Investigating the Spinal Cord Atrophy Measurement on MRI from Two Aspects: Physiological Variations and Longitudinal Measurement Methods

by

Chunfang Wang

B.Eng., Huazhong University of Science and Technology, 2006
M.Eng., Huazhong University of Science and Technology, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Biomedical Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2014

Abstract

Spinal cord atrophy is a valuable biomarker in multiple sclerosis (MS) for its significant correlation with physical disability. Measurement of spinal cord atrophy on MRI may be possibly confounded by fluctuations in water content, and the high measurement variance in previous longitudinal studies can be possibly reduced by registration-based methods. In this thesis, we investigated the effect of change in water content due to hydration status on cord cross-sectional area (CSA) measurement, and the applicability of three registration-based methods for longitudinal cord atrophy measurement.

Our first hypothesis is that dehydration can decrease the cord CSA measurement on MRI. We found a mean decrease of 0.65% in CSA on scans collected from ten controls following a dehydration protocol using two independent cross-sectional CSA measurement methods. Our result demonstrates that change in water content of the cord is associated with measurable change in cord CSA.

The second main hypothesis is that registration-based methods can decrease the variance in longitudinal cord atrophy measurement by using the signal from multiple scans to improve robustness to image noise and artifacts and by regularization of the registration to constrain the degrees of freedom. We implemented three algorithms: boundary shift integral based on rigid registration, Jacobian integration based on deformable registration and scale factor computation based on constrained registration (composed of rigid and scale transformation). We evaluated the three registration-based methods by comparing them to two cross-sectional methods, as applied to three longitudinal data sets: 1) images with simulated cord atrophy; 2) images acquired in the dehydration study described above; and 3) images of 15 MS patients over a two-year interval. Our main result was that while registration-based

methods achieved more accurate results on simulation data sets and overall smaller measurement variance, they were not as sensitive, reporting no dehydration effect and smaller magnitude of patient cord atrophy. We argue that the limited spatial resolution of 1 mm of MR scans in our experiment is possibly the main reason and future studies of cord atrophy measurement using registration-based methods should be conducted on MR scans with a high spatial resolution such as 0.5 mm.

Preface

This thesis is an original and independent work by the author, Chunfang Wang. All of the experiments presented henceforth were conducted by Chunfang Wang in the MS/MRI Research Group at the University of British Columbia, Point Grey campus.

Chapter 2. An earlier version of the work presented in this chapter has been published in the journal *Spinal Cord*. MR scans for the dehydration study were collected at the MRI Research Centre, University of British Columbia, Point Grey campus. I performed all the cord cross-sectional area measurements, the statistical analyses and the manuscript composition.

> C. Wang, R.C. Tam, E. Mackie, D.K.B. Li, and A.L. Traboulsee. Dehydration affects spinal cord cross-sectional area measurement on MRI in healthy subjects. *Spinal Cord*, vol. 52, pp. $616-620$, 2014

Chapter 3. The c++ code for performing deformable registration using symmetric diffeomorphic demons algorithm is a refinement and extension based on the open source ITK implementation of this algorithm released in The Insight Journal. Dr. Roger Tam proposed the concept for the work presented in this chapter, and I was the lead investigator, responsible for all the algorithm implementation, data collection and test analyses.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

MS multiple sclerosis

EDSS expanded disability status scale

RRMS relapsing-remitting multiple sclerosis

PPMS primary progressive multiple sclerosis

NAWM normal appearing white matter

NAGM normal appearing grey matter

COV coefficient of variation

ROI region of interest

CSF cerebrospinal fluid

PV partial volume

CSA cross-sectional area

BSI boundary shift integral

JI Jacobian integration

DoF degree of freedom

SF scale factor

SD standard deviation

AD Alzheimer's disease

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Roger Tam for his patience, guidance and support throughout my studies in UBC. I could not have finished this thesis without you.

Thanks to Dr. David Li and Dr. Anthony Traboulsee for their valuable clinical opinions on my experiment. A special thank to Dr. David Li for his continuous support and insightful advice on my research.

Thanks to Dr. Burkhard Madler, Dr. Corree Laule and Trudy Harris for their assistance in developing the scanning protocol. Thanks to the volunteers for their time and effort, and also thanks to Emilie Mackie for recruiting the volunteers for the dehydration experiment.

I am very thankful to my thesis examiners: Dr. Purang Abolmaesumi and Dr. Rafeef Abugharbieh for providing their valuable time to serve on my examination committee.

I want to thank my friends in the MS/MRI Research Group: Dr. Lisa Tang, Youngjin Yoo, Tom Brosch and Saurabh Garg. You guys always help me and encourage me along this journey, and your friendships are highly appreciated.

Finally, I would like to thank my mama and papa, my younger brother and my boyfriend for their continuous love and support.

Chapter 1

Introduction

1.1 Background

1.1.1 Multiple Sclerosis

Pathological basis and clinical course of multiple sclerosis

Multiple Sclerosis (MS) is a chronic disorder in the central nervous system, which involves inflammatory demyelination and neuroaxonal degeneration. It causes focal lesions in the white and grey matter and diffuse unevenly distributed changes in the normal appearing white matter (NAWM) and grey matter (NAGM) in the brain and spinal cord [36]. With the inflammatory demyelination disseminated in the central nervous system, the lesions develop in association with the breakdown of the blood-brain barrier, leading to acute breakdown of myelin with a degree of axonal destruction [76]. These acute inflammatory lesions lead to acute relapses of neurological deficit as a result of conduction block due to the loss of myelin. With time going on, the axonal loss evolves to be the main pathological substrate and the neuroaxonal degeneration leads to progressive disabilities over time.

MS takes several clinical forms, with new symptoms occurring either in discrete attacks (relapsing forms) or accumulating over time (progressive forms). About 85% of cases begin with relapsing-remitting (RRMS) course, suffering from relapses (attacks of symptom flare-ups) before a slow remission (period of recovery). There is usually good recovery from such relapses as a result of resolution of inflammation, remyelination and cortical adaptation [76]. In 15% of the patients, the onset of MS is one of progressively increasing and irreversible disability, primary progressive MS (PPMS). Within next 6 to 10 years, about 65% of RRMS patients enter the secondary progressive

(SPMS) phase and are subjected to gradual progression of physical disability and cognitive impairment [107]. There is also a relatively rare type of MS called progressive-relapsing (PRMS) in which patients experience steadily worsening symptoms and attacks during the period of remission [79, 107].

MS is the most common cause of neurological disability in young adults, with a prevalence that ranges between 2 and 150 per 100,000 and varying widely in different regions [76]. Canada has one of the highest rates of MS in the world, and an estimated 100,000 Canadians have MS [1]. MS is typically diagnosed based on the presenting symptoms in combination with supporting MRIs. There is no cure of MS, and treatments attempt to return function after an attack, prevent new attacks and prevent disability. There are ten disease-modifying therapies approved by Health Canada to date [2].

Imaging biomarkers in MS

MRI has an important role for the assessment of patients with MS, because of its sensitivity to MS-related abnormalities and correlation to pathological changes [36]. Lesion based measures, typically the number and extent of T2-hyperintense lesions and T1-enhancing lesions after Gadolinium enhancing administration in the brain are very useful to diagnosis the disease and predict further evolution in the early phase of MS [34, 76].

Beside lesions, quantitative MRI discloses the presence of abnormalities in the NAWM and NAGM before the development of lesions. Reduction of the magnetization transfer ratio value [90], increase in the mean diffusivity on diffusion tensor MRI [82, 88] and decrease of the concentration of N-acetyl-aspartate on MR spectroscopy [80] in these regions have been used in MS research studies and they are associated with the cognitive impairment.

Measuring irreversible tissue loss in the brain and spinal cord on MRI, which represents the overall destructive pathological changes in the central nervous system has become an area with increasing interest. Atrophy of the white or grey matter in MS reflects the overall axonal and neuronal loss and is well associated with clinical disability and cognitive deterioration [36]. Atrophy based measures, like global whole brain atrophy, regional grey matter

atrophy and spinal cord atrophy, are among the most widely used measures in disease monitoring and treatment trials. Both brain atrophy and spinal cord atrophy have been closely associated with disability in established MS [34]. In patients with different MS subphenotypes, brain volume quantified from T1-weighted MR images decreases on average by about 0.7%-1.0% per year [77]. Spinal cord atrophy quantified from T2 and T1 spine images correlates well with measures of disability [36]. In a review study by Barkhof et al.[9] which compared many of the MRI biomarkers that have been used to track neuroprotection and repair after treatment in MS clinical trials, they found that atrophy of whole brain volume and spinal cord on serial MRI are able to demonstrate established and probable response to treatment, respectively, over the course of one year [9].

1.1.2 Involvement of Spinal Cord Atrophy in MS

Spinal cord atrophy, especially the cervical cord atrophy, is thought to have strong effect on the locomotor disability in MS. The cervical cord is found to be significantly smaller in patients with progressive MS, and a strong association of spinal cord area and clinical disabilities measured by Expanded Disability Status Scale (EDSS) score [60] has been demonstrated [46, 59, 61, 65, 66]. Cohen et al. [25] found that the cross sectional area (CSA) of cervical cord most strongly correlates with EDSS (r = -0.52, $p = 0.02$) in MS, when comparing with other MR measures like cord lesion volume, cerebral grey matter volume, cerebral white matter volume, whole brain volume and whole brain lesion volume. A study of 117 patients with SPMS [44] found that only cervical cord CSA correlates with EDSS score. A recent study on 440 patients of a mixed cohort of different MS subtypes [73] reported that cervical cord CSA was the most significant MR imaging parameter for explaining physical disability, as measured with the EDSS score.

The relative clinical importance of cervical cord atrophy in MS may reflect both anatomical and pathological considerations. The cervical cord contains all the descending corticospinal fibers which are destined from motor targets in the trunk, arms and legs [25]. It is a cross-road for all cerebrospinal

descending and spinocerebellar ascending pathways. Pathological changes to the small cervical cord disproportionally affects a myriad of central nervous system functions. A histopathological study by De Luca et al. [28] found that there were significant reductions in cord area, axon density and fibre diameter in the corticospinal and sensory tracts and these pathological abnormalities relate closely to functional disabilities. Pathological studies of spinal cord atrophy suggested that it is the overall axonal degeneration that is responsible for spinal cord atrophy in MS rather than the tissue loss inside the individual lesions [33].

Accelerated atrophy occurs in the spinal cord in MS at all stages of the disease, from presentation with clinically isolated symptoms to advanced progressive forms [19, 77]. Previous cross-sectional studies have found that upper cervical cord CSA is significantly smaller in patients with SPMS and PPMS compared with healthy controls, and correlates with EDSS [35, 46, 61, 65, 66]. The estimated annual atrophy rate of CSA is reported to be around 1.6% from longitudinal cord atrophy studies in patients with SPMS [43, 46] and around 5% in patients with PPMS [4]. While the spinal cord atrophy in progressive MS is evident, the detection of cord atrophy in RRMS has been more elusive. Multiple cross-sectional studies have found that cord CSA is not reduced in RRMS patients compared to controls [14, 15, 59, 75**?**] and that could be used to separate progressive MS and RRMS patients [15], while longitudinal studies were able to detect the cervical cord atrophy in RRMS patients over three years [87] and over two years [72].

Longitudinal studies provide valuable information with higher statistical power than cross-sectional studies, because they measure within subject difference and can overcome the population variance in cord size. However, longitudinal studies on cord atrophy are limited with only a handful studies conducted [4, 43, 46, 65, 87, 97]. Rashid et al. found a decrease in CSA over 3 years in early RRMS [87], while most cross-sectional studies in this subgroup [15, 75, 86] have shown no significant difference in CSA between RRMS patients and control groups. Longitudinal studies using a mixed cohort of PPMS, SPMS and RRMS [4, 67, 89, 97] observed significant cord atrophy over the study duration, but correlation between decrease in CSA

and EDSS progression was only presented in one study [67]. Furthermore, high scan-rescan variability observed greatly lowered the statistical power of the results in previous longitudinal studies.

Quantification of spinal cord atrophy is a useful biomarker for monitoring disease progression and therapeutic drug effects in MS. Lin et al. reported that the change in CSA was significantly related to changes in clinical disability in a cohort of the interferon β-1a(Rebif) treatment trial over four years [67]. Kalkers et al. found that neuroprotective agents riluzole appear to be more effective in reducing the rate of cervical cord atrophy in the short term [53]. Lucas et al. examined the intrathecal injection of triamcinolone acetonide therapy outcome in progressive MS using upper cervical cord atrophy and showed a negative correlation between the degree of cord atrophy and treatment benefit [71]. Spinal cord atrophy is thought to be more important than lesion measures in clinical trials when the therapy aim is to prevent disability, especially in progressive MS [77].

1.1.3 Spinal Cord Atrophy Measurement Methods

Measurement of cord atrophy is usually performed on T1-weighted MR scans with a 3D acquisition sequence. The measurement is conducted at the C2 to C5 level, since significant decreases of cord volume are mainly observed in the upper cervical region rather than in the lower thoracic and lumbar areas [59]. Because of the simplicity of the cylinder cord shape, the process of monitoring cord atrophy can be reduced to performing 2D cross sectional area measurement. The average CSA of a normal cervical cord is about $80\,\mathrm{mm}^2$ [66].

There has been a range of techniques proposed to measure cord CSA on MRI. Manual outlining on axially acquired gradient echo images were initially produced to estimate cord CSA [57]. A sequential two-year longitudinal study of 60 MS patients established a scan-rescan coefficient of variation (COV) of 6%, rendering this method unsuitable for serial monitoring [69].

Losseff et al. proposed a semi-automated intensity-based contouring algorithm to delineate the cord to measure the CSA [70]. The method is applied

to a short length of cord of five slices with the most caudal slice located at the C2/C3 intervertebral disc. On each axial image, the operator manually defines a region of interest (ROI) which locates the cord to estimate the mean cord intensity, and then defines another ROI which locates the CSF to estimate the mean CSF intensity. Voxels affected by partial volume averaging at cord–CSF interface would have intensity midway between the mean cord and CSF intensities on average. Thus, the boundary is detected by thresholding and region growing with the provided seed point at the interface [99]. The scan-rescan COV for an experienced operator is around 0.79% [70]. The accuracy of this method is limited by a systematic overestimation of 4.5%-10% [99] because they do not measure the partial volume issue on the cord boundary. This method has been employed in a number of cross-sectional and longitudinal studies [5, 43, 97] and is able to detect significant reductions in CSA in SPMS patients over 12 months [4, 43, 97].

Coulon et al. developed an automatic surface-based segmentation algorithm to obtain the segmentation of the cord and then estimate the cord volume and CSA [26]. The algorithm optimizes a B-spline surface model fitting to the cord such that intensity gradient is maximized globally while maintaining a smooth tube-like shape constraint to the detected cord edge [99]. With the surface model obtained, it is possible to automatically correct for alignment of the cord cross-section relative to the acquisition plane, and for curvature of the cord. However, this method provides less repeatable results with scan-rescan COV of 1.3%. Furthermore, there is an overall underestimation since the procedure tended to ignore voxels affected by partial volume. Hickman et al. employed this method in their study and found that the computed CSA measure at C2/C3 intervertebral disc had significant decrease over one year in a mixed cohort of PPMS, SPMS and RRMS patients [48].

Tench et al. used edge detection to identify the cord–CSF boundary [100], and for the first time addressed the computation of partial volume. For each axial slice image between C1 and C2 level, a Sobel edge detector is applied with non-maximal suppression to locate all the edges. On axial slice images where the cord has been successfully isolated by the detected

edges, the operator places a seed point and the cord region is segmented by region growing. All the voxels strictly within the cord are assigned an area value equal to the pixel size; while for voxels on the boundary, they are partial volumed by the cord and surrounding CSF and each contributes a fraction f of the pixel size to the total CSA, which is computed by the local intensities. The total CSA for each slice is then corrected for cord inclination by the cosine of the angle between the cord axis and vertical axis. The scan-rescan COV of Tench's method is reported to be 0.55%. There have been two clinical studies using this method [86, 100].

Horsfield et al. proposed to parametrize the cord cylinder-like surface by its center line and the radii. They realized the segmentation of the cord over an extended length with rather few user inputs. For each axial slice image, the cord is considered as a polygonal shape and the radii are considered as a periodically varying function of θ, which is the angle of the radii subtended to the positive x axis. The segmentation process then uses the intensity gradient to update the radii, and refines the center-line to be the centre of the area on each slice. The optimal surface model is obtained by optimizing the function using a multi-scale approach. Measurement of the average CSA is then derived as the volume divided by the length, where the length and volume can both computed by integrals of the parametric model. They integrated their algorithm in Jim software package (Version 5.0, Xinapse System, Northants, United Kingdom; `http://www.xinapse.com/home.php`), and a number of recent cord atrophy studies were conducted using Jim [25, 59, 89, 103].

Chen et al. presented a fully automated spinal cord segmentation algorithm which combines deformable registration with topology preserving intensity classification. Their method firstly align an intensity atlas to the target image to be segmented by deformable registration, and apply the deformation on a topology atlas and statistical atlas associated with the intensity atlas, which provides the initialization for the segmentation. The cord segmentation is achieved by iteratively evolving the topology atlas to convergence [22]. The result of cord atrophy measurement using this method is consistent with the results reported by Horsfield et al [50]. This method has been applied by Oh et al. in two spinal cord atrophy studies [81, 82].

Recently De Leener et al. developed a robust, accurate and automatic spinal cord segmentation algorithm based on the propagation of a deformable model [27]. The algorithm firstly detects the spinal cord position and orientation using a circular Hough transform on multiple axial slices and builds an initial elliptical tubular mesh. Then a low-resolution deformable model is propagated along the spinal cord with a local contrast-to-noise adaptation at each iteration. Finally, a refinement process and a global deformation are applied on the propagated mesh to provide an accurate segmentation of the spinal cord. This method can manage MR images with poor contrast between the spinal cord and CSF by adjusting constraints in the deformable model. Results suggested that the achieved accuracy by this method was higher than the manual segmentation, and was slightly higher to the accuracy obtained by Horsfield's method [50].

With the computed spinal cord segmentation, spinal cord atrophy is either defined as a measure of change in CSA relative to an age-matched normal control population in cross-sectional studies, or a measure of change in CSA over a period of time in longitudinal studies. In cross-sectional studies, total intracranial volume [47, 75], thecal sac absolute volume [47], the largest skull cross-sectional area [102], and the length of the cord [47, 59, 81] have been used to normalize the cord volume to remove inter-subject variations in cord size in cross-sectional studies. Normalization by the length of the cord has been demonstrated to improve the ability to detect group difference and strength the clinical-radiological correlations [47, 59, 81]. In longitudinal studies, the cord CSA or volume measurement is usually carried out at all time points and then the measurement at each time pint is subtracted to estimate the amount of atrophy over the scanning interval.

1.2 Thesis Motivation

There are many challenges in evaluating cervical cord atrophy measurement on MRI, like the small size of the cord, limited resolution of the spine MR images, involuntary patient motion introduced image artifacts and physiological fluctuations like water content change. Spinal cord is a small structure

with a diameter around 8 mm, and on current MR images using a spatial resolution of 1 mm, the cord is bounded to a region size of less than 100 voxels and about 30% of cord area is partial volumed with the surrounding CSF [75]. Furthermore, the magnitude of the cord annual atrophy rate in MS is small with around -1.6% per year reported in SPMS patients. These require the measurement method with high sensitivity and precision, because small absolute errors in the measurement methods can translate into large relative errors in the results and thus make the small cord changes difficult to detect. For example, the inconsistent findings of reduced CSA in RRMS patients compared to control subjects may be attributed to the variations in water content of the cord, which mask out the changes in CSA due to cord tissue loss. Moreover, in the few longitudinal cord atrophy studies conducted, the limited scan-rescan reproducibility (with large COV) and high measurement variance of the segmentation-based methods hindered the statistical value of their results. We want to address these problems from two aspects: to investigate the variations due to water content in cord CSA measurement and to develop registration-based methods for longitudinal cord atrophy measurement.

1.2.1 The Effect of Change in Water Content on Spinal Cord CSA Measurement on MRI

Change in tissue water content can have significant effects on the spinal cord volume, which has a high composition of water. In MS, edema associated with acute lesions and anti-inflammatory therapy can change the water content of the cord, and therefore change the cord volume. In normal conditions, the body hydration level makes the total body weight fluctuate by approximately 3% [106], which can potentially change the water content of the spinal cord. This is particular relevant to the cord volume or CSA measurement on MRI, because the MRI signal is primary derived from the hydrogen atoms in water [45]. Change in water content of the cord will be represented as change over the cord region on spine MR images, which can eventually affect the CSA measurement on the spine MR images.

However, it is unknown how does the change in water content affect cord CSA measurement and whether the effect would cause substantial variation in the spinal cord atrophy studies. No previous studies has been conducted to investigate the effect of change in water content on cord CSA measurement. Considering the small magnitude of cord atrophy rate in MS, the question of how does the change in water content affect the CSA measurement requires examination. In this thesis, we aim to investigate the dehydration effect on cord CSA measurement. Dehydration caused by restricted fluid intake has been reported to decrease the whole brain volume as measured on MRI by up to 0.55% [31, 94]. Based on the decrease in brain volume following dehydration from previous studies [31, 55, 56, 98], we hypothesize that dehydration would lead to a decrease in cord CSA measurement as well.

