



Published in final edited form as:

J Neuroimaging. 2012 July ; 22(3): e12–e19. doi:10.1111/j.1552-6569.2011.00629.x.

Approaches to normalization of spinal cord volume: application to multiple sclerosis

Brian C. Healy^{1,3,*,+}, Ashish Arora^{1,2,+}, Douglas L. Hayden³, Antonia Ceccarelli^{1,2}, Shahamat S. Tauhid^{1,2}, Mohit Neema^{1,2}, and Rohit Bakshi^{1,2,*}

¹Department of Neurology, Partners Multiple Sclerosis Center, Brigham and Women's Hospital, Harvard Medical School

²Laboratory of Neuroimaging Research, Partners Multiple Sclerosis Center, Brigham and Women's Hospital, Harvard Medical School

³Biostatistics Center, Massachusetts General Hospital

Abstract

Background and purpose—To determine the proper method for normalization of spinal cord volume.

Materials and Methods—A group of 34 multiple sclerosis (MS) patients (28 relapsing and 6 progressive) and 15 healthy controls had whole spinal cord 3 mm thick T2-weighted axial fast spin-echo MRI images obtained at 3T. For each participant, four volumes were measured (C2-3 volume, cervical cord volume, thoracic cord volume and whole cord volume). The volumes were normalized by the number of slices and three potential measures of body size (intracranial volume (ICV), body mass index, and body surface area) using the proportional method.

Results—All raw volumes and volumes normalized by number of slices or ICV were significantly lower in progressive MS patients compared to relapsing MS patients/healthy controls ($p < 0.05$). In addition, C2-3 volume and cervical cord volume were significantly correlated with EDSS score ($p < 0.05$). All regional volumes showed high intercorrelation, and normalization by the number of slices significantly increased some correlations. Regarding reliability, whole cord volume regardless of normalization technique had lower coefficient of variation than C2-3 volume.

Conclusions—Since normalization factor had limited impact on reliability and the ability to detect differences, normalization by the number of slices is recommended.

Keywords

3 Tesla imaging; multiple sclerosis; normalization; spinal cord atrophy

Introduction

Spinal cord atrophy is recognized as a common and clinically relevant aspect of the disease process in patients with multiple sclerosis (MS).[1] Significant differences between MS patients and healthy controls have been observed in many MRI studies.[2–5] In addition, patients with progressive forms of the disease have more spinal cord atrophy than patients

*CORRESPONDENCE: Brian Healy PhD and Rohit Bakshi, M.D., Laboratory for Neuroimaging Research, Brigham and Women's Hospital, One Brookline Place, Suite 602, Brookline, MA 02445; bchealy@partners.org and rbakshi@bwh.harvard.edu; phone: 617-732-6568; fax: 617-732-6540.

+ Authors contributed equally to this paper

with relapsing-remitting (RR) MS.[6] The majority of studies have focused on the upper cervical cord due to technical as well as scientific reasons. From a technical perspective, both image acquisition and segmentation of the upper cervical region are more efficient than other parts of the cord or the whole cord. From a scientific perspective, the upper cervical cord is a logical surrogate of overall cord degeneration because lesions are more common in the cervical than thoracic cord [7] and destructive processes in the thoracic cord should be reflected by Wallerian degeneration in the cervical cord. However, the correlations between cervical cord atrophy and physical disability have been only mild to moderate in MS patients, suggesting the need for more reliable surrogates of the underlying destructive processes.

An additional challenge in the measurement of spinal cord atrophy is the uncertainty and lack of formal guidelines regarding the benefit and proper method of normalization of spinal cord volumes obtained from MRI scans. One goal of normalization is to improve power for group comparisons by removing differences between patients that are unrelated to the actual group effect. A common example in neuroimaging is normalizing the brain volume by a measure of the intracranial volume (ICV) because it is assumed that a patient with a larger ICV than another patient also has a larger brain volume, but this increase is unrelated to the disease. This normalization is either done via regression adjustment or a proportional method.[8] In one previous methodologic investigation of normalization for the spinal cord, three measures of the cervical cord volume were compared in terms of reliability and the ability to discriminate patients with MS and healthy controls: 1) non-normalized (raw), 2) normalized to the volume of the thecal sac, and 3) normalized to the intracranial volume.[5] In all cases, normalization was accomplished by dividing the cervical cord volume by the normalizing factor (i.e. proportional method). The authors advocated using raw volume because this approach had the best reliability and showed the most sensitivity. In another study, three other approaches were tested for normalization of the cervical cord: 1) the lumbar enlargement cord area, 2) the maximum intracranial cross-sectional area, and 3) the mid-sagittal intracranial area.[9] The authors advocated normalization by the lumbar enlargement cord area, a measure having the advantage of not requiring an accompanying brain scan for normalization. However, despite these two studies, there remain several unmet needs we sought to address in the present study.

The goals of the present study were to extend the previous investigations of spinal cord volume normalization in several ways. First, we extended normalization studies to the full spinal cord including characterization of the cervical, thoracic and whole spinal cord volume. Second, we acquired MRI data on a 3T platform, whereas previous studies were at lower field strengths. Third, we investigated normalization to body size measures not previously utilized as normalization factors for spinal cord volumes. To accomplish these goals, we compared normalization approaches in terms of their reliability, the ability to discriminate between MS patients and controls, correlation with clinical outcomes, and correlation between MRI measures.

