Longitudinal spinal cord atrophy in multiple sclerosis using the generalised boundary shift integral

RUNNING HEAD: GBSI for spinal cord atrophy

Marcello Moccia[*1-2], MD, PhD; Ferran Prados*[*1,3-5], PhD; Massimo Filippi[6,7], MD, PhD; Maria A Rocca[6,7], MD, PhD; Paola Valsasina[6], MD, PhD; Wallace J Brownlee[1], MD, PhD; Chiara Zecca[8], MD; Antonio Gallo[9], MD, PhD; Alex Rovira[10], MD; Achim Gass[11], MD; Jacqueline Palace[12], MD, PhD; Carsten Lukas[13], MD, PhD; Hugo Vrenken[14], PhD; Sebastien Ourselin[15], PhD; Claudia A.M. Gandini Wheeler-Kingshott[1,16-17], PhD; Olga Ciccarelli[1,4], MD, PhD; Frederik Barkhof[1,3-4,14] MD, PhD for the MAGNIMS Study Group**.*

***Authors contributed equally.**

****Members of the MAGNIMS Study Group Steering Committee are reported in the Supplementary Material 2.**

1. Queen Square MS Centre, Department of Neuroinflammation, UCL Queen Square Institute of Neurology, Faculty of Brain Sciences,
2. Multiple Sclerosis Clinical Care and Research Centre, Department of Neurosciences, Federico II University, Naples, Italy
3. Centre for Medical Image Computing (CMIC), Department of Medical Physics and Bioengineering,
4. National Institute for Health Research (NIHR), University College London Hospitals (UCLH) Biomedical Research Centre,

5. Universitat Oberta de Catalunya, Barcelona, Spain
6. Neuroimaging Research Unit, Institute of Experimental Neurology, Division of Neuroscience, San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy.
7. Department of Neurology, San Raffaele Scientific Institute, Milan, Italy.
8. Neurocenter of Southern Switzerland, Ospedale Regionale di Lugano, Lugano, Switzerland.
9. 3T-MRI Research Center, Department of Advanced Medical and Surgical Sciences, University of Campania Luigi Vanvitelli, Naples, Italy.
10. Section of Neuroradiology, Department of Radiology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain.
11. Department of Neurology, Universitätsmedizin Mannheim, University of Heidelberg, Mannheim, Germany.
12. Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, Oxford, United Kingdom.
13. St. Josef Hospital, Ruhr University, Bochum, Germany.
14. Department of Radiology and Nuclear Medicine, VU University Medical Center, Amsterdam, The Netherlands.
15. Department of Imaging and Biomedical Engineering, King’s College London, United Kingdom.
16. Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy.
17. Brain MRI 3T Research Center, IRCCS Mondino Foundation, Pavia, Italy.

CORRESPONDECE TO: Frederik Barkhof

Queen Square MS Centre, Department of Neuroinflammation

UCL Queen Square Institute of Neurology

Faculty of Brain Sciences

Email: f.barkhof@ucl.ac.uk

NUMBER OF CHARACTERS IN THE TITLE/RUNNING HEAD: 100/28

NUMBER OF WORDS IN ABSTRACT/INTRODUCTION/DISCUSSION/MANUSCRIPT: 259/280/1096/3126

NUMBER OF FIGURES/COLOR FIGURES/TABLE: 5/1/2.

Longitudinal spinal cord atrophy in multiple sclerosis using the generalised boundary shift integral

ABSTRACT

Objectives. Spinal cord atrophy is a clinically relevant feature of multiple sclerosis (MS), but longitudinal assessments on MRI using segmentation-based methods suffer from measurement variability, especially in multicentre studies. We compared the generalised boundary shift integral (GBSI), a registration-based method, with standard segmentation-based method.

Methods. Baseline and 1-year spinal cord 3DT1-weighted images (1mm isotropic) were obtained from 282 patients (52 clinically isolated syndrome (CIS), 196 relapsing-remitting MS (RRMS), 34 progressive MS (PMS)), and 82 controls from eight MAGNIMS sites, on multi-manufacturer and multi-field strength scans. Spinal Cord Toolbox was used for C2-5 segmentation and cross-sectional area (CSA) calculation. After cord straightening and registration, GBSI measured atrophy based on the probabilistic boundary-shift region-of-interest. CSA and GBSI percent annual volume change was calculated.

Results. GBSI provided similar rates of atrophy, but reduced measurement variability than CSA in all MS subtypes (CIS: -0.95±2.11% vs. -1.19±3.67%; RRMS: -1.74±2.57% vs. -1.74±4.02%; PMS: -2.29±2.40% vs. -1.29±3.20%), and healthy controls (0.02±2.39% vs. -0.56 ±3.77%). GBSI performed better than CSA in differentiating healthy controls from CIS (AUC=0.66 vs. 0.53; p=0.03), RRMS (AUC=0.73 vs. 0.59; p<0.001), PMS (AUC=0.77 vs. 0.53; p<0.001), and patients with disability progression from patients without progression (AUC=0.59 versus 0.50; p=0.04). Sample size to detect 60% treatment effect on spinal cord atrophy over one year was lower for GBSI than CSA (CIS: 106 vs. 830; RRMS: 95 vs. 335; PMS: 44 vs. 215) (power=80%; alpha=5%).

Interpretation. The registration-based method (GBSI) allowed better separation between MS patients and healthy controls and improved statistical power, when compared with conventional segmentation-based method (CSA), though still far from perfect.

INTRODUCTION

Spinal cord atrophy on MRI is a marker of neurodegeneration in multiple sclerosis (MS),[1,2] and is one of the main substrates of long-term disease progression.[3–11] Spinal cord atrophy progresses faster than brain atrophy (1.7%/year vs 0.4-0.6%/year), is greater in progressive MS than in the relapsing forms of MS, and predicts disability.[5,12,13] It is crucial to obtain an accurate and precise longitudinal measurement of spinal cord atrophy, because it could be used to monitor disease progression and become a primary outcome measure in phase 2 clinical trials with neuroprotective therapies, not only in MS, but also in other neurodegenerative disorders.[14–17]

Spinal cord atrophy is conventionally estimated with segmentation-based methods (e.g., cervical cord cross-sectional area (CSA)), applied to volumetric spinal cord images,[18] that measure cord characteristics at each time point; indirect longitudinal atrophy measurements are obtained by numerical subtraction, with relatively-low reproducibility and responsiveness to change.[6,19] On the contrary, longitudinal brain atrophy measurements are nowadays based on registration-based techniques that significantly reduce measurement noise.[20]

Based on this, we evaluated a registration-based method used for brain atrophy, to measure longitudinal spinal cord atrophy; in this method, named generalised boundary shift integral (GBSI), atrophy is directly measured from the probabilistic boundary-shift region-of-interest, adaptively estimated between two time points.[21–23] Therefore, in the present multicentre, multi-manufacturer and multi-field strength scan study, we aimed to: (1) compare measurements of spinal cord atrophy obtained using GBSI with those obtained with conventional CSA (automatic segmentation with the Spinal Cord Toolbox); (2) explore associations between GBSI- and CSA-derived spinal cord measurements and MS clinical features; and (3) estimate the sample size needed to detect changes in spinal cord atrophy over one year using GBSI and CSA.

