AJNR

Validation of a New Semiautomated Segmentation Pipeline Based on the Spinal Cord Toolbox DeepSeg Algorithm to Estimate the Cervical Canal Area

N. Mongay-Ochoa, D. Pareto, M. Alberich, M. Tintore, X. Montalban, À. Rovira and J. Sastre-Garriga

Validation of a New Semiautomated Segmentation Pipeline Based on the Spinal Cord Toolbox DeepSeg Algorithm to Estimate the Cervical Canal Area

N. Mongay-Ochoa, D. Pareto, M. Alberich, M. Tintore, X. Montalban, À. Rovira, and J. Sastre-Garriga

ABSTRACT

BACKGROUND AND PURPOSE: As in the brain reserve concept, a larger cervical canal area may also protect against disability. In this context, a semiautomated pipeline has been developed to obtain quantitative estimations of the cervical canal area. The aim of the study was to validate the pipeline, to evaluate the consistency of the cervical canal area measurements during a 1-year period, and to compare cervical canal area estimations obtained from brain and cervical MRI acquisitions.

MATERIALS AND METHODS: Eight healthy controls and 18 patients with MS underwent baseline and follow-up 3T brain and cervical spine sagittal 3D MPRAGE. The cervical canal area was measured in all acquisitions, and estimations obtained with the proposed pipeline were compared with manual segmentations performed by 1 evaluator using the Dice similarity coefficient. The cervical canal area estimations obtained on baseline and follow-up T1WI were compared; brain and cervical cord acquisitions were also compared using the individual and average intraclass correlation coefficients.

RESULTS: The agreement between the manual cervical canal area masks and the masks provided by the proposed pipeline was excellent, with a mean Dice similarity coefficient mean of 0.90 (range, 0.73–0.97). The cervical canal area estimations obtained from baseline and follow-up scans showed a good level of concordance (intraclass correlation coefficient = 0.76; 95% CI, 0.44–0.88); estimations obtained from brain and cervical MRIs also had good agreement (intraclass correlation coefficient = 0.77; 95% CI, 0.45–0.90).

CONCLUSIONS: The proposed pipeline is a reliable tool to estimate the cervical canal area. The cervical canal area is a stable measure across time; moreover, when cervical sequences are not available, the cervical canal area could be estimated using brain T1WI.

ABBREVIATIONS: CCaA = cervical canal area; FA = flip angle; GT = ground truth; HC = healthy controls; ICC = intraclass correlation coefficient; LoA = limits of agreement; pwMS = patients with multiple sclerosis; SCT = Spinal Cord Toolbox; SD = standard deviation

In patients with MS, the progression of neurologic disability cannot be explained only by the accumulation of brain white matter lesions.[1] Because neurodegenerative damage of the cervical cord is present in most patients with MS,[2] recent work has demonstrated the value of cervical cord atrophy as an independent prognostic factor for disability.[3]

In homology to the brain reserve concept, which implies that individuals with a larger premorbid brain (estimated using total intracranial volume as a proxy of maximal lifetime brain growth) have a lower risk of MS-related cognitive and physical impairment,[4] a larger cervical canal area (CCaA), which may be taken as a proxy for maximal lifetime spinal cord growth, may also protect against disability.[5]

In this context, a semiautomated pipeline has been developed to obtain quantitative estimations of the CCaA based on brain and cervical 3D T1WI, using the Spinal Cord Toolbox (SCT; https://www.nitrc.org/projects/sct/). To validate the reproducibility of the proposed pipeline, we compared CCaA measurements obtained with the SCT with those obtained with the manual ground truth (GT), both in healthy controls (HC) and patients with MS. Then, the performance of the pipeline was evaluated by assessing the CCaA at baseline and 1-year follow-up (scan-rescan test) and evaluating CCaA measurements obtained with brain and cervical T1WI.