Materials and methods

Subjects

Our study consisted of 34 MS patients and 15 healthy controls. Patient and control demographic/clinical characteristics are shown in Table 1. MS patients were prospectively recruited from the Partners MS Center clinic population. All MS patients had no relapses or corticosteroid use within the four weeks prior to study entry, and had not initiated any disease modifying therapy within the six months prior to study entry, to avoid any confounding “pseudoatrophy” effects on CNS volume.[10, 11] Patients were enrolled consecutively. Of the 34 MS patients, the majority were RR patients (n=26), but some were

classified as a clinically isolated syndrome with lesions on brain MRI scans that were suggestive of MS (CIS, n=2), primary progressive (PP, n=2), or secondary progressive (SP, n=4) MS, based on established guidelines.[12] Healthy controls were recruited using an IRB-approved advertisement as previously described.[13] This Health Insurance Portability and Accountability Act (HIPAA)-compliant study included approval by an institutional review board (IRB) and informed consent.

MRI acquisition

All subjects underwent whole spinal cord imaging on the same 3T MRI scanner (General Electric Whole Body Scanner, Milwaukee, WI) using the same scanning protocol throughout the study as previously described.[7] Spinal MRI was obtained with a spinal phased array coil using 8 channels at 20 mT/m maximal gradient strength. Axial T2 fast spin-echo images had the following parameters: field of view 24×19 cm, matrix size 256×256 , slice thickness 3 mm with no gap, repetition time (TR) 6116.66 ms, echo time (TE) 110.24 ms, echo train length 12, number of signal averages 2, flip angle 90 degrees, pixel size $0.937 \text{ mm} \times 0.937 \text{ mm}$, and total acquisition time was 22.5–32.0 minutes. One hundred fifty to 200 axial slices were acquired on each subject to cover the whole spinal cord. The whole spinal cord was defined from the foramen magnum through the inferior extent of the conus medullaris level. In a previous optimization experiment, T1- and T2 weighted spin-echo images showed similar reliability and validity for assessing cord atrophy in MS [14]; thus for efficiency's sake, we chose T2-weighted images going forward based on the ability to measure lesions and atrophy from the same sequence. Example scans are shown in Figure 1.

For the purposes of determining ICV (the size of the intracranial cavity), coronal 3D Modified Driven Equilibrium Fourier Transform (MDEFT) whole brain images were acquired at 3T with the following protocol: TR = 7.9 ms, TE = 3.14 ms, flip angle = 15° , voxel size = $0.938 \times 0.938 \times 1.5$ mm, 124 slices (no gap). The total acquisition time for this scan was ~ 7 minutes.

MRI analysis

MRI scans were analyzed using the software package Jim (version 3.0, Xinapse Systems, Northants, UK; <http://www.xinapse.com>). Scans were anonymized and randomized prior to analysis. An edge-finding tool based on local thresholding was applied to each axial slice to identify the spinal cord contour. Manual adjustments were applied where necessary. The volume of each region (cervical, thoracic, and whole) was calculated by measuring the area of each slice and combining across all slices. To calculate the C2-3 volume, we localized the cord at the C2-3 vertebral level. The region of interest on the cord was assessed on three slices at the C2-3 level. The volume of the C2-3 region of the spinal cord was calculated by taking an average of the volume of the three slices based on the contoured ROI. Spinal cord lesion data were not felt to be relevant to the present objectives and thus will be reported in separate studies.

ICV was determined from the MDEFT images by manual skull stripping of each slice using the Jim software.

Clinical outcomes

Within three months of MRI, all patients had a complete clinical exam with an experienced neurologist at our MS Center. During the clinical exam, the neurologist classified the patient's disease category (relapsing-remitting, secondary progressive, primary progressive or clinically isolated syndrome suggestive of MS) and rated the patient's disability using the Expanded Disability Status Scale (EDSS) [15]. In addition, each patient completed the timed

25 foot walk (T25FW) [16]; patients who could not complete the walk were given a time of 99 seconds.

Factors used for normalization

The volumes in each of the four regions, C2-3, cervical, thoracic and whole cord, were analyzed using the following eight strategies of normalization: (1) no normalization, (2) number of slices, (3) ICV, (4) number of slices and ICV, (5) body mass index (BMI), (6) number of slices and BMI, (7) body surface area (BSA), [17] and (8) number of slices and BSA. The first four strategies will be considered the main strategies of interest, but the remaining four are described as well. Normalization by the ICV was considered due to the moderate correlation between the ICV and each of the spinal cord volumes ($r > 0.5$ for Pearson correlation between ICV and all volumes other than C2-3). The formula for BMI is

$BMI = \frac{weight_{kg}}{height_m^2}$, and the formula for BSA is $BSA = \left(\frac{weight_{kg} * height_{cm}}{3600} \right)^{0.5}$. BSA showed moderate correlations with the spinal cord volumes ($0.2 < r < 0.35$ for all comparisons), but BMI showed limited correlation to any spinal cord volumes ($r < 0.1$ for all comparison). Since each patient had three slices in the C2-3 region, the results are presented for the average volume across the slices only (i.e. normalization by the number of slices)..

Reliability

A subset of participants (two patients and three controls) contributed to our investigation of the reliability of normalization factors. These patients had either multiple analyses of the same scans from a single reader or multiple analyses from different readers to determine the intra-rater and inter-rater reliability. The intra-rater and inter-rater reliability of each normalization method was calculated using the average within patient coefficient of variation ($COV = SD/mean * 100\%$) and the intraclass correlation coefficient (ICC) based on a single reader repeating the spinal volume measures (A.A) and based on two observers repeating the spinal volume measures (A.A and M.N.). The ICV was recalculated for the calculation of the intra-rater and inter-rater reliability of measures normalized by this factor. For the interrater ICC, the ICC for consistency was reported. [18] Given the BMI and BSA did not change between the two readings, the effect of these normalizing factors on reliability is not reported.