METHODS

Study design and population

This is a multicentre, retrospective study, conducted on prospectively collected data from Queen Square MS Centre (Magnetic Resonance Imaging in Multiple Sclerosis (MAGNIMS) Collaboration (www.magnims.eu) centres. Overall, we included 327 MS patients and 96 healthy controls. The London and MAGNIMS cohorts have been reported in previous publications.[3,18,24–26]

Eligibility criteria were: (i) diagnosis of clinically isolated syndrome or MS according to the 2010 McDonald Criteria;[27] (ii) healthy controls without history of neurological or psychiatric disorders; (iii) the presence of at least two volumetric MRI scans (interval between scans was collected), acquired with isotropic voxel of 1x1x1mm^3; (iv) information on: Expanded Disability Status Scale (EDSS) score at each time point,[28] subtype of MS (CIS, RRMS and PMS (including both primary and secondary progressive MS)),[27] age, gender and disease duration (time from clinical onset to baseline MRI).

Each participant had provided a written consent for research within each centre. The final protocol for the analysis of pseudo-anonymized scans, acquired independently and prospectively in each centre, was approved by the European MAGNIMS collaboration and by the local ethics committees.

MRI acquisition and processing

Dedicated cervical spinal cord 3DT1-weighted images (1x1x1mm^3) were analysed. Images were acquired in 8 MAGNIMS sites on 1.5 T and 3 T scanners, from different manufacturers and with different MRI parameters (**Supplementary Material 1**).

For calculating CSA and GBSI, masks of C2-5 spinal cord level were obtained for images acquired at each time point with Spinal Cord Toolbox, using the routine method known as *PropSeg* (version 3.1.1) (**FIG 1**).[29] C2-5 CSA was obtained by averaging all cord cross-sectional areas. For GBSI, we followed the previously described steps.[22,23,30] Briefly, after straightening the cord at both time points, a 3D symmetric and inverse-consistent rigid-only (9 DOF) registration to the half-way space between baseline and follow-up images was performed; masks were resampled to the same space using linear interpolation and registered to the halfway space. This method does not generate any bias between the baseline and follow-up images as the exact same image processing pipeline is applied to both time points. The probabilistic boundary-shift region-of-interest was then adaptively estimated from baseline and follow-up cord segmentations. The GBSI integral was finally computed (**FIG 1**).

Percent CSA annual change between time points was calculated using the following formula: 100*[(CSA at follow-up – CSA at baseline)/CSA at baseline)/years between baseline and follow-up scans]. Similarly, percent GBSI annual change was calculated using the following formula: GBSI/ years between baseline and follow-up scans. Of note, spinal cord atrophy based on CSA is measured as the average of the slice-wise 2D edge detection over a fixed cord section/length, hence effectively it is a 3D volume divided by its height; whilst GBSI measures the voxel-by-voxel difference in intensities between the baseline and the follow-up images, as a 3D volume change. We segmented the spinal cord over a fixed cord length (C2-5), and we used the same masks to compute CSA and GBSI, thus making the atrophy measurement effectively computed over the same boundary/region for both measurements.[19]

To determine measurement precision, we analysed test-retest data from 9 healthy controls and 9 MS patients that were re-scanned after having been removed from the scanner and repositioned between the scans during the same visit at the Queen Square MS Centre, .[31] Images were acquired and processed as described above (**FIG 1**). Measurement error between scans was quantified by the coefficient of variation and the median absolute deviation.

Statistical analyses

Means, medians and proportions of demographics, clinical features and MRI measures (percentage spinal cord changes obtained with GBSI and CSA) were calculated for patients (and their subgroups) and healthy controls. Differences were evaluated with *t*-test, Mann–Whitney test, χ^2 test or Fisher's exact test, as appropriate. GBSI and CSA atrophy progression measurements were compared using a paired t-test.

Linear regression models were employed to estimate spinal cord atrophy changes (with GBSI and CSA) in different disease phenotypes, when compared with healthy controls (used as reference group); age, sex, site of acquisition and disease duration were included as covariates (results are presented as adjusted coefficients (Coeff), 95% confidence intervals (95%CI), and p-values). Then, we used different disease phenotypes as reference group in the linear regression models, to perform direct comparisons between different disease phenotypes (e.g., CIS, RRMS, PMS).

To compare GBSI and CSA in their ability to predict different clinical variables, we employed logistic regression models to estimate associations between percentage spinal cord atrophy change obtained using GBSI and CSA (independent continuous variable), and different binary clinical variables (dependent variable) (e.g., disease subtypes/healthy controls, EDSS progression (which was defined as 1 point change if baseline EDSS≤5.5, and 0.5 point if ≥6.0); age, sex, site of acquisition and disease duration were included as covariates. Results are presented as odds ratios (OR), 95%CI and p-values. Based on this, we obtained areas-under-the-curve (AUC), using GBSI and CSA, in turn, as the main explanatory variables; we used bootstrap resampling (1000 repetitions) to calculate pointwise confidence intervals for the ROC curve.

Sample sizes required for a hypothetical clinical trial evaluating a neuroprotective medication over one year were estimated using CSA and GBSI. Sample size was computed using the formula $n = \frac{2(Z_{\alpha}+Z_{1-\beta})^2\sigma^2}{\Delta^2}$, where *n* is the required sample size per treatment arm in 1:1 controlled trials, Z_{α} and $Z_{1-\beta}$ are constant (set at 5% alpha-error and 80% power, respectively), σ is the standard deviation (from each disease phenotype), and Δ is the estimated effect size.[15,20] Effect size was derived from adjusted beta-coefficients at linear regression models, estimating spinal cord loss in MS patients, when compared with physiologic loss in controls. As such, we assumed that 100% treatment effect is theoretically reached when the spinal cord atrophy change in patients is equal to that observed in healthy controls. From there, with a conservative approach, we hypothesized a number of effect sizes (e.g., 30%, 60% and 90%), that were smaller than the observed difference between MS cases and physiologic spinal cord loss in controls.

Stata 15.0 was used for data processing and analysis. Results were considered statistically significant when associated with p values <0.05.

RESULTS

Measurement error

Among healthy controls and MS patients scanned twice during the same session, spinal cord atrophy would be expected to be zero. Both GBSI and CSA indeed had overall atrophy rates tending towards zero (0.28±3.62% and -0.17±5.39%, respectively), indicating no systematic bias. However, GBSI presented with lower random measurement error, when compared with CSA, reflected by a clearly lower coefficient of variation (12.92% and 31.70%) and slightly lower median absolute deviation (0.21% and 0.25%).

