Spinal cord atrophy as a primary outcome measure in phase II trials of progressive multiple sclerosis

N.Cawley[1], C. Tur[1], F. Prados[1,2], D. Plantone[1], H. Kearney[1], K. Abdel – Aziz[1], S. Ourselin[2], C.A.M Gandini Wheeler – Kingshott[1], D. H Miller[1,3], A.J. Thompson[1,3], O. Ciccarelli[1,3]

1 –Queen Square MS Centre, UCL Institute of Neurology, London, United Kingdom

2 - Translational Imaging Group, Centre for Medical Image Computing (CMIC), Department of Medical Physics and Bioengineering, UCL, London, United Kingdom

3 – NIHR University College London Hospitals Biomedical Research Centre, London, United Kingdom

Corresponding author and reprint request to:

Niamh Cawley

Department of Neuroinflammation

UCL Institute of Neurology,

1st floor, Russell Square House

10-12 Russel Square

London WC1B 5EH

UK

E-mail: n.cawley@ucl.ac.uk

Tel: +44 (0)2034484307

Abstract

Objectives: To measure the development of spinal cord (SC) atrophy over one year in patients with progressive MS (PMS) and determine the sample sizes required to demonstrate a reduction in SC cross-sectional area (SC-CSA) as an outcome measure in clinical trials.

Methods: Forty-four PMS patients (26 primary-progressive MS (PPMS), 18 secondary-progressive MS (SPMS) and 29 healthy controls (HCs) were studied at baseline and 12 months. SC-CSA was measured using the 3D-fast field echo sequences acquired at 3T and the active surface model. Multiple linear regressions were used to investigate changes in imaging measurements.

Results: PPMS patients had shorter disease duration, lower EDSS, and larger SC-CSA than SPMS. All patients together showed a significantly greater decrease in percentage SC-CSA change than HCs, which was driven by the PPMS. All patients deteriorated over one year, but no association was found between percentage SC-CSA change and clinical changes. The sample size per arm required to detect a 50% treatment effect over one year, at 80% power, was 57 for PPMS and 546 for SPMS.

Conclusion: SC-CSA may become an outcome measure in trials of PPMS patients, when they are at an early stage of the disease, have moderate disability and modest SC atrophy.

Introduction

There is an urgent need for treatments that slow progression in progressive MS. Phase II trials rely on biomarkers that are more sensitive to therapeutic effects than clinical measures. A putative biomarker for phase II trials is cervical cord atrophy (or reduction in spinal cord cross sectional area), which reflects axonal loss[1] and is an independent determinant of disability[2], Studies in MS subjects with a long disease duration found that cord atrophy is related to disability independently of brain lesion load and atrophy.[3] Additionally, the rate of atrophy in the cervical cord is greater than that seen in the brain, which is currently used as an outcome measure in progressive MS trials (~1.5% vs. 0.5%-1.0% per year).[4,5] To date, only a few clinical trials have used spinal cord atrophy as an exploratory endpoint.[6,7] This reflects the numerous challenges of detecting minimal changes in a small structure in a multi-centre setting, where there may differences in imaging protocols and scanners.

In this study, we considered the potential of using spinal cord atrophy as an endpoint in one year, phase II, single-centre, neuroprotective clinical trials in progressive MS. The aims of this study were to: (i) measure the progression of spinal cord atrophy over one year and evaluate its relationship with physical disability in progressive MS; and (ii) determine sample sizes required to demonstrate a reduction in spinal cord atrophy as a primary outcome measure in a phase II clinical trial. For each aim, we retrospectively combined our two recent cohorts of early PPMS and well-established SPMS into a single progressive MS cohort, as these two phenotypes are thought to share more similarities than differences in their clinical, pathological and imaging characteristics. Secondly, we looked at each group independently, adjusting for brain atrophy development and changes in brain lesion loads.

