Improving Accuracy of Brainstem MRI Volumetry: Effects of Age and Sex, and Normalization Strategies

Laura Sander[1,2], Antal Horvath[3], Simon Pezold[3], Simon Andermatt[3], Michael Amann[4], Tim Sinnecker[1,2,4], Maria J. Wendebourg[1,2], Eva Kesenheimer[1,2], Özgür Yaldizli[1,2], Ludwig Kappos[1,2], Cristina Granziera[1,2], Jens Wuerfel[4], Philippe Cattin[3] and Regina Schlaeger[1,2*]

[1] Neurologic Clinic and Policlinic, Departments of Medicine and Clinical Research, University Hospital Basel and University of Basel, Basel, Switzerland, [2] Translational Imaging in Neurology (ThINK) Basel, Department of Biomedical Engineering, University of Basel, Basel, Switzerland, [3] Department of Biomedical Engineering, Center for Medical Image Analysis & Navigation (CIAN), University of Basel, Allschwil, Switzerland, [4] Department of Biomedical Engineering, Medical Image Analysis Center (MIAC AG) and qbig, University of Basel, Basel, Switzerland

Background: Brainstem-mediated functions are impaired in neurodegenerative diseases and aging. Atrophy can be visualized by MRI. This study investigates extrinsic sources of brainstem volume variability, intrinsic sources of anatomical variability, and the influence of age and sex on the brainstem volumes in healthy subjects. We aimed to develop efficient normalization strategies to reduce the effects of intrinsic anatomic variability on brainstem volumetry.

Methods: Brainstem segmentation was performed from MPRAGE data using our deep-learning-based brainstem segmentation algorithm MD-GRU. The extrinsic variability of brainstem volume assessments across scanners and protocols was investigated in two groups comprising 11 (median age 33.3 years, 7 women) and 22 healthy subjects (median age 27.6 years, 50% women) scanned twice and compared using Dice scores. Intrinsic anatomical inter-individual variability and age and sex effects on brainstem volumes were assessed in segmentations of 110 healthy subjects (median age 30.9 years, range 18–72 years, 53.6% women) acquired on 1.5T (45%) and 3T (55%) scanners. The association between brainstem volumes and predefined anatomical covariates was studied using Pearson correlations. Anatomical variables with associations of $|r| > 0.30$ as well as the variables age and sex were used to construct normalization models using backward selection. The effect of the resulting normalization models was assessed by % relative standard deviation reduction and by comparing the inter-individual variability of the normalized brainstem volumes to the non-normalized values using paired t- tests with Bonferroni correction.

Results: The extrinsic variability of brainstem volumetry across different field strengths and imaging protocols was low (Dice scores > 0.94). Mean inter-individual variability/SD of total brainstem volumes was 9.8%/7.36. A normalization based on either total intracranial volume (TICV), TICV and age, or v-scale significantly reduced the inter-individual variability of total brainstem volumes compared to non-normalized volumes and similarly reduced the relative standard deviation by about 35%.

Conclusion: The extrinsic variability of the novel brainstem segmentation method MD-GRU across different scanners and imaging protocols is very low. Anatomic inter-individual variability of brainstem volumes is substantial. This study presents efficient normalization models for variability reduction in brainstem volumetry in healthy subjects.

Keywords: brainstem, covariate, normalization, segmentation, volumetry, MD-GRU

INTRODUCTION

The brainstem as the anatomical and functional link between the cerebrum, the cerebellum, and the spinal cord is a vitally important structure, playing a key role in controlling respiratory and cardiac function, defense reflexes, and awareness. From cranial to caudal, the brainstem is divided into the three substructures mesencephalon, pons, and medulla oblongata. It carries white matter tracts to and from the cerebrum, the spinal cord, and the cerebellum, multiple cranial nerves and reticular nuclei (Nieuwenhuys, 1985; Naidich et al., 2009). While the mesencephalon plays an important role mainly in oculomotor, optic, and acoustic function, the pons contains important white matter tracts as well as cranial nerve nuclei for facial sensory and motor functions (Basinger and Hogg, 2020). The medulla oblongata regulates respiratory function and contains important reflex centers e.g., for coughing and swallowing (Bolser et al., 2015; Ikeda et al., 2017).

Studying the brainstem is crucial for our understanding of both physiologic neurological function and neurological diseases.