Study population

Longitudinal spinal cord scans from 327 MS patients and 96 healthy controls were collected. Forthy-five patients' scans and 14 healthy controls' scans were excluded because of either Spinal Cord Toolbox failure, likely as a consequence of poor contrast (n=13), wrong voxel size (n=26), wrong acquisition parameters (n=10), and artifacts (n=8), or GBSI failure (n=2), mostly from the eldest cohorts acquired using 1.5 T scanner (e.g., CIS cohort) (see **FIG 2** for the study flow diagram). Therefore, scans from 282 patients and 82 healthy controls acquired in 8 MAGNIMS centres were included in the analysis (see **Supplementary Material 1** for the number of patients per centre and acquisition parameters). The demographic and clinical features of patients and healthy controls are given in **Table 1**.

Spinal cord atrophy obtained with CSA and GBSI

On paired t-test, the percentage spinal cord changes obtained with CSA were similar to those obtained with GBSI (p=0.55).

On linear regression models adjusted for age, sex, site of acquisition and disease duration, using CSA as main variable of interest, there was a significant decrease in the percentage spinal cord change over one year between RRMS (-1.74±4.02%) and healthy controls (-0.56±3.77%) (Coeff=-1.45; 95%CI=-2.81, -0.10; p=0.03), but not between CIS (-1.19±3.67%) and healthy controls (Coeff=-0.84; 95%CI=-2.49, 0.80; p=0.31), nor between PMS (-1.29±3.20%) and healthy controls (Coeff=-1.45; 95%CI=-3.70, 0.80; p=0.21) (**FIG 3A**). There were no differences in percentage CSA change over one year between CIS and RRMS (Coeff=-0.61; 95%CI=-1.95, 0.72; p=0.37), nor between CIS and PMS (Coeff=-0.60; 95%CI=-2.73, 1.53; p=0.58), nor between RRMS and PMS (Coeff=-0.00; 95%CI=-1.75, 1.73; p=0.99), when adjusted for the same covariates.

When using GBSI, overall, the rates of spinal cord decline were similar (or even higher among PMS) to those obtained with CSA, but the standard deviations of the measurements were smaller. On linear regression models adjusted for age, sex, site of acquisition and disease duration, using GBSI as main variable of interest, there was a significant decrease in the percentage spinal cord change obtained with GBSI between healthy controls (0.02±2.39%) when compared to: CIS (-0.95±2.11%) (Coeff=-1.37; 95%CI=-2.40, -0.33; p=0.01), RRMS (-1.74±2.57%) (Coeff=-1.84; 95%CI=-2.70, -0.99; p<0.01), and PMS (-2.29±2.40%) (Coeff=-2.44; 95%CI=-3.87, -1.02; p<0.01) (**FIG 3B**). Similarly to the findings obtained with CSA, no differences were detected in GBSI percent annual reduction between CIS and RRMS (Coeff=-0.48; 95%CI=-1.32/0.37; p=0.27), nor between CIS and PMS (Coeff=-1.08; 95%CI=-2.43, 0.27; p=0.12), nor between RRMS and PMS (Coeff=0.60; 95%CI=-0.60, 1.70; p=0.28), when adjusted for the same covariates.

Clinical correlates of CSA and GBSI.

On logistic regression models adjusted by age, sex, site of acquisition and disease duration, RRMS had higher probability of spinal cord atrophy progression on CSA, when compared with healthy controls. On GBSI, all MS subtypes (CIS, RRMS and PMS patients) had higher probability of spinal cord atrophy progression, when compared with controls. Also, on GBSI, MS patients with EDSS progression had higher probability of spinal cord atrophy progression, than those without (**Table 2**).

CIS patients were better differentiated from controls using GBSI (AUC=0.66; 95%CI=0.57, 0.75), than CSA (AUC=0.53; 95%CI=0.43, 0.63) (p=0.03; **FIG 4A**). RRMS patients were better differentiated from controls using GBSI (AUC=0.73; 95%CI=0.66, 0.80), than CSA (AUC=0.59; 95%CI=0.52, 0.66) (p<0.01; **FIG 4B**). PMS patients were better differentiated from controls using GBSI (AUC=0.77; 95%CI=0.68, 0.86), than CSA (AUC=0.53; 95%CI=0.45, 0.64) (p<0.01; **FIG 4C**). Patients with EDSS progression (n=76) were better differentiated from those without EDSS progression (n=206) using GBSI (AUC=0.59; 95%CI=0.52-0.66) than CSA (AUC=0.50; 95%CI=0.43-0.58) (p=0.04; **FIG 4D**).

Sample size estimates for a neuroprotective clinical trial using GBSI and CSA

The minimum sample sizes per arm required to detect a 60% treatment effect in one year clinical trial (i.e., a 60% reduction in percentage spinal cord change in MS cases when compared with physiologic spinal cord loss in controls, adjusted for age, sex, site of acquisition and disease duration) were lower for GBSI compared with CSA (CIS: 106 vs. 830; RRMS: 95 vs. 335; PMS: 44 vs. 215) (power=80%, alpha=5%). Similar results were obtained when estimating the sample size required to detect different treatment effects (**FIG 5**).

DISCUSSION

There have been few clinical trials and observational studies in MS that used spinal cord atrophy as an outcome measure, because of the large sample size required when using the available CSA method.[16,19,32,33] In the present study we applied a standard semi-automatic pipeline for spinal cord segmentation using the Spinal Cord Toolbox, and, then, a fully automated registration-based technique (GBSI) for spinal cord atrophy to a large, multicentre, multi-manufacturer and multi-field strength scan cohort, derived from longitudinal observational studies. The rates of spinal cord loss over one year obtained with GBSI were similar to those obtained with CSA, but they were associated with lower variability, greater ability to distinguish between MS patients and controls, and more robust clinical correlates, thereby holding promise for future MS research on spinal cord imaging. Use of GBSI yielded increased statistical power to detect treatment changes, suggesting that future treatment trials –particularly those testing neuroprotective agents– could include spinal cord atrophy as a primary outcome measure.

CSA and GBSI provided similar rates of spinal cord atrophy in each MS subtype, but CSA yielded a larger variability (standard deviation), when compared with GBSI (e.g., in RRMS ±4.02% vs. ±2.57%, respectively), implying that GBSI measurements are more precise. Similarly, when images were acquired with a scan-rescan fashion, GBSI presented with lower coefficient of variation, when compared with CSA, indicating higher precision; the median absolute deviation (0.21%) was similar to what has been previously described in the validation of a registration-based method for brain atrophy (0.15-0.20% for the Structural Image Evaluation using Normalisation of Atrophy (SIENA)).[34] Of note, a higher percentage of spinal cord decline (and a lower standard deviation) was found in PMS when using GBSI than when using CSA (-2.29±2.40% vs -1.29±3.20%), and this was probably due to the relatively small sample in this MS subtype, further highlighting the higher measurement precision of GBSI. The smaller variability in GBSI-derived measurements may be due to the ability of GBSI to deal with partial volume effects, which can lead to the inclusion of tissue outside of the area of interest with subsequent segmentation errors, and to the variability when calculating the absolute cross-sectional areas. These partial volume effects have less influence on GBSI boundary contours, with a consequent smaller variability of the measurements obtained with this technique.[21,22,35]