MATERIALS AND METHODS

Data Acquisition

An initial set of 10 HC and 21 patients with MS underwent baseline and follow-up brain and cervical spine sagittal 3D MPRAGE. All MRI scans were acquired in a 3T system (Tim Trio; Siemens)

using the following acquisition parameters: TR = 2300 ms, TE = 2.98 ms, TI = 900 ms, flip angle = 9°, voxel size = 1 × 1 × 1mm³; brain FOV = 240 × 256 × 176, cervical FOV = 240 × 25 × 128. Additionally, all subjects underwent a brain 2D FLAIR scan (TR = 9000 ms, TE = 93 ms, TI = 2500 ms, flip angle = 120°, voxel size = 0.49 × 0.49 × 3.0 mm³). The positioning protocol was the same across all subjects. The project was approved by the local ethics committee, and subjects signed an informed consent.

Image Processing

The CCaA was measured in all acquisitions using the following in-house pipeline based on the SCT (Version 5.0.1):[6] First, a segmentation of the cervical cord was obtained with the DeepSeg algorithm. Then, the posterior tip of the C2–C3 intervertebral disc was manually labeled by 2 evaluators (a neurologist with a 7 years' experience and an MRI technician with 11 years' experience). The output from the DeepSeg algorithm, along with these manual intervertebral disc landmarks, was used to normalize the images to the PAM50 atlas,[7] an unbiased multimodal MRI template of the full spinal cord (C1–L2 vertebral level) and brainstem where several spinal cord structures have been predefined. Previously, a spinal canal template covering from C1 to C5 was created by our research group in the same space as the PAM50 atlas and was added to the predefined structures (PAM50_41; Online Supplemental Data). A spinal canal segmentation mask was created in the same space as the atlas and added to the predefined structures, including the spinal canal template. Then, the images were normalized using the inverse normalization matrix, as proposed by SCT, and finally, the spinal canal mask was transferred from the atlas space to the native space (Fig 1).

Additionally, the total intracranial volume was assessed in all subjects using the T1WI sequences with statistical parametric mapping software (SPM; http://www.fil.ion.ucl.ac.uk/spm/software/spm12); the lesion volume was estimated using 2D FLAIR MRI with the Lesion Segmentation Toolbox, included in the SPM software (https://www.applied-statistics.de/lst.html).

Statistical Analysis

CCaA was then estimated as the mean cross-sectional area across either 5, 11, or 17 slices centered on the C2–C3 intervertebral disc, representing the 3 groups of comparisons. Anatomically, 5 slices usually cover the C2–C3 cervical disc, 11 slices cover from the lower margin of C2 to the upper margin of C3; and 17 slices cover from the odontoid basis to the midpoint of the posterior arch of C3 (a certain intersubject variability is detected in those limits according to the individual anatomy).

To identify outlier CCaA estimations, we removed all measures with a value beyond 1.5 times the interquartile range.[8]

Then, CCaA estimations in HC and patients with MS were compared by a multivariable regression model adjusted for age and sex; CCaA estimations from baseline and follow-up cervical cord scans and from brain and cervical MRIs were also compared using a paired t test.

To assess the reproducibility of the proposed pipeline, we compared the CCaA estimations obtained from the cervical cord and brain T1WI at 2 different time points with the proposed pipeline manual segmentations performed by 1 evaluator, considered the

GT, using the Dice similarity coefficient.[9] In addition, a second evaluator manually outlined the CCaA to assess the interoperator variability. Additionally, we compared the CCaA mean obtained with the manual GT at baseline for the cervical cord and brain scans using a paired t test. The GT, considered the reference value, was measured at the midpoint of C2–C3.

Finally, CCaA estimations obtained on baseline and follow-up cervical cord T1WI were compared; brain and cervical cord acquisitions were also compared using the individual and average intraclass correlation coefficient (ICC)[10] and the Bland-Altman method with their limits of agreement (LoA). Statistical analysis was performed with STATA 16.1 software (StataCorp). Before we performed a t test, the normal distribution of different variables was evaluated using the Shapiro-Wilk test, and the homogeneity of variances was determined by the Levene test. To appraise assumptions of linear regression, we checked the normality of residuals using the Shapiro-Wilk test; homoscedasticity was evaluated with the Breusch-Pagan test; independence of observations was determined using the Durbin-Watson test; and collinearity was assessed by the variance inflation factor. The P value for significance was set at $P < .05$.