Statistical analysis

To investigate the reliability and precision of the measures using different normalization techniques, the COV and ICC were calculated for the C2-3 and whole cord volume normalized by the ICV and number of slices. For the purposes of group comparisons, due to the small sample size, CIS/RR patients and SP/PP patients were combined. Although ideally these individual phenotypes represent separate and distinct groups, we wanted to maximize the statistical power to explore group effects by this pooling of data. Given the approximate normality of the data, a linear regression model adjusting for age was used to compare the normalized volumes in the three groups: healthy controls, relapsing MS patients (CIS/RR) and progressive MS patients (SP/PP). If significant differences between the groups were observed, pairwise comparisons between the groups were also completed with Bonferroni correction to adjust for multiple comparisons. In addition, the correlation between the normalized volumes and clinical outcomes (EDSS, T25FW and disease duration) were calculated using a Spearman's correlation coefficient; for the T25FW, patients who could not walk were given the highest rank. The correlation between spinal cord volume measures was also calculated using Pearson's correlation coefficient, and the change in correlation coefficients before and after normalization was investigated using the approach of Meng. [19]

Results

Reliability

The intra-rater and inter-rater reliability of each of the main normalization approaches for the whole spinal cord and C2-3 volume is presented in Table 2. The reliability of all measures of the whole cord was excellent, and the normalized measures had similar reliability compared to the raw volumes in part due the high reliability in the ICV and slice number measures in our sample. The C2-3 had worse reliability, especially in the inter-rater comparison. In particular, the mean COV for the inter-rater comparison of the C2-3 increased by more than 4–5 fold compared to the whole spinal cord volume measures. Even in the intra-rater comparison, the mean COV for the C2-3 increased by more than 2 fold compared to the other spinal cord volume measures.

Clinical-MRI associations

Group differences—The comparisons of the two groups of MS patients and healthy controls using normalization by number of slices and ICV are presented in Tables 3 and 4. All regions showed statistically significant differences between the three groups in age-adjusted analyses for all normalization factors. The differences between groups were driven by differences between the progressive (SP/PP) and relapsing (CIS/RR) or control groups (Table 4). In particular, the progressive patients had significantly lower volumes than healthy controls in each region and significantly lower volumes than relapsing patients in all but the thoracic region for some normalizations. In terms of comparing normalization techniques, normalization by the number of slices always improved the ability to discriminate between the groups, but normalization by the ICV never led to improved results. When normalization to both ICV and number of slices was compared to normalization to the number of slices only, the p-values for differences between groups was higher for normalization to ICV and number of slices. Therefore, adding ICV as a normalization factor provided no benefit in terms of detecting group differences.

For the group comparisons, normalization by the BMI decreased the ability to distinguish between groups compared to raw volumes (Supplementary Table 1). This result showed that normalizing using the proportional approach by an unimportant variable can lead to worse results because the normalization factor added variability to the outcome. Normalization by the BSA also decreased the ability to distinguish groups, but the change was small (Supplementary Table 1).

MRI-clinical correlations—The correlations between the clinical measures and each of the main normalized volumes (raw and normalized by number of slices/ICV) are presented in Table 5. C2-3 volume had the highest correlation with the EDSS and T25FW in the MS patients among all spinal cord volume measures ($p < 0.02$ for all correlations, Figure 2); normalization to ICV slightly lowered the correlations in both cases. For the cervical cord, all measures significantly correlated with EDSS score but only the volume normalized to number of slices significantly correlated with T25FW. For the thoracic cord, all correlations were mild, and none of the MRI-clinical correlations reached statistical significance. For the whole cord, the only MRI-clinical correlations that reached statistical significance were between volume normalized to ICV (with or without number of slices) and EDSS score. When comparing normalization to the number of slices alone to normalization to the number of slices and ICV, the number of slices alone led to slightly higher correlations for the C2-3 and cervical cord, but slightly lower correlations with the whole cord. Thus, there was no net benefit of adding ICV to the normalization provided by the number of slices. No significant correlations were observed in our sample between spinal cord volumes and disease duration ($p > 0.05$ for all comparisons; data not shown). Normalization to BMI and BSA did not

noticeably improve the correlations with EDSS or disease duration for any of the spinal cord volumes compared to the raw values or other normalization approaches, but correlations were improved and statistically significant with timed 25 foot walk (Supplementary Table 2).

MRI-MRI associations

Each of the spinal cord volume measures showed high correlations with the other measures. In particular, the C2-3 volume was highly correlated with the raw cervical cord volume ($r = 0.79$), thoracic cord volume ($r = 0.73$) and whole cord volume ($r = 0.79$). Interestingly, the correlation increased when each spinal cord volume was normalized by the number of slices ($r_{C2-3 \text{ vs } CC} = 0.90$, $r_{C2-3 \text{ vs } TC} = 0.76$, $r_{C2-3 \text{ vs } WC} = 0.84$ -Figure 3), and this normalization led to a significant improvement in the correlation between the cervical cord volume and C2-3 volume ($p=0.006$, Meng's test). The cervical and thoracic cord volumes were highly correlated using raw volumes ($r=0.82$), and the correlation increased slightly after normalization by the number of slices ($r=0.88$). Normalization by the ICV led to a slight decrease in the strength of the correlations among the spinal cord measures compared to the raw volumes (data not shown). Thus, as was seen in group comparisons and clinical-MRI correlations, there was no net benefit of adding ICV to the normalization provided by the number of slices. Given the poor performance of BMI and BSA in group differences and EDSS correlations, these normalization factors were not considered in this analysis.

