**Longitudinal spinal cord atrophy in multiple sclerosis using the generalised boundary shift integral on multicentre and multi-field strength setting**

*Marcello Moccia1-2, MD, PhD; Ferran Prados1,3-5, PhD; Massimo Filippi6,7, MD, PhD; Maria A Rocca6,7, MD, PhD; Paola Valsasina6, MD, PhD; Wallace J Brownlee1, MD, PhD; Chiara Zecca8, MD; Antonio Gallo9, MD, PhD; Alex Rovira10, MD; Achim Gass11, MD; Jacqueline Palace12, MD, PhD; Carsten Lukas13, MD, PhD; Hugo Vrenken14, PhD; Sebastien Ourselin15, PhD; Claudia A.M. Gandini Wheeler‐Kingshott1,16-17, PhD; Olga Ciccarelli1,4, MD, PhD; Frederik Barkhof1,3-4,14 MD, PhD; the MAGNIMS Study Group.*

1. Queen Square MS Centre, Department of Neuroinflammation, UCL Queen Square Institute of Neurology, Faculty of Brain Sciences, University College London, London, United Kingdom
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, University College London, London, United Kingdom
4. National Institute for Health Research (NIHR), University College London Hospitals (UCLH) Biomedical Research Centre, United Kingdom
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.

**ABSTRACT**

*Background*. 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. We compared the generalised boundary shift integral (GBSI), a registration-based method recently translated from the brain to the cord, with a standard segmentation-based method.

*Methods*. Baseline and 1-year spinal cord 3DT1-weighted images (1mm isotropic) were obtained from 364 individuals (52 clinically isolated syndrome, 196 relapsing-remitting MS (RRMS), 34 progressive MS (PMS), and 82 healthy controls) from eight MAGNIMS sites, on multi-manufacturer and multi-field strength scans. Spinal Cord Toolbox was used for C2-5 segmentation and subsequent cross-sectional area (CSA) calculation. The GBSI pipeline included cord straightening and registration, and atrophy measurement was based on the probabilistic boundary-shift region-of-interest. CSA and GBSI percent annual volume change was calculated and compared between groups.

*Results*. GBSI provided overall similar rates of atrophy, but reduced measurement variability when compared with 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 disability 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 compared to CSA in all MS subtypes (CIS: 106 vs. 830; RRMS: 95 vs. 335; PMS: 44 vs. 215) (power=80%; alpha=5%).

*Conclusions*. GBSI provided more precise spinal cord annual atrophy rates than CSA, leading to increased group separation and improved statistical power.

**ABSTRACT WORD COUNT**: 273.

**WORD COUNT**: 2791.

**TABLES/FIGURES/SUPPLEMENTARY MATERIAL**: 2/5/1.

**BACKGROUND**

Spinal cord atrophy on MRI is a marker of neurodegeneration in multiple sclerosis (MS) (Brex *et al.*, 2001; Gass *et al.*, 2015), and is one of the main substrates of long-term disease progression (Brownlee *et al.*, 2017; Ciccarelli *et al.*, 2019; Kearney *et al.*, 2015; Kearney, Rocca, *et al.*, 2014; Lukas *et al.*, 2013, 2015; Moccia *et al.*, 2019; Tsagkas *et al.*, 2018a, 2018b). 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 (Bonati *et al.*, 2011; Casserly *et al.*, 2018; Kearney, Rocca, *et al.*, 2014). 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 (Cawley *et al.*, 2018; Moccia *et al.*, 2017; Zaratin *et al.*, 2016; Ziegler *et al.*, 2018).

Spinal cord atrophy is conventionally estimated with segmentation-based methods (e.g., cervical cord cross-sectional area (CSA)), applied to volumetric spinal cord images (Kearney, Yiannakas, *et al.*, 2014), 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 (Moccia *et al.*, 2019; Prados and Barkhof, 2018). On the contrary, longitudinal brain atrophy measurements are nowadays based on registration-based techniques that significantly reduce measurement noise (Altmann *et al.*, 2009).

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 (Freeborough and Fox, 1997; Prados *et al.*, 2015, 2016a). 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 (University College London, London, United Kingdom) and 7 additional 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 (Brownlee *et al.*, 2017; Kearney *et al.*, 2013; Kearney, Yiannakas, *et al.*, 2014; Lukas *et al.*, 2018; Rocca *et al.*, 2018).

Eligibility criteria were: (i) diagnosis of clinically isolated syndrome or MS according to the 2010 McDonald Criteria (Polman *et al.*, 2011); (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 1x1x1mm3; (iv) information on: Expanded Disability Status Scale (EDSS) score at each time point (Kurtzke, 1983); subtype of MS (CIS, RRMS and PMS (including both primary and secondary progressive MS)) (Polman *et al.*, 2011); 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 (1x1x1mm3) 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) (**Figure 2**) (De Leener *et al.*, 2017). C2-5 CSA was obtained by averaging all cord cross-sectional areas. For GBSI, we followed the previously described steps (Prados *et al.*, 2015, 2016b, 2016a). In summary, 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 (**Figure 2**).

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.