RESULTS

The proposed pipeline failed in only 3 subjects when using 17 slices to obtain the mean CCaA, because the position of the brain scan was too high and did not cover the upper segment of the cervical cord completely.

After we removed 2 HC and 3 patients with MS, the final cohort included CCaA estimations from 8 HC and 18 patients with MS. Clinical and MRI data are shown in the Table. After we evaluated assumptions of linear regression (Shapiro-Wilk test, $P = .80$; Levene test, $P = .74$; Breusch-Pagan test, $P = .94$; Durbin-Watson test, $P = .84$; and variance inflation factor = 1.07), age- and sex-adjusted linear regression models confirmed that there were no significant differences in the CCaA between HC and patients with MS, estimated in both the cervical cord (mean absolute difference = 0.33 mm², $\beta = 0.10$, $P = .54$) and brain acquisitions (mean absolute difference = 2.18 mm², $\beta = 0.36$, $P = .14$). Consequently, to perform the statistical analysis between different sequences with a larger sample size, we considered HC and patients with MS as a single group (26 subjects).

In the assessment of the reproducibility of the proposed pipeline, the degree of overlap between the CCaA masks generated by the proposed pipeline and the manual GT was excellent with a Dice similarity coefficient mean of 0.90 (range, 0.73–0.97). The distribution across the 4 different acquisitions is shown in Fig 2. Agreement between the 2 evaluators was also excellent, with a Dice similarity coefficient of 0.95 (range, 0.78–1). Furthermore, we did not find significant differences when comparing CCaA estimations obtained with the pipeline and the GT by a t test, either at the baseline cervical cord T1WI (mean absolute difference = 9.56 mm², t [25] = 1.77, $P = .09$) or brain T1WI (mean absolute difference = 6.35 mm², $t[25] = 0.82$, $P = .42$).