Discussion

In this paper, several approaches for normalization of spinal cord volumes were considered. The normalization factor had a limited effect on the strength of correlations and the ability to detect group differences among progressive and relapsing MS patients or controls. The most consistently effective approach was normalization to the number of slices, which removed differences across patients due to length of the cord, and this normalization factor has the advantage of being easily collected at the time of analysis. We note that all scans in our study used the same slice thickness so that the number of slices alone could be used for normalization. If varied slice thicknesses were used, normalization by number of slices and slice thickness would be necessary for comparisons across patients. Since the proportional method is used, normalization by the number of slices corresponds to calculating the average volume per slice over the entire region of interest, which is an intuitively clear way to express spinal cord volume in MS patients.

In terms of normalizing for body size, a commonly used approach is to normalize by the ICV or some other measure of head size as discussed previously.[5, 9] In our study, normalization by the ICV often decreased the effect size between groups, while any improvements were generally minor. Since normalization by the ICV requires the addition of head imaging, this type of normalization does not appear to be fruitful. Our conclusion is consistent with previous work that showed normalization to the ICV or related measures of head size such as the maximum intracranial cross-sectional area (MICA)/mid-sagittal intracranial area (ICA), did not improve discrimination between groups.[5, 9] Normalization to two other measures of body size (BMI and BSA) led to worse performance in our study. However, simply normalizing to the number of slices was adequate to obtain good reliability and maximize effect sizes in group comparisons.

Our study directly compared the performance of C2-3 and three measures of spinal cord volume in terms of reliability and clinical correlations. Although the C2-3 showed better clinical correlations, the lower of reliability of this measure shows one potential disadvantage of this measure. The reliability of the C2-3 region in our study is lower than previous studies, but the use of a smaller number of slices for the C2-3 region led to a

greater possibility of error compared to the regional and whole cord measures. Newly available automated techniques have improved the reliability of C2-3 measurements so this limitation may not be a concern in future studies.[6]

Another important aspect of normalization is the use of a proportional approach, as in this paper, versus a residual approach. The differences between these approaches have been the subject of comparison for normalization of the brain.[8] The conclusions in brain studies have shown the residual approach can offer improvements under certain scenarios, but this model also requires additional assumptions and cannot be used for single measurements. In addition, each approach assumes a slightly different model for the relationship between the normalization factor and the volume of interest in terms of the error structure. Since our results showed normalization by the number of slices was optimal, the proportional approach provided a natural interpretation for the normalized volume since it is the average volume across slices. When a residual approach was used for group comparisons, the results were generally unchanged compared to the results for the proportional method (data not shown).

Our study has several limitations that warrant further discussion. First, our dataset had a small sample size, especially in the progressive group. Larger datasets including longitudinal data are needed to fully understand the true relationship between normalization factors and spinal cord volume. A second limitation is the lack of a gold standard to assess which normalization approach is superior. In the absence of such a gold standard, our approach was to compare normalizations based on several characteristics, but this approach can only provide guidance regarding normalization, rather than a definitive answer as to which approach is best. Another limitation of our study is that we did not collect information regarding the lumbar enlargement cord area, which was shown to perform well in previous work.[9] Therefore, we could not assess how normalization by this factor would have performed in our data.

In conclusion, our results show that significant differences in all regions of the spinal cord were observed between progressive MS patients and relapsing MS patients or controls. These differences persisted regardless of normalization measure or technique. The most robust and practical normalization measure is the number of slices only, and dividing the volumes by the number of slices provides an easily interpretable quantity for comparison.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding source: This research was funded in part by grants received from the NIH (1 R01 NS055083-01) and the National MS Society (NMSS - RG 3705-A-1).

References

1. Lin X, Tench CR, Evangelou N, Jaspan T, Constantinescu CS. Measurement of spinal cord atrophy in multiple sclerosis. *J Neuroimaging*. 2004; 14(3 Suppl):20S–26S. [PubMed: 15228756]
2. Lin X, Tench CR, Turner B, Blumhardt LD, Constantinescu CS. 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 beta-1a (Rebif) treatment trial. *J Neurol Neurosurg Psychiatry*. 2003; 74(8):1090–1094. [PubMed: 12876240]