1.2.2 Registration-based Atrophy Measurement Methods for Longitudinal Cord Atrophy Studies

Scans from multiple time points can be used to directly measure the change in spinal cord volume or area without an accurate estimate of the absolute size at each time point. Changes can be directly assessed between serial scans using registration-based change analysis methods. This strategy has been proven successfully for reducing measurement variance in longitudinal brain atrophy measures compared to segmentation-based methods.

Boundary shift integral (BSI) [38, 64], and SIENA [96] employ the rigid registration and approximate the brain volume change by measuring the intensity difference (BSI) or intensity profile distance (SIENA) between each corresponding pair of edge voxels of the rigidly registered baseline and repeat images. These two techniques can significantly reduce the variance from segmentation errors by assessing the changes directly using intensity information. Voxel morphometry-based method Jacobian integration (JI) employs deformable registration and quantifies the change at each voxel by its Jacobian determinant of the deformation field obtained from deformable registration between the rigidly registered baseline and repeat images. Integration of the Jacobian determinant at each voxel over the object region

is used to estimate volume change. These registration-based methods have been shown to be more robust and accurate than segmentation-based methods by being less sensitive to image quality and imaging system changes and achieving a higher precision with smaller variance in brain atrophy studies [6, 32]. By increasing measurement precision and statistical power, samples sizes can be reduced, which in turn reduces the length and cost of clinical trials.

Although registration-based methods have been widely used with great success to measure longitudinal brain volume atrophy, none of the longitudinal spinal cord methods uses a registration-based longitudinal change analysis strategy. Current cord atrophy measurement methods as reviewed in Section 1.1.3 all require a segmentation of the spinal cord for each subject in the study. As most of the segmentation-based methods are prone to segmentation errors that can be of the order of the amount of cord atrophy seen in MS (with detectable percentage change of 1.16% for Horsfield's method and of 0.27% for Tench's method), we could not help to think whether similar registration-based atrophy measurement methods would be applicable to spinal cord atrophy measurement in longitudinal studies.

In this thesis, we applied three registration-based measurement methods to measure longitudinal cord atrophy. Our hypothesis is that the registration-based measurement methods can reduce the variability in the longitudinal cord atrophy measurement. The cord CSA is inherently small with 30% of its area is partial volumed with the surrounding CSF on MR scans with a spatial resolution of 1 mm. The variability in the measurement using segmentation-based cross-sectional methods are mostly derived from the image noise and the partial volume region. With regularization using different types of transformation, registration-based measurement methods may counteract the extend of partial volume as well as enhance the signal to noise by directly looking at the difference.

1.3 Thesis Contributions

1.3.1 Dehydration Effect on Cord CSA Measurement

We designed the scanning protocol with the radiologists and collected T1-weighted MR scans from 10 volunteer subjects at four time points with a dehydration and rehydration protocol. We implemented the CSA measurement method based on Tench's method and test two modifications for partial volume computation. We measured the CSA on all the MR scans at four time points using Tench's method with three different partial volume computation approaches. We also used the Horsfield method which is integrated in Jim software to segment the cord and measure the CSA on all the scans. We calculated the percentage change in CSA from baseline to each subsequent time point for all the measurements obtained by Tench method and Horsfield method. Then we used statistical analyses (one-tailed Wilcoxon rank test) to assess the significance of the changes to determine the dehydration effect. We found that dehydration does have significant effect on the cord CSA measurement. A significant decrease in CSA after dehydration was observed in the CSA measurement obtained using Tench method (one-tailed Wilcoxon rank test $p = 0.018$) and there was a similar magnitude of decrease in CSA measure obtained using Horsfield method which was close to significance (one-tailed Wilcoxon rank test $p = 0.052$). A mean decrease of 0.65% in CSA was observed after dehydration in the results of both methods, which was consistent with the results from previous studies of dehydration effect on brain volume measurement.

1.3.2 Registration-based Cord Atrophy Measurement Methods

The idea of the registration-based cord atrophy measurement method is to estimate the change in cord size by the intensity differences between corresponding voxels assessed from registration. We explored ways to improve the sensitivity of the longitudinal atrophy measurement methods using registration with different levels of regularization in forms of three types of

transformation: rigid registration, deformable registration and constrained registration.

First, we used rigid registration to align the input images and calculated the boundary shift integral to estimate the cord atrophy. The boundary shift integral algorithm assumes that a change in cord volume is associated with an exact shift in the cord–CSF boundary. Measurement of the cord atrophy can be estimated by the cord boundary shift, which can be computed by the integral of intensity differences within specified intensity window between corresponding voxels over the cord boundary.

Second, we used deformable registration seeking to improve the sensitivity and calculated the integration of Jacobian determinants of voxels over the cord region in the deformation field to estimate the cord atrophy. The deformation field obtained from deformable registration can be used to visualize the structural change between the baseline and follow-up images, and is therefore used to quantify the local change by the Jacobian determinant of each voxel. Integration of the Jacobian determinants of the voxels over the cord region is calculated to estimate the cord atrophy.

Third, we used constrained registration with three parameters (two translations and one scaling factor) seeking to improve the robustness of the measurement by adding constraints in the registration. The uniform scale factor in the x and y axes obtained from constrained registration is used to estimate the change in cord size.

We evaluated these three registration-based methods on the following test data sets, 1) two sets of scan pairs with simulated atrophy created by scaling to quantify the measurement precision, 2) the scan–rescan pairs to quantify the measurement reproducibility, 3) the dehydration scan pairs with demonstrated dehydration effect and an MS patient data set with reported cord atrophy over a two-year interval to quantify the measurement sensitivity. We compared the results obtained by three registration-based methods on the test data sets to the results obtained by two segmentation-based methods [50, 100], which are currently utilized as standard approaches in spinal cord atrophy studies in MS.

The three registration-based methods obtained accurate results on the

data set with simulated atrophy. On the data set with rigid transformation and simulated atrophy, the errors in change rates computed by the three registration-based methods are with significant smaller variance than that of the two segmentation-based methods. The scan-rescan reproducibility computed by the three registration-based methods were comparable to that of the two segmentation-based methods. However, registration-based methods were not able to detect the dehydration effect (0.65% decrease in cord volume) on the dehydration scan pairs. On the MS patient data set, although these registration-based methods detected significant cord atrophy over a two-year interval with smaller measurement variance, they were not as sensitive as the two segmentation-based methods, reporting much smaller magnitudes of cord atrophy rate.

To the best of our knowledge, it is the first time that the applicability of registration-based methods to measure longitudinal spinal cord atrophy is investigated. Although registration-based methods achieved smaller measurement variance than the segmentation-based methods on the test data sets, they are not as sensitive. We argue that the limited spatial resolution of 1 mm and the inherently small size of the cord are probably the main reasons for the limited sensitivity of registration-based methods. MR scans with a spatial resolution of less than 1 mm, and which are able to differentiate the grey matter and white matter over the cord region, are required for future studies on longitudinal spinal cord atrophy measurement using registration-based methods.

1.4 Thesis Outline

An outline of this thesis is listed as follows:

Chapter2 We assessed the dehydration effect to the cord CSA measurement on MR scans acquired with a dehydration protocol using two cord CSA measurement methods. Significance analysis of the results and discussion about the dehydration effect are presented in this chapter.

Chapter3 We presented three registration-based atrophy measurement techniques and their results on the test data sets. Discussion of each method and its results are also provided.

Chapter4 We summarized this thesis and highlighted the main conclusions that can be drawn from our experiments.

Chapter 2

Dehydration Decreases the Cord Cross-Sectional Area Measurement on MRI

2.1 Introduction

MRI is commonly used to measure spinal cord atrophy in studies of neurodegenerative diseases such as MS [66, 70, 105]. Atrophy of the spinal cord, in particular the cervical cord [59], has been shown to contribute to physical disability in MS [33, 66]. Most previous MS cord studies have used the correlation between the CSA of the cord and the EDSS score, a measure of locomotor disability in MS patients, as an indicator of the strength of the relationship between atrophy and disability [66, 105]. Significant correlations between CSA and EDSS have been found in PPMS and SPMS patients in previous cross-sectional studies and longitudinal studies [4, 47, 52, 67, 91, 102]. In mixed cohorts of progressive MS and RRMS patients, cord CSA has been found to be significantly smaller compared to healthy control subjects [47, 66, 70, 102]. However, there have been contradictory findings in patients with early MS. A number of studies have found that cord CSA is not reduced in RRMS patients [15, 59, 75] compared with controls and could be used to separate progressive and RRMS patients [15], while others have shown that cord atrophy is detectable in RRMS patients [87, 97].

The inconsistent findings of cord involvement in early MS may be due to many reasons. The average CSA of a normal adult cervical cord is around

$80\,\mathrm{mm}^2$. Most current MRI studies of the cord use a spatial resolution of $1\,\mathrm{mm}^2$ in each plane, which is around 1% of the cord CSA. Since atrophy is a slow process, with an annual atrophy rate of around -1.6% observed in SPMS patients [43], a high methodological sensitivity is essential to accurately estimate the true rate of change on current MRI data.

Variation in water content may probably confound the CSA measurement on MRI, because MRI signals are primarily derived from the hydrogen atoms in water [45]. Change in water content due to hydration status can affect brain morphology as observed on MRI [29, 55, 56]. Duning et al. reported that a cohort of 20 healthy volunteers showed a significant decrease in brain volume of 0.55% (SD = 0.69) after dehydration by restricted fluid intake for 16 hours and an increase of 0.72% (SD = 0.21) after rapid rehydration [31]. Kempton et al. observed a significant increase in ventricular volume following dehydration via a thermal exercise protocol in two studies of seven [56] and ten [55] healthy subjects. Dehydration can potentially affect the size of the spinal cord, which, similar to the brain, also has high water content.

Considering the small magnitude of cord atrophy in diseases such as MS, the question of whether cord CSA measurement on MRI is susceptible to dehydration requires examination. The goal of this study is to estimate how much variation in CSA can be expected due to dehydration to the degree that would not be considered unusual in daily functioning.

2.2 Materials and Methods

2.2.1 Subjects and MRI Procedure

The subjects recruited for this study are 10 volunteers, aged 21 to 32, with no symptoms of neurological disorders or spine problems. The subjects gave informed consent in accordance with institutional regulations. Images were acquired using a Philips Achieva 3T MRI scanner (Philips Medical Systems, Best, The Netherlands) with a dedicated cervical spine receiver coil. The sequence is a sagittal 3D T1-weighted turbo field echo sequence with parameters: TR = 8.206 to 8.290 ms, TE = 3.790 to 3.834 ms, flip angle = 8°, pixel

spacing = 0.976×0.976 mm, slice thickness = 1.000 mm, and dimensions = $256 \times 256 \times 60$ pixels. Due to the inclusion of other cord sequences in the same session, there was insufficient time to acquire brain scans, which would have necessitated a coil change.

2.2.2 Dehydration Protocol

We employed a similar dehydration and rehydration protocol to that used by Duning et al. to study the effect on whole brain volume [31]. For each subject, MR scans were obtained at four time points over two days: 1) baseline, 2) rescan after one hour, 3) the next morning after fasting for at least 14 hours, 4) after drinking 1.5L of water over the course of one hour. The subjects were asked not to exercise strenuously during the two days of their participation in the study.

2.2.3 MR Image Analysis

The cord CSA in each scan was measured using two established semi-automatic methods. One is an in-house method that is a modified version of the technique by Tench et al. [100]. The other is an independent method by Horsfield et al. [50] which we used for cross-validation.

Modified Tench method

Similar to the Tench approach, the user interacts with our in-house software by marking the region of the cord to be measured in a sagittal view, then segmenting a number of consecutive axial slices while guided by an edge map that in most cases includes a well-defined contour of the cord, as shown in Figure 2.1(b). The operator places a seed point inside the cord on each slice and initiates a region growing process that is bounded by the contour, as shown in Figure 2.1(c). In the present study, we used a single sagittal landmark on the most inferior and posterior point of the C2/C3 intervertebral disc, and the eight slices superior to the landmark were used to compute an average CSA.

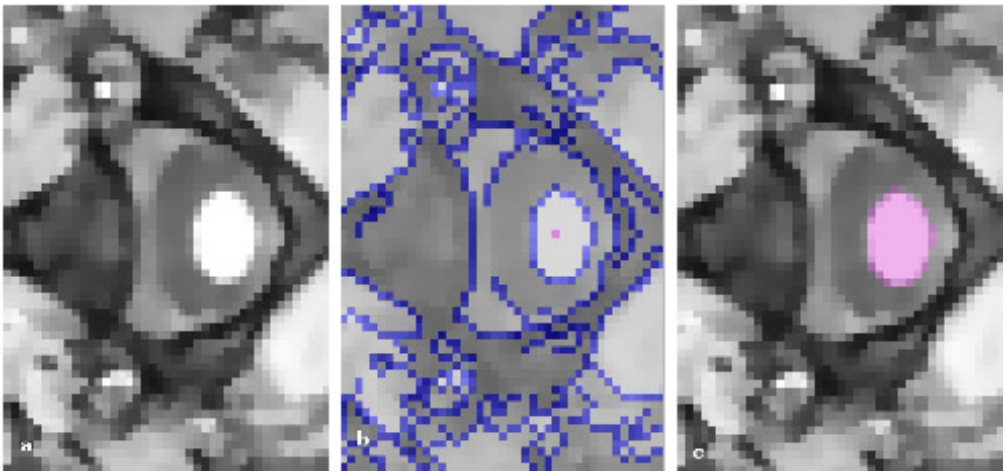


Figure 2.1: a) An axial image of the cervical cord surrounded by CSF. b) Edge map which includes a well-defined contour of the cord. The user-placed seed point is shown. c) Segmented cord region bounded by the edge contour.

We have made a number of improvements over the basic implementation of Tench method to allow the user to have greater control over the edge map and to improve the robustness of the partial volume computation. We use a Canny edge detector [21] which incorporates noise reduction, suppression of gradients that are not local maximum, and hysteresis thresholding with two thresholds. Hysteresis thresholding works by first applying a higher, more restrictive threshold to extract the strongest edges in the image, which typically includes a good part of the cord boundary, then applying a second, lower threshold that is used exclusively to link the strong edges already found. Our software interface allows the operator to interactively adjust the higher threshold, and the lower threshold is internally set to half of the upper threshold, and the resulting edge image is immediately displayed to the user as an overlay. This simple procedure effectively removes spurious edges within and around the cord that can interfere with region-growing and partial volume computation.

Partial volume computation

For each segmented axial slice, all of the pixels strictly inside the cord are assigned an area value equal to the pixel size ($1 \times 0.976\,\text{mm} = 0.976\,\text{mm}^2$) while each boundary pixel is assigned a value that is a fraction of the pixel size, modeled as partial volume (PV) between the cord and surrounding cerebrospinal fluid(CSF). The PV fraction f is calculated using Equation (2.1)

$$f = (I_{\text{edge}} - I_{\text{CSF}})/(I_{\text{cord}} - I_{\text{CSF}}) \tag{2.1}$$

where I_{edge} is the intensity of the boundary pixel, I_{CSF} and I_{cord} are the intensities of the CSF and cord. The fraction f is then multiplied by the pixel size to obtain the contribution of the boundary pixel to the slice area.

We tried three approaches to estimate the intensity of CSF I_{CSF} and the intensity of the cord I_{cord} in Equation (2.1). The first approach, denoted by PV1, is to interpolate the image and estimate I_{CSF} and I_{cord} at a distance of one voxel width from the edge voxel along their gradient directions on both sides. The second approach, denoted by PV2, is to interpolate the image and estimate I_{CSF} and I_{cord} by four sample points neighboring the edge voxel along their gradient directions on both sides, as shown in Figure 2.2. The first point is located at a distance of one voxel width from the edge voxel along its gradient direction, and three more sample locations are computed from the first point, with one further along the line of the strongest gradient, and the other two perpendicular to it. If a sample location falls on an edge pixel as determined by the Canny detector, it is unlikely to be a "pure" sample and is therefore discarded. The third approach, denoted by PV3, uses the median intensity value of the cord region on each slice to approximate I_{CSF} and the median intensity value of the CSF region on each slice to approximate I_{CSF}; the cord region is determined by the one pixel eroded region of the cord segmentation and the CSF region is determined by the one pixel dilated region of the cord segmentation minus itself.

The CSA for each slice is calculated by summing the contributions from the interior and boundary pixels. A correction factor is then applied to the CSA value to compensate for the fact that the cord is rarely perfectly

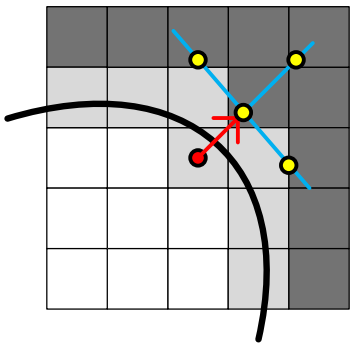


Figure 2.2: Four CSA sample points (yellow) in a T-shaped neighborhood of the boundary pixel (red). The mean intensity of the CSF sample points is used with the mean intensity of four cord sample points, computed similarly with a neighborhood inside the cord, to compute a partial volume estimate for the border pixel.

perpendicular to the axial image plane, resulting in an overestimation of the area. Using the same procedure preformed by Tench et al. [100], the correction factor is the cosine of the angle between the medial axis of the cord, as estimated by fitting a straight line through the centers of the segmented slices, and the perpendicular to the axial plane.

Horsfield method

In the method by Horsfield et al.[50], the operator first performs angle correction by rotating the volume so that its edges are parallel to the cord in the region of interest (C1-C2 in the current study). Then the operator places landmarks at C1-C2 region, and on every 10^{th} slice in-between. The software then uses these landmarks to automatically initialize a 3D surface and segments the spinal cord by fitting the surface to the image. There are three key parameters in the algorithm related to cord size and shape, including nominal cord diameter, number of shape coefficients, and order of longitudinal variation, which can be used to customize the fitting process. We used the default values in our analysis because the study is of healthy subjects and the segmentation appeared to be visually accurate.

The total cervical cord volume was calculated by the software as an

integral on the final fitted surface model. The average cervical cord CSA was then derived as the total volume divided by the length of the segmented region.

Statistical analysis

The percentage changes in CSA from baseline to the other three time points were calculated for each patient, and the mean changes between from the baseline to the other three time points were used to determine the scan-rescan reproducibility, the dehydration effect and rehydration effect. The statistical significances of the changes were assessed using one-tailed Wilcoxon rank test, with $p < 0.05$ as the threshold.

2.3 Results

The cervical cord CSA of the 10 subjects at four time points were assessed by the modified Tench method using three different PV computation approaches (PV1, PV2 and PV3 described in Section 1.1.3). The change rates from the baseline to the other three time points calculated by PV1, PV2 and PV3 are explained in Table 2.1. The scan-rescan coefficient of variation (COV) of the results using PV2 and PV1 were 0.634% and 0.616%, respectively. The scan-rescan COV of the results using PV3 was 0.736%, higher than that of PV1 and PV2. The results of PV1 and PV2 both demonstrated significant decrease in cord CSA after dehydration with one-tailed Wilcoxon rank test $p = 0.010$ and $p = 0.018$, respectively. Results of PV3 did not demonstrate any significant change in cord CSA after dehydration (one-tailed Wilcoxon rank test $p = 0.080$). Since the results of PV2 had slightly smaller scan-rescan COV than the results of PV1, we chose PV2 as the partial volume computation approach used in our in-house software.

In the results of our in-house software (Tench method using PV2), the cervical cord CSA of the 10 subjects at baseline were within the range of $69.13 - 92.12\,\mathrm{mm}^2$. The scan-rescan coefficient of variations was 0.616%. After dehydration, we observed a mean decrease of -0.654% (SD = 0.778, p

Table 2.1: Statistical results of the percentage changes in CSA computed by Tench method using three different PV computation approaches (PV1, PV2 and PV3). A significant reduction in cord CSA is observed after dehydration, with a return to the baseline-equivalent after rehydration.

	Methods	Mean change % (SD)	One-tailed Wilcoxon test (>0)	One-tailed Wilcoxon test (<0)
Δscan-rescan	PV1	-0.256(0.821)	$p = 0.052$	$p = 0.958$
	PV2	-0.217(0.794)	$p = 0.080$	$p = 0.934$
	PV3	-0.084(0.989)	$p = 0.652$	$p = 0.385$
Δbaseline to dehydrated	PV1	-0.695(0.801)	$p =$ **0.010**	$p = 0.958$
	PV2	-0.654(0.778)	$p =$ **0.018**	$p = 0.986$
	PV3	-0.465(0.891)	$p = 0.080$	$p = 0.935$
Δbaseline to rehydrated	PV1	0.073(1.354)	$p = 0.722$	$p = 0.312$
	PV2	0.121(1.318)	$p = 0.721$	$p = 0.313$
	PV3	0.251(1.213)	$p = 0.813$	$p = 0.216$

Table 2.2: Statistical results of the percentage changes in CSA computed by our in-house method (Tench method using PV2) and Jim software (Horsfield method). For both measurement methods, a decrease in cord CSA is observed after dehydration, with a return to the baseline-equivalent after rehydration.