When we compared CCaA estimations obtained from baseline and 1-year follow-up cervical cord MRIs, the highest agreement was obtained with 11 and 17 slices (ICC = 0.76; 95% CI, 0.44–0.88, and ICC = 0.78; 95% CI, 0.56–0.90, respectively).
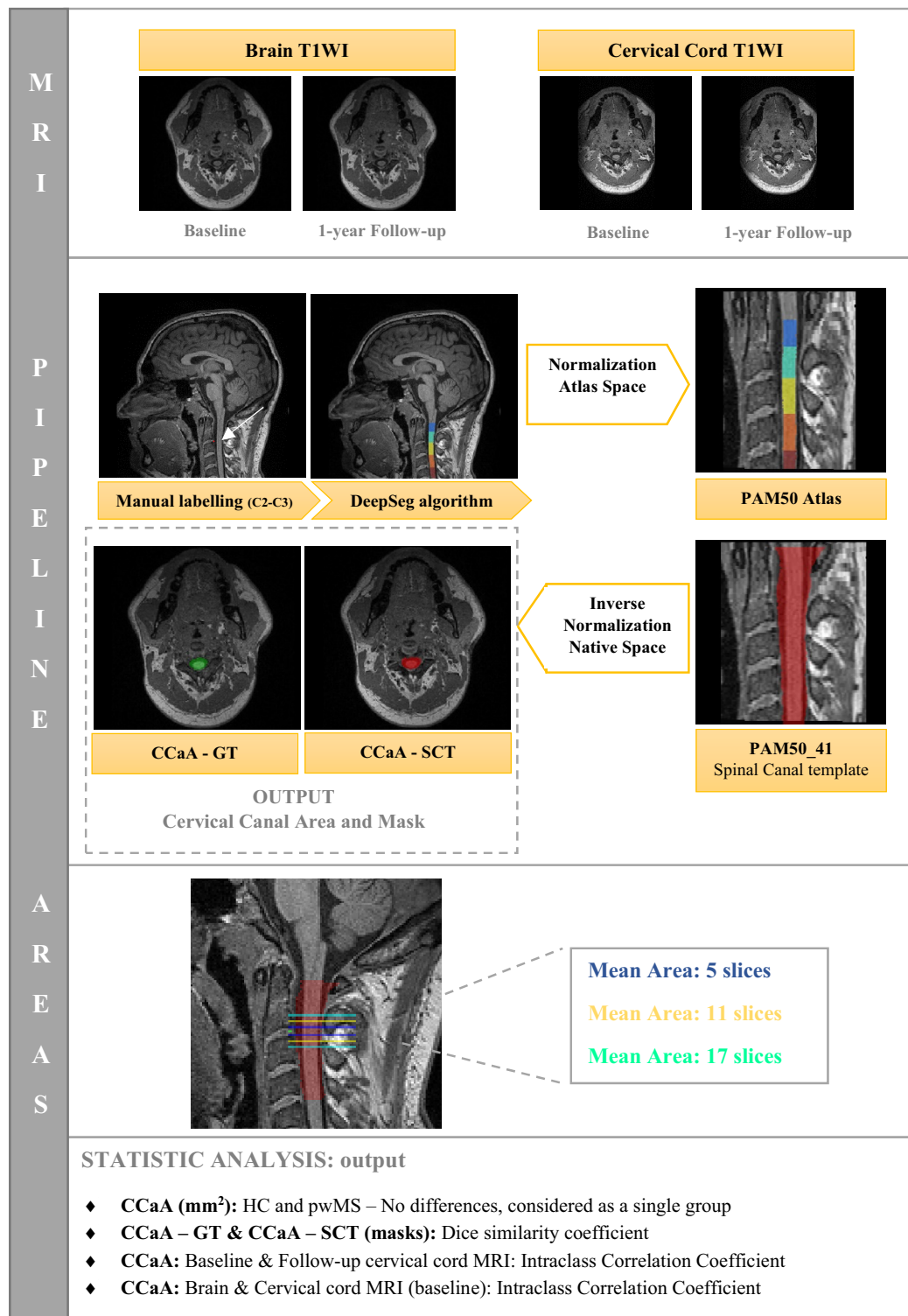
FIG 1. Graphical representation of the proposed pipeline to estimate the cervical canal, including the MRI sequences, a flowchart, the assessment of the mean cervical canal area across the different number of slices, and the statistical analysis performed.

Average ICCs are represented in Fig 3, and they are consistently higher than individual ICCs. Estimations of the CCaA with 17 and 11 slices were also highly similar when using the Bland-Altman method, in contrast to LoA obtained with 5 slices, with a narrower and better-centered LoA (Fig 4, left side). When comparing CCaA estimations obtained from cervical cord T1WI

acquisitions at baseline (mean = 218.37 [SD, 5.02] mm^2) and follow-up (mean = 217.09 [SD, 5.62] mm^2), we did not find significant differences (mean absolute paired difference = 1.28 mm^2, $t[25] = 1.22$, $P = .23$).

CCaA estimations obtained from brain and cervical cord MRIs had a high agreement, independent of the number of slices

Demographic, clinical, and radiologic characteristics of HC and patients with MS

		Patients with MS	
	HC (*n* = 8)	(*n* = 18)	*P* Value[a]
Sex (female) (No.) (%)	5 (62.5%)	11 (61.1%)	.97
Age (mean) (SD) (yr)	30.89 (1.44)	33.84 (1.98)	.36
CCaA (mean) (SD) (mm$_2$)			
Cervical MRI acquisition	218.15 (4.84)	218.47 (5.23)	.73
Brain MRI acquisition	214.57 (3.97)	216.75 (3.47)	.48
Total intracranial volume (mean) (SD) (mL)	1422.3 (0.10)	1392.9 (0.12)	.55
T2 lesion volume (mean) (SD) (mm^3)	–	2.31 (4.09)	–

Note: —Dash indicates not information available; HC: healthy controls; pwMS, patients with multiple sclerosis; SD, standard deviation.