**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, and different binary clinical variables (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. Of note, CSA and GBSI measured 2D area and 3D volume change, respectively; however, we restricted the region of analysis to C2-5 segments and used CSA masks to compute the GBSI. Sample size was computed using the formula , where *n* is the required sample size per treatment arm in 1:1 controlled trials, *Zα* and *Z1-β* are constant (set at 5% alpha-error and 80% power, respectively), *σ* is the standard deviation (from each disease phenotype), and *Δ* the estimated effect size (Altmann *et al.*, 2009; Cawley *et al.*, 2018). Effect size was derived from adjusted beta-coefficients at linear regression models. Different treatment effects were hypothesized (e.g., 30%, 60% and 90%). 100% treatment effect is theoretically reached when the spinal cord atrophy change in patients is equal to that observed in healthy controls.

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

**RESULTS**

Longitudinal spinal cord scans from 327 MS patients and 96 healthy controls were collected, but 45 patients’ scans and 14 healthy controls’ scans were excluded because of poor contrast, wrong voxel size, wrong acquisition parameters, and artifacts, mostly present in the eldest cohorts acquired using 1.5 T scanner (e.g., CIS cohort) (see **Figure 1** 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), (**Figure 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) (**Figure 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; **Figure 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; **Figure 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; **Figure 4C**). Patients with EDSS progression were better differentiated from those without EDSS progression 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; **Figure 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, adjusted for age, sex site of acquisition and disease duration) were lower for GBSI compared to 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 (**Figure 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 (Moccia *et al.*, 2017; Prados and Barkhof, 2018; Tur *et al.*, 2018; Yaldizli *et al.*, 2015). 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. Of note, a higher percentage of spinal cord decline (and a lower standard deviation) in PMS was found 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 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 (Freeborough and Fox, 1997; Leung *et al.*, 2010; Prados *et al.*, 2015).

We found that GBSI allowed a four-fold smaller (and therefore achievable) sample size than that obtained with CSA. 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 (Altmann *et al.*, 2009; Anderson *et al.*, 2007; Van Den Elskamp *et al.*, 2010; Healy *et al.*, 2009). However, monitoring spinal cord atrophy could be more clinically relevant than brain atrophy, due to its robust clinical correlates (Brownlee *et al.*, 2017; Kearney *et al.*, 2015; Kearney, Rocca, *et al.*, 2014). 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 (Brownlee *et al.*, 2017; Kearney *et al.*, 2015; Kearney, Rocca, *et al.*, 2014). The risk of disability progression increases with the rate of spinal cord atrophy (Tsagkas *et al.*, 2018a), accounting for up to 77% of motor disability, as measured by the EDSS (Brownlee *et al.*, 2017; Kearney *et al.*, 2015; Kearney, Rocca, *et al.*, 2014). 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.

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 longer-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 (Aymerich *et al.*, 2018). 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) (Casserly *et al.*, 2018), and the number of participants per arm was consistent with that obtained by other similar studies (Altmann *et al.*, 2009; Anderson *et al.*, 2007; Healy *et al.*, 2009). A possible caveat of our study is that we were only able to use 86% of the patients whose scans were originally collected, even though performed in experienced imaging centres, suggesting that good quality images (e.g., dedicated spinal cord 3DT1 images with 1mm isotropic voxel) are needed for spinal cord measurements. However, the strength of this study is the performance of this measure when including data from 8 different sites using varied protocols and field strength. How and whether these results translate to real-world clinical practice needs to be further tested. Also, the possibility to derive spinal cord GBSI measurements from brain scans needs to be explored (Liu *et al.*, 2016; Lukas *et al.*, 2018; Prados and Barkhof, 2018; Tsagkas *et al.*, 2018a).

In conclusion, in the present longitudinal multicentre, multi-manufacturer and multi-field strength scan study, measurements of spinal cord atrophy with GBSI were similar to those obtained with CSA, but were 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. 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**

**MAGNIMS Study Group Steering Committee Members are Christian Enzinger** (*co-chair*) (Department of Neurology, Medical University of Graz, Graz, Austria), **Maria A Rocca** (*co-chair*) (Neuroimaging Research Unit, Institute of Experimental Neurology, Division of Neuroscience, San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy), **Frederik Barkhof** (Department of Radiology and Nuclear Medicine, VU University Medical Center, Amsterdam, The Netherlands), **Olga Ciccarelli** (Queen Square MS Centre, Department of Neuroinflammation, UCL Queen Square Institute of Neurology, Faculty of Brain Sciences, University College London, London, United Kingdom), **Nicola De Stefano** (Department of Medicine, Surgery and Neuroscience, University of Siena, Siena, Italy), **Massimo Filippi** (Neuroimaging Research Unit, Institute of Experimental Neurology, Division of Neuroscience, San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy), **Jette Frederiksen** (The MS Clinic, Department of Neurology, University of Copenhagen, Glostrup Hospital, Denmark), **Claudio Gasperini** (Department of Neurology and Psychiatry, University of Rome Sapienza, Rome, Italy), **Ludwig Kappos** (Department of Neurology, University Hospital, Kantonsspital, Basel, Switzerland), **Jaqueline Palace** (Centre for Functional Magnetic Resonance Imaging of the Brain, Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, Oxford, United Kingdom), **Alex Rovira** (MR Unit and Section of Neuroradiology, Department of Radiology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain), **Jaume Sastre-Garriga** (Department of Neurology/Neuroimmunology, Multiple Sclerosis Centre of Catalonia (CEMCAT), Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain), **Hugo Vrenken** (Department of Radiology and Nuclear Medicine, VU University Medical Center, Amsterdam, The Netherlands), **Tarek Yousry** (NMR Research Unit, Department of Neuroinflammation, UCL Queen Square Institute of Neurology, Faculty of Brain Sciences, University College London, London, United Kingdom).