	Method	Mean change % (SD)	One-tailed Wilcoxon test (>0)	One-tailed Wilcoxon test (<0)
Δscan-rescan	in-house	-0.207(0.794)	$p = 0.080$	$p = 0.934$
	Jim	-0.452(1.197)	$p = 0.161$	$p = 0.862$
Δbaseline to dehydrated	in-house	-0.654(0.778)	$p =$ **0.018**	$p = 0.986$
	Jim	-0.650(1.071)	$p =$ **0.052**	$p = 0.958$
Δbaseline to rehydrated	in-house	0.121(1.318)	$p = 0.721$	$p = 0.313$
	Jim	0.057(1.129)	$p = 0.539$	$p = 0.500$
Δdehydrated to rehydrated	in-house	0.782(1.208)	$p = 0.967$	$p =$ **0.042**
	Jim	0.715(0.939)	$p = 0.990$	$p =$ **0.014**

= 0.018 for one-tailed Wilcoxon rank test) in cord CSA as shown in Figure 2.3 and Table 2.2. After the rehydration procedure, the mean cord CSA was not significantly different from baseline (mean change = 0.121%, SD = 1.318). However, the mean change in CSA between the dehydrated and rehydrated states was significant (mean change = 0.782%, SD = 1.208, $p = 0.042$ for one-tailed Wilcoxon rank test).

Measured using the Jim software (Horsfield method), the cervical cord CSAs at baseline were within the range of $72.48 - 91.28\,\mathrm{mm}^2$. The scan-rescan COV of the results of Horsfield method was 0.899%. We observed a mean change in CSA of -0.650% (SD = 1.071, $p = 0.052$ for one-tailed Wilcoxon rank test) after dehydration compared with baseline, as shown in Figure 2.4 and Table 2.2. After rehydration, the mean CSA measurement increased by 0.715% (SD = 0.939, $p = 0.014$ for one-tailed Wilcoxon rank test) compared to the dehydrated state, but was not significantly different from baseline.

2.4 Discussion

We investigated the effect of mild to moderate dehydration and rehydration on CSA measurement of the cervical spinal cord in healthy subjects. We have observed a decrease in cervical cord CSA after fasting for an overnight period that would not be considered unusual in daily functioning.

We used two independent CSA measurement methods in our analysis to account for any bias introduced by either method. The two methods agreed well on the mean change in cervical cord CSA observed after dehydration (-0.654% for the Tench method and -0.650% for the Horsfield method), which is similar to the change of -0.550% observed by Duning et al. [31] in their study on dehydration effect to whole brain volume. In addition, the two methods agreed well on the mean change between the dehydrated and rehydrated states (0.782% for the Tench method and 0.715% for the Horsfield method), which is similar to the 0.720% increase in whole brain volume that they found.

We tried two partial volume computation approaches other than the one

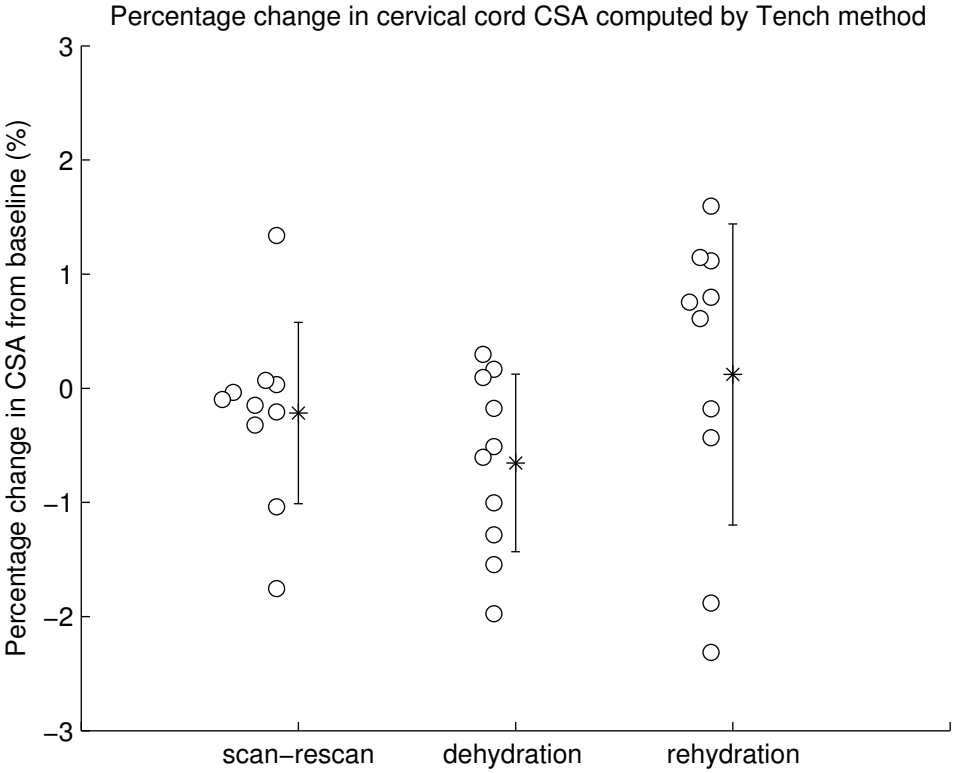


Figure 2.3: The percentage change in cervical cord CSA from baseline to the other three time points computed by our in-house software (Tench method). Each circle represent an individual subject, and the star and error bar represent the means and the standard deviations. Dehydration and rehydration appear to affect CSA measurements.

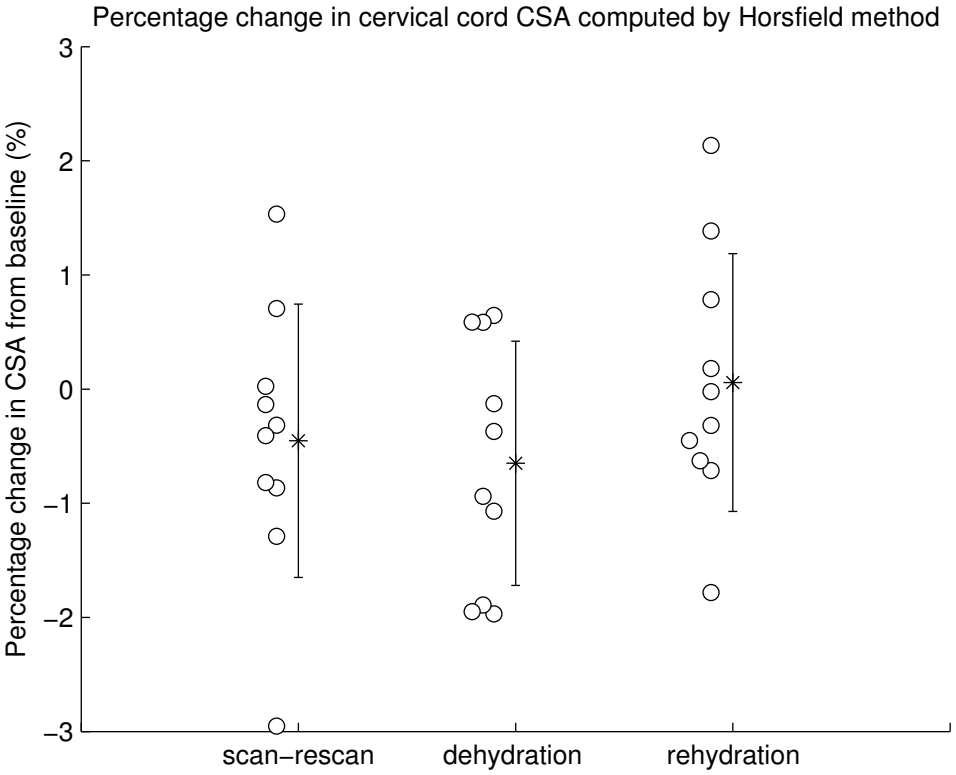


Figure 2.4: The percentage change in cervical cord CSA from baseline to the other three time points using software Jim (Horsfield method). Each circle represent an individual subject, and the star and error bar represent the means and the standard deviations. Dehydration and rehydration appear to affect CSA measurements.

explained in Tench et al.'s paper to estimate the CSF intensity I_{CSF} and the cord intensity I_{cord} in Equation(2.1), seeking to improve the accuracy of the calculated PV values. Our experiments on the PV computation emphasize that the accuracy of PV values which contribute to around 30% of the cord CSA is essential to the precision and sensitivity of the cord CSA measurement. The results of PV computation approach using median intensity of voxels over the CSF and cord region to estimate I_{CSF} and I_{cord} on each slice did not detect any dehydration effect. The results of the other two PV computation approaches which locally estimate the cord intensity and CSF intensity demonstrated significant dehydration effect. The CSF intensity and cord intensity in Equation(2.1) should be estimated locally for each edge voxel to compute the partial volume fraction.

The brain and spinal cord are directly connected and have similar mechanisms for their regulation of water balance [13, 93], so it is reasonable to speculate that the current results reflect the cord exhibiting an incomplete compensation to fluid deficiency, similar to what has been observed in the brain. Overall, our results lend further evidence that hydration status can affect volumetric measures of the central nervous system on MRI.

However, there are a number of limitations in our study, including the small size of the cord, the reproducibility of the measurement methods, and the small sample size. To understand these limitations, it is helpful to examine the precision when using each measurement method. Horsfield et al. estimated the minimum area change detectable by their method using the following equation: minimum detectable change = group mean $\times$ intra-scan COV $\times$ 1.96 = 0.87 mm^2 for their study. Alternatively, the detectable percentage change can be estimated by multiplying the intra-scan COV by 1.96 to obtain the change that can be detected with 95% confidence, which yields 1.16% and 0.27% for the Horsfield and Tench methods, respectively, using values from their published studies. As the intra-scan COV is image- and operator-dependent, we also estimated the precision of the two methods with our current data, which resulted in detectable percentage changes of 1.08% and 0.09% for the Horsfield and Tench methods, respectively. Our estimated change due to dehydration was 0.65% for both methods, which meets the 95%

confidence threshold of the Tench method, but is below that of the Horsfield method, which favors the former method but does not preclude the possibility of the latter method detecting a systematic change. Given that the two methods agree well on both the effects of dehydration and rehydration, and that at least one is confirmed to have the neccessary sensitivity on this data, we conclude that the change observed was likely real. It should be noted that because the Tench method was our own implementation, the operator was very familiar with the software, which may explain the better reproducibility, but otherwise, these results do not indicate that either method is superior.

Another limitation is that no brain scans (due to limited scanning time) were collected and a firm conclusion cannot be made about whether similar magnitudes of change can be expected in both structures. In addition, while the Duning study [31] found that mean brain volume increased beyond baseline after rehydration, the mean cord volume in our study was significantly increased only when compared to the dehydrated state, and not to baseline. A related confounding factor is the effect of brain volume changes on cord position. The cord could in theory be shifted rostrally due to the shrinkage of the brain after dehydration and because the cord does not have fixed landmarks (one could conceivably use the peripheral nerves, but they are very small and quite far apart), it would be very difficult to ensure that exactly the same level of cord is being measured. However, given that the dehydration effect on brain volume is likely to be less than 1%, the cord shift is likely to be correspondingly small, and in the absence of local injury, the cord diameter varies smoothly over its length, so we expect the effect of cord shift to be minor.

While there is increasing evidence that hydration status is a confounding factor in the volumetric analysis of MRIs, there is little information on how to correct for such fluctuations and whether this is even possible. Traditional methods for monitoring hydration status, such as urinalysis, are unlikely to be reliable due to the complex nature of the body mechanisms for water balance, which involves multiple systems whose health can change over time, even in normal aging. Nonetheless, studies of the hydration effect on brain

and cord measures are valuable for improving the understanding of study results that may be affected by changes in water content. The results of our current study have particular implications for studies of spinal cord disorders that involve an inflammatory response. For example, previous cross-sectional studies have shown that cord volume in MS patients can be increased (with varying levels of statistical significance) when compared to healthy controls, especially in early MS [59, 75]. These findings are somewhat unintuitive for a neurodegenerative disease, but are hypothesized to reflect the presence of inflammation and associated edema, which can induce a temporary increase in cord volume. Our current results help to bolster that hypothesis by demonstrating that measurable volume changes are associated with fluctuations in water content. In a longitudinal study of MS patients who had a spinal cord-related relapse [24], the patients showed a decline in cervical spinal cord area of approximately 0.7% monthly during the follow-up period of six months, even though they were improving clinically, which may be attributable to a resolution of inflammation and edema. Our current results show that changes of that magnitude can occur in a short period of time, and that frequent cord scanning after an acute episode can be a potentially useful method for monitoring edema. This is especially true given that certain MS therapies, such as natalizumzab, have been shown in human subjects to have a pseudoatrophy effect on the brain [78, 92, 106] and in preclinical studies to have a strong anti-inflammatory effect on the cord.

In conclusion, we have demonstrated that hydration status affects spinal cord CSA measurements on MRI and should be considered a source of variability in clinical studies of spinal cord atrophy. Our results also support the use of frequent MRI scanning to monitor conditions that may involve changes in water content, such as the inflammation and edema associated with acute spinal cord pathology.

Chapter 3

Registration-based Cord Atrophy Measurement Methods

3.1 Methods

Registration-based atrophy measurement methods, which directly quantify the change in volume between registered image pairs by registration, have been proven to be more precise and sensitive than segmentation-based methods in longitudinal whole brain atrophy measurement. However, none of the methods for accessing longitudinal cord atrophy uses a registration-based change analysis strategy. Current methods used to assess spinal cord atrophy in patients with MS, as reviewed in Chapter 1 Section 1.1.3 , are all cross-sectional methods based on the cord segmentation. In this chapter, we described the details of the pre-processing and three registration-based atrophy measurement methods that we implemented for longitudinal spinal cord atrophy measurement with the aim of improving the sensitivity and precision of the measurement: boundary shift integral, Jacobian integration and scale factor based on constrained registration (composed of rigid and scale transformation).

3.1.1 Preprocessing

Rigid Registration

The baseline scan and follow-up scan need to be positionally registered for registration-based measurement methods to yield meaningful results, be-

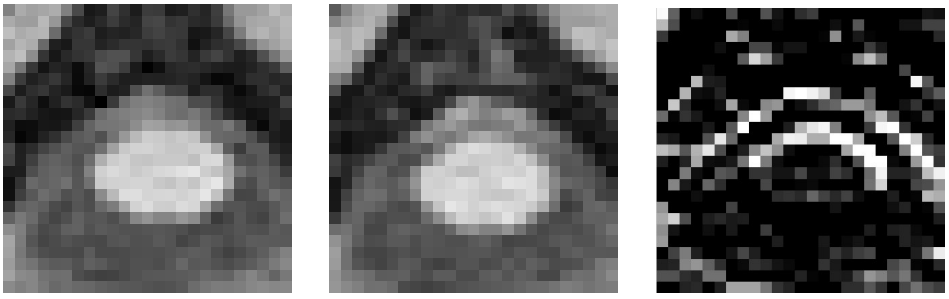
cause the intensity change on MR scans derived from real cord tissue loss can be small compared with the differences caused by patient movement and cord mismatch. For example, there is rigid transformation and 2% simulated atrophy created by scaling between Figures 3.1 (a) and 3.1 (b). Because of the spatial mismatch, the intensity changes due to the scaling atrophy are masked out by the intensity changes brought by rigid transformation, as shown on their difference image Figure 3.1 (c). After rigid registration, these two images are resampled as shown in Figures 3.1 (d) and 3.1 (e). The difference image of these two registered images is shown in Figure 3.1 (f), which exclusively shows the intensity changes due to scaling atrophy and makes it possible for the change analysis to accurately assess the simulated atrophy.

The steps to register the baseline scan and follow-up scan in our experiments are described as follows.

Firstly, the baseline and follow-up scans were segmented using our inhouse software described in Chapter 2 Section 1.1.3 to obtain the cord segmentation which contains 11 slices above and 5 slices below the C2/C3 intervertebra disc, with partial volume values on the cord boundary. The binary baseline segmentation $S^{(1)}$ and follow-up segmentation $S^{(2)}$ of the size of $25 \times 25 \times 17$ voxels were created by cropping and using the threshold of 0.6 on the partial volume region. The baseline and follow-up scans were also cropped to be $25 \times 25 \times 17$ axial subvolumes $I^{(1)}$ and $I^{(2)}$, respectively, using the cord segmentations $S^{(1)}$ and $S^{(2)}$.

We assumed that rigid transformation is able to align the cord region we examined (around a length of 16 mm) although the cord moves articulately with the vertebra. The baseline segmentation $S^{(1)}$ was dilated by one voxel slice by slice as labeled by the yellow line in Figure 3.3 (a). The dilated cord region in the baseline image was used to provide a mask for computing the mean square intensity difference cost function in the rigid registration to align the baseline image $I^{(1)}$ and follow-up image $I^{(2)}$. The resulting rigid transformation T was split into forward transformation T_{fwd} and backward transformation T_{bwd}. The baseline image $I^{(1)}$ and baseline segmentation $S^{(1)}$ were resampled using the backward transformation T_{bwd} to obtain the registered baseline image $I_{\text{r}}^{(1)}$ and registered baseline segmentation $S_{\text{r}}^{(1)}$. The

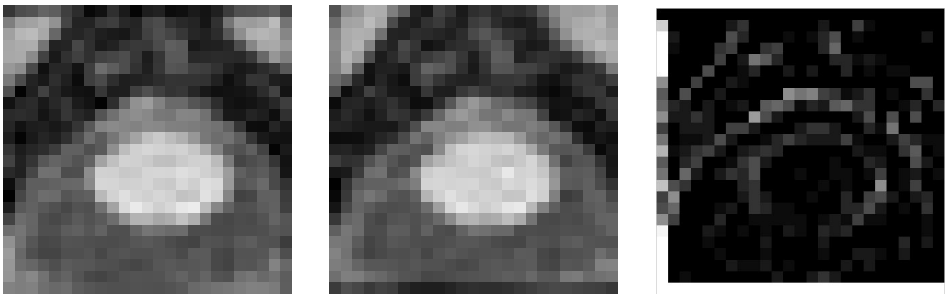
follow-up image $I^{(2)}$ and follow-up segmentation $S^{(2)}$ were resampled using the forward transformation T_{fwd} to obtain the registered follow-up image $I_{\text{r}}^{(2)}$ and registered follow-up segmentation $S_{\text{r}}^{(2)}$. This resampling procedure transforms the baseline and follow-up images to a position that is halfway between them to ensure that the two images being compared undergo equivalent processing steps.



(a) An axial slice of the baseline image

(b) An axial slice of the follow-up image

(c) The difference image of the baseline and follow-up images



(d) An axial slice of the registered baseline image

(e) An axial slice of the registered follow-up image

(f) The difference image of the registered baseline and follow-up images

Figure 3.1: Illustration of the difference image of the baseline and follow-up images before and after rigid registration. The baseline image (a) and follow-up image (b) are created from one control MR scan by scaling and rigid transformation. After rigid registration, the difference image (f) exclusively shows the intensity difference introduced by scaling atrophy.

Intensity Normalization

In order to maximize the accuracy of registration-based change analysis, the intensity of the same tissue on the baseline scan and the follow-up scan need to be as similar as possible. We performed the intensity mapping using the mean intensities of the CSF and cord on the registered baseline and follow-up images to correct for intensity and contrast differences. Firstly the registered baseline segmentation $S_{\mathrm{r}}^{(1)}$ and follow-up segmentation $S_{\mathrm{r}}^{(2)}$ were converted to binary images $S_{\mathrm{r\,b}}^{(1)}$ and $S_{\mathrm{r\,b}}^{(2)}$ using a threshold of 255*20%. The CSF mask region was defined as the binary segmentation dilated by one voxel minus itself $((S_{\mathrm{r\,b}}^{(i)} \oplus B) \setminus S_{\mathrm{r\,b}}^{(i)}$, where $i \in \{1, 2\})$. We utilized the morphological operator B, consisting of the origin and its nearest four neighbors in two dimensions. We calculated the mean and standard deviation of the intensities over the CSF region on the registered baseline and follow-up images, which are $(\mu_{\mathrm{CSF}}^{(1)}, \sigma_{\mathrm{CSF}}^{(1)})$ for $I_{\mathrm{r}}^{(1)}$ and $(\mu_{\mathrm{CSF}}^{(2)}, \sigma_{\mathrm{CSF}}^{(2)})$ for $I_{\mathrm{r}}^{(2)}$. Secondly the cord mask region is defined as the binary segmentation eroded by one voxel $(S_{\mathrm{r\,b}}^{(i)} \ominus B$, where $i \in \{1, 2\})$. We calculated the mean and standard deviation of the intensities over the cord mask region on the registered baseline and follow-up images, which are $(\mu_{\mathrm{cord}}^{(1)}, \sigma_{\mathrm{cord}}^{(1)})$ for $I_{\mathrm{r}}^{(1)}$ and $(\mu_{\mathrm{cord}}^{(2)}, \sigma_{\mathrm{cord}}^{(2)})$ for $I_{\mathrm{r}}^{(2)}$.