We found that GBSI allowed a four-fold smaller (and therefore achievable) sample size than that obtained with CSA. In a recent review of previous clinical trials in MS,[9] differences between treated and untreated patients ranged from 0.4% to 1.8%/year spinal cord loss on CSA, in line with a roughly 60% treatment effect (corresponding to 0.7%/year spinal cord loss). Overall, the sample size estimates for spinal cord atrophy measurements with GBSI are of the same order of magnitude as those for brain atrophy obtained with registration-based methods.[20,36–38] However, monitoring spinal cord atrophy could be more clinically relevant than brain atrophy, due to its robust clinical correlates.[3–5] Spinal cord atrophy occurs since the early stages of MS (e.g., CIS), is more obvious in progressive patients than relapsing types of MS, and progresses faster than brain atrophy.[3–5] The risk of disability progression increases with the rate of spinal cord atrophy,[7] accounting for up to 77% of motor disability, as measured by the EDSS.[3–5] In line with this, the percentage spinal cord decline measured with GBSI was associated with EDSS progression and with disease subtype, and performed better than CSA-derived spinal cord atrophy in detecting more disabled patients. Of note, differentiation between healthy controls and MS patients (especially progressive MS) revealed better performance of GBSI when compared with CSA, whilst differentiation between patients with and without EDSS progression showed similar GBSI and CSA measurements, with CSA tending to random prediction of results (AUC=0.50) and GBSI presenting with statistically higher but not particularly better predictive value (AUC=0.59). The lack of separation in our study might be due to the limited proportion of patients presenting with 3-month confirmed disability progression over 1-year observation time, and these findings should be confirmed with a longer follow-up in order to avoid false positive and negative results.

Limitations of the present study include the short follow-up duration (1.6 years in MS patients), meaning that we could not assess the association between spinal cord atrophy and long-term disability progression. Previous studies have shown that spinal cord atrophy predicts disease progression and conversion from CIS to RRMS over a long follow-up period.[39] However, we have *a priori* set our study duration at 1 year in order to obtain estimates for clinical trials and short-term observational studies; of note, we obtained annualized atrophy rates in line with pooled estimates from 94 studies (1.78%/year),[13] and the number of participants per arm was consistent with that obtained by other similar studies.[20,36,38] We segmented the spinal cord using *PropSeg* within the Spinal Cord Toolbox. The *PropSeg* tool has already been demonstrated as a robust, accurate and fast segmentation method of the spinal cord in both MS patients and controls, compared with other segmentation methods (e.g., semi-manual active surface method using JIM).[31,40,41] Also, the possibility to derive spinal cord GBSI measurements from brain scans needs to be explored.[7,19,25,42] A possible caveat of our study is that good quality images (e.g., dedicated spinal cord 3DT1 images with 1mm isotropic voxel) are needed for spinal cord atrophy measurements. We were only able to use 86% of the patients whose scans were originally collected in eight experienced imaging centres, using slight variations in MRI protocols and field strength. However, there is an initiative to mitigate the problem of setting-up a standard high-quality image protocol for spinal cord MRI (http://www.spinalcordmri.org - Protocols). In this project, a consensus acquisition protocol has been developed and tested in about 30 different sites across the world, for different vendors (GE, Siemens and Philips). This protocol is easy to apply and facilitates the adoption curve of spinal cord MRI acquisitions. Also, latest software improvements, such as machine-learning based segmentation methods (e.g., *DeepSeg* within the Spinal Cord Toolbox), could further facilitate the spinal cord image processing.

In conclusion, in the present longitudinal multicentre, multi-manufacturer and multi-field strength scan study, GBSI and CSA provided similar rates of spinal cord atrophy, but the registration-based method (GBSI) was associated with lower variability, providing smaller sample size estimates as well as a higher significance at differentiating between different MS subtypes and patients with disability progression, compared with segmentation-based method (CSA), though still far from perfect. This study provides evidence that GBSI should be considered as a precise and reliable tool for calculating MS-related spinal cord atrophy in clinical trials and in observational datasets.

ACKNOWLEDGEMENTS

The present research was conducted thanks to the 2017 ECTRIMS-MAGNIMS fellowship, awarded to Marcello Moccia, the non-clinical postdoctoral Guarantors of Brain fellowship, awarded to Ferran Prados, the German Federal Ministry for Education and Research, BMBF, German Competence Network Multiple Sclerosis (KKNMS), grant no. 01GI1601I and grant no. 01GI0914, awarded to Carsten Lukas, and the NIHR UCLH Biomedical Research Centre, supporting Olga Ciccarelli.

AUTHOR CONTRIBUTIONS

MM, FP, MF, HV, SO, CAMGWK, OC and FB contributed to conception and design of the study.

MM, FP, MF, MAR, PV, WJB, CZ, AGal, AR, AGas, JP, CL, SO and CAMGWK contributed to acquisition and analysis of data.

MM, FP, HV, CAMGWK, OC and FB contributed to drafting a significant portion of the manuscript or figures.

Members of the MAGNIMS Study Group Steering Committee are reported in the **Supplementary Material 2**.

POTENTIAL CONFLICTS OF INTEREST

Authors have nothing to report related to this work.

REFERENCES

1. Brex P, Leary S, O’Riordan J, et al. Measurement of spinal cord area in clinically isolated syndromes suggestive of multiple sclerosis. J Neurol Neurosurg Psychiatry. 2001;70(4):544–547.
2. Gass A, Rocca MA, Agosta F, et al. MRI monitoring of pathological changes in the spinal cord in patients with multiple sclerosis. Lancet Neurol. 2015;14(4):443–454.
3. Brownlee WJ, Altmann DR, Da Mota PA, et al. Association of asymptomatic spinal cord lesions and atrophy with disability 5 years after a clinically isolated syndrome. Mult. Scler. 2017;23(5):665–674.
4. Kearney H, Schneider T, Yiannakas MC, et al. Spinal cord grey matter abnormalities are associated with secondary progression and Physical disability in multiple sclerosis. J. Neurol. Neurosurg. Psychiatry 2015;86(6):608–614.
5. Kearney H, Rocca M, Valsasina P, et al. Magnetic resonance imaging correlates of physical disability in relapse onset multiple sclerosis of long disease duration. Mult. Scler. 2014;20(1):72–80.
6. Moccia M, Ruggieri S, Ianniello A, et al. Advances in spinal cord imaging in multiple sclerosis. Ther. Adv. Neurol. Disord. 2019; 12:1756286419840593.
7. Tsagkas C, Magon S, Gaetano L, et al. Spinal cord volume loss. A marker of disease progression in multiple sclerosis. Neurology 2018;91(4):e349–e358.
8. Tsagkas C, Magon S, Gaetano L, et al. Preferential spinal cord volume loss in primary progressive multiple sclerosis. Mult. Scler. 2018;1352458518775006.
9. Ciccarelli O, Cohen J, Reingold S, et al. Spinal Cord Involvement in Multiple Sclerosis and Neuromyelitis Optica Spectrum Disorders. Lancet Neurol. 2019;18(2):185–197.
10. Lukas C, Sombekke M, Bellenberg B, et al. Relevance of spinal cord abnormalities to clinical

disability in multiple sclerosis: MR imaging findings in a large cohort of patients. Radiology 2013;269(2):542–552.

