneurol.2015.0993

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., … Matthews, P. M. (2004). Advances in functional and structural MR image

analysis and implementation as FSL. *NeuroImage*, *23*, S208–S219.

https://doi.org/10.1016/J.NEUROIMAGE.2004.07.051

Smith, S. M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P. M., Federico, A., & De Stefano, N. (2002). Accurate, Robust, and Automated Longitudinal and Cross-Sectional Brain Change Analysis. *NeuroImage*, *17*(1), 479–489. https://doi.org/10.1006/NIMG.2002.1040

Sormani, M. P., Arnold, D. L., & De Stefano, N. (2014). Treatment effect on brain atrophy correlates with treatment effect on disability in multiple sclerosis. *Annals of Neurology*, *75*(1), 43–49. https://doi.org/10.1002/ana.24018

Steenwijk, M. D., Geurts, J. J. G., Daams, M., Tijms, B. M., Wink, A. M., Balk, L. J., … Pouwels, P. J. W. (2016). Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. *Brain*, *139*(1), 115–126. https://doi.org/10.1093/brain/awv337

Stroman, P. W., Wheeler-Kingshott, C., Bacon, M., Schwab, J. M., Bosma, R., Brooks, J., … Tracey, I. (2014). The current state-of-the-art of spinal cord imaging: Methods. *NeuroImage*, *84*, 1070–1081. https://doi.org/10.1016/j.neuroimage.2013.04.124

Suh, J. W., & Wyatt, C. L. (2006). Deformable Registration of Prone and Supine Colons for CT Colonography. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society* (Vol. 1, pp. 1997–2000). IEEE. https://doi.org/10.1109/IEMBS.2006.260249

Taso, M., Le Troter, A., Sdika, M., Cohen-Adad, J., Arnoux, P.-J., Guye, M., … Callot, V. (2015). A reliable spatially normalized template of the human spinal cord — Applications to automated white matter/gray matter segmentation and tensor-based morphometry (TBM)

mapping of gray matter alterations occurring with age. *NeuroImage*, *117*, 20–28. https://doi.org/10.1016/j.neuroimage.2015.05.034

Terao, S., Sobue, G., Hashizume, Y., Li, M., Inagaki, T., & Mitsuma, T. (1996). Age-related changes in human spinal ventral horn cells with special reference to the loss of small neurons in the intermediate zone: a quantitative analysis. *Acta Neuropathologica*, *92*(2), 109–114. https://doi.org/10.1007/s004010050497

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., … Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology*, *17*(2), 162–173. https://doi.org/10.1016/S1474-4422(17)30470-2

Valsasina, P., Aboulwafa, M., Preziosa, P., Messina, R., Falini, A., Comi, G., … Rocca, M. A. (2018). Cervical Cord T1-weighted Hypointense Lesions at MR Imaging in Multiple Sclerosis: Relationship to Cord Atrophy and Disability. *Radiology*, *288*(1), 234–244. https://doi.org/10.1148/radiol.2018172311

Valsasina, P., Rocca, M. A., Horsfield, M. A., Absinta, M., Messina, R., Caputo, D., … Filippi, M. (2013). Regional Cervical Cord Atrophy and Disability in Multiple Sclerosis: A Voxel-based Analysis. *Radiology*, *266*(3), 853–861. https://doi.org/10.1148/radiol.12120813

Vercauteren, T., Pennec, X., Perchant, A., & Ayache, N. (2008). Symmetric Log-Domain Diffeomorphic Registration: A Demons-Based Approach (pp. 754–761). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-85988-8_90

Wheeler-Kingshott, C. A., Stroman, P. W., Schwab, J. M., Bacon, M., Bosma, R., Brooks, J., … Tracey, I. (2014). The current state-of-the-art of spinal cord imaging: applications.

NeuroImage, *84*, 1082–93. https://doi.org/10.1016/j.neuroimage.2013.07.014

Yiannakas, M. C., Kearney, H., Samson, R. S., Chard, D. T., Ciccarelli, O., Miller, D. H., &
Wheeler-Kingshott, C. A. M. (2012). Feasibility of grey matter and white matter
segmentation of the upper cervical cord in vivo: A pilot study with application to
magnetisation transfer measurements. *NeuroImage*, *63*(3), 1054–1059.
https://doi.org/10.1016/j.neuroimage.2012.07.048

Yiannakas, M. C., Mustafa, A. M., De Leener, B., Kearney, H., Tur, C., Altmann, D. R., …
Gandini Wheeler-Kingshott, C. A. M. (2016). Fully automated segmentation of the cervical
cord from T1-weighted MRI using PropSeg: Application to multiple sclerosis. *NeuroImage:
Clinical*, *10*, 71–77. https://doi.org/10.1016/J.NICL.2015.11.001

Zhang, Y. (2012). MRI texture analysis in multiple sclerosis. *International Journal of
Biomedical Imaging*, *2012*, 762804. https://doi.org/10.1155/2012/762804

Zhang, Y., Moore, G. R. W., Laule, C., Bjarnason, T. A., Kozlowski, P., Traboulsee, A., & Li,
D. K. B. (2013). Pathological correlates of magnetic resonance imaging texture
heterogeneity in multiple sclerosis. *Annals of Neurology*, *74*(1), 91–99.
https://doi.org/10.1002/ana.23867