We performed a line fitting from the intensities of the follow-up image to the intensities of the baseline image using their two corresponding mean intensities $(\mu_{\mathrm{cord}}^{(2)}, \mu_{\mathrm{cord}}^{(1)})$ and $(\mu_{\mathrm{CSF}}^{(2)}, \mu_{\mathrm{CSF}}^{(1)})$. The mapping equation was defined by $y = ax + b$, where $a = (\mu_{\mathrm{cord}}^{(1)} - \mu_{\mathrm{CSF}}^{(1)})/(\mu_{\mathrm{cord}}^{(2)} - \mu_{\mathrm{CSF}}^{(2)}), b = (\mu_{\mathrm{CSF}}^{(1)}\mu_{\mathrm{cord}}^{(2)} - \mu_{\mathrm{CSF}}^{(2)}\mu_{\mathrm{cord}}^{(1)})/(\mu_{\mathrm{cord}}^{(2)} - \mu_{\mathrm{CSF}}^{(2)})$. The intensities of the registered follow-up image $I_{\mathrm{r}}^{(2)}$ were then normalized using the mapping equation to create the normalized registered follow-up image $I_{\mathrm{r\,n}}^{(2)}$.

The registered and normalized baseline and follow-up images $I_{\mathrm{r}}^{(1)}$ and $I_{\mathrm{r\,n}}^{(2)}$ were passed into the change analysis stage. Three different registration-based methods were implemented to estimate the change in cord volume with input images of $I_{\mathrm{r}}^{(1)}$ and $I_{\mathrm{r\,n}}^{(2)}$, as explained individually in the following three sections.

3.1.2 Boundary Shift Integral

The boundary shift integral (BSI) is a widely recognized technique firstly proposed by Fox et al. [38] to measure atrophy directly from the difference image of the registered serial MR images. It has been successfully used to measure the volume change in the whole brain [38, 64], the hippocampus [10, 12, 74], the caudate [49] and the ventricles [85]. Results from these applications have shown that the BSI algorithm is able to detect the difference in the atrophy rate in these tissues to distinguish the patient and healthy groups in a range of neurological disorders including Alzheimer's disease (AD) [11, 12, 18] and MS [6]. The rate of cerebral atrophy in patients with MS was 3 times that of aged matched controls over a 1-year period in a previous study using BSI [39]. Furthermore, the whole brain atrophy assessed by BSI has been used as an outcome measure in therapeutic intervention studies for AD [37].

The boundary shift integral algorithm assumes that a change in volume of a soft tissue object must be associated with an exact shift in the boundary of that object. The shift of the tissue boundary results in an exactly equivalent shift of the signal which is constructed from the MR samples [41]. Hence, if the baseline scan and follow-up scan are registered, in the area around the boundary of the registered scans I_{base} and I_{reg}, the intensities of $I_{\text{base}}(x, y, z)$ and $I_{\text{reg}}(x, y, z)$ should shift by an amount corresponding to the position shift; this permits the precise measurements of boundary shifts by determining intensity shifts in the boundary region. The change in volume can thus be estimated by computing the integral of all of the boundary shifts.

If $i_{\text{base}}(x)$ is the MR signal along the cord boundary at the location x of the baseline scan and $i_{\text{reg}}(x)$ is the MR signal at location x of a registered follow-up scan on which there has been a boundary shift of Δw from the baseline, then these two MR signals can be related by $i_{\text{reg}}(x) = i_{\text{base}}(x+\Delta w)$ in the region of the cord boundary [41]. Moreover, if the intensity changes monotonically across the cord boundary, then $i_{\text{base}}(x)$ and $i_{\text{reg}}(x)$ will take the form shown in Figure 3.2. We can therefore define inverse functions $x_{\text{base}}(i)$ and $x_{\text{reg}}(i)$, related by $x_{\text{reg}}(i) = x_{\text{base}}(i) - \Delta w$.

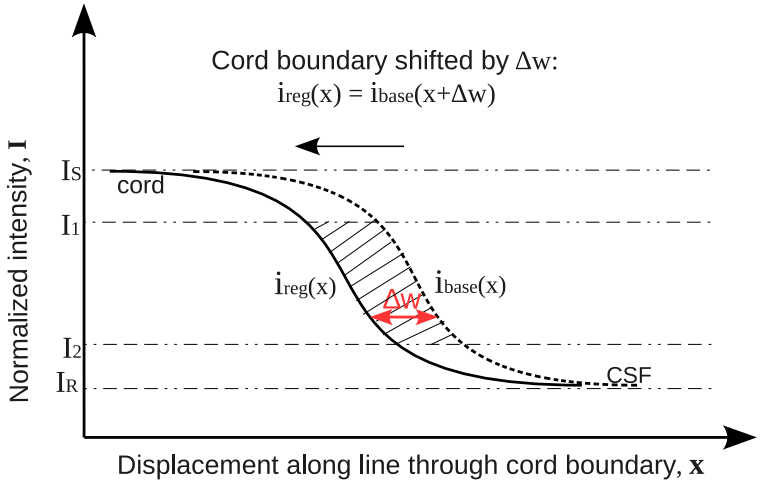


Figure 3.2: Example of an idealized one dimensional cord boundary shift between the intensity $i_{\text{base}}(x)$ along x axis on baseline scan, and the intensity $i_{\text{reg}}(x)$ along x axis on registered follow-up scan. An estimate of the shift along x axis, Δw, may be obtained as the shaded area divided by the intensity range $(I_1 - I_2)$. This strategy can be extended to three dimensions to estimate the cord volume loss Δv.

A simple estimate of Δw can be obtained using $\Delta w = x_{\text{base}}(i) - x_{\text{reg}}(i)$, where i may be any value within the intensity range of the cord boundary region $[I_{\text{R}}, I_{\text{S}}]$. In 3D T1 weighted spine MR images, the cord is brighter while the CSF is darker, thus I_{R} is the intensity on the CSF side of the boundary and I_{S} is the intensity on the cord side of the boundary. A more robust estimated can be obtained by averaging the estimates of Δw over an intensity range $[I_2, I_1]$, as shown in Equation (3.1).

$$\Delta w = \frac{1}{I_1 - I_2} \int_{I_2}^{I_1} (x_{\text{base}}(i) - x_{\text{reg}}(i)) \mathrm{d}i \tag{3.1}$$

where $I_{\text{R}} \leq I_2 < I_1 \leq I_{\text{S}}$.

Equation (3.1) can alternatively be expressed as an integral with respect to x over the boundary, as written in Equation (3.2). Equation (3.1) and

Equation (3.2) are equivalent by considering that both integrals evaluate the area of the shaded region in Figure 3.2.

$$\Delta w = \frac{1}{I_1 - I_2} \int_{\text{boundary}} (\text{Clip}(i_{\text{base}}(x), I_1, I_2) - \text{Clip}(i_{\text{reg}}(x), I_1, I_2))\text{d}x \quad (3.2)$$

where $I_{\text{R}} \leq I_2 < I_1 \leq I_{\text{S}}$, and $\text{Clip}(a, I_1, I_2) = \begin{cases} I_2 & a < I_2 \\ a & I_2 \leq a \leq I_1 \\ I_1 & a > I_1 \end{cases}$

If we extend this strategy to three-dimensions and determine the integral numerically by evaluating the integrand at small sampling intervals, the volume change can be calculated as shown in Equation (3.3)

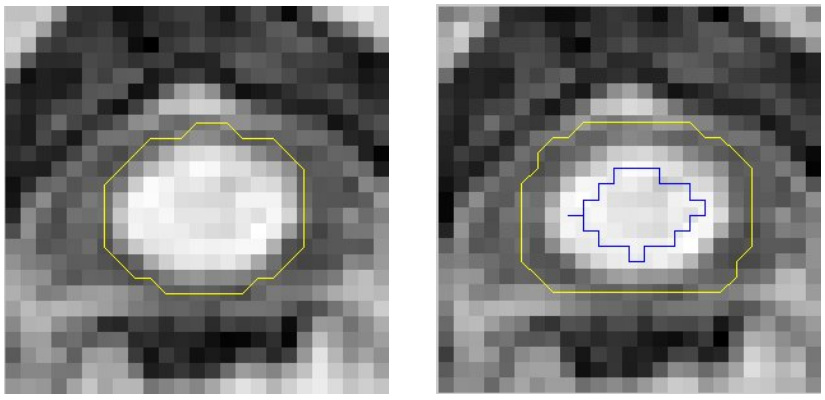
$$\Delta v = \frac{K}{I_1 - I_2} \times \sum_{x,y,z \in E} (\text{Clip}(I_{\text{base}}(x, y, z), I_1, I_2) - \text{Clip}(I_{\text{reg}}(x, y, z), I_1, I_2)) \quad (3.3)$$

where K is the unit voxel volume, E is the set of voxels in the border region of the cord, $I_{\text{base}}(x, y, z)$ and $I_{\text{reg}}(x, y, z)$ are the voxel intensities on the registered baseline and follow-up scans at (x, y, z), and the intensity range of the integral $[I_2, I_1]$ is referred to as the intensity window.

In our application to the cord, $I_{\text{r}}^{(1)}$ and $I_{\text{r n}}^{(2)}$ are the registered baseline and follow-up images, and $S_{\text{r}}^{(1)}$ and $S_{\text{r}}^{(2)}$ are the corresponding segmentation images. The resampled segmentation images $S_{\text{r}}^{(1)}$ and $S_{\text{r}}^{(2)}$ were converted to binary images $S_{\text{r b}}^{(1)}$ and $S_{\text{r b}}^{(2)}$ using a threshold of 255*50%. We defined the border region E as the set of voxels that are members of the union of $S_{\text{r b}}^{(1)}$ and $S_{\text{r b}}^{(2)}$ dilated by an operator B slice by slice but not members of the intersection of $S_{\text{r b}}^{(1)}$ and $S_{\text{r b}}^{(2)}$ eroded by an operator B slice by slice, as explained in Equation (3.4). We utilized the morphological operator B, consisting of the origin and its nearest four neighbors in two dimensions. The region between the yellow line and blue line on Figure 3.3(b) shows the border region E created by Equation (3.4) overlaid on the registered baseline image $I_{\text{r}}^{(1)}$, where the outer boundary of the border region is labeled by the yellow line and the inner boundary of the border region is labeled by the

blue line.

$$E = ((S^{(1)}_{\mathrm{r\,b}} \cup S^{(2)}_{\mathrm{r\,b}}) \oplus B) \setminus ((S^{(1)}_{\mathrm{r\,b}} \cap S^{(2)}_{\mathrm{r\,b}}) \ominus B) \tag{3.4}$$



(a) The boundary of the dilated cord segmentation of a baseline cropped cord image $I^{(1)}$ is labeled by the yellow line. The dilated baseline cord segmentation is used as a mask for the rigid registration between the baseline image $I^{(1)}$ and repeat image $I^{(2)}$.

(b) The border region is overlaid on the registered baseline cord image $I^{(1)}_{\mathrm{r}}$. The outer boundary of the border region is labeled by the yellow line. The inner boundary of the border region is labeled by the blue line. Boundary shift integral is computed over the border region.

Figure 3.3: Illustration of the dilated cord segmentation labeled on the baseline cord image $I^{(1)}$, and the border region labeled on the registered baseline cord image $I^{(1)}_{\mathrm{r}}$.

Intensity Window Selection

The evaluation of BSI requires the appropriate selection of an intensity window. The intensity window $[I_2, I_1]$ should be selected such that it falls entirely within the intensity transitions associated with the boundaries of a structure. Fox et al. [38] selected the intensity window for applying the BSI to whole brain atrophy measurement by comparing simulated and measured volumes of brain loss over a range of intensity window parameter values. Boyes et al. [17] improved the accuracy of BSI results by determining the intensity window parameters based on comparing the BSI to segmented volume differences for a range of windowing parameters. Hobbs et al. [49] used

the mean and standard deviation of different tissues involved to automatically calculate the intensity window. They took the median of all the specific optima of the subjects in their data set to be the optimal intensity window in their application using BSI to estimate caudate atrophy in a cohort study. Leung et al. [64] proposed determining the intensity window by the mean and standard deviation of CSF and GM which are estimated by the k-means clustering in the border region of the brain in their application to measure whole brain atrophy.

For our application to compute the cord BSI, we adopted the strategy Hobbs et al. proposed for automatic intensity window selection. In order to capture most of the tissue-type change between the cord and CSF, it was desirable to ignore changes within the same tissue type and maximize the changes between different tissue types. Therefore, the intensity window for the cord boundary shift computation was chosen to be $[\mu^{(1)}_{\text{CSF}} + \sigma^{(1)}_{\text{CSF}}, \mu^{(1)}_{\text{cord}} - \sigma^{(1)}_{\text{cord}}]$, where $\mu^{(1)}_{\text{CSF}}$ and $\sigma^{(1)}_{\text{CSF}}$ are the mean and standard deviation of intensities of the voxels in the CSF region on the registered baseline image $I^{(1)}_{\text{r}}$, and $\mu^{(1)}_{\text{cord}}$ and $\sigma^{(1)}_{\text{cord}}$ are the mean and standard deviation of intensities of the voxels in the cord region on $I^{(1)}_{\text{r}}$.

With the selected intensity window, BSI was computed using Equation (3.3) over the border region between $I^{(1)}_{\text{r}}$ and $I^{(2)}_{\text{r n}}$ on eight slices above the C2/C3 intervertebra disc landmark. Percentage change rate in cord volume was calculated by dividing the computed BSI by the cord volume of the baseline scan over eight slices.

3.1.3 Jacobian Integration

The deformation field obtained from deformable registration makes it possible to visualize structural changes that occur over time by deforming a subject's baseline scan onto their subsequent scans, and to statistically quantify local changes [62]. Based on the analysis of the Jacobian determinant of the deformation field obtained, temporal changes due to growth or atrophy can be identified [51]. This technique has been widely applied to assess the atrophy of whole brain as well as regional areas of the brain using different

forms of deformable registration [11, 42, 101].

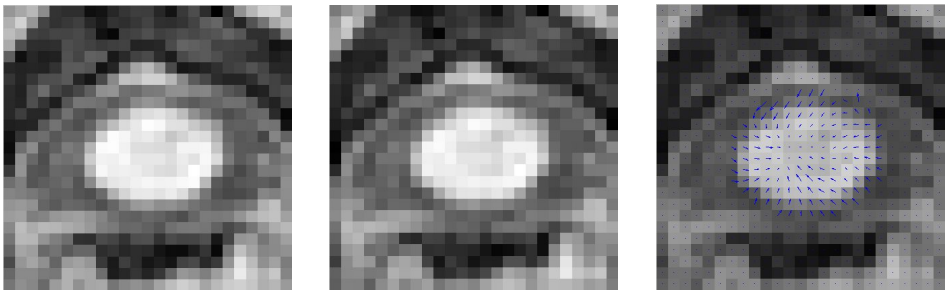
The quantification of the amount of warping applied at each voxel by the deformation field $T(x, y, z)$ can be locally derived from the Jacobian matrix $\nabla T(x, y, z)$ of the deformation in terms of determinant $\det(\nabla T(x, y, z))$ [23]. A Jacobian matrix $\nabla T(x, y, z)$ is obtained for each voxel by taking the secondary derivative of the deformation field $T(x, y, z)$, as defined in Equation (3.5).

$$\nabla T(x,y,z) = \begin{pmatrix} \frac{\partial T_x(x,y,z)}{\partial x} & \frac{\partial T_x(x,y,z)}{\partial y} & \frac{\partial T_x(x,y,z)}{\partial z} \\ \frac{\partial T_y(x,y,z)}{\partial x} & \frac{\partial T_y(x,y,z)}{\partial y} & \frac{\partial T_y(x,y,z)}{\partial z} \\ \frac{\partial T_z(x,y,z)}{\partial x} & \frac{\partial T_z(x,y,z)}{\partial y} & \frac{\partial T_z(x,y,z)}{\partial z} \end{pmatrix} \tag{3.5}$$

The determinant $\det(\nabla T(x, y, z))$, denoted by $\det(J)$, represents an expansion if $\det(J) > 1$ or a contraction if $\det(J) < 1$ at each voxel. The change in voxel size after deformation calculated by $(\det(J) - 1)$ can then be integrated over a specified region to obtain an approximation to the total volume change in this region. The estimated total volume change divided by the number of voxels in this region yields an average measure of volume change rate over this region.

Figures 3.4 (a) and 3.4 (b) show two images created from a spine MRI of a control subject, with simulated atrophy between them created by scaling. The baseline image Figure 3.4 (a) is created by applying a scale factor of 1.0075 on both the x and y axes of a spine MRI from a control subject to obtain 1.5% explansion in volume, and the follow-up image Figure 3.4 (b) is created by applying a scale factor of 0.9925 on both the x and y axes to the same MR scan to obtain 1.5% shrunk in volume. In this way, 3% simulated scaling atrophy is created between the baseline image and follow-up image. Figure 3.4 (c) shows the deformation field, which is obtained from deformable registration between the two images shown in Figures 3.4 (a) and 3.4 (b), overlaid on the baseline image over the cord region. There is obvious shrinkage of the cord region, illustrated by the deformation vectors pointing into the cord region on Figure 3.4 (c).

The deformable registration aims to find the displacement $s(p)$ at each



(a) An axial slice of the baseline image

(b) An axial slice of the follow-up image

(c) The deformation vectors of the deformation field over the cord region overlaid on the baseline image.

Figure 3.4: Illustration of the deformation field generated by images with 3% simulated atrophy. The baseline image (a) and the follow-up image (b) are created from one control MR scan by scaling to create 3% atrophy between them. After deformable registration, the deformation field over the cord region in (c) demonstrates the expected shrinkage of the cord, with the majority of the deformation vectors pointing into the cord.

voxel p in order to capture the shape change (similarity) as well as constrain the smoothness of the deformation field (regularization) from the moving image to the fixed image. The optimization of the deformable mapping can be summarized briefly as:

1. Compute the update field u by minimizing the similarity function
2. Update the deformation field s with u
3. Regularize the deformation field s

The non-rigid registration algorithm employed in our experiment is the symmetric log-domain diffeomorphic demons proposed by Vercauteren et al. [104]. The energy function to be minimized is defined in Equation (3.6),

$$E(F, M, s) = \frac{1}{\sigma_i^2}\mathrm{Sim}(F, M, s) + \frac{1}{\sigma_T^2}\mathrm{Reg}(s) \tag{3.6}$$

where F is the fixed image, M is the moving image, s is the deformation

field, $\mathrm{Sim}(F, M, s) = \frac{1}{2}\|F - M \circ s\|^2$, $\mathrm{Reg}(s) = \|\nabla s\|^2$, σ_i measures the local intensity noise, and σ_T controls the amount of regularization we need.

In order to register F and M, we need to optimize Equation (3.6) over a given space. The diffeomorphic demons algorithm performs the optimization in a space of diffeomorphisms to enforce invertibility by defining an exponential mapping from the vector space to diffeomorphisms [104]. At each iteration, the algorithm computes an update step u by minimizing the correspondence energy $E_{\mathrm{s}}^{\mathrm{corr}}(u) = \|F - M \circ (s \circ \exp(u))\|^2 + \|u\|^2$ with respect to u using a Newton method and then maps it to the space of diffeomorphisms through the exponential $\exp(u)$. Updating the deformation field s with the update step u is thus in the form of $s \leftarrow s \circ \exp(u)$ [104].

In order to easily compute the inverse of the spatial transformation, the complete spatial transformation s is encoded through the exponential of a smooth stationary velocity field v where $s = \exp(v)$ [7, 104], thus the inverse of the spatial transformation s^{-1} can be obtained by $\exp(-v)$ in the log domain. The update step $s \leftarrow s \circ \exp(u)$ is cast into the form of $v \leftarrow \log(\exp(v) \circ \exp(u))$ [104]. The updated transformation $\log(\exp(v) \circ \exp(u))$ can then be approximated by the first two terms of the Baker-Campbell-Hausdorff formula $v \leftarrow v + u + \frac{1}{2}[v, u]$, where $[v, u]$ is the Lie bracket: $[v, u] = Jac(v) \times u - Jac(u) \times v$ [16].

To make the registration framework symmetric, the global energy is symmetrized by $s_{\mathrm{opt}} = \arg\min_s(E(I_0, I_1, s) + E(I_1, I_0, s^{-1}))$, where $s^{-1} = \exp(-v)$ in the log domain, and E is defined by Equation (3.6), I_0 and I_1 are the two input images of the registration. The symmetric update is the mean of the forward and backward update steps in each iteration [30]. The registration is optimized using a multi-resolution scheme.