11. Lukas C, Knol DDL, Sombekke MMH, et al. Cervical spinal cord volume loss is related to clinical disability progression in multiple sclerosis. J Neurol Neurosurg Psychiatry. 2015;86(4):410–418.

12. Bonati U, Fisniku LK, Altmann DR, et al. Cervical cord and brain grey matter atrophy independently associate with long-term MS disability. J. Neurol. Neurosurg. Psychiatry 2011;82(4):471–472.

13. Casserly C, Seyman EE, Alcaide-Leon P, et al. Spinal Cord Atrophy in Multiple Sclerosis: A Systematic Review and Meta-Analysis. J. Neuroimaging 2018;28(6):556–586.

14. Zaratin P, Comi G, Coetzee T, et al. Progressive MS Alliance Industry Forum: Maximizing Collective Impact To Enable Drug Development. Trends Pharmacol. Sci. 2016;37(10):808–810.

15. Cawley N, Tur C, Prados F, et al. Spinal cord atrophy as a primary outcome measure in phase II trials of progressive multiple sclerosis. Mult. Scler. 2018;24(7):932–941.

16. Moccia M, de Stefano N, Barkhof F. Imaging outcomes measures for progressive multiple sclerosis trials. Mult. Scler. 2017;23(12):1614–1626.

17. Ziegler G, Grabher P, Thompson A, et al. Progressive neurodegeneration following spinal cord injury. Neurology 2018;90(14):e1257–e1266.

18. Kearney H, Yiannakas M, Abdel-Aziz K, et al. Improved MRI quantification of spinal cord atrophy in multiple sclerosis. J. Magn. Reson. Imaging 2014;39(3):617–623.

19. Prados F, Barkhof F. Spinal cord atrophy rates. Ready for prime time in multiple sclerosis clinical trials? Neurology 2018;91(4):157–158.

20. Altmann DR, Jasperse B, Barkhof F, et al. Sample sizes for brain atrophy outcomes in trials for

secondary progressive multiple sclerosis. Neurology 2009;72(7):595–601.

21. Freeborough P, Fox N. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. IEEE Trans. Med. Imaging 1997;16(5):623–629.

22. Prados F, Cardoso MJ, Leung KK, et al. Measuring brain atrophy with a generalized formulation of the boundary shift integral. Neurobiol Aging. 2015;36(S1):S81–S90.

23. Prados F, Yiannakas M, Cardoso M, et al. Computing spinal cord atrophy using the Boundary Shift Integral: a more powerful outcome measure for clinical trials? ECTRIMS 2016;

24. Kearney H, Miszkiel KA, Yiannakas MC, et al. A pilot MRI study of white and grey matter involvement by multiple sclerosis spinal cord lesions. Mult. Scler. Relat. Disord. 2013;2(2):103–108.

25. Lukas C, Prados F, Valsasina P, et al. Quantification of spinal cord atrophy in MS: which software, which vertebral level, spinal cord or brain MRI? A multi-centric, longitudinal comparison of three different volumetric approaches. Mult. Scler. 2018;24(Suppl 2):88–89.

26. Rocca M, Valsasina P, Meani A, et al. Cranio-caudal patterns of cervical cord atrophy progression in MS according to disease phenotype and clinical worsening: a multicenter study. Mult. Scler. 2018;24(Suppl 2):102–103.

27. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. Ann Neurol. 2011;69:292–302.

28. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983;33:1444–1452.

29. De Leener B, Lévy S, Dupont SM, et al. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. Neuroimage 2017;145:24–43.

30. Prados F, Yiannakas M, Cardoso M, et al. Atrophy computation in the spinal cord using the

Boundary Shift Integral. ISMRM 2016.

31. Yiannakas MC, Mustafa AM, De Leener B, et al. Fully automated segmentation of the cervical cord from T1-weighted MRI using PropSeg: Application to multiple sclerosis. NeuroImage Clin. 2015;10:71–77.

32. Yaldizli Ö, MacManus D, Stutters J, et al. Brain and cervical spinal cord atrophy in primary progressive multiple sclerosis: results from a placebo-controlled phase III trial (INFORMS). Mult. Scler. 2015;22(S11):30–31.

33. Tur C, Moccia M, Barkhof F, et al. Assessing treatment outcomes in multiple sclerosis trials and in the clinical setting. Nat. Rev. Neurol. 2018;14(2):75–93.

34. Smith SM, Zhang Y, Jenkinson M, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage 2002;17(1):479–89.

35. Leung K, Clarkson M, Bartlett J, et al. Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection. Neuroimage 2010;50(2):516–523.

36. Healy B, Valsasina P, Filippi M, Bakshi R. Sample size requirements for treatment effects using gray matter, white matter and whole brain volume in relapsing-remitting multiple sclerosis. J Neurol Neurosurg Psychiatry. 2009;80(11):1218–23.

37. Van Den Elskamp IJ, Boden B, Dattola V, et al. Cerebral atrophy as outcome measure in short-term phase 2 clinical trials in multiple sclerosis. Neuroradiology 2010;52(10):875–881.

38. Anderson VM, Fernando KTM, Davies GR, et al. Cerebral atrophy measurement in clinically isolated syndromes and relapsing remitting multiple sclerosis: A comparison of registration-based methods. J. Neuroimaging 2007;17(1):61–68.

39. Aymerich FX, Auger C, Alonso J, et al. Cervical Cord Atrophy and Long-Term Disease Progression in Patients with Primary-Progressive Multiple Sclerosis. AJNR. Am. J. Neuroradiol.

2018;39(2):399–404.

40. De Leener B, Cohen-Adad J, Kadoury S. Automatic Segmentation of the Spinal Cord and Spinal Canal Coupled with Vertebral Labeling. IEEE Trans. Med. Imaging 2015;34(8):1705–1718.

41. De Leener B, Kadoury S, Cohen-Adad J. Robust, accurate and fast automatic segmentation of the spinal cord. Neuroimage 2014;98:528–536.

42. Liu XY, Lukas XC, Steenwijk XMD, et al. Multicenter Validation of Mean Upper Cervical Cord Area Measurements from Head 3D T1-Weighted MR Imaging in Patients with Multiple Sclerosis. AJNR. Am. J. Neuroradiol. 2016;37:749–754.

43. McCoy DB, Dupont SM, Gros C, et al. Convolutional neural network–based automated segmentation of the spinal cord and contusion injury: Deep learning biomarker correlates of motor impairment in acute spinal cord injury. AJNR Am. J. Neuroradiol. 2019;40(4):737–744.

TABLE 1. Demographic and clinical features.