Methods

Subjects

We retrospectively collected data of patients with SPMS and PPMS who participated in our previous studies,[8,9] and then analysed them separately. Inclusion criteria were a diagnosis of PPMS[10] or SPMS, and aged between 18 and 65 years. We did not specifically recruit patients who showed a steady progression as the major cause of increasing disability in the preceding 2 years. Healthy controls (HCs) were also recruited. All subjects were invited to come back for a follow-up after 1 year. Patients were clinically assessed on the day of the MRI at each time point.

All subjects recruited into the study provided written informed consent prior to taking part in the study, which was approved by the local research ethics committee.

Clinical Assessments

At the time of the MRI, all patients were clinically assessed with the Expanded Disability Status Scale (EDSS),[11] the 9-Hole Peg Test (9-HPT) and timed 25-Foot Walk Test (TWT).[12] Z-scores were calculated for the 9-HPT and TWT.[13] Mean grip strength from the upper and lower limbs were measured using the Jamar hydraulic dynamometer (Sammons Preston. Incorporated, Bolingbrook, IL, USA).[14] American Spinal Injury Association (ASIA) motor and sensory scores[15] were also recorded.

MRI acquisition

All subjects were scanned at 3T using a Philips Achieva MRI system at baseline and at 1-year follow-up, using the manufacturer's 16 – channel neurovascular coil (Philips Healthcare Systems, Best, Netherlands). The cervical cord was imaged in the axial plane, perpendicular

to the longitudinal axis of the cord, to measure the mean spinal cord cross sectional area (SC-CSA). The imaging volume was centred on the C2/3 intervertebral disc, using a fat – suppressed 3D slab-selective fast field echo (FFE) sequence [TR=23ms; TE=5ms; flip angle α 7º; FOV=240x180mm^2; voxel size=0.5x0.5x5 mm^3; NEX=8; 11 axial contiguous slices].

For calculation of brain volumes, a 3D T1-weighted magnetisation-prepared gradient-echo sequence was used [TR=6.9ms; TE=3.1ms; TI=824ms; flip angle α 8°; FOV= 256x256mm^2; voxel size=1x1x1mm^3; NEX=1; 180 sagittal contiguous slices].

For calculation of brain T2 lesion volumes, PD/T2 weighted images were acquired using a dual-echo TSE sequence [TR=3500ms; TE=15/85ms; flip angle α 90°; FOV=240x180mm^2; voxel size=1x1x3mm^3; NEX=1; 50 axial contiguous slices].

MRI analysis

Spinal cord area measurement

The active surface model (ASM) was used to obtain the SC-CSA (algorithm parameters: number of fourier coefficients=27, polynomial order=7).[16] The ASM involves placing a predetermined size and shape region-of-interest (ROI) in the centre of the cord on each slice. The rater who placed this ROI was blinded to the subject identity and clinical characteristics, but was aware of the time points. The program then uses intensity gradient information to calculate the radius and centre of each slice.[16] The centre line is refined from the initial user estimate and segmentation then involves a multistage approach allowing greater complexity of the cord radius to be calculated.[16] The cord outline is generated automatically for each slice, and then the mean CSA obtained by averaging the area of the

three contiguous 5mm axial slices, centred on C2/3 disc, was calculated (**Figure 1**). Finally, the percentage change in SC-CSA between the two time points was calculated.

Brain atrophy

After lesion-filling,[17] images were segmented using STEPS,[18] which is a multi-atlas segmentation propagation technique. STEPS was used inside SIENA in conjunction with BET and helps to avoid bias due to the amount of peripheral CSF. STEPS computed the brain tissue mask and BET computed the skull. Then, the transformation to the half-way space between both time points was calculated; the Percent Brain Volume Change (PBVC) between two time-points was obtained[19].

Brain and spinal cord lesions

Brain T2-lesion volume (T2LV) was calculated by outlining lesions on T2-weighted images using a semi-automated edge finding tool (JIMv.6.0). The presence (or absence) of T2-hyperintense lesions at C2-C3 cord level (i.e., within the 15mm segment of cord used to calculate the SC-CSA) was recorded by reviewing the axial FFE images acquired at baseline and 1-year follow-up. The contrast between lesions and non-lesional tissue was too low to allow lesion contouring. All SC segmentations were visually inspected to ensure that the visible lesions didn't affect the CSA.