Physiologic aging is associated with morphological changes throughout the brain, especially with atrophy of the cortex, deep gray matter structures (Walhovd et al., 2011), and the cerebellum (Woodruff-Pak et al., 2010). Brainstem-mediated functions such as cardiovascular reflexes (Vita et al., 1986), swallowing (Sura et al., 2012), gaze stability (Anson et al., 2016), and auditory function (Skoe et al., 2015) can be altered with increasing age.

Several neurodegenerative diseases as e.g., Alzheimer's (Grinberg et al., 2009) and Parkinson's disease (Grinberg et al., 2009), progressive supranuclear palsy (Williams and Lees, 2009), multisystem atrophy (Ghorayeb et al., 2002), amyotrophic lateral sclerosis (Warabi et al., 2017), and multiple sclerosis (Noseworthy et al., 2000) can affect the brainstem or its substructures. Preferential involvement of distinct brainstem substructures has been observed in several neurodegenerative diseases e.g., mesencephalic atrophy in patients with progressive supranuclear palsy (Cosottini et al., 2007). Progressive degeneration of brainstem structures in these diseases can result in autonomic dysfunction, dysphagia, dysarthria, and other symptoms interfering with the patient's quality of life and potentially also survival (Grinberg et al., 2010; Kim et al., 2018).

Brainstem tissue loss – acquired either during aging or neurodegenerative diseases – can be visualized and quantified by MRI *in vivo*, offering a potential as diagnostic, prognostic, or therapeutic marker in these diseases.

A recently published deep-learning-based algorithm provided accurate, highly reproducible, and robust brainstem segmentation in healthy subjects (HS) and patients with Alzheimer's disease and multiple sclerosis (Andermatt et al., 2016, 2018; Sander et al., 2019).

Despite its successful application in patients, volumetry of the brainstem and its substructures has not yet been systematically assessed in HS. Brainstem volume assessments are subject to inter-individual variation, e.g., due to head size, head position in the scanner, sex or body height. Volume normalization reduces the physiologic inter-individual measurement variation due to individual anatomical effects, ideally without interfering with measurements related to possible disease processes. This allows for better statistical comparison between two inhomogeneous groups, such as healthy controls and patients. A frequently used normalization parameter for brain volumes is the FreeSurfer-derived total intracranial volume (TICV; Whitwell et al., 2001) and SIENAX-derived volumetric scaling factor (v-scale; Fein et al., 2004). So far, normalization covariates for brainstem segmentation have not yet been investigated and relevant normalization factors for brainstem volumetry are not known.

Using a novel fully-automated deep-learning-based segmentation approach, the objectives of this study were to assess:

a) the extrinsic variability of brainstem volumes depending on different scanners, field strengths, and acquisition protocols,
b) the intrinsic anatomical variability and the influence of age and sex on the brainstem and its substructure volumes in HS, and
c) the effects of normalization models on variability in brainstem volumetry.

MATERIALS AND METHODS

Brainstem Segmentation

Brainstem volumes were assessed using a recently published fully-automated segmentation approach based on multi-dimensional gated recurrent units (MD-GRU). The deep-learning-based algorithm provides accurate, robust, and reproducible segmentations of the brainstem and its substructures (Andermatt et al., 2016, 2018; Sander et al., 2019). All segmentations were visually inspected.

Image Acquisition for Assessing the Extrinsic Variability of Brainstem Volumes by Different Scanners and Protocols

Eleven HS (mean age 36.9 years, median age 33.3 years, range 24–56 years, SD 11.2, 7 women) were scanned twice within the

same day on two different scanners (Magnetom Prisma and Skyra, Siemens Healthineers, Erlangen, Germany). Acquisition parameters for all scans were TR = 2,300 ms, TI = 900 ms, TE = 3 ms, α = 9°, spatial resolution of 1 × 1 × 1mm³. Brainstem volumes were obtained from 3D high-resolution T1w MR imaging data (MPRAGE).

Additional 22 HS (mean age 29.7 years, median age 27.6 years, range 20–51 years, SD 8.1, 50% women) were assessed with two different MR protocols, each on the same 1.5T Magnetom Avanto scanner (Siemens Healthineers, Erlangen, Germany). Acquisition parameters of the MPRAGE scans for the first protocol were TR = 2,700 ms, TI = 950 ms, TE = 5 ms, α = 8°, spatial resolution of 1 × 1 × 1 mm³, for the second protocol TR = 2,080 ms, TI = 1,100 ms, TE = 3 ms, α = 15°, spatial resolution of 0.98 × 0.98 × 1 mm³. Brainstem segmentation volumes were obtained for the two protocols.