Table shows demographic and clinical features of MS and controls. P-values are shown from *t*-test, Mann–Whitney test, χ^2 test or Fisher's exact test, as appropriate (*p<0.05).

		MS *(n=282)*	**Controls** *(n=82)*	*p-values*
Age, *years*		38.2 ± 11.2	36.6 ± 12.5	*0.339*
Sex, *female*		169 (59.8%)	46 (55.7%)	*0.590*
Interval between scans, *years*		1.6 ± 1.1	1.2 ± 0.7	*0.001**
Disease duration, *years*		5.9±8.2		
Disease subtype	*CIS*	52 (18.4%)		
	RRMS	196 (69.5%)		
	PMS	34 (12.1%)		
EDSS at baseline		1.5 (0-7.5)		
EDSS at follow-up		2.0 (0-8.0)		
Patients with EDSS progression		74 (26.2%)		

FIG 1. Spinal cord segmentation and GBSI.

Spinal Cord Toolbox was used for spinal cord segmentation. C2-3, C3-4, and C4-5 reference points were set manually (**A**). Representative images of semi-automatic spinal cord segmentation output are shown (sagittal and, in the inset, axial views) (**B**). Afterwards, baseline (**C**/**D**) and follow-up (**E**/**F)** spinal cord images were straightened, and, ultimately, registered to the halfway space. Intensity changes in the vicinity of the cord boundaries were estimated for generalized boundary shift integral (GBSI) calculation (**G**/**H**).

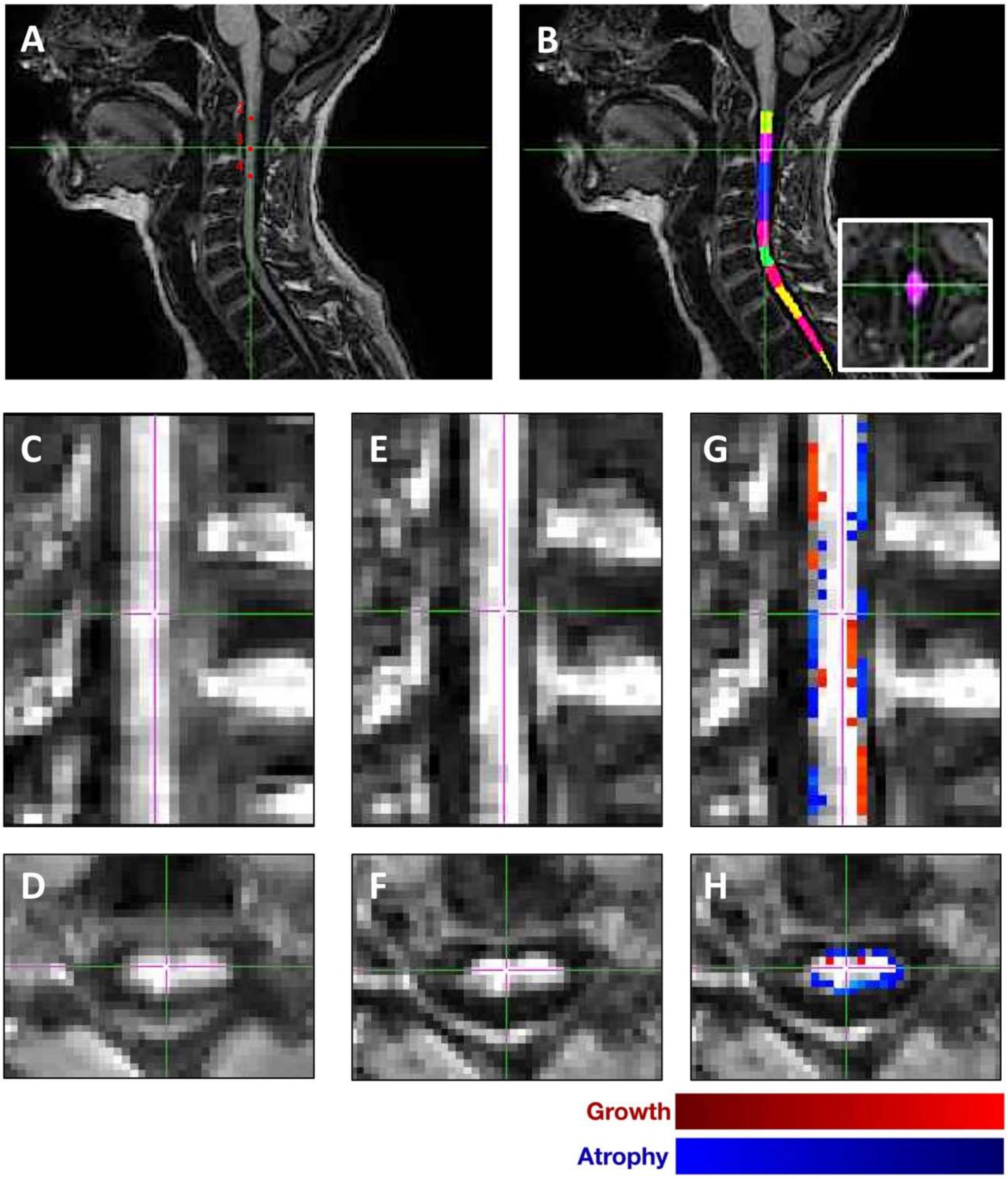


FIG 2. Study flow diagram.

Figure shows MS cases and controls from MAGNIMS and UCL Queen Square Institute of Neurology cohorts; scanner filed strength is reported. Exclusion rate from the original cohort is shown, as a consequence of poor contrast, wrong voxel size, wrong acquisition parameters, and artifacts, mostly present in the eldest cohorts acquired using 1.5T scanners (e.g., CIS cohort). MAGNIMS: Magnetic Resonance Imaging in Multiple Sclerosis; QS: Queen Square; CIS: clinically isolated syndrome; SC: spinal cord; RRMS: relapsing-remitting multiple sclerosis; PMS: progressive multiple sclerosis; GBSI: generalized boundary shift integral.