Statistical analysis

All analyses were performed in Stata 13.1 (Stata Corporation, College Station, Texas, USA).

Paired t-tests were used to assess changes in clinical measures over time and two sample t-tests were used to investigate differences in variables at baseline between subject groups.

Age was centred in order to get more realistic estimates in the linear regression models.

To assess differences in the percentage change in SC-CSA over time between all patients together (and each patient group separately) and HCs the following regression equation was used:

Percentage SC-CSA change = beta0 (% change in healthy controls) + beta1*SPMS-type + beta2*PPMS-type + beta3*centered-age + beta4*sex

(beta1*SPMS-type and beta2*PPMS-type were substituted with beta*all-MS-patients for the model which included all patients together).

The same models were then re-run by correcting for PBVC.

A similar statistics was done when analysing the changes over time in brain volume by swapping % SC-CSA change with PBVC and correcting for brain T2LV.

To assess the association between clinical changes and percentage SC-CSA change the following equation was used for each patient group separately and then for all patients together:

Changes in clinical data (EDSS or 9-HPT or TWT or grip strength or ASIA motor and sensory scores) = beta0 + beta1*percentage SC-CSA change in each patient group + beta2*centered-age + beta3*sex

As evident by the equations above, all models were adjusted for age and gender. Because of the very small number of patients without spinal cord lesions at C2/C3, no statistical test has been carried out to test the impact of lesions on SC-CSA measurement.

Statistical significance was considered when p values were below 0.05.

Sample size calculations for 30%, 50% and 100% treatment effects on percentage SC-CSA change (and PBVC) were computed (with 80% power at 5% statistical significance) using the age- and gender-adjusted differences between each patient group and HCs. The SD of the percentage SC-CSA change (and PBVC) within each group was used for the sample size calculations. 100% treatment effect is theoretically reached when the percentage SC-CSA change in patients (or PBVC in patients) is equal to that observed in HCs.

Results

Participant demographics and characteristics

Forty-four patients with progressive MS (26 PPMS and 18 SPMS), and 29 HCs were studied at baseline and one year. Patient demographics, clinical and radiological characteristics are summarised in **Table 1.**

There was a significant difference in age between all progressive MS patients and HCs ($p<0.001$). At baseline, PPMS patients had significantly shorter disease duration ($p<0.0001$), lower EDSS ($p<0.004$), lower brain T2LV ($p<0.001$) and larger SC-CSA ($p<0.02$) than SPMS patients. The duration of progressive disease was similar among groups (**Table 1**).

There was a significant increase in brain T2LV over 12 months in PPMS ($p<0.002$) and SPMS ($p<0.012$). At baseline, all SPMS patients (N=18) and 20 PPMS patients (20/26, 77%) had spinal cord lesions at C2-C3. At follow-up, two PPMS patients showed new lesions at this level (22/26 (85%).

All patients together progressed significantly over one year on EDSS ($p<0.001$) (**Table 1**), 9-HPT ($p<0.001$), TWT ($p=0.053$), grip strength ($p=0.001$) and ASIA-motor scores ($p=0.015$) (**Table 2**). While PPMS patients showed significant progression in all these scores (except TWT), SPMS significantly progressed only on EDSS and 9-HPT (all $p<0.05$) (**Supplemental Material**).

Spinal cord atrophy

PMS patients and each progressive phenotype had a smaller spinal cord area at both baseline and one-year follow-up than HCs (**Table 1**).

Taken together, all patients showed a significantly greater decrease in percentage SC-CSA change over time than HCs (adjusted difference: -2.56, 95% Confidence Intervals (CIs) -4.35,

-0.76, p=0.006), when adjusted for age and gender. When looking at the phenotypes, only PPMS showed a greater decrease in percentage SC-CSA change than HCs (-2.95, 95%CIs -4.88, -1.02, p=0.003). There was no difference in the percentage SC-CSA change between SPMS and HCs (-1.78, 95%CIs -4.07, -0.51, p=0.13). In this model, neither age nor gender was significant. Both patients groups showed a significant reduction in percentage SC-CSA change over time, whilst HCs did not (**Table 1**).