[a] Values correspond to univariate comparisons using parametric and non-parametric tests, as convenience.
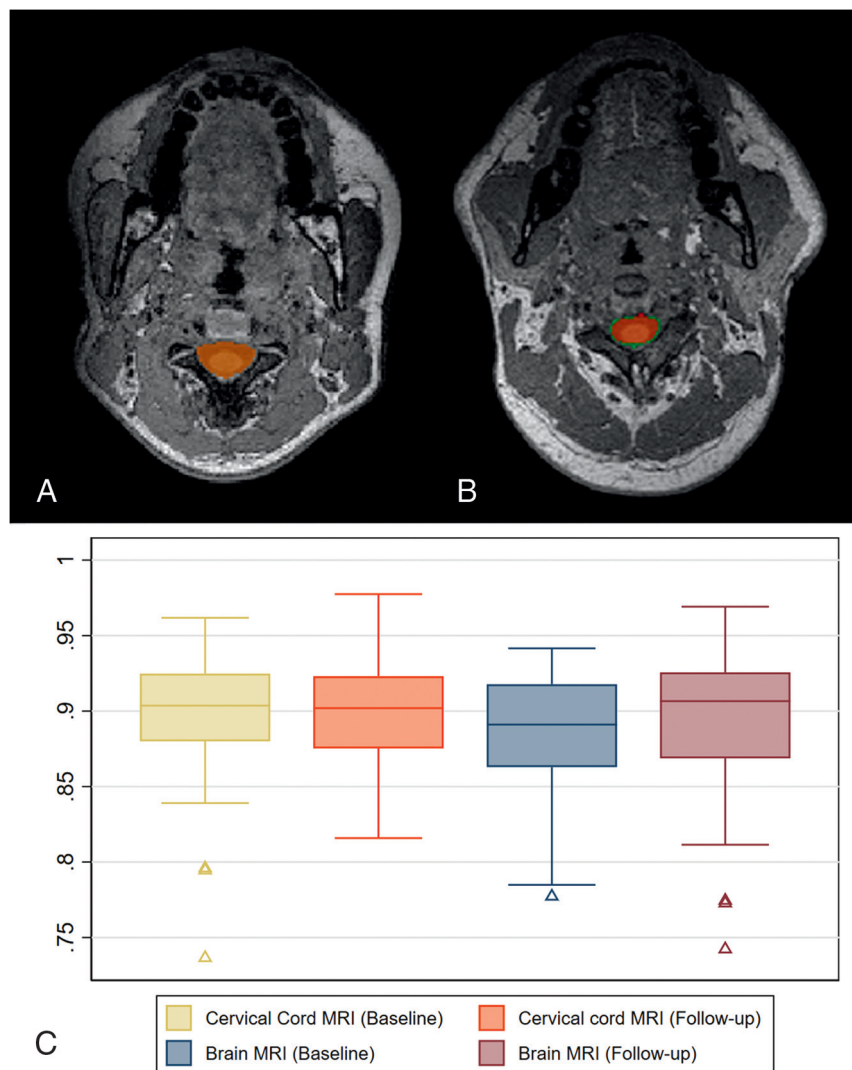

FIG 2. CCaA masks obtained with the proposed pipeline (green) versus the manual segmentation (red) in a patient with MS. *A,* Spinal MRI acquisition shows a Dice similarity coefficient of 0.92. *B,* Brain MRI acquisition shows a Dice similarity coefficient of 0.88. *C,* Distribution of Dice similarity coefficients between CCaA masks from the in-house pipeline and the GT across the 4 acquisitions, both in HC and patients with MS.

used to estimate the CCaA (Fig 3). However, the Bland-Altman method showed a better agreement with CCaA estimations of 17 and 11 slices, than with those obtained with 5 slices (Fig 4). When analyzing absolute means, we found minimal-but-significant differences between CCaA estimations from brain (mean = 216.07 [SD, 3.7] mm^2) and cervical MRIs (mean = 218 [SD, 5.0]2 mm^2) (mean absolute paired difference = 2.30, $t[25]$ = 2.97, *P* = .006).