Finally, MPRAGE images of the same 22 HS acquired on a 1.5T Avanto (acquisition parameters TR = 2,080 ms, TI = 1,100 ms, TE = 3 ms, α = 15°, spatial resolution of 0.98 × 0.98 × 1 mm³) and in addition on a 3T Skyra (TR = 2,300 ms, TI = 900 ms, TE = 3 ms, α = 9°, spatial resolution of 1 × 1 × 1 mm³, parallel imaging, acceleration factor 2) were analyzed to compare brainstem segmentations from both assessments.

Image Acquisition for Assessing the Intrinsic Anatomical Variability

3D high-resolution T1weighted MR imaging data (MPRAGE) were obtained from 110 HS (mean age 34.9 years, median age 30.9 years, range 18–72 years, SD 12.8, 53.6% women) on 1.5T (45%) and 3T (55%) scanners with two different protocols, respectively.

Briefly, acquisition parameters for the 1.5T Magnetom Avanto scanner (Siemens Healthineers, Erlangen, Germany) were TR = 2,080 ms, TI = 1,100 ms, TE = 3 ms, α = 15°, spatial resolution of 0.98 × 0.98 × 1 mm³ (Bendfeldt et al., 2009; Weier et al., 2014) and TR = 2,700 ms, TI = 950 ms, TE = 5 ms, α = 8°, spatial resolution of 1 × 1 × 1 mm³. For the 3T Prisma scanner (Siemens Healthineers, Erlangen, Germany) acquisition parameters were TR = 1,680 ms, TI = 900 ms, TE = 2.5 ms, α = 8°, spatial resolution of 1 × 1 × 1 mm³ and TR = 2,300 ms, TI = 900 ms, TE = 3 ms, α = 9°, spatial resolution of 1 × 1 × 1 mm³, acceleration factor 2.

Written informed consent was obtained from all participants mentioned above.

All brainstem segmentations were visually inspected for anatomic accuracy.

Statistical Analyses

Statistical analyses were performed using JMP Pro 14 and SPSS 25.

Assessing the Extrinsic Variability of Brainstem Volumes

Dice coefficients were each calculated comparing brainstem segmentations obtained in the same individual (a) on different scanners (1.5T vs. 3T), (b) on the same scanner but with different protocols, and (c) on different scanners and protocols.

Assessing the Intrinsic Variability of Brainstem Volumes

To assess the inter-individual variability, the respective deviation from the group mean was calculated for each subject as (measured volume – mean volume)/mean volume.

Age and Sex Effects

Differences in brainstem/brainstem substructure volumes between men and women were assessed using linear regression analysis with (a) age as covariate, as well as in a sensitivity analysis with (b) field strength, (c) acquisition protocol, and (d) TICV as additional covariates, respectively. Correction for multiple testing (4 analyses) was performed using the Bonferroni correction, adjusting the level of significance to $p < 0.05/4$.

The associations between age and brainstem/brainstem substructure volumes were assessed using linear regression analysis covarying for sex. Differences of brainstem volumes in younger vs. older persons (below vs. above the group mean) were assessed using linear regression analysis with (a) field strength and (b) acquisition protocol as additional covariates, respectively.

Assessing Effects of Different Normalization Models on Reduction of Anatomical Variability

Several skull-derived and intracerebral anatomical structures were assessed regarding their potential role as covariates for brainstem volumes. As skull-based parameters, foramen magnum diameter, nasion-opisthion, basion-opisthion (Mc Rae's line), dens length, dens-opisthion, and brainstem angle were assessed (**Figure 1**) using the open-source software 3D Slicer 4.8.1 (www.slicer.org) and GIMP (www.gimp.org). Additionally, v-scale by SIENAX (http://www.fmrib.ox.ac.uk/analysis/research/siena), TICV, gray matter (GM), and white matter (WM) volumes, as well as total brain volume (BV) by FreeSurfer (http://surfer.nmr.mgh.harvard.edu/) were determined.

The associations of these parameters with brainstem volume were first assessed using Pearson correlation coefficients. To correct for multiple tests, Bonferroni correction was performed with a correction factor of $n = 12$ (11 anatomical variables and age) ($p < 0.05/12$).

Only those anatomical metrics showing a significant association with all brainstem and brainstem substructure volumes with a Pearson correlation coefficient of $|r| > 0.30$ (Cohen, 1988), respectively, were considered as potential normalization covariates in further analyses.