No associations were found between % change in SC-CSA and clinical changes over one year in all progressive MS patients together and in each patient group separately.

The inclusion of PBVC in the models did not affect these results.

Brain atrophy

PBVC was significantly greater in the progressive MS cohort than HCs over the follow-up (adjusted difference= -1.17, 95%CIs -1.78, -0.56, p<0.001), when adjusting for age and gender. Looking at the individual phenotype, the PPMS patients showed a significant greater PBVC than HCs (-1.23, 95%CIs -1.89, -0.58, p<0.001); similarly, the SPMS group showed a significantly greater PBVC than HCs (-1.02, 95%CIs -1.83, 0.02, p=0.016). There was not significant brain volume loss in HCs over time (**Table 1).** When correcting for brain T2LV, the results did not change.

Sample size calculations for a neuroprotective clinical trial

The sample size calculations for all patients together using the percentage SC-CSA change (adjusted for age and gender), in a 12-month neuroprotective clinical trial, demonstrated that the minimum sample size per arm required to detect a 50% treatment effect (that is a 50% reduction in % change in SC-CSA) (5% significance, 80% power), was 146 per arm; when

the two patient groups were analysed separately, the minimum sample size per arm reduced to 57 in PPMS and increased to 546 in SPMS (**Table 3**).

To complete the analysis, we calculated the sample size using the PBVC in the progressive MS cohort as well as in each patient group. We found that the minimum sample size per arm required to detect the same 50% treatment effect, at 80% power and 5% significance, in all patients together was 51 per arm; when the two cohorts were analysed separately, the minimum sample size required per arm was 57 in PPMS and 33 in SPMS (**Table 3**).

Discussion

Our single-centre study showed a significant development of cord atrophy over 1 year in progressive MS when compared to HCs. The PPMS subgroup, who had shorter disease duration, lower EDSS, and larger SC-CSA compared with SPMS, showed the greatest reduction in percentage SC-CSA change over one year. Sample size calculations gave realistically achievable numbers of patients per arm in PPMS when using percentage SC-CSA change, which were similar to the numbers obtained with PBVC, though this was not the case in SPMS. Thus, SC-CSA shows potential to be a feasible outcome measure, in addition to brain atrophy, for inclusion in single-centre, placebo-controlled neuroprotective trials in PPMS, when they are at an early stage of the disease, have moderate disability and modest SC atrophy. We will discuss each of these results in turn.

We found a significant difference in percentage SC-CSA change between PPMS patients and HCs, but not between SPMS patients and HCs. The SPMS subjects were older, had longer disease duration and were more disabled than those with PPMS, they had more marked spinal cord atrophy at study baseline and had a lower percentage change in SC-CSA than PPMS. The PPMS patients had less cord atrophy at baseline and therefore had more to lose over time, while the critical loss of axons may have already occurred in the SPMS subgroup. This raises the possibility that cord atrophy may not be a useful biomarker in SPMS, especially if they have very long disease duration and marked cord atrophy at study entry. The better performance of cord atrophy measures in the PPMS subgroup compared to the SPMS subgroup may be important in understanding progressive MS. Interestingly, the length of progressive disease appeared to be similar between SPMS and PPMS; however, it

is often difficult to establish the onset of a progressive disease. A long term, prospective study that follows people from RRMS through to SPMS would be required to clarify the temporal dynamics of spinal cord atrophy in relapse-onset MS.

In this study, all SPMS patients and the vast majority of PPMS patients had spinal cord lesions at C2-C3 level. Therefore, we could not do any statistics to investigate the role of spinal cord lesions in the development of cord atrophy in this study and the relationship between cord lesions and SC-CSA measurements. However, the fact that the ASM algorithm constructs a 3D cylindrical model over all slices helps to minimise the effect that the lesions may have on the shape of the cord, with consequent effect on the CSA measurements. A previous voxel-wise morphometry study of the cervical cord showed that the spatial distribution of cord atrophy and cord lesions was not the same for any phenotype, suggesting that cord atrophy is not necessarily the local result of focal demyelination[20]. A post-mortem study found that lesions in the cervical cord do not play a major role in local atrophy in the cervical cord[21]. On the other hand, a significant decline in cord area may occur after a focal MS lesion[22], and that asymptomatic spinal cord lesions at baseline, change in cord lesion number and the development of spinal cord atrophy, contribute to the development of future disability in clinically isolated syndrome[23].