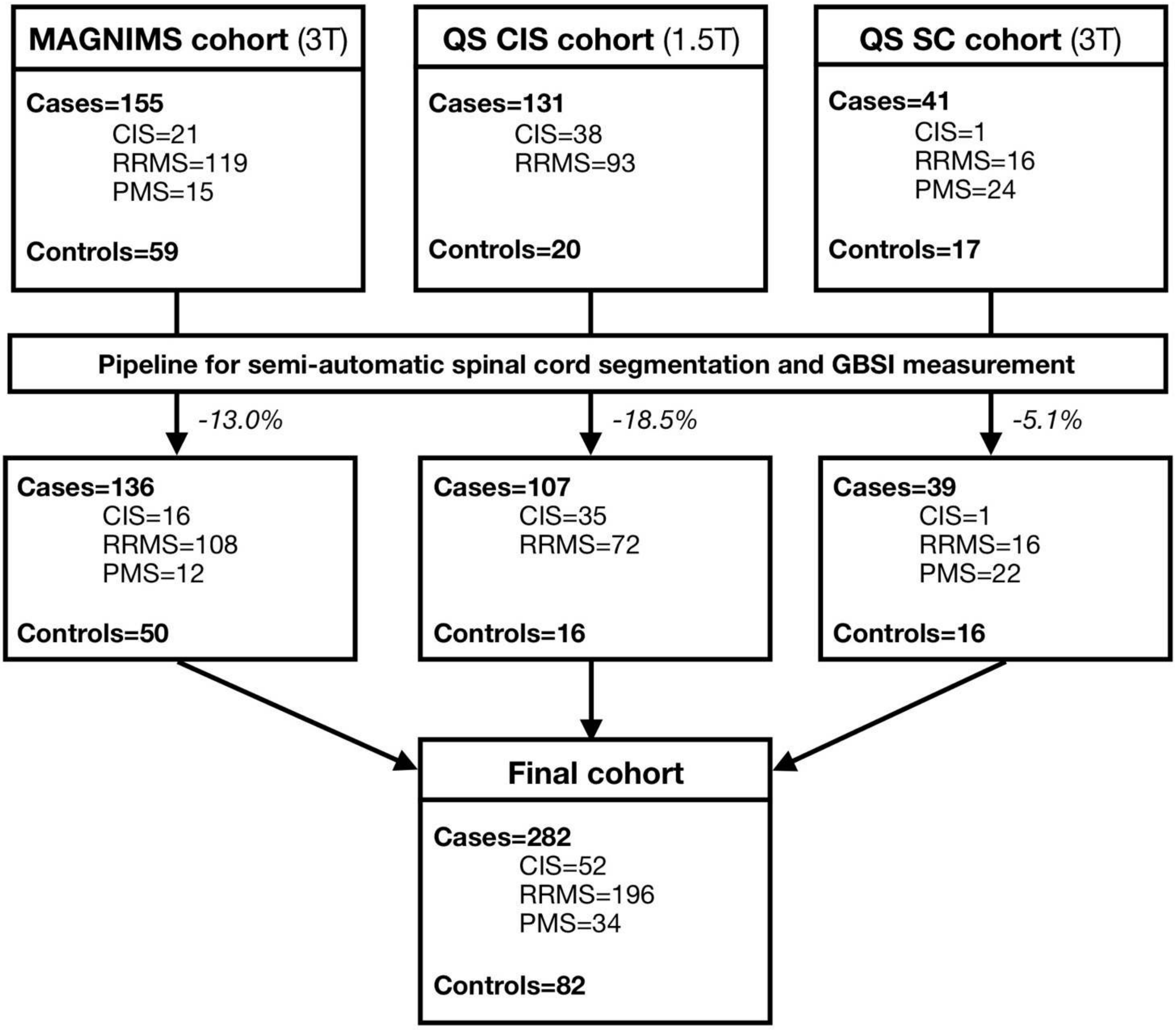


FIG 3. Box-and-Whisker plot of CSA and GBSI measurements

Box-and-Whisker plot for 1-year percentage change with cross-sectional spinal cord area (CSA) (**A**) and generalized boundary shift integral (GBSI) (**B**) in healthy controls, clinically isolated syndrome (CIS), relapsing-remitting multiple sclerosis (RRMS) and progressive multiple sclerosis (PMS). Coefficients (Coeff) and p-values are reported from linear regression models using healthy controls as reference group (age, sex, disease duration and site of MRI acquisition were used as covariates). Standard deviation is also reported for each group.

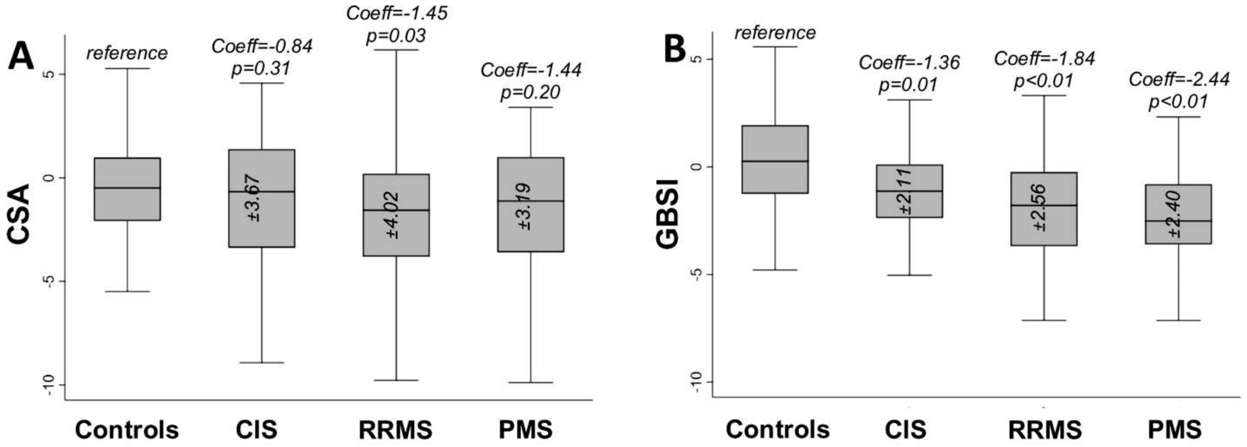
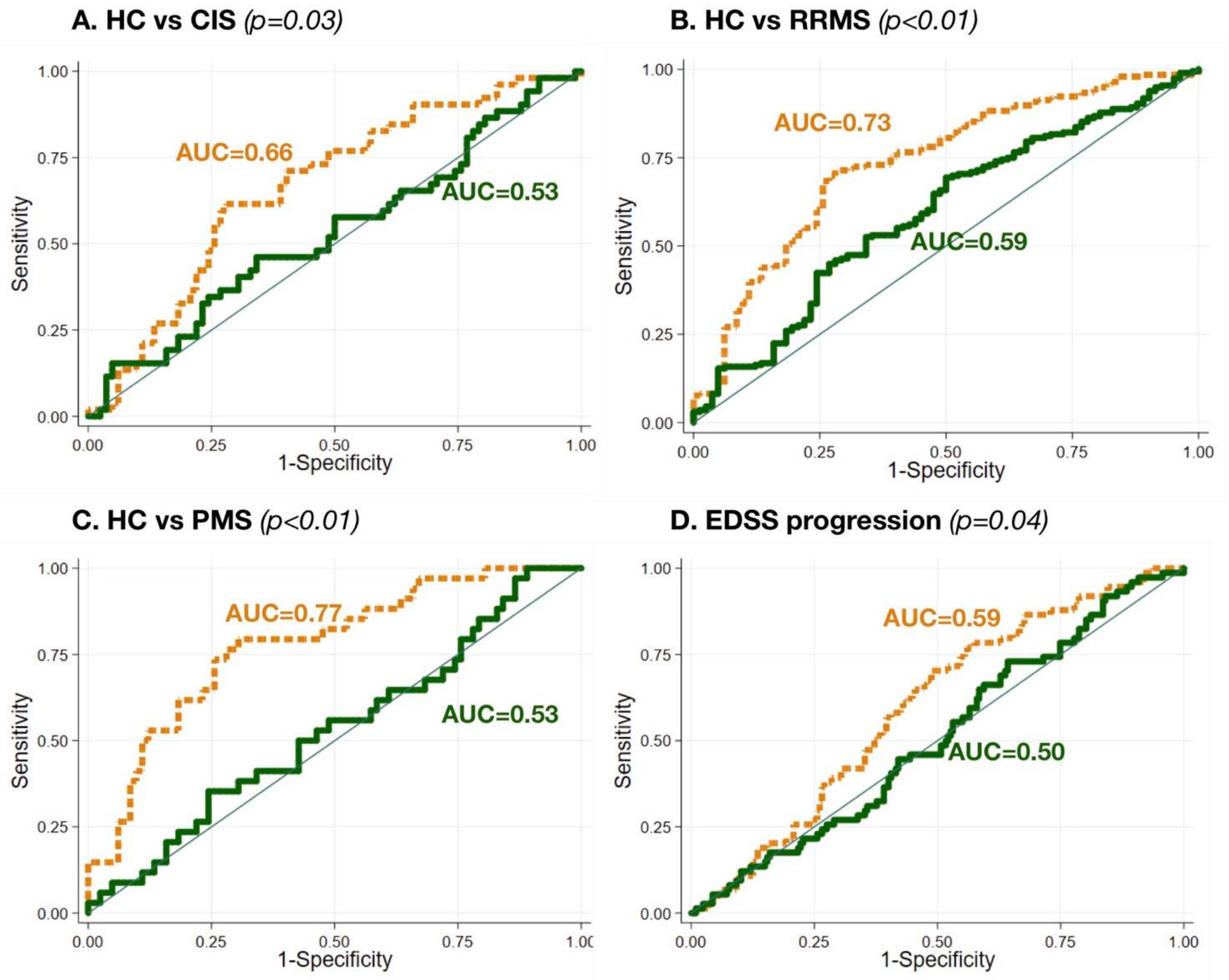