3. Gilmore CP, DeLuca GC, Bo L, Owens T, Lowe J, Esiri MM, et al. Spinal cord atrophy in multiple sclerosis caused by white matter volume loss. *Arch Neurol*. 2005; 62(12):1859–1862. [PubMed: 16344343]
4. Brex PA, Leary SM, O'Riordan JI, Miszkiet KA, Plant GT, Thompson AJ, et al. Measurement of spinal cord area in clinically isolated syndromes suggestive of multiple sclerosis. *J Neurol Neurosurg Psychiatry*. 2001; 70(4):544–547. [PubMed: 11254786]
5. Zivadinov R, Banas AC, Yella V, Abdelrahman N, Weinstock-Guttman B, Dwyer MG. Comparison of three different methods for measurement of cervical cord atrophy in multiple sclerosis. *AJNR Am J Neuroradiol*. 2008; 29(2):319–325. [PubMed: 17974604]
6. Horsfield MA, Sala S, Neema M, Absinta M, Bakshi A, Sormani MP, et al. Rapid semiautomatic segmentation of the spinal cord from magnetic resonance images: application in multiple sclerosis. *Neuroimage*. 2010; 50(2):446–455. [PubMed: 20060481]
7. Stankiewicz JM, Neema M, Alsop DC, Healy BC, Arora A, Buckle GJ, et al. Spinal cord lesions and clinical status in multiple sclerosis: A 1.5 T and 3 T MRI study. *J Neurol Sci*. 2009; 279(1–2): 99–105. [PubMed: 19178916]
8. Sanfilippo MP, Benedict RH, Zivadinov R, Bakshi R. Correction for intracranial volume in analysis of whole brain atrophy in multiple sclerosis: the proportion vs. residual method. *Neuroimage*. 2004; 22(4):1732–1743. [PubMed: 15275929]
9. Song F, Huan Y, Yin H, Ge Y, Wei G, Chang Y, et al. Normalized upper cervical spinal cord atrophy in multiple sclerosis. *J Neuroimaging*. 2008; 18(3):320–327. [PubMed: 18318794]
10. Bermel RA, Bakshi R. The measurement and clinical relevance of brain atrophy in multiple sclerosis. *Lancet Neurol*. 2006; 5(2):158–170. [PubMed: 16426992]
11. Khoury S, Bakshi R. Cerebral pseudoatrophy or real atrophy after therapy in multiple sclerosis. *Ann Neurol*. 2010; 68(6):778–779. [PubMed: 21194148]
12. Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Ann Neurol*. 2005; 58(6):840–846. [PubMed: 16283615]
13. Neema M, Guss ZD, Stankiewicz JM, Arora A, Healy BC, Bakshi R. Normal findings on brain fluid-attenuated inversion recovery MR images at 3T. *AJNR Am J Neuroradiol*. 2009; 30(5):911–916. [PubMed: 19369605]
14. Arora A, Neema M, Stankiewicz J, et al. Regional and whole spinal cord atrophy in multiple sclerosis. *Neurology*. 2008; 70(Supplement 1):A464.
15. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983; 33(11):1444–1452. [PubMed: 6685237]
16. Fischer JS, Rudick RA, Cutter GR, Reingold SC. The Multiple Sclerosis Functional Composite Measure (MSFC): an integrated approach to MS clinical outcome assessment. National MS Society Clinical Outcomes Assessment Task Force. *Mult Scler*. 1999; 5(4):244–250. [PubMed: 10467383]
17. Mosteller RD. Simplified calculation of body-surface area. *N Engl J Med*. 1987; 317(17):1098. [PubMed: 3657876]
18. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996; 1(1):30–46.
19. Meng X-L, Rosenthal R, Rubin DB. Comparing correlated correlation coefficients. *Psychol Bull*. 1992; 111(1):172–175.

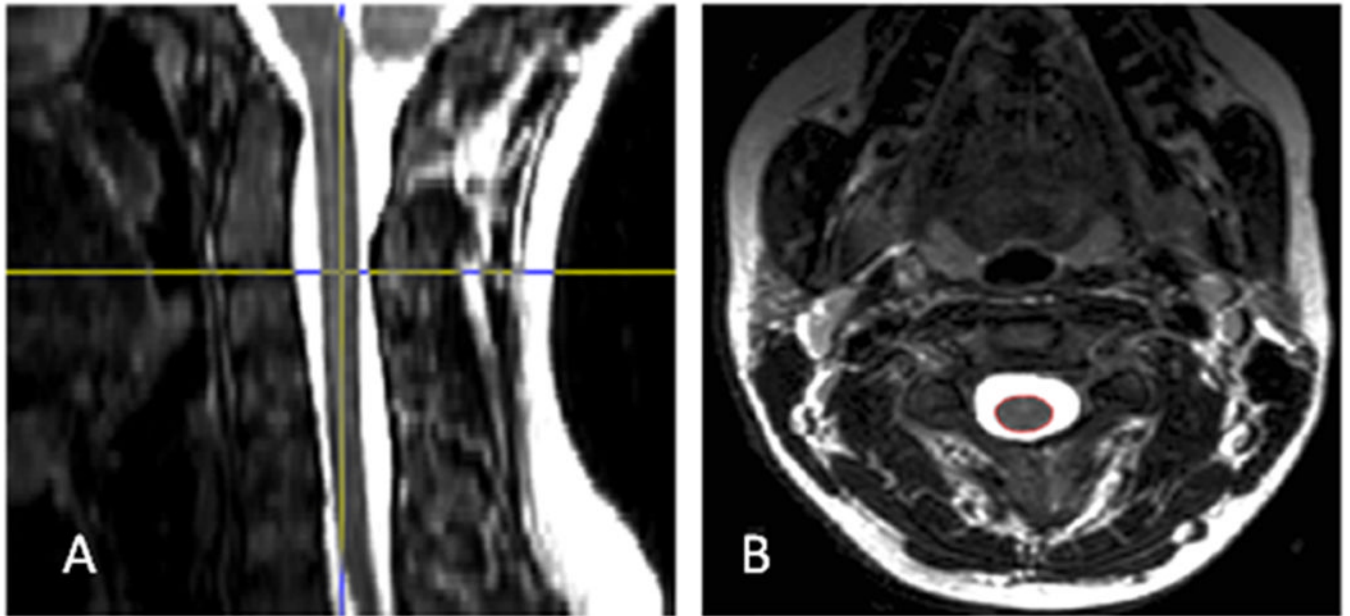


Figure 1. Representative spinal cord image

Representative MRI images: The figure shows the native axial T2-weighted image (B) and the reconstructed sagittal view of the data set provided by a reslicing orthogonal tool (A). The localizing lines in the sagittal view were used to identify the anatomic level of the axial slices. Then using an edge finding tool based on local thresholding, the spinal cord contour was identified, resulting in a region of interest (in red) in the axial image.