The progressive MS cohort significantly progressed in a number of clinical measures over 1 year, including EDSS, ASIA-motor scores, 9-HPT and grip strength. However, there were no significant associations between percentage SC-CSA change and change in clinical scores in all patients together and in each group separately. This may be due to the short follow-up

period and the small sample size of the study, which makes it difficult to detect significant associations. Also, the patients recruited into this study were not recruited based on clinical progression in the year prior to recruitment, as this was a retrospective study and patients were not selected on the basis of this criterion.

The overall sample size calculations for the percentage SC-CSA change gave achievable numbers of patients per arm in all patients together, but this was clearly driven by the PPMS group. The sample size calculations for 50% treatment effect in the SPMS patients gave higher numbers per arm (N=546) than PPMS (N=57). This highlights the importance of recruiting subjects with early PPMS into clinical trials and raises the question if cord atrophy behaves differently in SPMS as a result of the longer disease duration (rather than the duration of progressive disease), and the more marked spinal cord atrophy at baseline. It is likely that these numbers could be further reduced by recruiting only subjects with a history of disease progression in the previous 12 months, and by recruiting patients who showed a steady progression as the major cause of increasing disability in the preceding 2 years history.

The ASM has been used in one large cross-sectional multicentre study in MS;[24] this study confirmed the stability of this measure among different centers, but the stability of longitudinal SCA needs to be explored and validated before being used in multicenter trials. Since the post-processing and analysis are not complex, we think that the methods can be adopted by any center with a special interest in spinal cord imaging. However, the challenges in assessing small absolute changes longitudinally in the size of a small structure should not be underestimated and in the real world setting of large, multicentre, long term,

phase III clinical trials, methodological variability includes the use of multiple acquisition sequences and multiple scanner models, software and hardware upgrades and changes in scanners during follow up. Such methodological factors may require larger sample sizes than we have estimated from our single centre/scanner study.[7] Notwithstanding this, the importance of good image quality and analysis method are crucial for measuring spinal cord atrophy in order to obtain acceptable sample size estimates.

An interesting observation of our single-centre study is that sample size calculations for brain atrophy gave similar results as those for spinal cord atrophy only in the PPMS group (N=57 for 50% treatment effects when using both imaging outcome measures), whilst the numbers per arm were much higher for spinal cord atrophy (N=546) than brain atrophy (N=33) in the SPMS group. This suggests that spinal cord atrophy has the potential to be used as an outcome measure in clinical trials with therapies that target neurodegeneration in addition to brain atrophy in PPMS.

Technical considerations and limitations

The active surface model was originally developed for near isotropic 3D data. Overall, the effect of using 2D rather than 3D data is unknown, but is likely to be negligible.

From a technical perspective, only the CSA of three slices centred at C2-C3 (which represents a short segment of the cervical cord (15mm)) was used in the analysis. However, the cervical cord was imaged using a neck collar in the axial-oblique plane (i.e. slices perpendicular to the cord) with the centre of the imaging volume positioned at the level of C2-3 intervertebral disc, in order to ensure minimal variability and slice-angulation. Our CSA measurements are made considering the slice in the middle of C2-C3 intervertebral disc level and one slice above and one below[25].