**FUNDING AND CONFLICT OF INTEREST**

Marcello Moccia has received the 2017 ECTRIMS-MAGNIMS fellowship, and honoraria for speaking and/or consulting activities from Biogen, Merck and Sanofi-Genzyme.

Ferran Prados has a non-clinical postdoctoral Guarantors of Brain fellowship.

Massimo Filippi is Editor-in-Chief of the Journal of Neurology; has received honoraria for speaking and/or consulting activities from Biogen Idec, Merck-Serono, Novartis, Teva Pharmaceutical Industries; and has received research support from Biogen Idec, Merck-Serono, Novartis, Teva Pharmaceutical Industries, Roche, Italian Ministry of Health, Fondazione Italiana Sclerosi Multipla, and ARiSLA (Fondazione Italiana di Ricerca per la SLA).

Mara A Rocca has received honoraria for speaking from Biogen Idec, Novartis, Genzyme, Sanofi-Aventis, Teva, Merck Serono, and Roche and has received research support from the Italian Ministry of Health and Fondazione Italiana Sclerosi Multipla.

Paola Valsasina has received honoraria for speaking from ExceMed.

Chiara Zecca has received honoraria for speaking and/or consulting activities for Almirall, Abbvie, Biogen Idec, Merck Serono, Novartis, Roche, Sanofi and Teva.

Antonio Gallo received honoraria for speaking and/or consulting activities from Biogen Idec, Merck Serono, Novartis, Roche, Sanofi, Teva and Mylan.

Jacqueline Palace has received support for scientific meetings and honoraria for consulting activities from Merck Serono, Biogen Idec, Novartis, Teva, Chugai Pharma and Bayer Schering, Alexion, Roche, Genzyme, MedImmune, EuroImmun, MedDay, Abide and ARGENX; has received research support from Merck Serono, Novartis, Biogen Idec, Teva, Abide and Bayer Schering; and has received research support from the MS society, Guthie Jackson Foundation, NIHR, Oxford Health Services Research Committee, EDEN, MRC, GMSI, and John Fell.

Hugo Vrenken has received research grants from MerckSerono, Novartis, Teva and Pfizer; has received honoraria for speaking from Novartis; and has received honoraria for consulting activities from Merck Serono.

Carsten Lukas has received an endowed professorship supported by the Novartis foundation; has received honoraria for speaking and/or consulting activities from Biogen-Idec, Bayer Schering, Novartis, Sanofi, Genzyme and TEVA; and has received research support from Merck-Serono and Novartis. The part of the study from Bochum was supported by the German Federal Ministry for Education and Research, BMBF, German Competence Network Multiple Sclerosis (KKNMS), grant no. 01GI1601I and grant no. 01GI0914.

Olga Ciccarelli is supported by the NIHR UCLH Biomedical Research Centre; has received honoraria for speaking and/or consulting activities from Novartis, Roche, Merck, Teva and Biogen.

Frederik Barkhof is supported by the NIHR UCLH Biomedical Research Centre.

**REFERENCES**

Altmann DR, Jasperse B, Barkhof F, Beckmann K, Filippi M, Kappos LD, et al. Sample sizes for brain atrophy outcomes in trials for secondary progressive multiple sclerosis. Neurology 2009; 72: 595–601.

Anderson VM, Fernando KTM, Davies GR, Rashid W, Frost C, Fox NC, et al. Cerebral atrophy measurement in clinically isolated syndromes and relapsing remitting multiple sclerosis: A comparison of registration-based methods. J. Neuroimaging 2007; 17: 61–68.

Aymerich FX, Auger C, Alonso J, Alberich M, Sastre-Garriga J, Tintoré M, et al. Cervical Cord Atrophy and Long-Term Disease Progression in Patients with Primary-Progressive Multiple Sclerosis. AJNR. Am. J. Neuroradiol. 2018; 39: 399–404.

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

Brex P, Leary S, O’Riordan J, Miszkiel K, Plant G, Thompson A, et al. Measurement of spinal cord area in clinically isolated syndromes suggestive of multiple sclerosis. J Neurol Neurosurg Psychiatry. 2001; 70: 544–547.

Brownlee WJ, Altmann DR, Da Mota PA, Swanton JK, Miszkiel KA, Gandini Wheeler-Kingshott CAM, et al. Association of asymptomatic spinal cord lesions and atrophy with disability 5 years after a clinically isolated syndrome. Mult. Scler. 2017; 23: 665–674.

Casserly C, Seyman EE, Alcaide-Leon P, Guenette M, Lyons C, Sankar S, et al. Spinal Cord Atrophy in Multiple Sclerosis: A Systematic Review and Meta-Analysis. J. Neuroimaging 2018; 28: 556–586.