To sum up, the multi-resolution symmetric log-domain diffeomorphic demons registration is performed using the following steps :

- Choose a starting deformation field v
- For each resolution level, resample the deformation field v obtained from the previous resolution level be the same size as the fixed image at the current resolution level and use the resampled deformation field

to initialize the registration

- Iterate until convergence

    1. Compute the forward demons force update field u^{forward} where I_0 plays the role of the fixed image F and I_1 plays the role of the moving image M
    2. Compute the backward demons force update field u^{backward} where I_1 plays the role of fixed image F and I_0 plays the role of the moving image M
    3. Compute the symmetric update step $u = \frac{1}{2}(u^{\text{forward}} - u^{\text{backward}})$
    4. Update the current deformation field v with the symmetric update step u and apply the diffusion-like regularization $v \leftarrow K_{\text{diff}} * (v + u + \frac{1}{2}[v, u])$, where K_{diff} is a Gaussian convolution kernel with the parameter σ set to 1.0
    5. Compute the mean square error (MSE) after applying the updated deformation field v and compare the current MSE value with the two MSE values obtained from the previous two iterations. If the current MSE value is not smaller than the previous two values, convergence is obtained; otherwise continue to iterate

In our application to compute the cord change rate using Jacobian integration (JI), we used this deformable registration algorithm to obtain the deformation field between the registered baseline image $I_{\text{r}}^{(1)}$ and follow-up image $I_{\text{r n}}^{(2)}$ in our experiments. Firstly the resampled baseline segmentation $S_{\text{r}}^{(1)}$ was converted to a binary image $S_{\text{r b}}^{(1)}$ using a threshold of 255*50% and $S_{\text{r b}}^{(1)}$ was dilated by two voxels slice by slice to create the dilated cord mask region. Then the registered baseline image $I_{\text{r}}^{(1)}$ and registered follow-up image $I_{\text{r n}}^{(2)}$ were masked by the dilated cord segmentation to create two images $I_{\text{r m}}^{(1)}$ and $I_{\text{r n m}}^{(2)}$ with voxels outside the dilated cord region being zero. The masked images $I_{\text{r m}}^{(1)}$ and $I_{\text{r n m}}^{(2)}$ were used as the fixed and moving images, respectively, in the deformable registration, and the numbers of iterations chosen to be 5 and 10 for the two resolution levels (full-image resolution and

half-image resolution), respectively. With the obtained deformation field, the volume change rate was calculated by the mean value of $(\det(J) - 1)$ over the core region $S^{(1)}_{\mathrm{rb}}$.

3.1.4 Scale Factor from 3-DoF Registration

As cord atrophy is a slow process, it is reasonable to assume that the cervical cord will not significantly change its cylindrical shape dramatically within the time frame of most clinical studies. Therefore, in contrast to measuring local changes and summing them up as described in the boundary shift integral and Jacobian integration, a more strongly regularized approach to measuring global changes in volume can be attempted by using a constrained rigid registration with scaling. The potential advantage of this approach is increased robustness with respect to local artifacts.

We defined a 3-DoF transformation T containing three parameters: tx, ty and $scalexy$, where $scalexy$ is the scale factor in the x and y axes, and tx and ty are the translations in the x and the y axes respectively. The transformation uses one scale factor $scalexy$ in the x and y axes to model a cord that changes size uniformly. We added the translations tx and ty to adjust for small translations that have not yet been corrected by the rigid registration. The change rate of the cord size can be approximated by the scale factor $scalexy$ obtained from the 3-DoF constrained registration.

The 3-DoF constrained registration is realized in the ITK registration framework using the mean square intensity difference as the similarity metric shown in Equation (3.7). A regular step gradient descent optimizer is used to compute the update step u for each parameter using the gradient of the energy function with respective to each parameter.

$$E = \|F - M \circ T\|^2 \tag{3.7}$$

The 3-DoF constrained registration is performed using the following steps:

- Choose an initial transformation for T

- Iterate until convergence

    1. Compute the gradient ∇E of the energy function
    2. Compute the update step u by multiplying the gradient ∇E with the step length L that was usually set to be the maximum step length, $u = L \times \nabla E$.
    3. Compare the direction of ∇E of current iteration and the direction of the gradient of the previous iteration. If the directions are opposite, reduce the update step u by a relaxing factor r which is set to be 0.2, $u = r \times L \times \nabla E$
    4. If the length of the update step u is larger than the minimum step length, update the transformation to the new transformation $T \leftarrow T + u$ and continue to iterate, otherwise stop and the convergence is obtained

In our application to measure cord change rate, the 3-DoF constrained registration was performed on the registered baseline and follow-up images, with $I_{\mathrm{r}}^{(1)}$ used as the fixed image and $I_{\mathrm{r\,n}}^{(2)}$ used as the moving image. The resampled baseline cord segmentation $S_{\mathrm{r}}^{(1)}$ was used as the mask with only voxels inside the mask region included when computing the mean square intensity difference cost function. With the scale factor (SF) obtained after the constrained registration, the change rate of cord volume between baseline and follow-up scans can be computed by $(scalexy^2 - 1)$.

3.1.5 Experiments Performed

In order to assess the sensitivity and precision of these three registration-based methods which are abbreviated by BSI, JI and SF, different tests were performed on the scan pairs which are outlined below.

Scaled scan pairs

We created a simulation data set with different levels of simulated cord atrophy created by scaling between the scan pairs, assuming that the cord

changes uniformly as it undergoes atrophy, which is not strictly true but is a useful approximation. We applied two scale factors (one leading to an increase in volume and the other one leading to a decrease in volume) on each baseline scan in the hydration data set (described in Section 2.2.2) to create two simulated images to eliminate the problem of different degrees of blurring between the two simulated cord images introduced by interpolation. The steps to create one simulated scan pair with 2% scaling atrophy from a control scan are explained as follows. Firstly, the control scan is enlarged by increasing the voxel size by a scale factor of 1.005 in the x and y axes to create a simulated cord image with 1% increase in volume with respect to the control scan. Secondly, the control scan is shrunk by decreasing the voxel size by a scale factor of 0.995 in both x and y axes to create the simulated cord image with 1% decrease in volume with respect to the control scan. Thus, a simplistic simulation of 2% cord atrophy, an amount typical of a SPMS patient over the course of one year [43] is generated between the scaled scan pair.

From each baseline scan, we created three scaled scan pairs with three different levels of change in volume (-1%, -2% and -3%) using the scale factors listed in Table 3.1. Measures of change were computed using BSI, JI and SF on the scaled scan pairs of each baseline scan in the hydration data set. The difference between the computed change rate and the ground truth value represents the error of the result computed by the method. We calculated the mean and standard deviation of the errors obtained by BSI, JI and SF to assess the performance of these three registration-based methods on this simulation data set.

Scaled scan pairs with rigid transformation

To additionally simulate changes in patient position between scans, we created a data set with both simulated atrophy and rigid transformation to simulate the cord tissue loss and spinal cord repositioning.

From the baseline scan of each subject in the hydration data set, we created four simulated scan pairs with four different levels of scaling change

Table 3.1: Scale factors applied to create scaled scan pairs with three different levels of change in cord volume between the simulated baseline image and follow-up image.

Simulated change rate	simulated baseline image (scale factor)	simulated follow-up image (scale factor)
-3%	1.5% (1.0075)	-1.5% (0.9925)
-2%	1.0% (1.005)	-1.0% (0.995)
-1%	0.5% (1.0025)	-0.5% (0.9975)

(0%, -1%, -2% and -3%) generated by the scale factors listed in Table 3.1 , and with rigid transformation randomly generated by rotation parameters in the range of -4 degrees and 4 degrees and translation parameters in the range of -2 mm and 2 mm.

Measures of change rate in cord volume were computed using BSI, JI and SF on the four scaled scan pairs with rigid transformation for each subject. We also measured the change rate in cord volume on this simulation data set using the Tench [100] and Horsfield [50] methods for comparison. The difference between the computed change rate and the ground truth value represents the error of the result computed by the method. We calculated the mean and standard deviation of these errors to assess the performance of the three registration-based methods and the two cross-sectional methods on this simulation data set.

Hydration data set

The hydration data set described in Chapter 2 Section 2.2.2 is composed of serial MR scans of 10 healthy subjects at four time points (named baseline, rescan, dehydrated and rehydrated). Cord volume change was measured using BSI, JI and SF on the scan pairs from the baseline to the other three time points as described below.

- Scan-rescan scan pairs
  The cord size is assumed to be constant between these two scans which

were taken one hour apart, so the calculated cord volume change rate should be zero and any departure from zeros is assumed to be scan-rescan variance. Means and SDs of the results computed by BSI, JI and SF methods on the scan-rescan scan pairs were compared with results of the Tench and Horsfield methods to see whether registration-based methods are able to improve the scan-rescan reproducibility.

- Dehydration scan pairs
  We computed the measures of change using BSI, JI and SF from the baseline to the dehydrated scans. A one-tailed Wilcoxon rank test was performed on the BSI, JI and SF measures on the dehydration scan pairs to see whether they are able to detect any dehydration effect, which was reported in the results of Tench and Horsfield methods (explained in Chapter 2) and whether they can improve the measurement sensitivity.

- Rehydration scan pairs
  There is a return in the cord CSA after rehydration and no significant change is observed from the baseline to the rehydrated scans in the results of Tench and Horsfield methods. We computed the measures of change using BSI, JI and SF on the rehydration scan pairs to see whether registration-based methods are able to yield similar results as the two cross-sectional methods and whether they can improve the measurement precision.

MS patient data set

The MS patient data set used in our experiment is composed of MR scans of 15 SPMS patients selected from a negative MS clinical trail. For each patient in the MS patient data set, there are two 3D T1 weighted MR images collected at two time points with a two-year interval.

There should be significant cord atrophy between the scan pairs with a two-year interval in this MS patient data, because SPMS patients are typically characterized by gradual progression of their disabilities and cognitive

impairment and steady annual cord atrophy rate has been reported in this subtype [43]. Two-year change rate in cord volume was measured using BSI, JI and SF as well as the Tench and Horsfield methods. A one-tailed Wilcoxon rank test was performed on the change rate computed on the scan pairs in the MS patient data set to see whether significant cord atrophy can be detected by these methods and whether registration-based methods are able to improve the sensitivity and precision compared with the segmentation-based methods in the MS patient data set.

3.2 Results

No failed rigid registration was detected in the preprocessing of all the scan pairs in our experiments. The rigid registration was evaluated by computing the Dice coefficient [58] of the resampled segmentation images $S_{\mathrm{r\,b}}^{(1)}$ and $S_{\mathrm{r\,b}}^{(2)}$. The computed Dice coefficients are in the range of 0.90 to 0.99, indicating that the resampled baseline and repeat scan pairs $I_{\mathrm{r}}^{(1)}$ and $I_{\mathrm{r}}^{(2)}$ are fairly rigid registered in our experiments.

3.2.1 Scaled Scan Pairs

Table 3.2 lists mean and SD of the differences of the percentage change rates computed by BSI, JI and SF to the ground truth values in the simulation scan pairs with three different degrees of scaling atrophy. The three registration-based methods all obtained accurate results (small mean errors) on this simulation data set with SF achieving the lowest magnitude of absolute errors.

3.2.2 Scaled Scan Pairs with Rigid Transformation

The percentage change rates computed by BSI, JI and SF as well as Tench and Horsfield methods for simulated scan pairs with random rigid transformation and four different levels of scaling atrophy (0%, -1%, -2% and -3%) are shown in four Figures 3.5 , 3.6 , 3.7 and 3.8 , respectively.

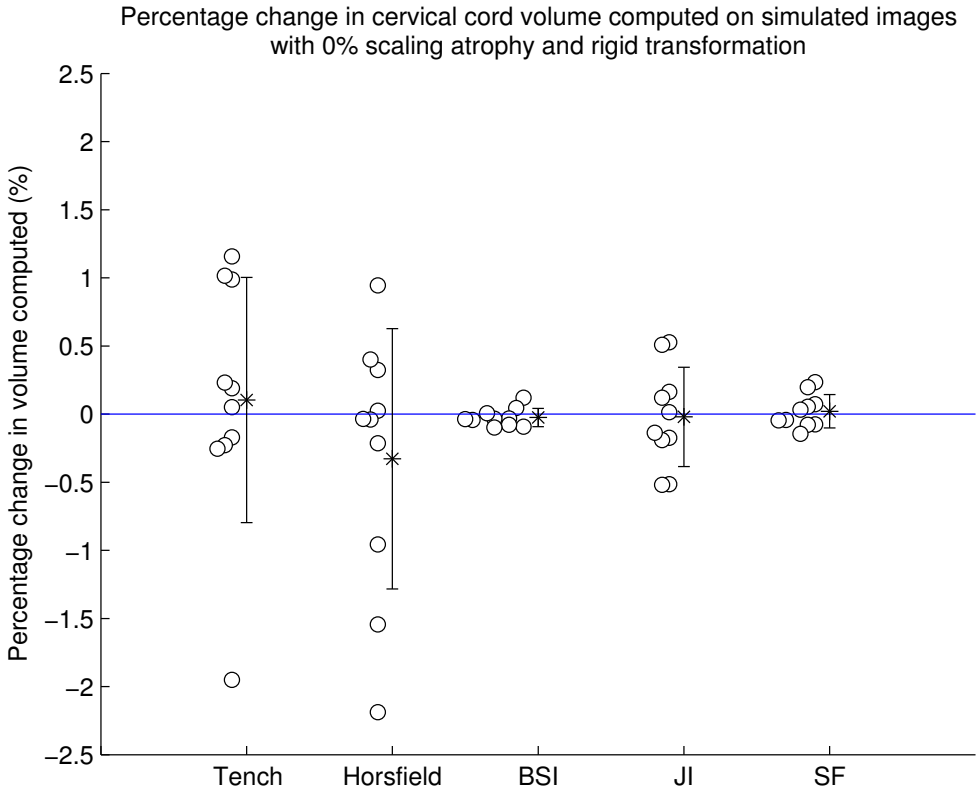


Figure 3.5: Percentage change rates computed by the three registration-based methods (BSI, JI and SF) along with the two segmentation-based methods (Tench and Horsfield) on simulated scan pairs with rigid transformation and no scaling change (0%). Each circle represents a single subject in the simulated scan pairs, and the stars and error bars represent the means and standard deviations of the measurements. The blue line represents the ground truth change rate between the two scans in the simulated scan pairs. Registration-based methods outperformed the segmentation-based methods with smaller mean errors to the ground truth and smaller measurement variance on this simulated scan pairs .

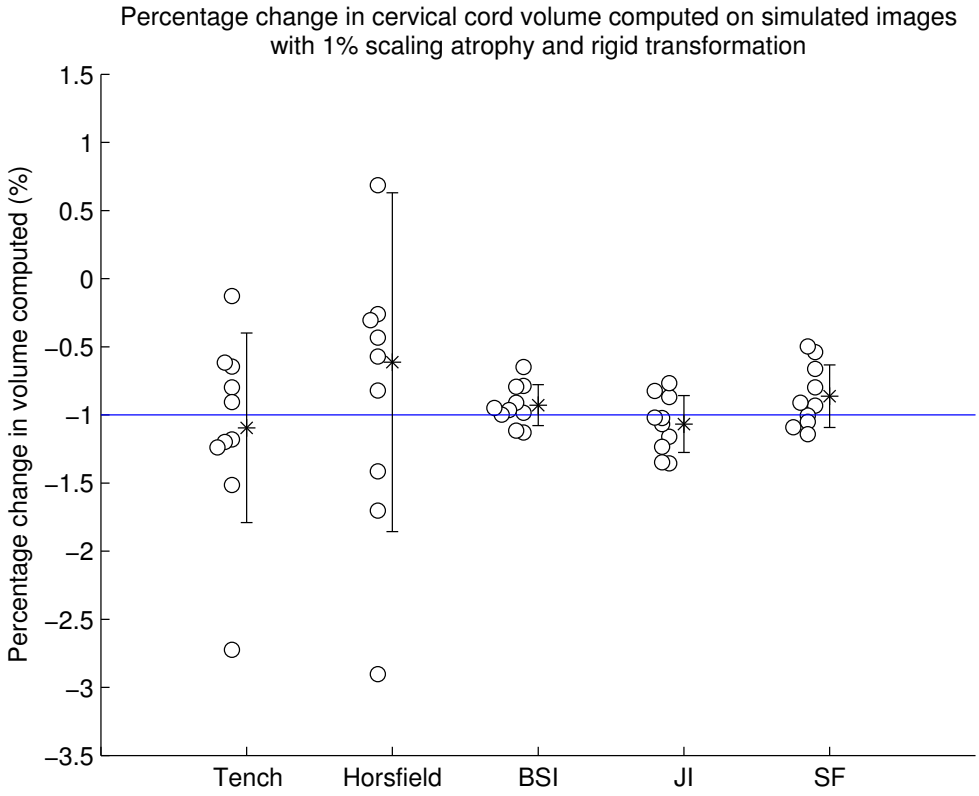


Figure 3.6: Percentage change rates computed by the three registration-based methods (BSI, JI and SF) along with the two segmentation-based methods (Tench and Horsfield) on simulated scan pairs with rigid transformation and 1% scaling change. Each circle represents a single subject in the simulated scan pairs, and the stars and error bars represent the means and standard deviations of the measurements. The blue line represents the ground truth change rate between the two scans in the simulated scan pairs. The results obtained by BSI and JI have smaller mean errors and smaller variance than the two segmentation-based methods. Although the mean of the Tench results is closer to the ground truth value -1% compared with the mean of the SF results, the standard deviation of SF results is much smaller. Registration-based methods are able to significantly improve the measurement variance on this simulation scan pairs.

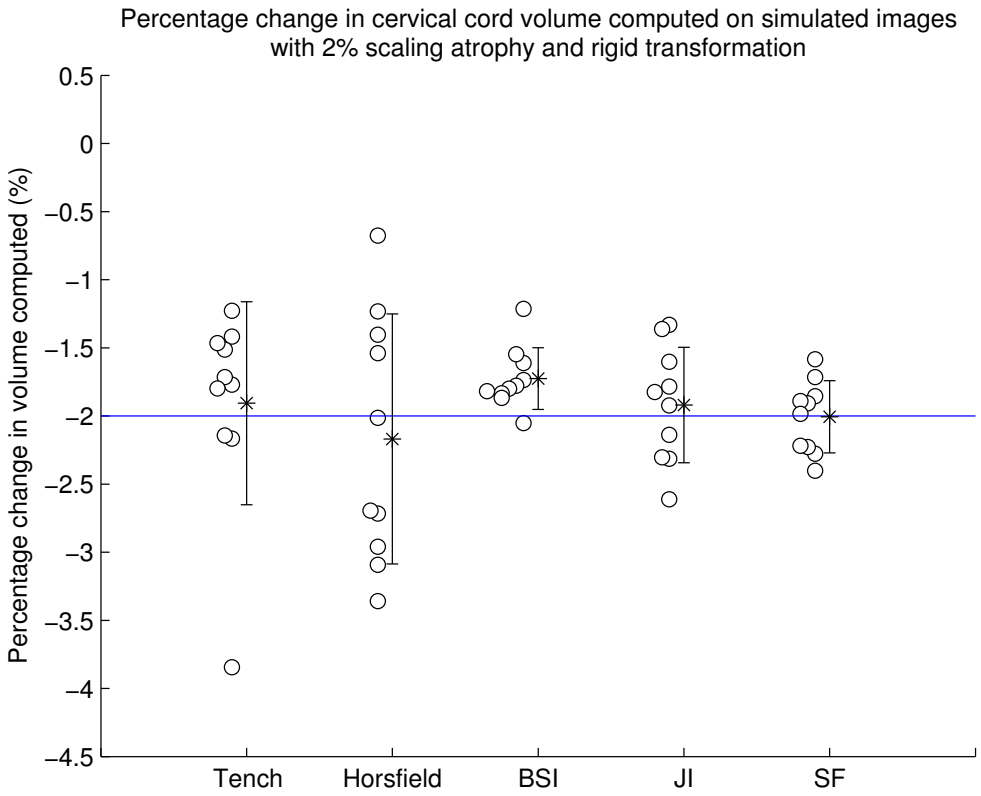


Figure 3.7: Percentage change rates computed by the three registration-based methods (BSI, JI and SF) along with the two segmentation-based methods (Tench and Horsfield) on simulated scan pairs with rigid transformation and 2% scaling change. Each circle represents a single subject in the simulated scan pairs, and the stars and error bars represent the means and standard deviations of the measurements. The blue line represents the ground truth change rate between the two scans in the simulated scan pairs. The JI and SF results are better than the results of the two segmentation-based methods with smaller mean error and smaller variance. Although the mean error of BSI results is larger than the two segmentation-based methods, the standard deviation of the BSI result is much smaller. Registration-based methods improve the measurement variance on this simulation scan pairs.