Previous studies have calculated the inter- and/or intra-rater assessment of the longitudinal change in cord area using the ASM. A previous study found intra- and inter- coefficients of variation values of 0.44% and 1.07% respectively using the ASM[16]. Another study found the intra-class correlation coefficient (ICC) for scan-rescan reliability using the active surface model of 0.992 in HCs and MS patients[26]. Kearney et al. found that the inter-observer ICCs of the active surface model were 0.993 and 0.995 (depending on the MRI sequence used), and the scan-rescan ICCs were between 0.978 and 0.981 (depending on the MRI sequence)[27]. We previously found that the coefficient of variation and the Dice score coefficient of a test-retest experiment repeated on 5 healthy controls 3 times on separate occasions were 0.8% and 0.965[28]. In the same paper, we demonstrated that the presence of lesions has a low impact on the reproducibility values[28]. In the present study, the correlation between the two cord area obtained at two different times points was 99% in controls, 97% in PPMS and 94% in SPMS, indirectly suggesting reproducible measurements.

To date only segmentation-based methods are available for measuring longitudinal change in spinal cord atrophy, unlike brain atrophy, where registration based methods are available, such as SIENA and Boundary Shift Integral (BSI). This is a limitation for measuring longitudinal change in spinal cord atrophy. We are currently developing and validating a BSI technique for the spinal cord, similar to that like used in the brain, which may improve spinal cord atrophy measurements in MS patients.

The MS patients and HCs are not well matched for age and gender in this study. To address this limitation, we have adjusted for age and gender in the regression models, which should remove any possible confounding effects of these two factors.

Conclusion

Our sample size calculations gave acceptable numbers of patients per arm, suggesting spinal cord atrophy shows potential to be a feasible outcome measure, in addition to brain atrophy, for single-centre, placebo-controlled neuroprotective trials in early PPMS, who are at an early stage of disease and have modest cord atrophy at baseline. Conversely, cord atrophy may not be a useful outcome measure in SPMS who have not showed a steady progression as the major cause of increasing disability in the preceding 2 years history, have long disease duration and marked spinal cord atrophy at study entry.

REFERENCES:

1. Bot JC, Blezer EL, Kamphorst W, et al. The spinal cord in multiple sclerosis: relationship of high-spatial-resolution quantitative MR imaging findings to histopathologic results. Radiology 2004, 233, 531-40.

2. Kearney H, Miller DH, Ciccarelli O. Spinal cord MRI in multiple sclerosis--diagnostic, prognostic and clinical value. Nat Rev Neurol. 2015 Jun;11(6):327-38.

3. Kearney H, Rocca MA, Valsasina P, et al. Magnetic resonance imaging correlates of physical disability in relapse onset multiple sclerosis of long disease duration. Mult Scler 2014, 20, 72-80.

4. Lukas C, Knol DL, Sombekke MH, et al. Cervical spinal cord volume loss is related to clinical disability progression in multiple sclerosis. J Neurol Neurosurg Psychiatry 2015, 86, 410-8.

5. De Stefano N, Giorgio A, Battaglini M, et al. Assessing brain atrophy rates in a large population of untreated multiple sclerosis subtypes. Neurology 2010, 74, 1868-76.

6. Kapoor R, Furby J, Hayton T, et al. Lamotrigine for neuroprotection in secondary progressive multiple sclerosis: a randomised, double-blind, placebo-controlled, parallel-group trial. Lancet Neurol 2010, 9, 681-8.

7. 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), ECTRIMS 2015.

8. Kearney H, Altmann DR, Samson RS, et al. Cervical cord lesion load is associated with disability independently from atrophy in MS. Neurology. 2015 Jan 27;84(4):367-73. PubMed PMID: 25540312. Epub 2014/12/30. eng.

9. Abdel-Aziz K, Schneider T, Solanky BS, et al. Evidence for early neurodegeneration in the cervical cord of patients with primary progressive multiple sclerosis. Brain: a journal of neurology. 2015 Apr 10. PubMed PMID: 25863355. Epub 2015/04/12. Eng.

10. Polman CH, Reingold SC, Edan G, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". Annals of neurology. 2005 Dec;58(6):840-6.

11. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983, 33, 1444-52.

12. Cutter GR, Baier ML, Rudick RA, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. Brain 1999, 122 (Pt 5), 871-82.