Cawley N, Tur C, Prados F, Plantone D, Kearney H, Abdel-Aziz K, et al. Spinal cord atrophy as a primary outcome measure in phase II trials of progressive multiple sclerosis. Mult. Scler. 2018; 24: 932–941.

Ciccarelli O, Cohen J, Reingold S, Weishenker B, International Conference on Spinal Cord Involvement and Imaging in Multiple Sclerosis and Neuromyelitis Optica Spectrum Disorders. Spinal Cord Involvement in Multiple Sclerosis and Neuromyelitis Optica Spectrum Disorders. Lancet Neurol. 2019; 18: 185–197.

Van Den Elskamp IJ, Boden B, Dattola V, Knol DL, Filippi M, Kappos L, et al. Cerebral atrophy as outcome measure in short-term phase 2 clinical trials in multiple sclerosis. Neuroradiology 2010; 52: 875–881.

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: 623–629.

Gass A, Rocca MA, Agosta F, Ciccarelli O, Chard D, Valsasina P, et al. MRI monitoring of pathological changes in the spinal cord in patients with multiple sclerosis. Lancet Neurol. 2015; 14: 443–454.

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: 1218–23.

Kearney H, Miszkiel KA, Yiannakas MC, Ciccarelli O, Miller DH. A pilot MRI study of white and grey matter involvement by multiple sclerosis spinal cord lesions. Mult. Scler. Relat. Disord. 2013; 2: 103–108.

Kearney H, Rocca M, Valsasina P, Balk L, Sastre-Garriga J, Reinhardt J, et al. Magnetic resonance imaging correlates of physical disability in relapse onset multiple sclerosis of long disease duration. Mult. Scler. 2014; 20: 72–80.

Kearney H, Schneider T, Yiannakas MC, Altmann DR, Wheeler-Kingshott CAMM, Ciccarelli O, et al. Spinal cord grey matter abnormalities are associated with secondary progression and Physical disability in multiple sclerosis. J. Neurol. Neurosurg. Psychiatry 2015; 86: 608–614.

Kearney H, Yiannakas M, Abdel-Aziz K, Wheeler-Kingshott C, Altmann D, Ciccarelli O, et al. Improved MRI quantification of spinal cord atrophy in multiple sclerosis. J. Magn. Reson. Imaging 2014; 39: 617–623.

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

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

Leung K, Clarkson M, Bartlett J, Clegg S, Jack CJ, Weiner M, 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: 516–523.

Liu XY, Lukas XC, Steenwijk XMD, Daams XM, Versteeg XA, Duan XY, 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.

Lukas C, Knol DDL, Sombekke MMH, Bellenberg B, Hahn HHK, Popescu V, et al. Cervical spinal cord volume loss is related to clinical disability progression in multiple sclerosis. J Neurol Neurosurg Psychiatry. 2015; 86: 410–418.

Lukas C, Prados F, Valsasina P, Parmar K, Brouwer I, Bellenberg B, 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: 88–89.

Lukas C, Sombekke M, Bellenberg B, Hahn H, Popescu V, Bendfeldt K, 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: 542–552.

Moccia M, Ruggieri S, Ianniello A, Toosy A, Pozzilli C, Ciccarelli O. Advances in spinal cord imaging in multiple sclerosis. Ther. Adv. Neurol. Disord. 2019

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

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

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

Prados F, Cardoso MJ, Leung KK, Cash DM, Modat M, Fox NC, et al. Measuring brain atrophy with a generalized formulation of the boundary shift integral. Neurobiol Aging. 2015; 36: S81–S90.

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

Prados F, Yiannakas M, Cardoso M, Grussu F, De Angelis F, Plantone D, et al. Atrophy computation in the spinal cord using the Boundary Shift Integral. ISMRM 2016b

Rocca M, Valsasina P, Meani A, Gobbi C, Zecca C, Rovira 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: 102–103.

Tsagkas C, Magon S, Gaetano L, Pezold S, Naegelin Y, Amann M, et al. Spinal cord volume loss. A marker of disease progression in multiple sclerosis. Neurology 2018a; 91: e349–e358.

Tsagkas C, Magon S, Gaetano L, Pezold S, Naegelin Y, Amann M, et al. Preferential spinal cord volume loss in primary progressive multiple sclerosis. Mult. Scler. 2018b: 1352458518775006.

Tur C, Moccia M, Barkhof F, Chataway J, Sastre-Garriga J, Thompson AJ, et al. Assessing treatment outcomes in multiple sclerosis trials and in the clinical setting. Nat. Rev. Neurol. 2018; 14: 75–93.

Yaldizli Ö, MacManus D, Stutters J, Häring D, Lublin F, Freedman M, 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: 30–31.

Zaratin P, Comi G, Coetzee T, Ramsey K, Smith K, Thompson A, et al. Progressive MS Alliance Industry Forum: Maximizing Collective Impact To Enable Drug Development. Trends Pharmacol. Sci. 2016; 37: 808–810.

Ziegler G, Grabher P, Thompson A, Altmann D, Hupp M, Ashburner J, et al. Progressive neurodegeneration following spinal cord injury. Neurology 2018; 90: e1257–e1266.