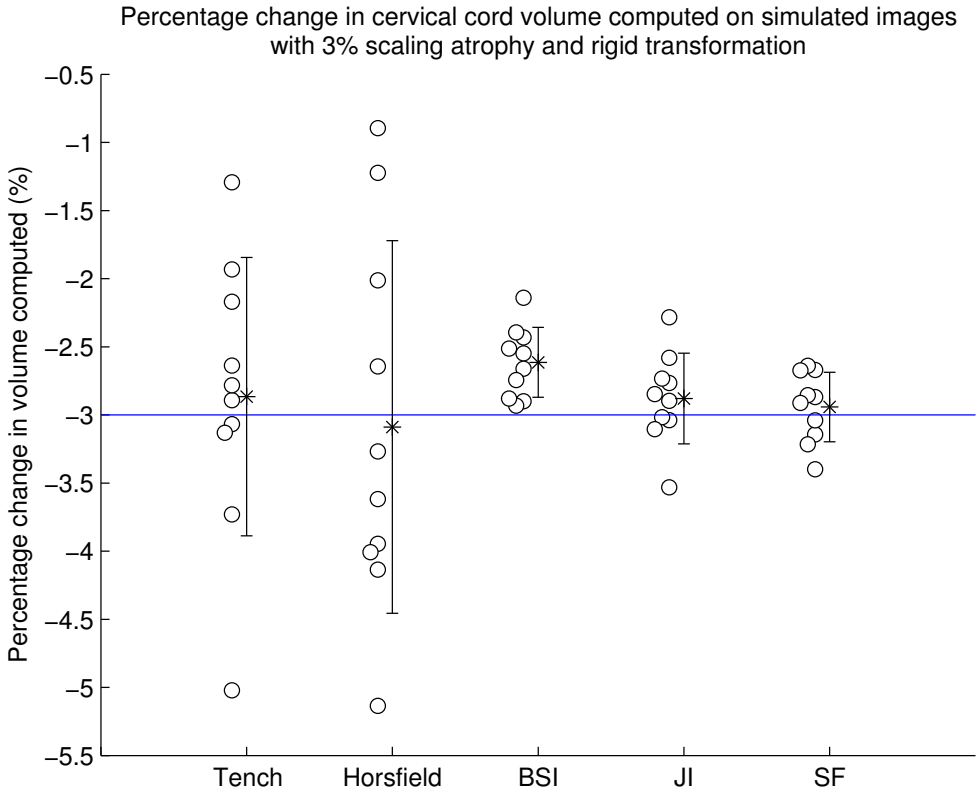


Figure 3.8: Percentage change rates computed by the three registration-based methods (BSI, JI and SF) along with the two segmentation-based methods (Tench and Horsfield) on simulated scan pairs with rigid transformation and 3% scaling change. Each circle represents a single subject in the simulated scan pairs, and the stars and error bars represent the means and standard deviations of the measurements. The blue line represents the ground truth change rate between the two scans in the simulated scan pairs. The measurement variance using BSI, JI and SF is significantly improved with much smaller SDs compared to the measurements using Tench and Horsfield methods.

Table 3.2: Mean and SD of the errors which are the differences between the percentage change rates computed by BSI, JI and SF on scaled scan pairs and their respective ground truth change rates.

	Mean (SD) of the errors (%)		
Methods	BSI	JI	SF
-3%	-0.13 (0.37)	0.01 (0.17)	0.01 (0.02)
-2%	-0.08 (0.26)	0.01 (0.06)	0.01 (0.05)
-1%	-0.07 (0.13)	-0.01 (0.07)	0.05 (0.05)

Table 3.3: Mean (SD) of the errors, which are the differences of the computed change rates to the ground truth values, on the simulation data set with rigid transformation and simulated atrophy. The measurement SDs using the three registration-based methods are smaller than the measurement SDs using the two segmentation-based methods.

	Mean (SD) of the errors (%)				
	Tench	Horsfield	BSI	JI	SF
-3%	0.13 (1.02)	-0.09 (1.37)	0.38 (0.25)	0.12 (0.33)	0.06 (0.25)
-2%	0.09 (0.75)	-0.17 (0.92)	0.27 (0.22)	0.08 (0.42)	-0.01 (0.26)
-1%	-0.09 (0.69)	-0.39 (1.24)	0.07 (0.14)	-0.06 (0.21)	0.13 (0.23)
0%	0.10 (1.02)	-0.33 (0.96)	-0.02 (0.07)	-0.02 (0.36)	0.02 (0.12)

The means and standard deviations of the errors of the results obtained by BSI, JI and SF as well as Tench and Horsfield methods on the four sets of simulated scan pairs are provided in Table 3.3 . The registration-based methods improved the measurement variance, achieving much smaller standard deviations than the two segmentation-based methods on this simulation data set. Their results support our hypothesis that the registration-based atrophy measurement methods are robust to the variability in cord segmentation and can eventually improve the precision of cord atrophy measurement by directly assessing the change.

3.2.3 Hydration Data Set

Table 3.4: Means and SDs of the percentage change rates computed by BSI, JI and SF along with the results of Tench and Horsfield methods on the scan-rescan pairs (labeled 1 in the table), dehydration scan pairs (labeled 2 in the table) and rehydration scan pairs (labeled 3 in the table). The three registration-based methods have comparable reproducibility on the scan-rescan scan pairs, compared to the results of the two segmentation-based methods. However, they did not detect any significant decrease in cord volume on the dehydration scan pairs which was detected by the two segmentation-based methods.

	Mean (SD) of the change rate(%)				
	Tench	Horsfield	BSI	JI	SF
1	-0.22 (0.79)	-0.40 (1.26)	-0.24 (0.77)	-0.26 (0.61)	-0.09 (0.73)
2	-0.65 (0.78)*	-0.65 (1.04)†	0.09 (0.68)	-0.34 (0.97)	-0.34 (0.71)
3	0.12 (1.32)	0.06 (1.13)	0.05 (0.70)	0.00 (0.73)	-0.01 (1.18)

*$p \leq 0.05$, † $p \leq 0.1$

The means and standard deviations of the percentage changes computed by each technique on the three sets of scan pairs in the hydration data set are presented in Table 3.4 .

On the scan-rescan scan pairs, the coefficients of variance (SD divided by the mean) of the three registration-based methods were comparable to that of two segmentation-based methods, indicating no obvious improvement by the registration-based methods in scan-rescan reproducibility.

On the dehydration scan pairs, there was no significant change in cord volume detected by BSI, JI and SF. Registration-based methods are not sensitive enough to detect the dehydration effect.

On the rehydration scan pairs, results of BSI, JI and SF showed no change from zero, which is similar to the results of Tench and Horsfield methods.

The results of each technique on the three sets of scan pairs in the hydration data set are shown in Figures 3.9, 3.10 and 3.11.

3.2.4 MS Patient Data Set

The mean (SD) percentage change rates quantified using each technique on the scan pairs with a two-year interval in the MS patient data set are presented in Table 3.5 and Figure 3.12.

Table 3.5: Mean and standard deviation (SD) of the percentage change rates in cord volume computed by BSI, JI and SF along with the results of Tench and Horsfield methods on the MS patient data set. All these methods detected significant cord atrophy on the scan pairs with a two-year interval. However, the magnitudes of cord atrophy detected by the registration-based methods were smaller compared with the cord atrophy detected by the two segmentation-based methods.

Mean (SD)(%)				
Tench	Horsfield	BSI	JI	SF
-2.53 (3.81)**	-2.13 (3.59)*	-0.74 (1.68)*	-1.00 (1.78)*	-1.45 (2.63)*

**p≤0.01, *p≤0.05

These methods all detected significant cord atrophy over two years with the percentage change rates computed by the three registration-based methods and two segmentation-based methods all significantly below zero ($p \leq 0.05$) in the one-tailed Wilcoxon rank test. The registration-based methods improved the measurement variance by obtaining smaller standard deviations. However, they do not seem to be as sensitive, reporting smaller magnitudes of atrophy compared with the atrophy detected by the segmentation-based methods. The cord atrophy previously reported in the literature is around -1.6% in SPMS patients per year and the magnitude of cord atrophy computed by the two segmentation-based methods over two years agrees reasonably well to the reported figure. In terms of correlation, the measurements using SF agreed well with the measurements using Tench method. The measurements using SF significantly correlated with the measurements using Tench (Spearman's correlation coefficient $r = 0.63$, $p = 0.01$).

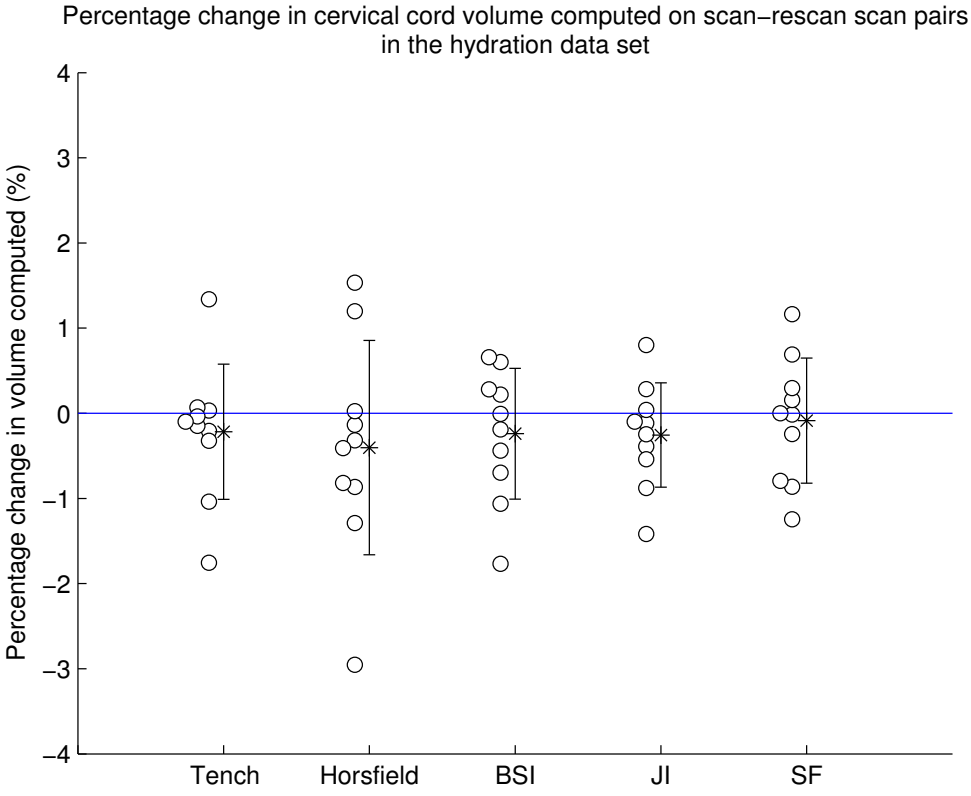


Figure 3.9: The percentage change in cord volume from baseline to the rescan time point computed by the three registration-based methods (BSI, JI and SF) and the two segmentation-based methods (Tench and Horsfield). Each circle represents a single subject in the hydration data set, and the stars and error bars represent the means and standard deviations. The scan-rescan variation computed by the three registration-based methods was comparable to that computed by Tench method with similar means and SDs.

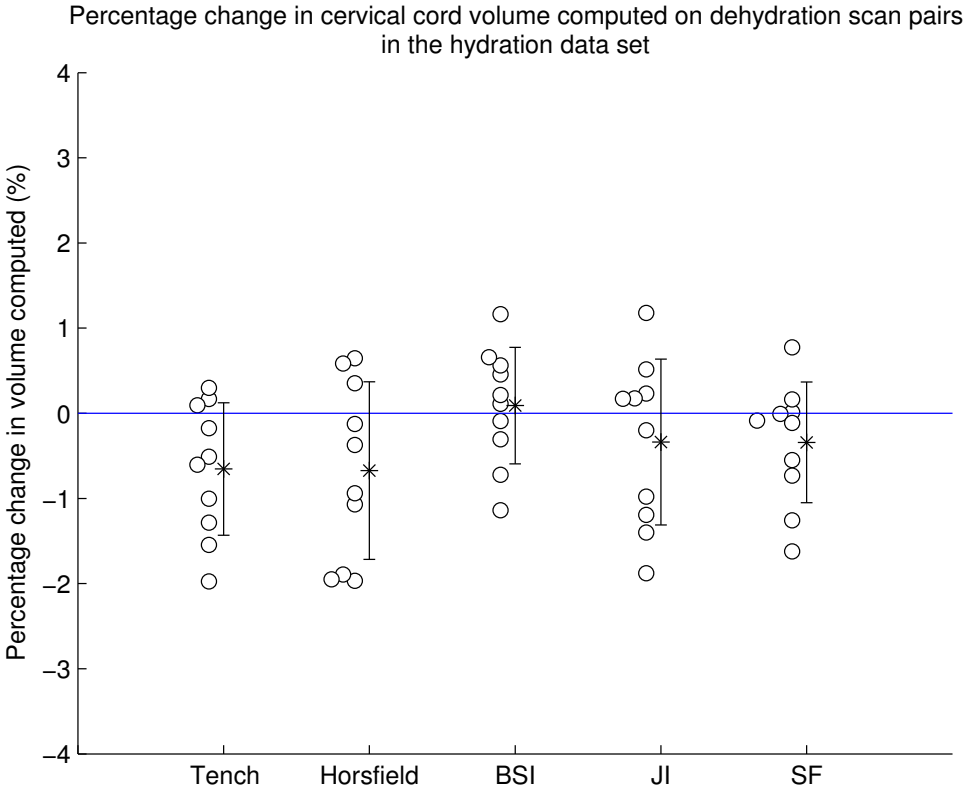


Figure 3.10: The percentage change in cord volume from baseline to the dehydration time point computed by the three registration-based methods (BSI, JI and SF) and the two segmentation-based methods (Tench and Horsfield). Each circle represents a single subject in the hydration data set, and the stars and error bars represent the means and standard deviations. The results of registration-based methods did not demonstrate any statistically significant decrease in cord volume following dehydration.

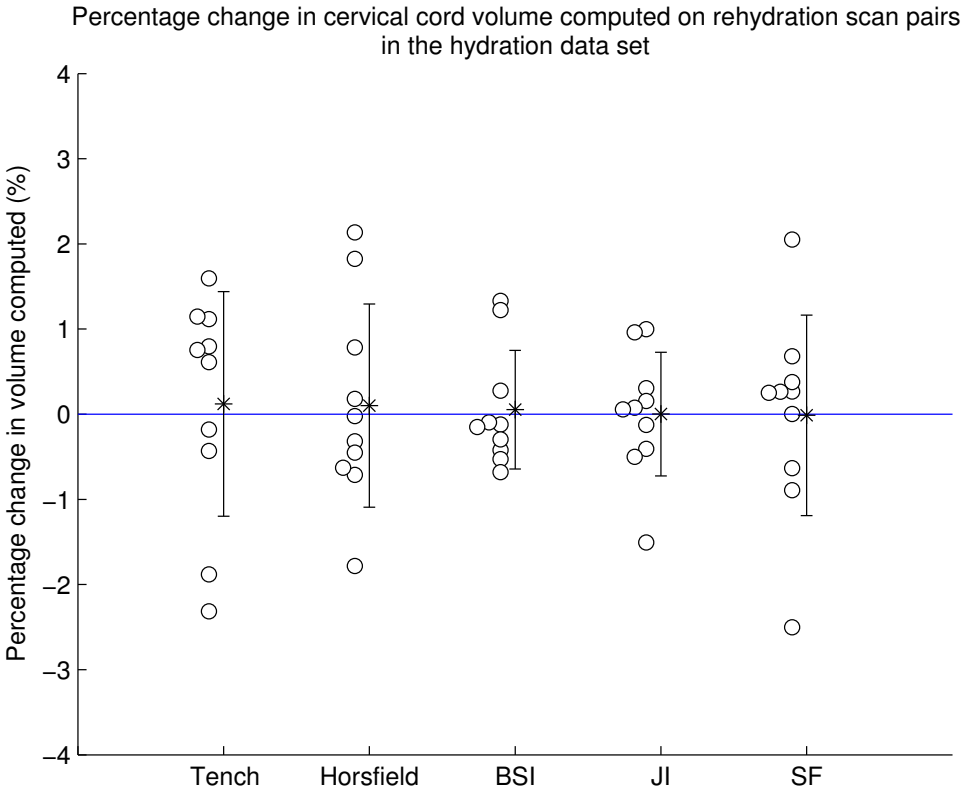


Figure 3.11: The percentage change in cord volume from baseline to the rehydration time point computed by the three registration-based methods (BSI, JI and SF) and the two segmentation-based methods (Tench and Horsfield). Each circle represents a single subject in the hydration data set, and the stars and error bars represent the means and standard deviations. No significant change in cord volume was detected by the registration-based methods from the baseline scan to the rehydrated scan, which is similar to the results of the two segmentation-based methods.

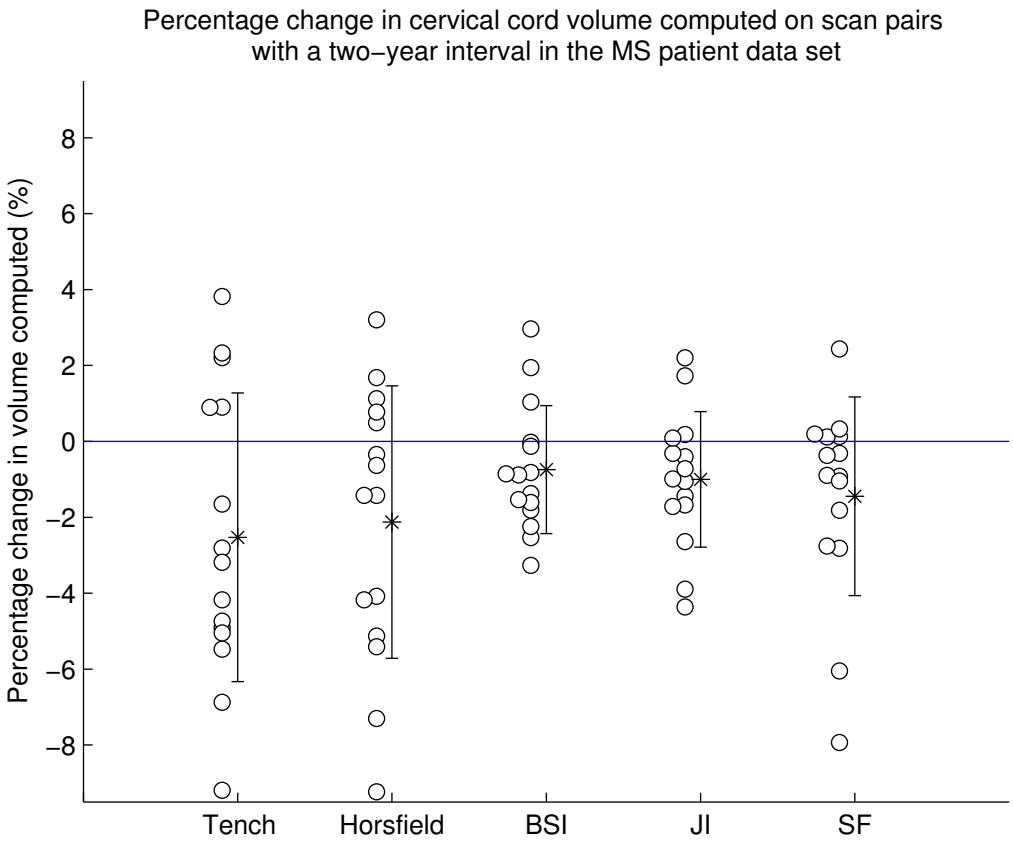


Figure 3.12: The percentage change in cord volume on the scan pairs with two years interval computed by the three registration-based methods (BSI, JI and SF) and the two segmentation-based methods (Tench and Horsfield). Each circle represents a single subject in the MS patient data set, and the stars and error bars represent the means and standard deviations. The means of the change rates computed by all the methods are below zero. The measurement variance is slightly improved in the results of registration-based methods. However, the magnitudes of atrophy rates computed by the registration-based methods are smaller than the magnitudes of atrophy rates computed by the segmentation-based methods.

3.3 Discussion

In this study, we implemented three registration-based atrophy measurement techniques to measure longitudinal spinal cord atrophy from locally registered serial MR images. We evaluated these three registration-based methods on the following test data sets, 1) two sets of scan pairs with simulated change to quantify the measurement precision, 2) the scan-rescan scan pairs to quantify the measurement reproducibility, 3) the dehydration scan pairs with demonstrated dehydration effect and SPMS patient scan pairs with reported cord atrophy to quantify the measurement sensitivity. We compared the results obtained by the registration-based methods on these test data sets to the results obtained by two segmentation-based methods that are currently utilized as standard approaches in spinal cord atrophy studies in MS.

Our experiments showed that the registration-based methods reduced measurement variance (overall smaller standard deviations) compared with the segmentation-based methods on all the test data sets and improved the measurement precision, achieving less errors on the two simulated data sets. However, the registration-based methods were not as sensitive as the segmentation-based methods, detecting no dehydration effect on the dehydration scan pairs and reporting reduced magnitude cord atrophy on the MS patient data set. We argue in the following sections that the limitation in sensitivity of registration-based methods on hydration data set and MS patient data set is probably due to the limited spatial resolution of the MR scans in our experiments and the inherently small size of the cord.