13. Fischer JS, Rudick RA, Cutter GR, et al. 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, 244-50.

14. Svens B, Lee H. Intra- and inter-instrument reliability of Grip-Strength Measurements: GripTrackTM and Jamar_ hand dynamometers. Br J Hand Ther 2005; 10: 47–55.

15. Maynard, FMJR, Bracken MB, Creasey G, et al. 1997. International Standards for Neurological and Functional Classification of Spinal Cord Injury. American Spinal Injury Association. Spinal Cord, 35, 266-74.

16. Horsfield MA, Sala S, Neema M, et al. Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: application in multiple sclerosis. Neuroimage 2010, 50, 446-55.

17. Chard DT, Jackson JS, Miller DH, et al. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. J Magn Reson Imaging 2010, 32, 223-8.

18. Cardoso MJ, Leung K, Modat M, et al. STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation. Medical image analysis. 2013 Aug;17(6):671-84.

19. Smith SM, De Stefano N, Jenkinson M, Matthews PM. Normalized accurate measurement of longitudinal brain change. Journal of computer assisted tomography. 2001 May-Jun;25(3):466-75. PubMed PMID: 11351200. Epub 2001/05/15. eng.

20. Rocca MA, Valsasina P, Damjanovic D, Horsfield MA, Mesaros S, Stosic-Opincal T, Drulovic J, Filippi M. Voxel-wise mapping of cervical cord damage in multiple sclerosis patients with different clinical phenotypes. J Neurol Neurosurg Psychiatry. 2013 Jan;84(1):35-41.

21. Evangelou N, DeLuca GC, Owens T, Esiri MM. Pathological study of spinal cord atrophy in multiple sclerosis suggests limited role of local lesions. Brain. 2005 Jan;128(Pt 1):29-34.

22. Ciccarelli O, Altmann DR, McLean MA, Wheeler-Kingshott CA, Wimpey K, Miller DH, Thompson AJ. Spinal cord repair in MS: does mitochondrial metabolism play a role? Neurology. 2010 Mar 2;74(9):721-7.

23. Brownlee WJ, Altmann DR, Alves Da Mota P, Swanton JK, Miszkiel KA, Wheeler-Kingshott CG, Ciccarelli O, Miller DH. Association of asymptomatic spinal cord lesions and atrophy with disability 5 years after a clinically isolated syndrome. Mult Scler. 2016 Aug 1. pii: 1352458516663034.

24. Rocca, M. A., Horsfield, M. A., Sala, S., Copetti, M., Valsasina, P., Mesaros, S., Martinelli, V., Caputo, D., Stosic-Opincal, T., Drulovic, J., Comi, G. & Filippi, M. 2011. A multicenter assessment of cervical cord atrophy among MS clinical phenotypes. Neurology, 76, 2096-102.

25. Yiannakas MC, Kearney H, Samson RS, Chard DT, Ciccarelli O, Miller DH, Wheeler-Kingshott CA. Feasibility of grey matter and white matter segmentation of the upper cervical cord in vivo: a pilot study with application to magnetisation transfer measurements. Neuroimage. 2012 Nov 15;63(3):1054-9.

26. Yiannakas MC, Mustafa AM, De Leener B, Kearney H, Tur C, Altmann DR, De Angelis F, Plantone D, Ciccarelli O, Miller DH, Cohen-Adad J, Gandini Wheeler-Kingshott CA. Fully automated segmentation of the cervical cord from T1-weighted MRI using PropSeg: Application to multiple sclerosis. Neuroimage Clin. 2015 Nov 10;10:71-7.

27. Kearney H, Yiannakas MC, Abdel-Aziz K, Wheeler-Kingshott CA, Altmann DR, Ciccarelli O, Miller DH. Improved MRI quantification of spinal cord atrophy in multiple sclerosis. J Magn Reson Imaging. 2014 Mar;39(3):617-23.

28. Prados F, Cardoso MJ, Yiannakas MC, Hoy LR, Tebaldi E, Kearney H, Liechti MD, Miller DH, Ciccarelli O, Wheeler-Kingshott CA, Ourselin S. Fully automated grey and white matter spinal cord segmentation. Sci Rep. 2016 Oct 27;6:36151.