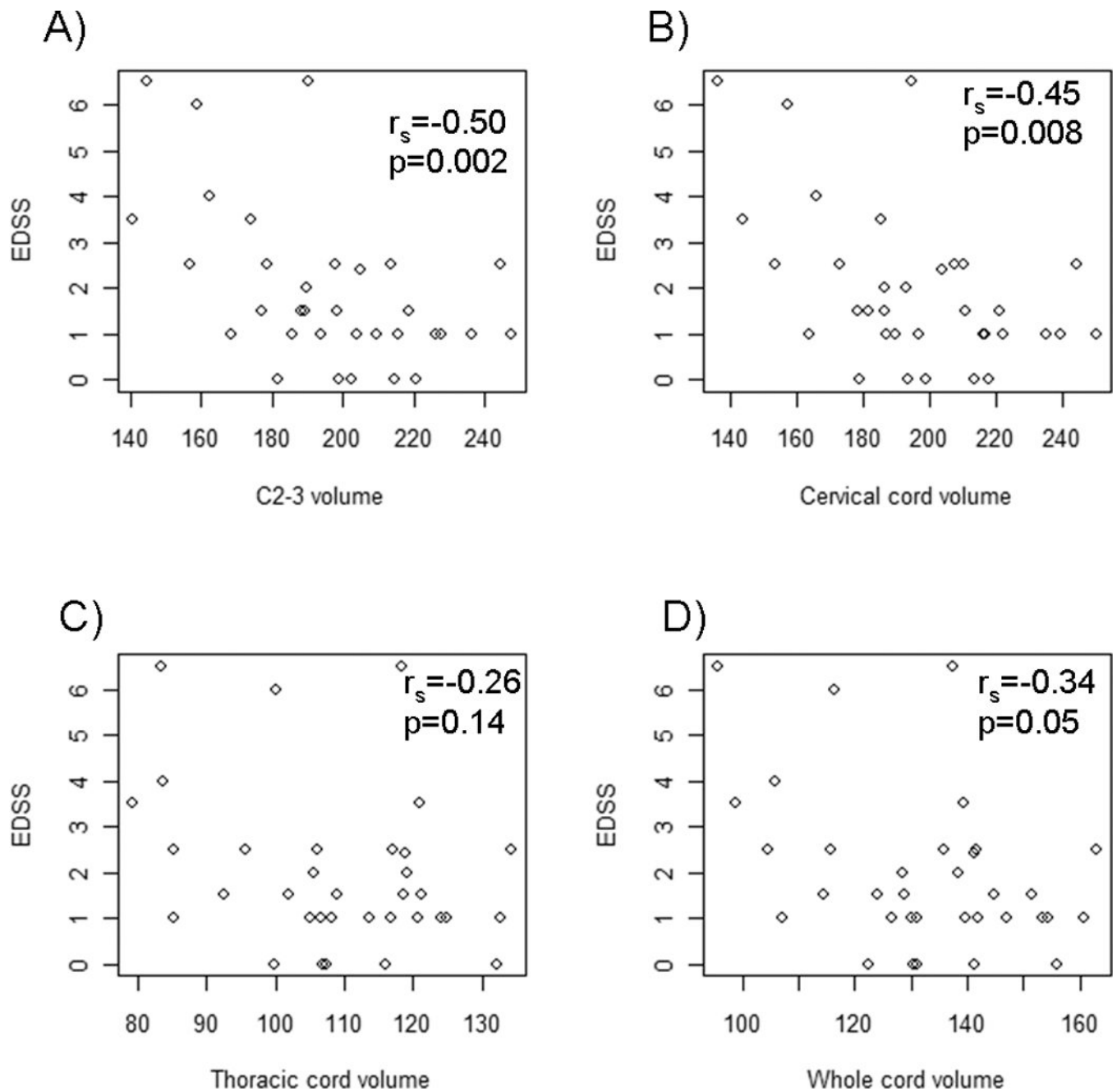


Figure 2. Comparison of regional cord volumes/whole cord volume and EDSS

All volumes (mm^3) were normalized by the number of slices. Spearman correlation coefficients between the two measures and p-values are reported in each graph. First row: Relationship between Expanded Disability Status Scale (EDSS) score and A) C2-3 or B) cervical cord volume. Second row: Relationship between EDSS and C) thoracic or D) whole cord volume.