We then performed a backward selection procedure starting with a model with total brainstem volume as outcome parameter, and all anatomical variables with a Pearson correlation coefficient of $|r| > 0.30$ in univariate analysis as well as age and sex as predictor variables.

This procedure was performed (a) with TICV and (b) with v-scale separately, as these parameters are co-linear.

The adjusted r^2 of the models resulting from the backward selection were reported, as well as of a further simplified model (considering simple application with preference for fewer and easy to measure covariates).

The normalization of brainstem volumes and its substructure volumes was then performed by using the following equation
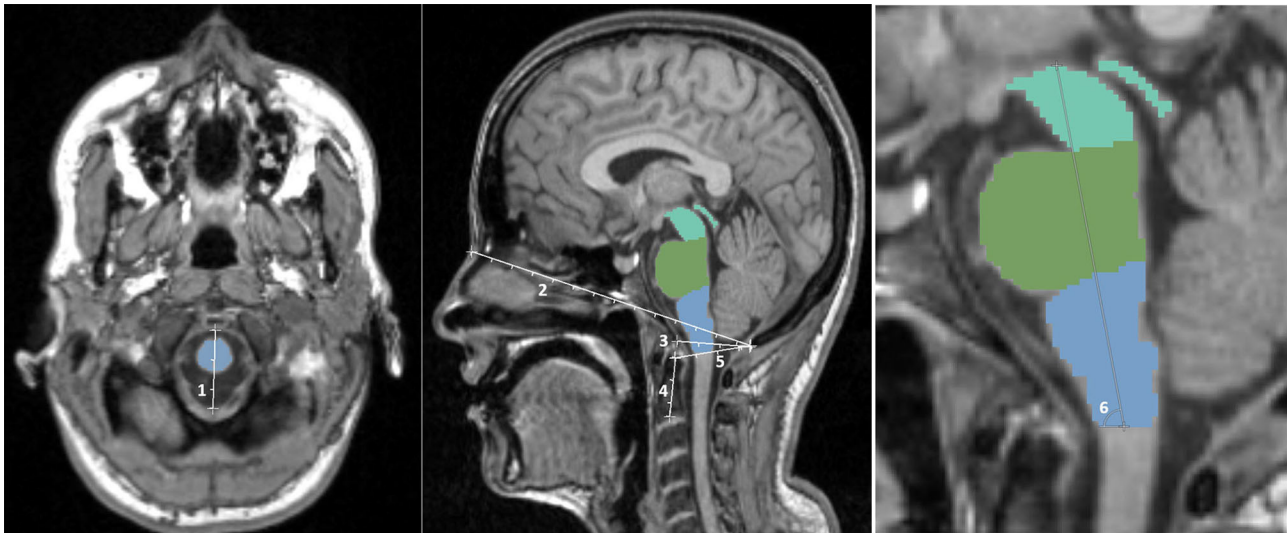
FIGURE 1 | Illustration of the skull-based parameters foramen magnum diameter (1), nasion-opisthion (2), basion-opisthion (3), dens length (4), dens-opisthion (5), and brainstem angle (6).

(Sanfilipo et al., 2004; Papinutto et al., 2019):

$$\text{Volume}_{\text{predicted}} = \text{Volume}_{\text{measured}} + a(X_{\text{mean}} - X_{\text{measured}})$$
$$+ b(Y_{\text{mean}} - Y_{\text{measured}}) + c(Z_{\text{mean}} - Z_{\text{measured}})$$

with a, b, c being the estimates (regression coefficients) obtained by the linear regression analysis and X, Y, Z their measured values.

To assess the performance of the different normalization models, the inter-individual variability of the normalized brainstem volumes of each model was first compared to the variability of the non-normalized brainstem volumes using paired-t-tests, with Bonferroni correction for multiple tests (3 models, $p < 0.05/3$).

In a second step, we compared the performance between the normalization models by comparing the inter-individual variability of the normalized brainstem volumes by a one-way ANOVA (analysis of variance).

The performance of the different normalization models was also expressed by the % relative standard deviation (%RSD) reductions of the predicted volumes to the %RSD of the non-normalized, measured volumes of the whole group ($n = 110$). The relative standard deviation (RSD) is the standard deviation divided by the mean volume.

RESULTS

Brainstem Segmentation
The automated brainstem segmentation approach yielded anatomically accurate results in all subjects in < 200 s/scan on an NVidia GeForce GTX 1080 GPU: All obtained brainstem segmentations were considered anatomically correct in its location and borders, when visually inspected, no manual correction was needed.