**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%) |  |  |

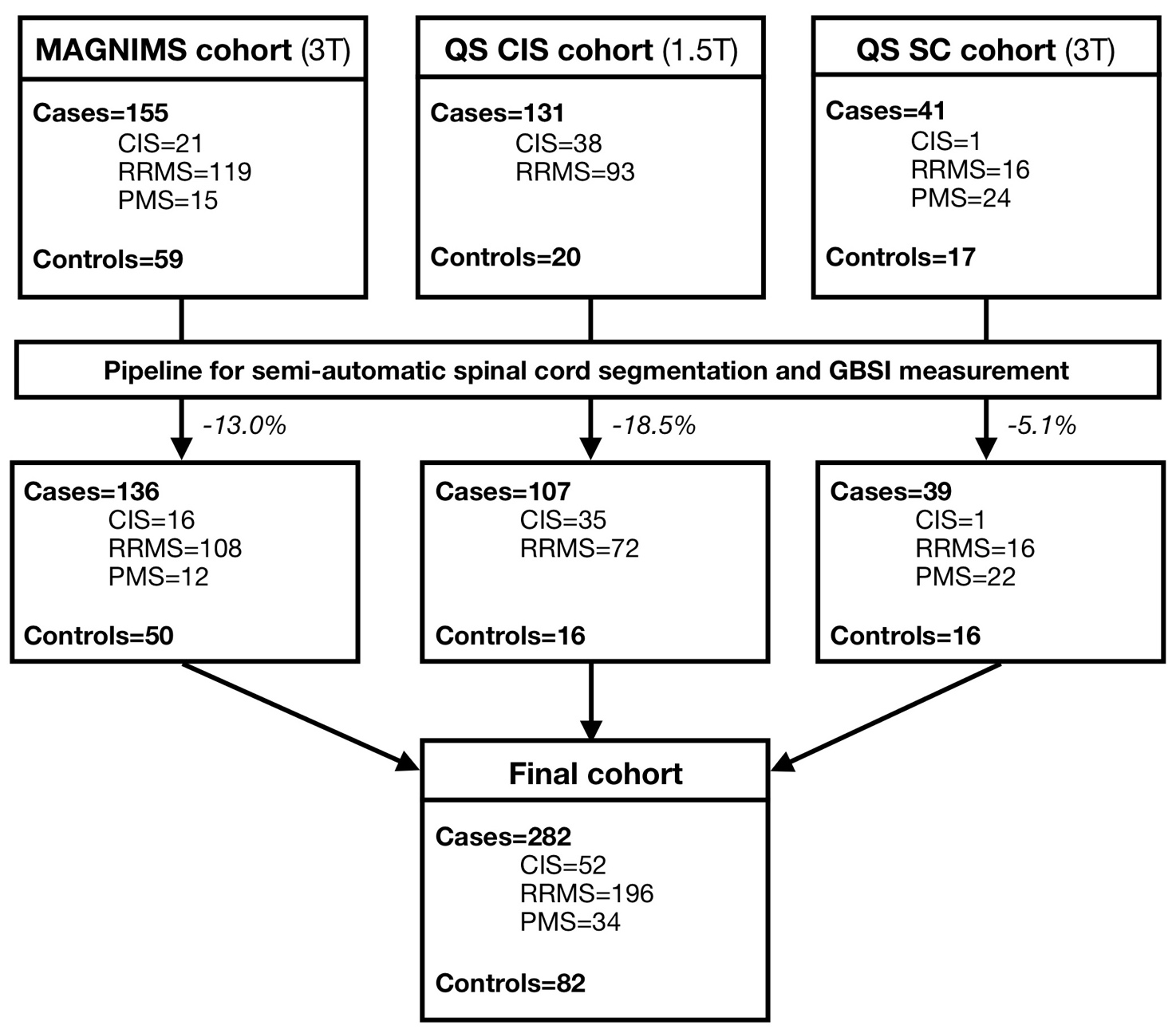
**TABLE 2. Clinical correlates of spinal cord atrophy with CSA and GBSI.**

Table shows associations between spinal cord atrophy (with CSA and GBSI), and different clinical variables. Odds ratio (OR), 95% confidence intervals (95%CI) and p-values are shown from logistic regression models; age, sex, site of acquisition and disease duration were included as covariates (\*p<0.05).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **CSA** | | | | **GBSI** | | | |
|  | **OR** | **95%CI** | | **p-values** | **OR** | **95%CI** | | **p-values** |
|  |  | *Lower* | *Upper* |  |  | *Lower* | *Upper* |  |
| ***CIS vs HC*** | -0.04 | -0.14 | 0.04 | 0.33 | -0.23 | -0.40 | -0.07 | <0.01\* |
| ***RRMS vs HC*** | -0.07 | -0.14 | -0.00 | 0.02\* | -0.31 | -0.42 | -0.19 | <0.01\* |
| ***PMS vs HC*** | -0.05 | -0.17 | 0.05 | 0.32 | -0.43 | -0.64 | -0.23 | <0.01\* |
| ***EDSS progression*** | -0.00 | -0.07 | 0.06 | 0.83 | -0.13 | 0.25 | -0.02 | 0.02\* |