FIG 4. ROC curves for CSA and GBSI in relation to clinical variables.

ROC curves for cross-sectional spinal cord area (CSA) (dashed line) and generalized boundary shift integral (GBSI) (solid line) in relation to differentiating healthy controls (n=82) from clinically isolated syndrome (CIS) (n=52) (**A**), relapsing-remitting multiple sclerosis (RRMS) (n=196) (**B**), and progressive multiple sclerosis (PMS) (n=34) (**C**), and patients with expanded disability status scale (EDSS) progression (n=74), from those without EDSS progression (n=208) (**D**). Area under the curve (AUC) and p-value are reported.



A. HC vs CIS *(p=0.03)*

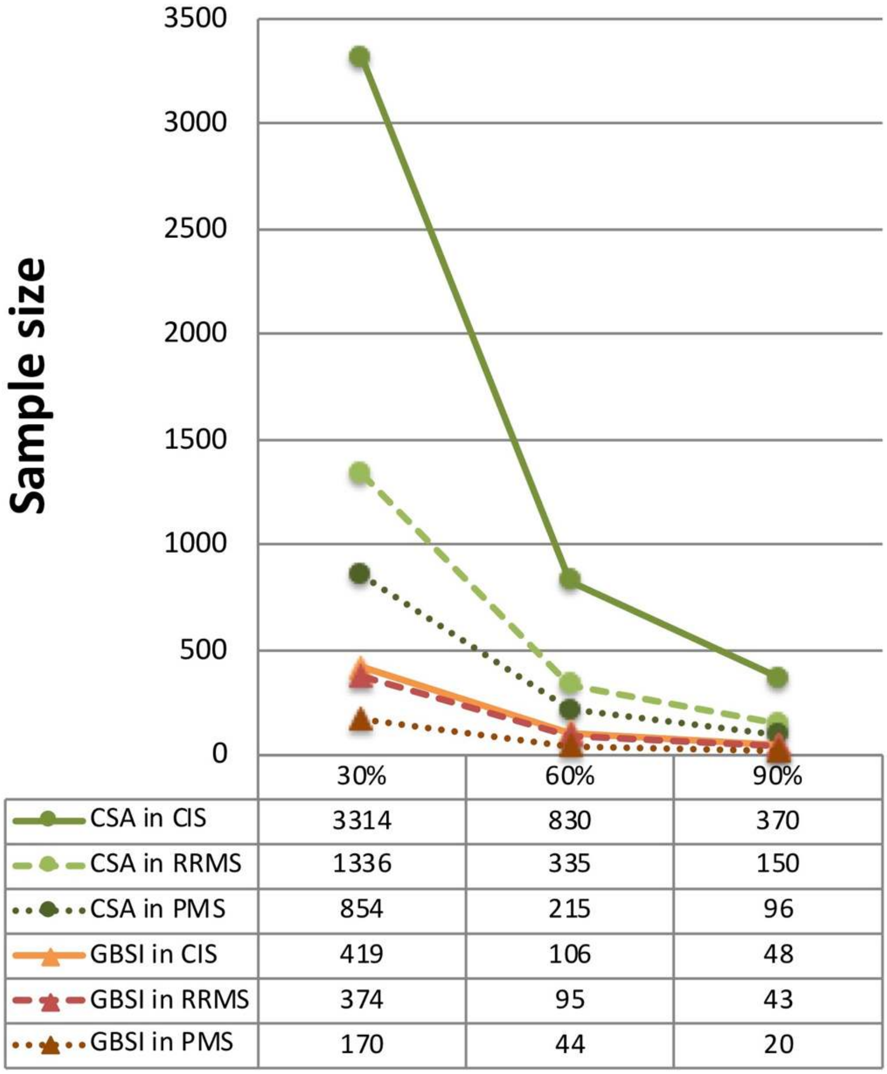
B. HC vs RRMS *(p<0.01)*

C. HC vs PMS *(p<0.01)*

D. EDSS progression *(p=0.04)*

FIG 5. Sample size estimates for CSA and GBSI.

Profile plot shows sample size estimates for cross-sectional spinal cord area (CSA) and generalized boundary shift integral (GBSI) in different disease phenotypes (clinically isolated syndrome (CIS), relapsing-remitting multiple sclerosis (RRMS), and progressive multiple sclerosis (PMS)). For sample size calculation, we included adjusted beta-coefficients from linear regression models, estimating spinal cord loss in MS patients, when compared with physiologic loss in controls, and standard deviation from each disease phenotype. Power was set at 80% and alpha at 5%. Different treatment effects were hypothesized (e.g., 30%, 60% and 90%), that were smaller than the observed difference between MS cases and physiologic spinal cord loss in controls.



	30%	60%	90%
CSA in CIS	3314	830	370
CSA in RRMS	1336	335	150
CSA in PMS	854	215	96
GBSI in CIS	419	106	48
GBSI in RRMS	374	95	43
GBSI in PMS	170	44	20