DISCUSSION

In the present study, we validated a semiautomated segmentation pipeline to estimate the CCaA on the basis of the SCT by comparing the generated masks with a manual GT. The overlap was excellent, and significant differences were not found when comparing both measurement methods, indicating that the proposed pipeline seems to appropriately measure the CCaA. Additionally, we have shown that the CCaA is stable for a 1-year period in all subjects. Finally, the CCaA could be properly estimated using either brain or cervical cord MRIs.

To our knowledge, CCaA variations across time have not been analyzed before, though changes were not expected a priori.[11] We verified its consistency during a 1-year period by assessing the measurement in baseline and follow-up cervical T1WIs. Consequently, the CCaA could be used in future studies as a proxy for the premorbid status of the spinal cord, because stability across time is a prerequisite for such use. Because it is usually done in other cervical cord area measurement methods,[12] we considered more appropriated to calculate the mean area over several sections instead of only in 1 section. An increase in the number of sections used would reduce the variability of the measurement, but in the case of the spinal canal, variability could increase because sections may cover a region where the canal area physiologically increases toward the foramen magnum. To check what number of sections would provide the best compromise, we calculated ICCs using CCaA estimations with the SCT across 5, 7, and 11 slices centered at the midpoint of the C2–C3 vertebral disc. The study showed a good level of concordance between time points, obtaining the highest individual ICC when using 11 and 17 slices for the analysis, compared with 5 slices. We considered that differences in the ICC between the number of slices are related to minor inaccuracies in subject repositioning; hence, the lower the number of slices used
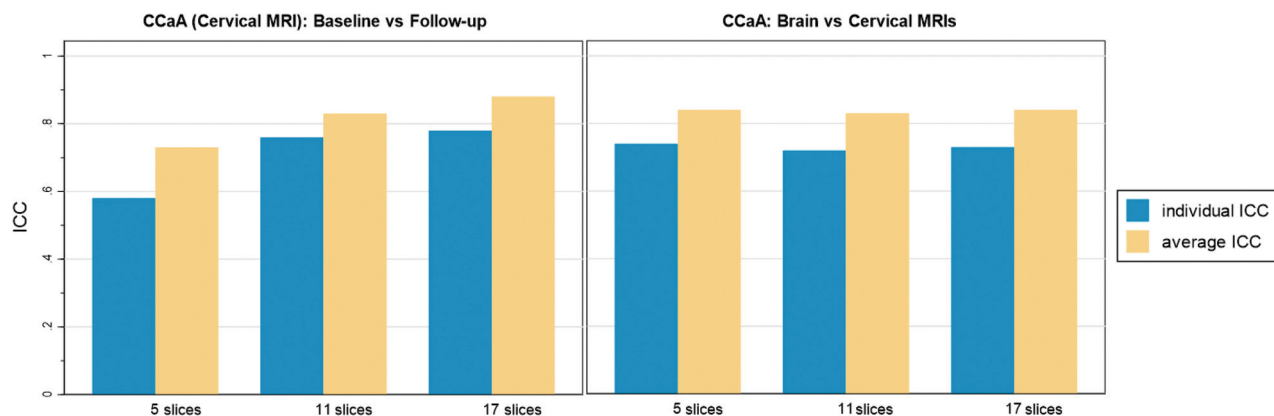
FIG 3. Representation of individual ICC (blue) and average (yellow) ICCs, calculated in 5, 11, and 17 slices. On the left, the ICC between baseline and follow-up cervical MRIs. On the right, degree of concordance of the CCaA analyzed in brain and cervical acquisitions.
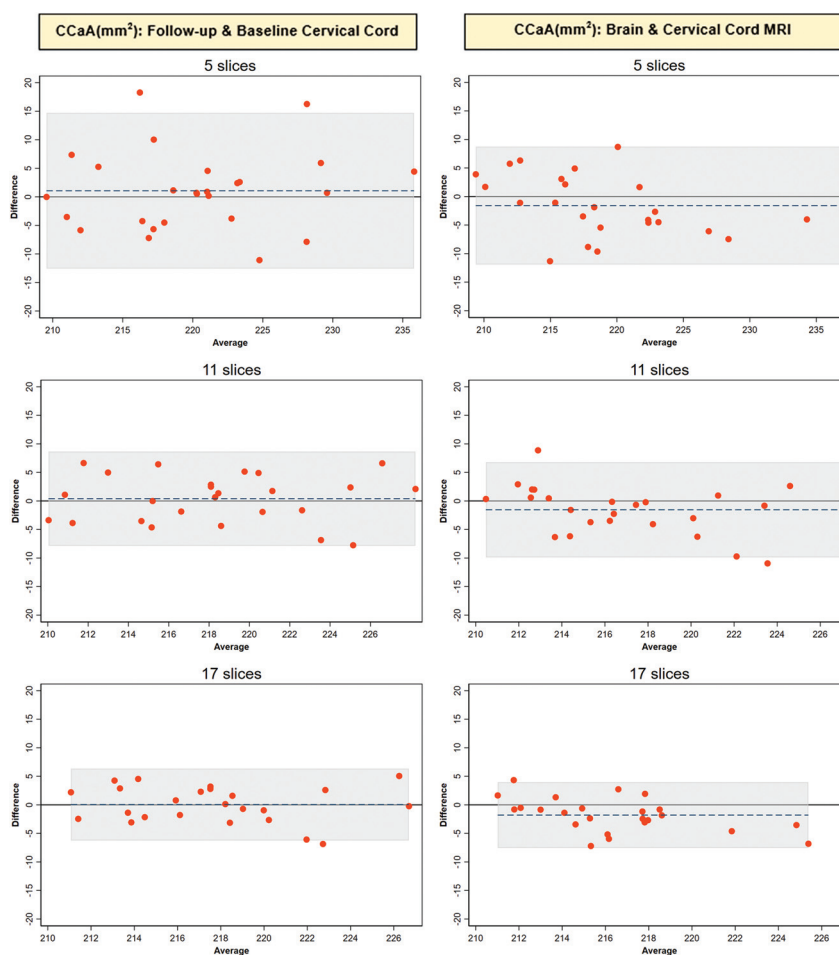