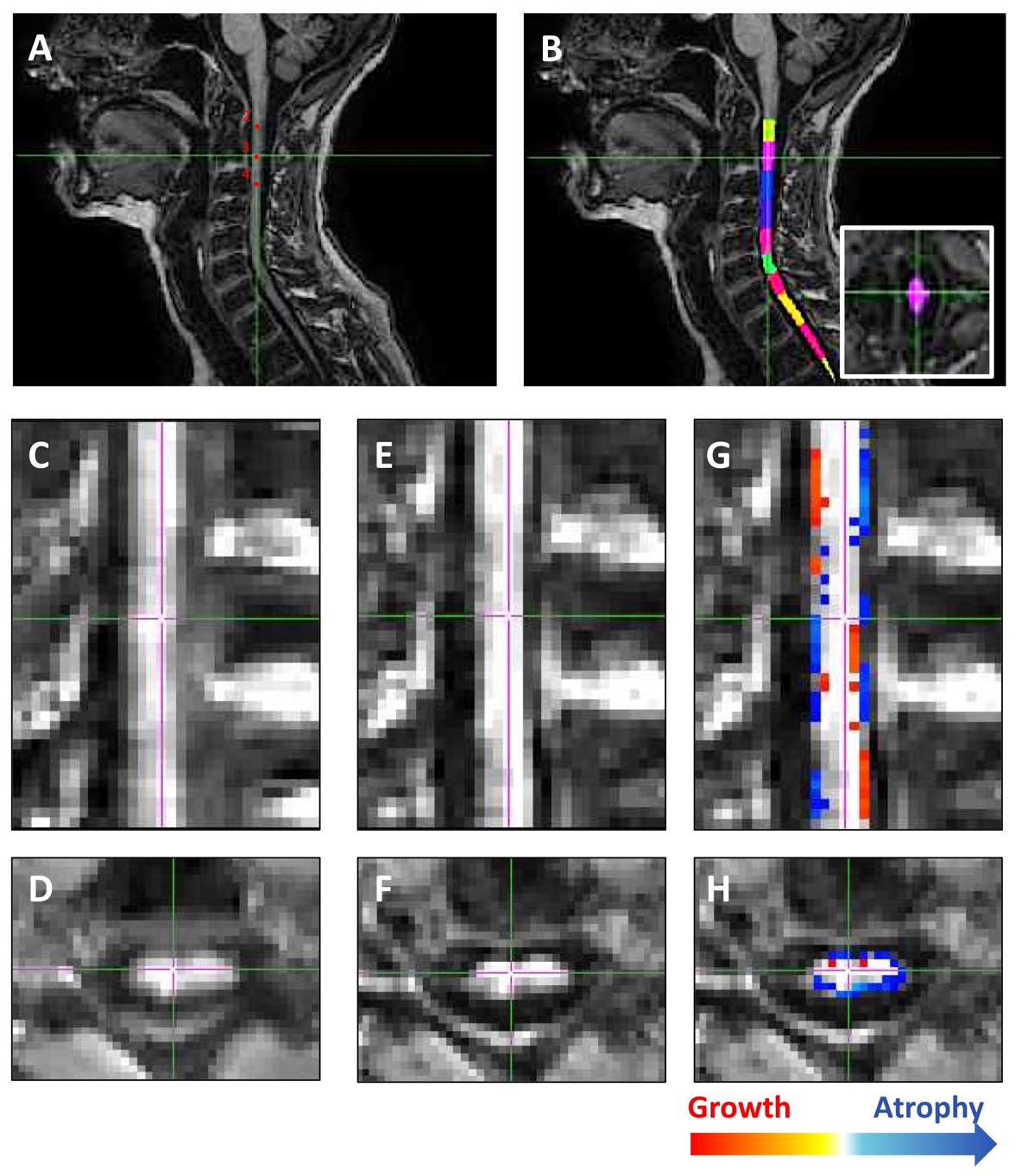
**FIGURE 1. Study flow diagram.**

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.



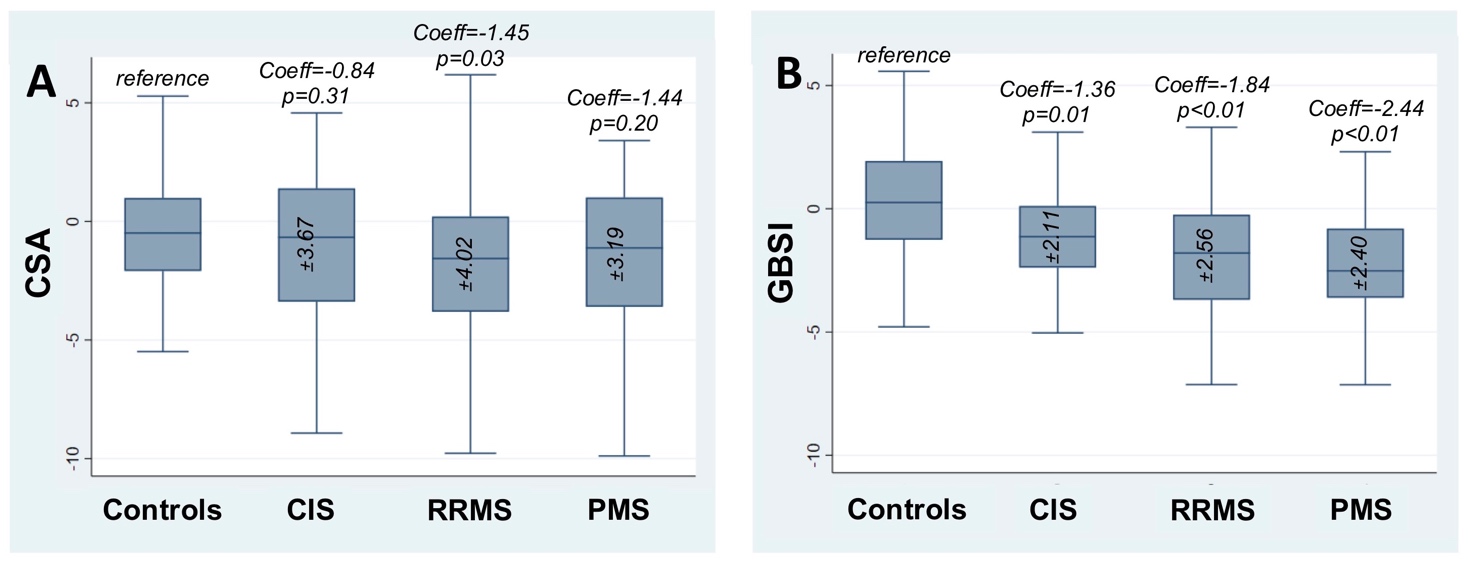
**FIGURE 2. Spinal cord segmentation and GBSI.**