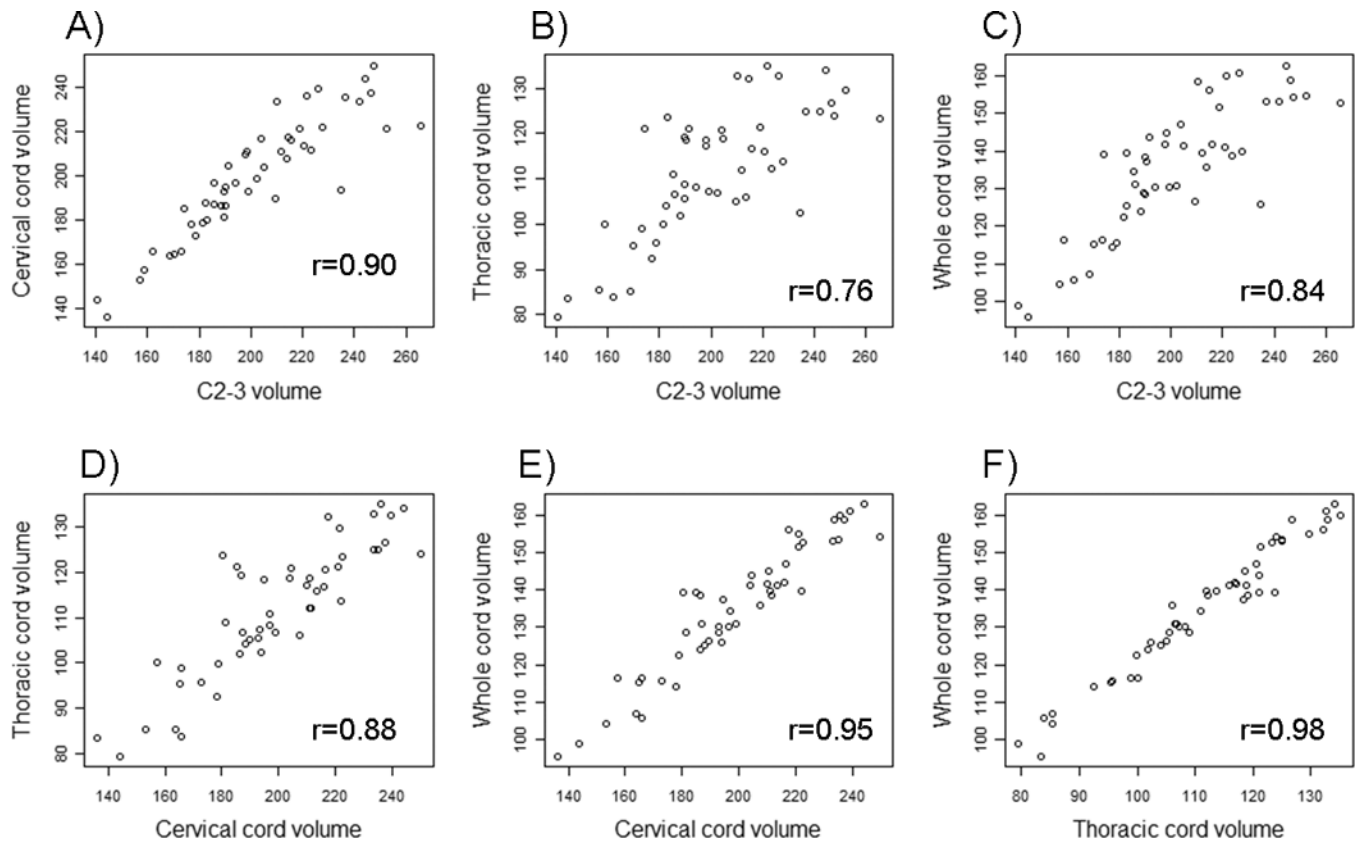


Figure 3. Comparison of regional cord volumes and whole cord volume

All volumes (mm^3) were normalized by the number of slices. Pearson correlation coefficients between the two measures are reported in each graph, and p-values for each correlation were less than 0.0001. First row: Relationship between C2-3 volume and A) cervical, B) thoracic, or C) whole cord volume. Second row: Relationship between D) cervical and thoracic cord volume, E) cervical and whole cord volume and F) thoracic and whole cord volume.

Table 1

Demographic and clinical characteristics of MS patients and healthy controls

Characteristic	MS patients	Healthy controls	p-value
N	34	15	
Gender (% female)	76.5	73.3	1
Age [years, mean \pm SD, (range)]	41.6 \pm 8.9, (25–55)	41.5 \pm 8.5, (30–53)	0.99
Disease duration [years, mean \pm SD, (range)]	8.37 \pm 8.61, (0.2, 30)	-	-
EDSS [mean (range)]	1.95 (0, 6.5)	-	-
MS disease category (CIS/RR/SP/PP)	2/26/4/2	-	-
Number of slices in cervical cord	37.8 \pm 3.7	38.2 \pm 3.6	0.75
Number of slices in thoracic cord	102.9 \pm 6.4	103.2 \pm 7.3	0.99
Number of slices in whole cord	140.7 \pm 8.6	141.4 \pm 9.9	0.85

MS= multiple sclerosis, EDSS=expanded disability status scale, CIS=clinically isolated syndrome, RR=relapsing-remitting, SP=secondary progressive, PP=primary progressive, (-)=not applicable

Table 2

Comparison of reliability of normalization techniques

	% COV intra-rater	ICC intra- rater	% COV inter-rater	ICC inter- rater
C2-3 spinal cord volume normalized to # of slices	1.24	0.939	4.22	0.184
C2-3 volume normalized to # of slices and ICV	1.36	0.987	4.03	0.902
Whole cord raw volume	0.52	0.996	0.66	0.987
Whole cord volume normalized to # of slices	0.61	0.988	0.74	0.982
Whole cord volume normalized to ICV	0.44	0.997	1.30	0.990
Whole cord volume normalized to ICV and number of slices	0.53	0.998	1.34	0.994
ICV	0.11	0.9998	1.42	0.999

COV=coefficient of variation, ICC=intraclass correlation coefficient for consistency, ICV=intracranial volume