3.3.1 Boundary Shift Integral

The technique BSI directly estimates the cord volume change between two registered images by calculating the intensity changes within specified intensity window at the cord–CSF boundary. The accuracy of BSI method depends on the chosen intensity window. Any intensity transitions that lie outside the window do not contribute to the BSI, resulting in an underestimation. Brain BSI was reported to underestimate around 0.71% of simulated

scaling atrophy in Boyes et al.'s study [18], and Camara et al. showed that brain BSI consistently underestimated atrophy by around 18% especially at higher level of atrophy on a cohort of realistic simulated images with known amounts of atrophy [20]. The caudate BSI also had a tendency to under-report change relative to the manual measures [49]. In our experiments, BSI underestimated the simulated scaling atrophy on the simulation data set and also underestimated the real cord atrophy on the MS patient data set, which can be probably explained by the limitation of the intensity window in BSI method. While this underestimate is significant, it is understood to be linear and does not affect the sensitivity of BSI results to differentiate AD subjects and healthy controls [18]. As a comparison group of control subjects with serial MR scans was not available in the MS patient data set, we do not know whether the underestimate of cord BSI would affect its sensitivity to differentiate the control and MS patient group.

The intensity mean and SD estimated on the limited samples of the cord voxels and CSF voxels (around 85 cord voxels and 35 CSF voxels on the axial slice of MR images with a spatial resolution of 1 mm), which are used to decide the intensity window, are easily affected by the noise or artifacts, resulting in miscalculation of cord BSI. While the effect of noise and artifacts also exists in brain BSI, it is likely to be cancelled out and not significant, because the whole brain region is bigger containing more voxels than the cord and the means and SDs estimated on a larger sample are more reliable. Furthermore, as the cord region we are examining is so small, any miscalculation in cord BSI will disproportionally affect the final result. Despite the fact that BSI has been successfully used to measure atrophy of small structures inside the brain like caudate nucleus and hippocampus, the mean (SD) of annual atrophy rate computed by BSI is 4.63(2.78)% for hippocampus volume of 147 patients with AD in Leung et al.'s study [63] and 2.90(1.60)% for caudate nucleus volume of 16 patients with Huntington's disease in Hobbs et al.'s study [49], which are both larger than the magnitude of dehydration effect (-0.65%) and two-year cord atrophy of SPMS patients (-2.13% to -2.53%) in our experiments. Moreover, the change quantified by BSI brought by common image non-idealities like image noise and contrast difference would

possibly exceed the disease effect of MS. Preboske et al. [85] pointed that the magnitudes of the error using BSI in measures of longitudinal brain atrophy that can result from commonly encountered image non-idealities can significantly exceed the disease effect which range from 1% to 2.78% per year for brain atrophy in AD.

To summarize, the precision and sensitivity of cord BSI on MR scans with a spatial resolution of 1 mm are limited due to these conditions. MR images with a spatial resolution less than 1 mm can probably overcome these limitations because there would be more voxels over the cord region providing more samples to estimate the mean and SD of the cord and CSF intensities. The effect of image noise and other image non-idealities on the cord BSI result would be lessened, which would improve the precision of cord BSI. MR scans with higher resolution are suggested for further studies using BSI in longitudinal cord atrophy measurement.

3.3.2 Jacobian Integration

The technique JI uses the Jacobian determinant values of deformation field obtained from deformable registration over the cord region to give an average estimate of cord change rate over time. The accuracy of JI depends on the deformable registration algorithm applied.

The deformation field for the small cord region obtained from the non-linear registration algorithm should be both plausible and smooth for the Jacobian determinants to yield meaningful results. On one hand, we are seeking to obtain a plausible registration to optimally align the images by maximizing the similarity measure. On the other hand, we need to regularize the registration results to provide robust and meaningful measures of anatomical changes, because the statistical power of studies using JI to quantify anatomical changes largely depends on the smoothness of the Jacobian determinant maps associated with the deformation [40, 68].

We chose the symmetric log-domain diffeomorphic demons algorithm as implemented in ITK for our cord deformable registration application because of its demonstrated theoretical advantages [104] and practical efficiency. The

symmetric log-domain diffeomorphic algorithm optimizes the displacement field using an efficient second-order minimization framework and provides diffeomorphic transformations that are smooth in terms of Jacobian determinants. Thanks to the open-source implementation of this algorithm [30], we can practically modify and debug its source code. In addition, the relative ease for us to find the optimal parameters, compared to another symmetric deformable registration algorithm SyN [8], is another practical advantage. Choosing a right combination of parameters for deformable registration is always an application-dependent problem. In the case of symmetric log-domain diffeomorphic demons, the most important parameter to be tuned is the sigma σ of the Gaussian kernel for regularization. We evaluated the JI results on the simulation data set with scaling atrophy over a range of σ values (between no regularization at all and $\sigma = 2.0$ with an interval of 0.2) and chose the one ($\sigma = 1.0$) which gave the lowest errors to be the parameter used in our experiments. Meanwhile, we used the change in mean square error merit values suggested by Peroni et al. [84] instead of the predefined iteration numbers in the original implementation to be the stopping criterion and the cord deformable registration in our experiments actually converged after no more than 10 iterations in most cases. On the contrary, there are more parameters to be tuned using SyN beside the smoothing sigma for the Gaussian kernel, like the weights for the mean square error and cross-correlation similarity metrics, the gradient step size, the time step for integration. It is hard to find the optimal parameters using SyN for our cord application and different parameter settings ended up with very different JI results.

The atrophy rates provided by JI were more accurate and less variable than those from the two segmentation-based methods on both simulation data sets with scaling. However, on the hydration data set and MS patient data set, JI results did not demonstrate to be superior in terms of sensitivity and precision, which is probably due to the inherent limitation of spine MRI images used in our experiments. The spinal cord is composed of white matter and grey matter. On MR images used in our experiments, which were T1-weighted images with a spatial resolution of 1 mm, these two internal structures are not differentiable, making it impossible for the deformable

registration algorithm to find structural meaningful correspondence inside the cord boundary. JI performed well on the simulation data sets because there are good intensity correspondences between the baseline and repeat images which were created from the same control scan. The reduced intensity correspondence over the cord region on the MR images in the hydration data set and MS patient data set resulted in the limited sensitivity to detect real cord atrophy using JI. Another reason that may underlie the underestimation of the cord atrophy on the MS patient data set could be the inclusion of the partial volume voxels in the region of interest to be integrated.

The fact that JI results critically rely on deformable registration emphasizes that we can not directly compare atrophy rates computed using different deformable registration methods. As shown in Camara et al.'s paper [20], JI using two different deformable registration algorithms (free-form deformable registration vs. fluid registration) to quantify realistic simulated atrophy in multiple regions of brain yielded significantly different results. In our attempt using SyN [8] as the deformation registration algorithm in our experiments, JI results detected significantly larger magnitude cord atrophy on the MS patient data set with mean (SD) of atrophy rate to be -3.03(3.72)% (one-tailed Wilcoxon test $p = 0.006$), and the measurements significantly correlated with the measurements using Tench method (Spearman's correlation coefficient $r = 0.61$, $p = 0.02$). However, the JI results using SyN did not show any improvement in measurement variance and also they did not show any improvement in measurement accuracy on the simulation data sets.

To summarize, the standard resolution of the MR images used in our experiments accounts for the limited sensitivity of the JI technique to quantify cord atrophy. Future cord atrophy studies using JI are suggested to be performed on MR images with higher resolution that are able to differentiate the GM and WM inside the cord. Additional validation of the non-linear registration algorithm should be performed, and parameters for the chosen deformable registration algorithm should be carefully selected to obtain meaningful JI results. Moreover, although we used a 0.5 probability threshold on the resampled baseline cord segmentation to create the binary

baseline cord mask as the region of interest within which Jacobian determinants are integrated, other probability thresholds should be investigated, as misclassified partial volume voxels will increase measurement errors.

3.3.3 Scale Factor from 3-DoF Registration

The technique SF uses the scale factor obtained from the 3-DoF constrained registration (2 translation tx and ty, and 1 scaling factor $scalexy$), where we constrain uniform scale factor $scalexy$ on x and y axes, to measure the global change in cord size. The accuracy of the cord atrophy measurement using SF depends on the scale factor obtained from optimization of the 3-DoF constrained registration.

In our experiments, SF achieved very accurate results on the simulation data set with scaling atrophy because the simulated atrophy generated was global scaling change and 3-DoF constrained registration measures the global change in cord size. On the simulation data set with rigid transformation and scaling atrophy, SF achieved more accurate results than the two segmentation-based methods, with smaller mean errors and smaller variance. On the dehydration scan pairs, SF was able to detect a mean decrease of 0.34% in cord volume with one-tailed Wilcoxon rank test p value of 0.12. On the MS patient data set, SF detected significant cord atrophy over two years with a mean atrophy rate of 1.45%, and the measurements of SF were significantly correlated with the measurements of Tench method with a systematic underestimation.

The underestimation in SF results can be explained by the uniform scale factor $scalexy$ on the x and y axes in the 3-DoF transformation. SF assumes that the cord changes uniformly on the x and y axes, which might not be true in real scenarios. Imposing a constraint of an equal scale change in the x and y dimensions regularized the optimization process, thereby reducing measurement variance; however, it also reduced sensitivity of this method, thus resulting in the underestimation. In addition, the interpolation done to resample the input images at each iteration also introduced artifacts in the merit function either by generating many local optima or by shifting

the global optimum [3]. These artifacts consequently increase the chance of registration converging to a wrong minimum, thereby producing inaccurate registration results.

The limited resolution of MR images used in our experiments also caused the limited sensitivity of SF. On T1 weighted MR images with a spatial resolution of 1 mm, the intensities inside the cord boundary are fuzzy because of the non-differentiable intensities of grey matter and white matter. The mean square error merit function which is computed over the small cord region (less than 100 voxels on each slice) is easily to be affected by image noise and artifacts, and interpolation artifacts. Thus, the optimization of the 3-DoF constrained registration is likely to converge to an undesirable minimum, degrading the sensitivity. However, on MR images with higher spatial resolution, there are more voxels over the partial volume region explicitly capturing the tissue loss and also the cord region inside the cord boundary would have good soft tissue contrast of the grey matter and white matter, providing better intensity correspondence for the registration. The merit function computed from a larger number of voxels would be robust to image noise and artifacts, and the negative effect of interpolation artifacts would be lessened. The optimization of the 3-DoF constrained registration would yield more accuracy estimation of scale change.

In summary, SF generated small measurement variance in longitudinal cord atrophy measurement but at the cost of lower sensitivity. Although SF tends to underestimate atrophy, we do not know whether this underestimation will affect its sensitivity in separating MS patient group and control group, since a comparison control group is not available in the MS patient data set. Experiments on MS patients with a comparison control group and larger sample sizes are suggested for future study to investigate longitudinal cord atrophy measurement using SF. Future studies using SF are recommended to be performed on high resolution MR images with resolution less than 1 mm in order to improve the sensitivity.

Chapter 4

Conclusion

4.1 Summary

We investigated the spinal cord atrophy measurement on MRI from two aspects. Firstly, change in water content due to hydration status affects the cord CSA measurement on MRI, whose signals are mainly derived from water. We conducted the experiment to assess the dehydration effect on cord CSA measurement. We designed the dehydration and rehydration protocol and collected the MR scans from ten healthy subjects at four time points. Two established cord CSA measurement methods were employed to measure the cord CSA on all the MR scans. Statistical analyses of the percentage change in CSA from baseline time point to each subsequent time point were performed to assess the significance of these changes to determine the dehydration effect. Results from the two methods agree well and we have observed a decrease in the cervical cord CSA by 0.65% after solid and liquid fasting for an overnight period that would not be considered unusual in routine research or clinical settings involving MRI. Our findings lend evidence that change in water content due to hydration status affects the spinal cord CSA measurement and should be considered a source of variability in clinical studies of spinal cord atrophy.

Secondly, registration-based methods were adapted for longitudinal cord atrophy measurement for the first time. Three registration-based methods (boundary shift integral, Jacobian integration and scale factor obtained from 3-DoF constrained registration), were implemented to measure the change in spinal cord volume on rigid registered serial MR images. These three methods were evaluated on two data sets with simulated atrophy, on the hydration data set with dehydration effect, and on an MS patient data set

with cord atrophy over a two-year interval. Their results were compared to the results obtained by two segmentation-based methods, which are currently utilized as standard approaches in spinal cord atrophy studies in MS.

Our experiments showed that registration-based methods reduce the measurement variance with smaller standard deviations of their measurements on all the data sets over the two segmentation-based methods. However, they were not sensitive enough to detect the dehydration effect on the dehydrated scan pairs and also detected a reduced magnitude of cord atrophy on the MS patient data set. We argue that their results with limited sensitivity are possibly due to the limited spatial resolution of 1 mm of MR scans in our experiment and the inherently small size of the cord. Registration-based methods estimate the change in cord size by assessing the intensity differences between corresponding voxels using registration with different levels of regularization. In all three methods examined, regardless of whether the registration is rigid registration (used in BSI), deformable registration (used in JI) and 3-DoF constrained registration (used in SF), the intensity differences of corresponding voxels over the cord region between the baseline and follow-up images were the fundamental information used in all three methods explored. However, on MR images used in our experiments, which were T1-weighted images with a spatial resolution of 1 mm, the cord is bounded to a region size of less than 100 voxels on each axial slice and lacks internal contrast between grey matter and white matter. The intensity window in BSI, which is determined by the intensity mean and standard deviation of the cord and surrounding CSF, is easily affected by image noise and artifacts, resulting in underestimation of cord BSI. The mean square error merit function in the deformable registration (used in JI) and constrained registration (used in SF) evaluated over the cord region with poor intensity correspondence and a limited number of voxels, is very sensitive to artifacts introduced by image noise and interpolation done to resample the input images at each iteration. These artifacts consequently generate many local optima or change the global optimum, increasing the chance of registration converging to an undesirable minimum and thereby resulting in the limited sensitivity.

Registration-based methods reduced the measurement variance by im-

posing different levels of regularization in registration but at the cost of less sensitivity. As the cord atrophy in MS is a slow process with an annual atrophy of around 1.6% in SPMS patients reported in previous literature, the registration-based methods applied on MR images with a spatial resolution of 1 mm are not able to meet the precision and sensitivity requirements for longitudinal spinal cord atrophy measurement. High resolution MR scans with a spatial resolution less than 1 mm are required to conduct further studies on spinal cord atrophy measurement using the three registration-based methods proposed.

4.2 Future work

The results presented in the dehydration study provide support for our hypothesis that change in water content does have a significant effect on the CSA measurement on MRI. However, there are limitations in our experiment and further investigation will improve our understanding of the effect of water content to the cord measurement on MRI. In future work, brain scans can be collected along with the spine scans to verify whether similar magnitude of change can be expected in both structures after dehydration and provide further information on the relationship between the dehydration effect on the MRI measurements of brain volume and cord volume. Our current studies use conventional T1 weighted MR images with a resolution of 1 mm. MR scans acquired using phase-sensitive inversion recovery (PSIR) imaging can provide good grey matter and white matter contrast over the cord region and have been used to investigate the spinal cord grey matter atrophy in MS [83, 95]. It would be interesting to investigate the dehydration effect on the volume of spinal cord white matter and spinal cord grey matter using PSIR scans.

In future studies investigating registration-based methods for longitudinal cord atrophy measurement, MR images with a spatial resolution below 1 mm, and which are able to differentiate the white matter and grey matter over the cord region, such as PSIR scans [54], are suggested to be collected and used to evaluate the precision and sensitivity of registration-based meth-

ods. A larger sample size of MS patients and a comparison control group are recommended to examine the sensitivity of the registration-based methods in separating the patient and healthy groups. In addition, automatic and accurate segmentation of the cord should be a direction to explore in the future. Deformable segmentation algorithms based on cord atlas can be a possible way to provide an accurate and automatic cord segmentation for cord atrophy analysis [27].

Bibliography

[1] `https://beta.mssociety.ca/about-ms`, assessed on Dec 22, 2014.

[2] `https://beta.mssociety.ca/about-ms/treatments`, assessed on Dec 22, 2014.

[3] I Aganj, BTT Yeo, MR Sabuncu, and B Fischl. On removing interpolation and resampling artifacts in rigid image registration. *Image Processing, IEEE Transactions on*, 22(2):816–827, 2013.

[4] F Agosta, M Absinta, MP Sormani, A Ghezzi, A Bertolotto, E Montanari, G Comi, and M Filippi. In vivo assessment of cervical cord damage in MS patients: a longitudinal diffusion tensor MRI study. *Brain*, 130(8):2211–2219, 2007.

[5] F Agosta, E Pagani, D Caputo, and M Filippi. Associations between cervical cord gray matter damage and disability in patients with multiple sclerosis. *Archives of Neurology*, 64(9):1302–1305, 2007.

[6] VM Anderson, K Fernando, GR Davies, W Rashid, C Frost, NC Fox, and DH Miller. Cerebral atrophy measurement in clinically isolated syndromes and relapsing remitting multiple sclerosis: a comparison of registration-based methods. *Journal of Neuroimaging*, 17(1):61–68, 2007.

[7] V Arsigny, O Commowick, X Pennec, and N Ayache. A log-euclidean framework for statistics on diffeomorphisms. *Medical Image Computing and Computer-Assisted Intervention*, 4190:924–931, 2006.

[8] BB Avants, CL Epstein, M Grossman, and JC Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008.

[9] F Barkhof, PA Calabresi, DH Miller, and SC Reingold. Imaging outcomes for neuroprotection and repair in multiple sclerosis trials. *Nature Reviews Neurology*, 5(5):256–266, 2009.

[10] J Barnes, RG Boyes, EB Lewis, JM Schott, C Frost, RI Scahill, and NC Fox. Automatic calculation of hippocampal atrophy rates using a hippocampal template and the boundary shift integral. *Neurobiology of Aging*, 28(11):1657–1663, 2007.

[11] J Barnes, J Foster, RG Boyes, T Pepple, EK Moore, JM Schott, C Frost, RI Scahill, and NC Fox. A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. *NeuroImage*, 40(4):1655–1671, 2008.

[12] J Barnes, RI Scahill, RG Boyes, C Frost, EB Lewis, CL Rossor, MN Rossor, and NC Fox. Differentiating AD from aging using semi-automated measurement of hippocampal atrophy rates. *NeuroImage*, 23(2):574–581, 2004.

[13] O Benga and VJ Huber. Brain water channel proteins in health and disease. *Molecular Aspects of Medicine*, 33(5):562–578, 2012.

[14] V Biberacher, CC Boucard, P Schmidt, C Engl, D Buck, A Berthele, MM Hoshi, C Zimmer, B Hemmer, and M Mühlau. Atrophy and structural variability of the upper cervical cord in early multiple sclerosis. *Multiple Sclerosis Journal*, 2014. In press.

[15] M Bieniek, DR Altmann, GR Davies, GT Ingle, W Rashid, J Sastre-Garriga, AJ Thompson, and DH Miller. Cord atrophy separates early primary progressive and relapsing remitting multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 77(9):1036–1039, 2006.

[16] M Bossa, M Hernandez, and S Olmos. Contributions to 3D diffeomorphic atlas estimation: Application to brain images. *Medical Image Computing and Computer-Assisted Intervention*, 4791:667–674, 2007.

[17] R Boyes, J Schott, C Frost, and N Fox. Increasing accuracy of atrophy measures from serial MR scans using parameter analysis of the boundary shift integral. *Medical Image Computing and Computer-Assisted Intervention*, 3217:1036–1037, 2004.

[18] RG Boyes, D Rueckert, P Aljabar, J Whitwell, JM Schott, DLG Hill, and NC Fox. Cerebral atrophy measurements using jacobian integration: comparison with the boundary shift integral. *NeuroImage*, 32(1):159–169, 2006.

[19] PA Brex, SM Leary, JI O'Riordan, KA Miszkiel, GT Plant, AJ Thompson, and DH Miller. Measurement of spinal cord area in clinically isolated syndromes suggestive of multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 70(4):544–547, 2001.

[20] O Camara, JA Schnabel, GR Ridgway, WR Crum, A Douiri, RI Scahill, DLG Hill, and NC Fox. Accuracy assessment of global and local atrophy measurement techniques with realistic simulated longitudinal alzheimer's disease images. *NeuroImage*, 42(2):696–709, 2008.

[21] J Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 8(6):679–698, 1986.

[22] M Chen, A Carass, J Oh, G Nair, DL Pham, DS Reich, and JL Prince. Automatic magnetic resonance spinal cord segmentation with topology constraints for variable fields of view. *NeuroImage*, 83:1051–1062, 2013.

[23] MK Chung, KJ Worsley, T Paus, C Cherif, DL Collins, JN Giedd, JL Rapoport, and AC Evans. A unified statistical approach to deformation-based morphometry. *NeuroImage*, 14(3):595–606, 2001.

[24] O Ciccarelli, DR Altmann, MA McLean, CA Wheeler-Kingshott, K Wimpey, DH Miller, and AJ Thompson. Spinal cord repair in MS does mitochondrial metabolism play a role. *Neurology*, 74(9):721–727, 2010.

[25] AB Cohen, M Neema, A Arora, E Dell'Oglio, RHB Benedict, S Tauhid, D Goldberg-Zimring, C Chavarro-Nieto, A Ceccarelli, JP Klein, et al. The relationships among MRI-defined spinal cord involvement, brain involvement, and disability in multiple sclerosis. *Journal of Neuroimaging*, 22(2):122–128, 2012.