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**).



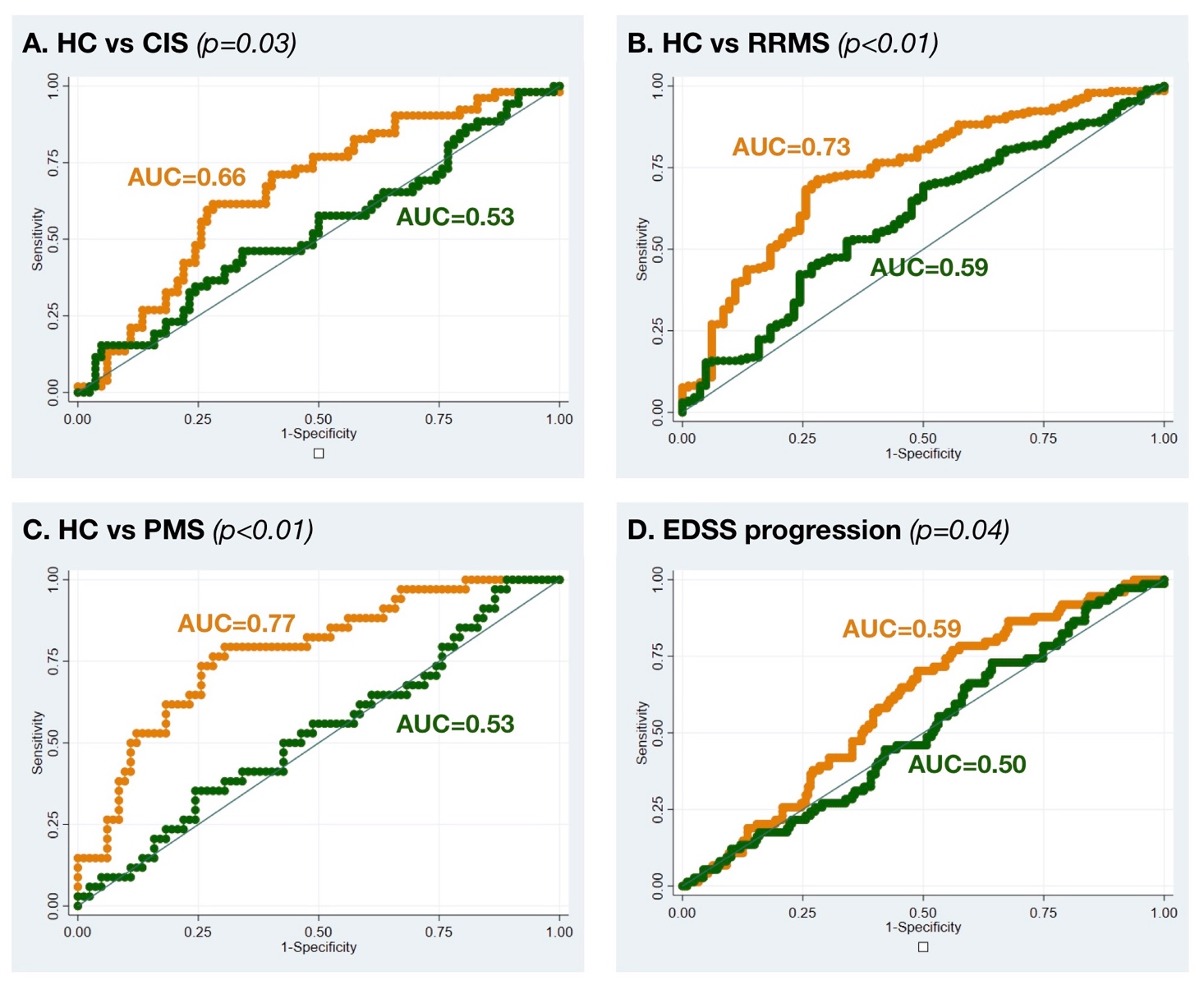
**FIGURE 3. Box-and-Whisker plot of CSA and GBSI measurements**

Box-and-Whisker plotfor 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).