FIG 4. Bland-Altman plots showing the agreement between CCaA estimations assessed in different numbers of slices. On the left, the agreement between baseline and follow-up cervical cord MRI is shown; on the right, between brain and cervical cord MRIs. Notice that the x-axis scale of the plot analyzing CCaA estimations on 5 slices is larger than the others.

to calculate the CCaA, the greater the variability found among patients.

Despite the spinal cord being located in the periphery of the FOV on brain T1WI, where gradient nonlinearity distortion effects are substantial,[13] it has already been proved that it is possible to reliably measure the cervical cord area using brain acquisitions.[14] Therefore, we tested the robustness of CCaA estimations obtained from brain and spine scans, obtaining good agreement between them. Similar ICC values across the different numbers of slices used to calculate the CCaA may be because no repositioning is needed between brain and spine acquisitions.

Overall, ICCs obtained were lower than those reported in other validation studies.[13,14] A possible explanation might be that the individual ICC has been reported instead of the average, which tends to minimize variations and provides higher ICCs. Moreover, although the degree of agreement between CCaA estimations from brain and cervical MRIs was not excellent and there were significant differences between both measurements, the mean difference was inferior to 3 mm^2 in the paired t test analysis. Therefore, results seem to suggest that brain CCaA estimations might be considered when dedicated cervical sequences are not available, though both acquisitions may not be fully interchangeable when analyzing the CCaA of a single subject, possibly because the cervical canal is differently located in the FOV of cervical cord and brain T1WIs. Because the in-house pipeline failed in 3 subjects when using 17 slices, it might be advisable to use the 11-slice approach, which provides similar reproducibility parameters.