[26] O Coulon, SJ Hickman, GJ Parker, GJ Barker, DH Miller, and SR Arridge. Quantification of spinal cord atrophy from magnetic resonance images via a B-spline active surface model. *Magnetic Resonance in Medicine*, 47(6):1176–1185, 2002.

[27] B De Leener, S Kadoury, and J Cohen-Adad. Robust, accurate and fast automatic segmentation of the spinal cord. *NeuroImage*, 98:528–536, 2014.

[28] GC DeLuca, GC Ebers, and MM Esiri. Axonal loss in multiple sclerosis: a pathological survey of the corticospinal and sensory tracts. *Brain*, 127(5):1009–1018, 2004.

[29] JM Dickson, HM Weavers, N Mitchell, EM Winter, ID Wilkinson, EJR Van Beek, JM Wild, and PD Griffiths. The effects of dehydration on brain volume-preliminary results. *International Journal of Sports Medicine*, 26(6):481–485, 2004.

[30] F Dru, T Vercauteren, et al. An ITK implementation of the symmetric log-domain diffeomorphic demons algorithm. *The Insight Journal*, 2009.

[31] T Duning, S Kloska, O Steinsträter, H Kugel, W Heindel, and S Knecht. Dehydration confounds the assessment of brain atrophy. *Neurology*, 64(3):548–550, 2005.

[32] F Durand-Dubief, B Belaroussi, JP Armspach, M Dufour, S Roggerone, S Vukusic, S Hannoun, D Sappey-Marinier, C Confavreux, and F Cotton. Reliability of longitudinal brain volume loss measurements between 2 sites in patients with multiple sclerosis: comparison of 7 quantification techniques. *American Journal of Neuroradiology*, 33(10):1918–1924, 2012.

[33] N Evangelou, GC DeLuca, T Owens, and MM Esiri. Pathological study of spinal cord atrophy in multiple sclerosis suggests limited role of local lesions. *Brain*, 128(1):29–34, 2005.

[34] M Filippi and F Agosta. Imaging biomarkers in multiple sclerosis. *Journal of Magnetic Resonance Imaging*, 31(4):770–788, 2010.

[35] M Filippi, A Campi, B Colombo, C Pereira, V Martinelli, C Baratti, and G Comi. A spinal cord MRI study of benign and secondary progressive multiple sclerosis. *Journal of Neurology*, 243(7):502–505, 1996.

[36] M Filippi, MA Rocca, F Barkhof, W Brück, JT Chen, G Comi, G DeLuca, N De Stefano, BJ Erickson, N Evangelou, et al. Association between pathological and MRI findings in multiple sclerosis. *The Lancet Neurology*, 11(4):349–360, 2012.

[37] NC Fox, RS Black, S Gilman, MN Rossor, SG Griffith, L Jenkins, M Koller, et al. Effects of a β immunization on MRI measures of cerebral volume in Alzheimer disease. *Neurology*, 64(9):1563–1572, 2005.

[38] NC Fox and PA Freeborough. Brain atrophy progression measured from registered serial MRI: validation and application to Alzheimer's disease. *Journal of Magnetic Resonance Imaging*, 7(6):1069–1075, 1997.

[39] NC Fox, R Jenkins, SM Leary, VL Stevenson, NA Losseff, WR Crum, RJ Harvey, MN Rossor, DH Miller, and AJ Thompson. Progressive cerebral atrophy in MS: a serial study using registered volumetric MRI. *Neurology*, 54(4):807–812, 2000.

[40] NC Fox, GR Ridgway, and JM Schott. Algorithms, atrophy and Alzheimer's disease: cautionary tales for clinical trials. *NeuroImage*, 57(1):15–18, 2011.

[41] PA Freeborough and NC Fox. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *Medical Imaging, IEEE Transactions on*, 16(5):623–629, 1997.

[42] PA Freeborough and NC Fox. Modeling brain deformations in Alzheimer disease by fluid registration of serial 3D MR images. *Journal of Computer Assisted Tomography*, 22(5):838–843, 1998.

[43] J Furby, T Hayton, D Altmann, R Brenner, J Chataway, KJ Smith, DH Miller, and R Kapoor. A longitudinal study of MRI-detected atrophy in secondary progressive multiple sclerosis. *Journal of Neurology*, 257(9):1508–1516, 2010.

[44] J Furby, T Hayton, V Anderson, D Altmann, R Brenner, J Chataway, RAC Hughes, KJ Smith, DH Miller, and R Kapoor. Magnetic resonance imaging measures of brain and spinal cord atrophy correlate with clinical impairment in secondary progressive multiple sclerosis. *Multiple Sclerosis*, 14(8):1068–1075, 2008.

[45] RH Hashemi, WG Bradley, and CJ Lisanti. Basic principles of MRI. In *MRI: The Basics*, chapter 2, page 16. Lippincott Williams & Wilkins, 2010.

[46] T Hayton, J Furby, KJ Smith, DR Altmann, R Brenner, J Chataway, K Hunter, DJ Tozer, DH Miller, and R Kapoor. Clinical and imaging correlates of the multiple sclerosis impact scale in secondary progressive multiple sclerosis. *Journal of Neurology*, 259(2):237–245, 2012.

[47] BC Healy, A Arora, DL Hayden, A Ceccarelli, SS Tauhid, M Neema, and R Bakshi. Approaches to normalization of spinal cord volume: application to multiple sclerosis. *Journal of Neuroimaging*, 22(3):e12–e19, 2012.

[48] SJ Hickman, O Coulon, GJM Parker, GJ Barker, VL Stevenson, DT Chard, SR Arridge, AJ Thompson, and DH Miller. Application of a B-spline active surface technique to the measurement of cervical cord volume in multiple sclerosis from three-dimensional MR images. *Journal of Magnetic Resonance Imaging*, 18(3):368–371, 2003.

[49] NZ Hobbs, S Henley, EJ Wild, KK Leung, C Frost, RA Barker, RI Scahill, J Barnes, SJ Tabrizi, and NC Fox. Automated quantification of caudate atrophy by local registration of serial MRI: evaluation and application in huntington's disease. *NeuroImage*, 47(4):1659–1665, 2009.

[50] MA Horsfield, S Sala, M Neema, M Absinta, A Bakshi, MP Sormani, MA Rocca, R Bakshi, and M Filippi. Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: application in multiple sclerosis. *NeuroImage*, 50(2):446–455, 2010.

[51] X Hua, B Gutman, CP Boyle, P Rajagopalan, AD Leow, I Yanovsky, AR Kumar, AW Toga, CR Jack, N Schuff, et al. Accurate measurement of brain changes in longitudinal MRI scans using tensor-based morphometry. *NeuroImage*, 57(1):5–14, 2011.

[52] GT Ingle, VL Stevenson, DH Miller, and AJ Thompson. Primary progressive multiple sclerosis: A 5-year clinical and MR study. *Brain*, 126(11):2528–2536, 2003.

[53] NF Kalkers, F Barkhof, E Bergers, R Van Schijndel, and CH Polman. The effect of the neuroprotective agent riluzole on MRI parameters in primary progressive multiple sclerosis: a pilot study. *Multiple Sclerosis*, 8(6):532–533, 2002.

[54] H Kearney, MC Yiannakas, K Abdel-Aziz, CAM Wheeler-Kingshott, DR Altmann, O Ciccarelli, and DH Miller. Improved MRI quantification of spinal cord atrophy in multiple sclerosis. *Journal of Magnetic Resonance Imaging*, 39(3):617–623, 2014.

[55] MJ Kempton, U Ettinger, R Foster, SCR Williams, GA Calvert, A Hampshire, FO Zelaya, RL O'Gorman, T McMorris, AM Owen,

et al. Dehydration affects brain structure and function in healthy adolescents. *Human Brain Mapping*, 32(1):71–79, 2011.

[56] MJ Kempton, U Ettinger, A Schmechtig, EM Winter, L Smith, T McMorris, ID Wilkinson, SCR Williams, and MS Smith. Effects of acute dehydration on brain morphology in healthy humans. *Human Brain Mapping*, 30(1):291–298, 2009.

[57] D Kidd, JW Thorpe, AJ Thompson, BE Kendall, IF Moseley, DG MacManus, WI McDonald, and DH Miller. Spinal cord MRI using multiarray coils and fast spin echo II: Findings in multiple sclerosis. *Neurology*, 43(12):2632–2632, 1993.

[58] A Klein, J Andersson, BA Ardekani, J Ashburner, B Avants, MC Chiang, GE Christensen, DL Collins, J Gee, P Hellier, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786–802, 2009.

[59] JP Klein, A Arora, M Neema, BC Healy, S Tauhid, D Goldberg-Zimring, C Chavarro-Nieto, JM Stankiewicz, AB Cohen, GJ Buckle, et al. A 3T MR imaging investigation of the topography of whole spinal cord atrophy in multiple sclerosis. *American Journal of Neuroradiology*, 32(6):1138–1142, 2011.

[60] JF Kurtzke. Rating neurologic impairment in multiple sclerosis an expanded disability status scale (EDSS). *Neurology*, 33(11):1444–1452, 1983.

[61] C Laule, IM Vavasour, Y Zhao, AL Traboulsee, J Oger, JD Vavasour, AL Mackay, and DKB Li. Two-year study of cervical cord volume and myelin water in primary progressive multiple sclerosis. *Multiple Sclerosis*, 16(6):670–677, 2010.

[62] AD Leow, AD Klunder, CR Jack Jr, AW Toga, AM Dale, MA Bernstein, PJ Britson, JL Gunter, CP Ward, JL Whitwell, et al. Longitudinal stability of MRI for mapping brain change using tensor-based morphometry. *NeuroImage*, 31(2):627–640, 2006.

[63] KK Leung, J Barnes, GR Ridgway, JW Bartlett, MJ Clarkson, K Macdonald, N Schuff, NC Fox, and S Ourselin. Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. *NeuroImage*, 51(4):1345–1359, 2010.

[64] KK Leung, MJ Clarkson, JW Bartlett, S Clegg, CR Jack, MW Weiner, NC Fox, and S Ourselin. Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection. *NeuroImage*, 50(2):516–523, 2010.

[65] X Lin, LD Blumhardt, and CS Constantinescu. The relationship of brain and cervical cord volume to disability in clinical subtypes of multiple sclerosis: a three-dimensional MRI study. *Acta Neurologica Scandinavica*, 108(6):401–406, 2003.

[66] X Lin, CR Tench, N Evangelou, T Jaspan, and CS Constantinescu. Measurement of spinal cord atrophy in multiple sclerosis. *Journal of Neuroimaging*, 14(S3):20–26, 2004.

[67] X Lin, CR Tench, B Turner, LD Blumhardt, and CS Constantinescu. Spinal cord atrophy and disability in multiple sclerosis over four years: application of a reproducible automated technique in monitoring disease progression in a cohort of the interferon β-1a (rebif) treatment trial. *Journal of Neurology, Neurosurgery & Psychiatry*, 74(8):1090–1094, 2003.

[68] M Lorenzi, N Ayache, GB Frisoni, and X Pennec. LCC-Demons: a robust and accurate diffeomorphic registration algorithm. *NeuroImage*, 81:470–483, 2013.

[69] NA Losseff and DH Miller. Measures of brain and spinal cord atrophy in multiple sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 64(S1):102–105, 1998.

[70] NA Losseff, SL Webb, JI O'riordan, R Page, L Wang, GJ Barker, PS Tofts, WI McDonald, DH Miller, and AJ Thompson. Spinal cord atrophy and disability in multiple sclerosis: A new reproducible and sensitive MRI method with potential to monitor disease progression. *Brain*, 119(3):701–708, 1996.

[71] C Lukas, B Bellenberg, HK Hahn, J Rexilius, R Drescher, K Hellwig, O Köster, and S Schimrigk. Benefit of repetitive intrathecal triamcinolone acetonide therapy in predominantly spinal multiple sclerosis: prediction by upper spinal cord atrophy. *Therapeutic Advances in Neurological Disorders*, 2(6):349–355, 2009.

[72] C Lukas, DL Knol, MH Sombekke, B Bellenberg, HK Hahn, V Popescu, K Weier, EW Radue, A Gass, L Kappos, et al. Cervical spinal cord volume loss is related to clinical disability progression in multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 2014. In press.

[73] C Lukas, MH Sombekke, B Bellenberg, HK Hahn, V Popescu, K Bendfeldt, EW Radue, A Gass, SJ Borgwardt, L Kappos, et al. Relevance of spinal cord abnormalities to clinical disability in multiple sclerosis: MR imaging findings in a large cohort of patients. *Radiology*, 269(2):542–552, 2013.

[74] KE Macdonald, KK Leung, JW Bartlett, M Blair, IB Malone, J Barnes, S Ourselin, and NC Fox. Automated template-based hippocampal segmentations from MRI: the effects of 1.5T or 3T field strength on accuracy. *Neuroinformatics*, 12(3):405–412, 2014.

[75] RS Mann, CS Constantinescu, and CR Tench. Upper cervical spinal cord cross-sectional area in relapsing remitting multiple sclerosis: Application of a new technique for measuring cross-sectional area on magnetic resonance images. *Journal of Magnetic Resonance Imaging*, 26(1):61–65, 2007.

[76] DH Miller. Biomarkers and surrogate outcomes in neurodegenerative disease: lessons from multiple sclerosis. *NeuroRx*, 1(2):284–294, 2004.

[77] DH Miller, F Barkhof, JA Frank, GJM Parker, and AJ Thompson. Measurement of atrophy in multiple sclerosis: pathological basis, methodological aspects and clinical relevance. *Brain*, 125(8):1676–1695, 2002.

[78] DH Miller, D Soon, KT Fernando, DG MacManus, GJ Barker, TA Yousry, E Fisher, PW O'Connor, JT Phillips, CH Polman, et al. MRI outcomes in a placebo-controlled trial of natalizumab in relapsing MS. *Neurology*, 68(17):1390–1401, 2007.

[79] A Minagar, EG Toledo, and JS Alexander. Central nervous system atrophy in multiple sclerosis: Mechanisms. In *Brian and Spinal Cord Atrophy in Multiple Sclerosis*, chapter II, pages 5–13. Nova Science Publishers, Inc., 2004.

[80] PA Narayana, TJ Doyle, D Lai, and JS Wolinsky. Serial proton magnetic resonance spectroscopic imaging, contrast-enhanced magnetic

resonance imaging, and quantitative lesion volumetry in multiple sclerosis. *Annals of Neurology*, 43(1):56–71, 1998.

[81] J Oh, M Seigo, S Saidha, E Sotirchos, K Zackowski, M Chen, J Prince, M Diener-West, PA Calabresi, and DS Reich. Spinal cord normalization in multiple sclerosis. *Journal of Neuroimaging*, 24(6):577–584, 2014.

[82] J Oh, K Zackowski, M Chen, S Newsome, S Saidha, SA Smith, M Diener-West, J Prince, CK Jones, PCM Van Zijl, et al. Multiparametric MRI correlates of sensorimotor function in the spinal cord in multiple sclerosis. *Multiple Sclerosis Journal*, 19(4):427–435, 2013.

[83] N Papinutto, R Schlaeger, V Panara, E Caverzasi, S Ahn, KJ Johnson, AH Zhu, WA Stern, G Laub, SL Hauser, et al. 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*, 2014. In press.

[84] M Peroni, P Golland, GC Sharp, and G Baroni. Stopping criteria for log-domain diffeomorphic demons registration: an experimental survey for radiotherapy application. *Technology in Cancer Research & Treatment*, 2013. In press.

[85] GM Preboske, JL Gunter, CP Ward, and Clifford R Jack Jr. Common MRI acquisition non-idealities significantly impact the output of the boundary shift integral method of measuring brain atrophy on serial MRI. *NeuroImage*, 30(4):1196–1202, 2006.

[86] W Rashid, GR Davies, DT Chard, CM Griffin, DR Altmann, R Gordon, R Kapoor, AJ Thompson, and DH Miller. Upper cervical cord area in early relapsing-remitting multiple sclerosis: Cross-sectional study of factors influencing cord size. *Journal of Magnetic Resonance Imaging*, 23(4):473–476, 2006.

[87] W Rashid, GR Davies, DT Chard, CM Griffin, DR Altmann, R Gordon, AJ Thompson, and DH Miller. Increasing cord atrophy in early relapsing-remitting multiple sclerosis: a 3 year study. *Journal of Neurology, Neurosurgery & Psychiatry*, 77(1):51–55, 2006.

[88] MA Rocca, M Cercignani, G Iannucci, G Comi, and M Filippi. Weekly diffusion-weighted imaging of normal-appearing white matter in MS. *Neurology*, 55(6):882–884, 2000.

[89] MA Rocca, P Valsasina, D Damjanovic, MA Horsfield, S Mesaros, T Stosic-Opincal, J Drulovic, and M Filippi. Voxel-wise mapping of cervical cord damage in multiple sclerosis patients with different clinical phenotypes. *Journal of Neurology, Neurosurgery & Psychiatry*, 84(1):35–41, 2013.

[90] S Ropele and F Fazekas. Magnetization transfer MR imaging in multiple sclerosis. *Neuroimaging Clinics of North America*, 19(1):27–36, 2009.

[91] M Rovaris, E Judica, J Sastre-Garriga, A Rovira, MP Sormani, B Benedetti, T Korteweg, N De Stefano, Z Khaleeli, X Montalban, et al. Large-scale, multicentre, quantitative MRI study of brain and cord damage in primary progressive multiple sclerosis. *Multiple Sclerosis*, 14(4):455–464, 2008.

[92] RA Rudick. Impact of disease-modifying therapies on brain and spinal cord atrophy in multiple sclerosis. *Journal of Neuroimaging*, 14(s3):54S–64S, 2004.

[93] S Saadoun and MC Papadopoulos. Aquaporin-4 in brain and spinal cord oedema. *Neuroscience*, 168(4):1036–1046, 2010.

[94] MP Sampat, BC Healy, DS Meier, E Dell'Oglio, M Liguori, CR Guttmann, et al. Disease modeling in multiple sclerosis: assessment and quantification of sources of variability in brain parenchymal fraction measurements. *NeuroImage*, 52(4):1367, 2010.

[95] R Schlaeger, N Papinutto, V Panara, C Bevan, IV Lobach, M Bucci, E Caverzasi, JM Gelfand, AJ Green, KM Jordan, et al. Spinal cord gray matter atrophy correlates with multiple sclerosis disability. *Annals of Neurology*, 76(4):568–580, 2014.

[96] SM Smith, Y Zhang, M Jenkinson, J Chen, PM Matthews, A Federico, N De Stefano, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage*, 17(1):479–489, 2002.

[97] VL Stevenson, SM Leary, NA Losseff, GJM Parker, GJ Barker, Y Husmani, DH Miller, and AJ Thompson. Spinal cord atrophy and disability in MS A longitudinal study. *Neurology*, 51(1):234–238, 1998.

[98] DP Streitbürger, HE Möller, M Tittgemeyer, M Hund-Georgiadis, ML Schroeter, and K Mueller. Investigating structural brain changes of dehydration using voxel-based morphometry. *PloS One*, 7(8):e44195, 2012.

[99] CR Tench and CS Constantinescu. Technical aspects of spinal cord atrophy measurement. In *Brian and Spinal Cord Atrophy in Multiple Sclerosis*, chapter IV, pages 63–71. Nova Science Publishers, Inc., 2004.

[100] CR Tench, PS Morgan, and CS Constantinescu. Measurement of cervical spinal cord cross-sectional area by MRI using edge detection and partial volume correction. *Journal of Magnetic Resonance Imaging*, 21(3):197–203, 2005.

[101] PM Thompson, KM Hayashi, G de Zubicaray, AL Janke, SE Rose, J Semple, D Herman, MS Hong, SS Dittmer, David M Doddrell, et al. Dynamics of gray matter loss in Alzheimer's disease. *The Journal of Neuroscience*, 23(3):994–1005, 2003.

[102] L Vaithianathar, CR Tench, PS Morgan, and CS Constantinescu. Magnetic resonance imaging of the cervical spinal cord in multiple sclerosis. *Journal of neurology*, 250(3):307–315, 2003.

[103] P Valsasina, MA Rocca, MA Horsfield, M Absinta, R Messina, D Caputo, G Comi, and M Filippi. Regional cervical cord atrophy and disability in multiple sclerosis: A voxel-based analysis. *Radiology*, 266(3):853–861, 2013.

[104] T Vercauteren, X Pennec, A Perchant, N Ayache, et al. Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage*, 45(1):S61–72, 2008.

[105] R Zivadinov and R Bakshi. Central nervous system atrophy and clinical status in multiple sclerosis. *Journal of Neuroimaging*, 14(s3):27S–35S, 2004.

[106] R Zivadinov, AT Reder, M Filippi, A Minagar, O Stüve, H Lassmann, MK Racke, MG Dwyer, EM Frohman, and O Khan. Mechanisms of action of disease-modifying agents and brain volume changes in multiple sclerosis. *Neurology*, 71(2):136–144, 2008.

[107] R Zivadinov, M Zorzon, and R Bakshi. Central nervous system atrophy in multiple sclerosis: Introduction. In *Brian and Spinal Cord Atrophy*

in Multiple Sclerosis, chapter I, pages 1–4. Nova Science Publishers, Inc., 2004.