Table 3

Group comparisons of spinal cord volumes among MS and healthy control groups

Region	Volume	Relapsing MS: mean (SD)	Progressive MS: Mean (SD)	Healthy control: Mean (SD)	P- value*
C2-3	Normalized to # of slices (mm ³)	202.9 (22.1)	161.9 (18.5)	212.9 (30.9)	0.0002
	Normalized to # of slices and ICV**	1.5*10 ⁻⁴ (1.8*10 ⁻⁵)	1.2*10 ⁻⁴ (1.5*10 ⁻⁵)	1.6*10 ⁻⁴ (2.1*10 ⁻⁵)	0.0025
Cervical cord	Raw (mm ³)	7691 (1162)	6065 (1125)	7876 (1015)	0.0024
	Normalized to ICV**	0.0057 (0.0007)	0.0046 (0.0006)	0.0057 (0.0006)	0.0025
	Normalized to # of slices (mm ³)	202.4 (24.0)	163.9 (22.9)	206.7 (24.6)	0.0003
	Normalized to # of slices and ICV**	1.5*10 ⁻⁴ (1.8*10 ⁻⁵)	1.2*10 ⁻⁴ (1.8*10 ⁻⁵)	1.5*10 ⁻⁴ (1.6*10 ⁻⁵)	0.0031
Thoracic cord	Raw (mm ³)	11580 (1713)	9811 (2176)	12048 (1442)	0.018
	Normalized to ICV**	0.0085 (0.0011)	0.0074 (0.0014)	0.0088 (0.0009)	0.034
	Normalized to # of slices (mm ³)	111.6 (13.0)	97.7 (18.5)	116.9 (12.7)	0.010
	Normalized to # of slices and ICV**	8.2*10 ⁻⁵ (9.0*10 ⁻⁶)	7.4*10 ⁻⁵ (1.2*10 ⁻⁵)	8.5*10 ⁻⁵ (8.1*10 ⁻⁶)	0.038
Whole cord	Raw (mm ³)	19271 (2770)	15877 (3100)	19924 (2263)	0.0052
	Normalized to ICV**	0.014 (0.0018)	0.012 (0.0018)	0.015 (0.0014)	0.0076
	Normalized to # of slices (mm ³)	135.9 (15.3)	115.5 (19.0)	141.1 (15.2)	0.0022
	Normalized to # of slices and ICV**	1.0*10 ⁻⁴ (1.1*10 ⁻⁵)	8.7*10 ⁻⁵ (1.3*10 ⁻⁵)	1.0*10 ⁻⁴ (9.5*10 ⁻⁶)	0.011

MS= multiple sclerosis, ICV=intracranial volume; Progressive MS= secondary or primary progressive; Relapsing MS =clinically isolated syndromes or relapsing-remitting,

* p-values are age-adjusted and are three group comparisons,

** unitless measure.

Table 4

Pairwise group comparisons of spinal cord volumes among progressive MS and relapsing MS or healthy control groups

Region	Normalization approach	PP/SP vs. controls: p-value	PP/SP vs. RR/CIS: p-value
C2-3	Normalized to # of slices (mm ³)	0.00011	0.00062
	Normalized to # of slices and ICV ^{**}	0.0020	0.0061
Cervical cord	Raw volume (mm ³)	0.0025	0.0035
	Normalized to ICV ^{**}	0.0031	0.0032
	Normalized to # of slices (mm ³)	0.0004	0.0005
	Normalized to #of slices and ICV ^{**}	0.0042	0.0036
Thoracic cord	Raw volume (mm ³)	0.015	0.040
	Normalized to ICV ^{**}	0.032	0.071
	Normalized to #of slices (mm ³)	0.0078	0.037
	Normalized to #of slices and ICV ^{**}	0.034	0.11
Whole cord	Raw volume (mm ³)	0.0046	0.010
	Normalized to ICV ^{**}	0.0072	0.013
	Normalized to #of slices (mm ³)	0.0017	0.0061
	Normalized to #of slices and ICV ^{**}	0.0098	0.023

ICV=intracranial volume, PP= primary progressive, SP=secondary progressive, RR=relapsing remitting, CIS=clinically isolated syndrome. Reported p-values are age-adjusted p-values from pairwise comparisons of groups with Bonferonni correction to account for multiple comparisons.

^{**} unitless measure.

Table 5

Correlations between MRI-based spinal cord volumes and neurologic disability

Region	Volume	Correlation (r_s) with EDSS	p- value	Correlation (r_s) with T25FW	p- value
C2-3	Normalized to # of slices	-0.509	0.002	-0.424	0.012
	Normalized to # of slices and ICV	-0.451	0.007	-0.400	0.019
Cervical cord	Raw	-0.401	0.019	-0.272	0.120
	Normalized to ICV	-0.428	0.012	-0.297	0.088
	Normalized to # of slices	-0.450	0.008	-0.352	0.041
	Normalized to # of slices and ICV	-0.420	0.013	-0.334	0.054
Thoracic cord	Raw	-0.296	0.089	-0.269	0.124
	Normalized to ICV	-0.298	0.087	-0.295	0.091
	Normalized to # of slices	-0.258	0.141	-0.229	0.193
	Normalized to # of slices and ICV	-0.286	0.102	-0.209	0.237
Whole cord	Raw	-0.330	0.057	-0.268	0.125
	Normalized to ICV	-0.373	0.030	-0.303	0.081
	Normalized to # of slices	-0.339	0.050	-0.230	0.191
	Normalized to # of slices and ICV	-0.354	0.040	-0.256	0.144

ICV=intracranial volume, EDSS=Expanded Disability Status Scale, T25FW=Timed 25 Foot Walk, r_s =Spearman's correlation coefficient.