Several limitations should be mentioned. First, the sample size is small and could explain the range of ICCs obtained. Second, our pipeline includes manual labeling of the C2–C3 intervertebral disc, which may be a limiting factor when dealing with large cohorts because it clearly increases processing time. Third, we performed the image acquisition with the same scanner and positioning protocol; therefore, we have not tested the pipeline under other conditions. Finally, we adjusted measurements by age and sex, but not height, because normalization using anthropometric parameters still remains controversial.[15,16]

CONCLUSIONS

This study validates a new semiautomated algorithm to estimate the CCaA based on the SCT. An excellent agreement was obtained between the manual segmentations and those provided by the pipeline. We used this algorithm to demonstrate the consistency of CCaA measurements across time, showing no changes during a 1-year period. Finally, results suggested that brain CCaA estimations might be considered when dedicated cervical sequences are not available.

ACKNOWLEDGMENTS

REFERENCES

1. Liu Z, Yaldizli Ö, Pardini M, et al. **Cervical cord area measurement using volumetric brain magnetic resonance imaging in multiple sclerosis**. *Mult Scler Relat Disord* 2015;4:52–57 CrossRef Medline
2. Eden D, Gros C, Badji A, et al. **Spatial distribution of multiple sclerosis lesions in the cervical spinal cord.** *Brain* 2019;142:633–46 CrossRef Medline
3. Arrambide G, Rovira A, Sastre-Garriga J, et al. **Spinal cord lesions: a modest contributor to diagnosis in clinically isolated syndromes but a relevant prognostic factor.** *Mult Scler* 2018;24:301–12 CrossRef Medline
4. Sumowski JF, Rocca MA, Leavitt VM, et al. **Brain reserve against physical disability progression over 5 years in multiple sclerosis.** *Neurology* 2016;86:2006–09 CrossRef Medline
5. Sastre-Garriga J, Rovira A, García-Vidal A, et al. **Spinal cord reserve in multiple sclerosis**. *J Neurol Neurosurg Psychiatry* 2023 Jan 23 [Epub ahead of print] CrossRef Medline
6. 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 CrossRef Medline
7. De Leener B, Fonov VS, Collins DL, et al. **PAM50: unbiased multimodal template of the brainstem and spinal cord aligned with the ICBM152 space.** *Neuroimage* 2018;165:170–79 CrossRef Medline
8. Tukey JW. *Exploratory Data Analysis.* Addison-Wesley; 1977:43–44
9. Zou KH, Warfield SK, Bharatha A, et al. **Statistical validation of image segmentation quality based on a spatial overlap index 1.** *Acad Radiol* 2004;11:178–89 CrossRef Medline
10. Koo TK, Li MY. **A guideline of selecting and reporting intraclass correlation coefficients for reliability research.** *J Chiropr Med* 2016;15:155–63 CrossRef Medline
11. Kato F, Yukawa Y, Suda K, et al. **Normal morphology, age-related changes and abnormal findings of the cervical spine, Part II: magnetic resonance imaging of over 1,200 asymptomatic subjects.** *Eur Spine J* 2012;21:1499–507 CrossRef Medline
12. Rocca MA, Valsasina P, Meani A, et al; MAGNIMS Study Group. **Clinically relevant cranio-caudal patterns of cervical cord atrophy evolution in MS.** *Neurology* 2019;93:e1852–66 CrossRef Medline
13. Papinutto N, Bakshi R, Bischof A, et al; North American Imaging in Multiple Sclerosis Cooperative (NAIMS). **Gradient nonlinearity effects on upper cervical spinal cord area measurement from 3D T1-weighted brain MRI acquisitions.** *Magn Reson Med* 2018;79:1595–601 CrossRef Medline
14. Liu Y, Lukas C, Steenwijk MD, 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;37:749–54 CrossRef Medline
15. Oh J, Seigo M, Saidha S, et al. **Spinal cord normalization in multiple sclerosis.** *J Neuroimaging* 2014;24:577–84 CrossRef Medline
16. Papinutto N, Schlaeger R, Panara V, et al. **Age, gender and normalization covariates for spinal cord gray matter and total cross-sectional areas at cervical and thoracic levels: a 2D phase sensitive inversion recovery imaging study.** *PLoS One* 2015;10:e0118576–18 CrossRef Medline