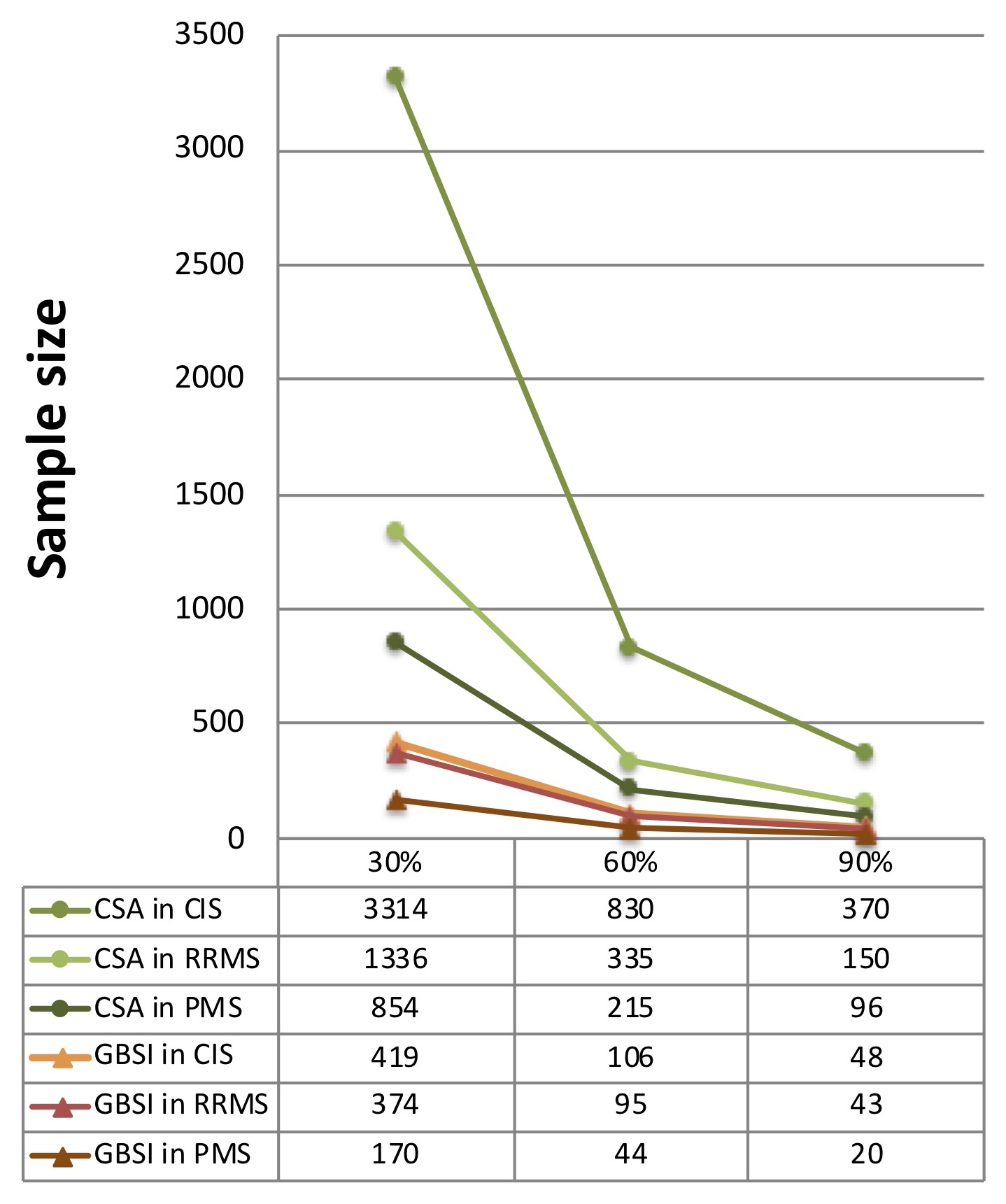
**FIGURE 4. ROC curves for CSA and GBSI in relation to clinical variables.**

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



**FIGURE 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, 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%).



**Supplementary Material 1. MAGNIMS sites and included patients.**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Site*** | ***Number of included individuals*** | ***Manufacturer, model and field strength*** | ***Acquisition parameters***  *(TR/TE/TI msec)* |
| **Barcelona** | 21 | Siemens Magnetom Trio 3T | 2300/3.2/900 |
| **Bochum** | 4 | Philips Achieva 3T | 8/”shortest”/”shortest” |
| **London** | 123 | GE Signa 1.5T | 15.6/4.2/450 |
|  | 55 | Philips Achieva 3T | 8/3.7/860 |
| **Lugano** | 43 | Siemens Magnetom Skyra 3T | 2300/5.1/1140 |
| **Mannheim** | 14 | Siemens Magnetom Skyra 3T | 1900/2.6/1000 |
| **Milan** | 49 | Philips Achieva 3T | 8/”shortest”/”shortest” |
| **Naples** | 28 | General Electric Signa HDtx 3T | 7.9/3.2/450 |
| **Oxford** | 15 | Siemens Magnetom Prisma 3T | 2300/3.59/900 |