Assessing the Influence of Different Scanners and Protocols on Brainstem Volume Variability
The results of brainstem segmentation comparisons from different scanners and protocols are shown in **Table 1**.

MD-GRU derived brainstem segmentations from scans of the same individual obtained on different 3T scanners (Prisma vs. Skyra) using the same acquisition protocol showed Dice scores between 0.95 and 0.98.

Similarly, Dice coefficients comparing segmentations from scans of the same individual using different imaging protocols as specified above on the same 1.5T Avanto scanner were between 0.94 and 0.97.

Dice coefficients of segmentation comparisons using different scanners (1.5T Avanto vs. 3T Skyra) and different protocols were similarly high (0.95–0.98).

Intrinsic Anatomical Variability
Inter-individual Variability of Brainstem Segmentations
Mean % inter-individual variability/SD for the different brainstem volumes were: 9.1/6.40 (mesencephalon), 11.2/8.01 (pons), 10.0/7.23 (medulla oblongata) and 9.8/7.36 (total brainstem).

Age and Sex Influence
Men had significantly larger unadjusted volumes of the total brainstem, mesencephalon, pons, and medulla oblongata (all $p < 0.0001$, respectively), compared to women with total brainstem volumes of 28274.0/2670.1 for men (mean [mm^3]/SD) vs. 24826.2/2824.1 (mean [mm^3]/SD) for women. However, after adjustment for age and TICV (to account for head size differences), men showed significantly larger medulla oblongata volumes compared to women (Appendix A in

TABLE 1 | Mean Dice scores, SD and 95%CI for the total brainstem, mesencephalon, pons, and medulla oblongata volumes comparing segmentations of the same individuals obtained from different scanners, different protocols, and different scanners and protocols.

	Comparison of different scanners	Comparison of different protocols	Comparison of different scanners and protocols
Brainstem			
Mean dice score	0.9736	0.9685	0.9737
SD	0.004	0.003	0.002
95%CI	0.9710–0.9762	0.9670–0.9699	0.9730–0.9743
Mesencephalon			
Mean dice score	0.9567	0.9501	0.9577
SD	0.009	0.007	0.005
95%CI	0.9516–0.9618	0.9473–0.9529	0.9557–0.9597
Pons			
Mean dice score	0.9756	0.9747	0.9760
SD	0.003	0.002	0.002
95%CI	0.9736–0.9776	0.9737–0.9757	0.9753–0.9767
Medulla oblongata			
Mean dice score	0.9496	0.9362	0.9525
SD	0.013	0.012	0.007
95%CI	0.9419–0.9573	0.9311–0.9413	0.9495–0.9554

TABLE 2 | Pearson correlation coefficients of the metrics and volumes of the total brainstem.

	Brainstem Volume	
Variable	p	**Pearson correlation coefficient**
Nasion-opisthion	**< 0.0001**	0.462
Dens length	**< 0.0001**	0.388
TICV	**< 0.0001**	0.748
v-scale	**< 0.0001**	−0.747
Age	0.213	0.120
WM volume	**< 0.0001**	0.788
GM volume	**< 0.0001**	0.559
BV	**< 0.0001**	0.723
Basion-opisthion	0.0139	0.234
Foramen magnum diameter	0.5978	0.051
Dens-opisthion	**0.0006**	0.324
Brainstem angle	0.0112	0.241

p-values surviving the Bonferroni correction (for 12 tests) < 0.0042 are bolded. TICV (total intracranial volume), WM (white matter), GM (grey matter), BV (brain volume).

TABLE 3 | Linear regression analysis with total brainstem volumes as outcome and normalization variables after backward selection.

	Brainstem		
	p	Adj. r^2	Estimate
Model 1a	< 0.0001	0.586	
TICV	< 0.0001		0.0149
Age	0.0036		46.7822
Model 1b	< 0.0001	0.556	
TICV	< 0.0001		0.0146
Model 2	< 0.0001	0.553	
v-scale	< 0.0001		−17,650.89

This finding was independent of additional adjustment for field strength or acquisition protocol (Appendix C in **Supplementary Material**).

Assessing Potential Normalization Models for Anatomical Variability Reduction

Table 2 reports the strength of the correlations of total brainstem volume with each of the investigated variables. Amongst these metrics, nasion-opisthion, dens length, TICV, v-scale, WM, GM, and BV (all normally distributed) showed a significant correlation with brainstem and all substructure volumes surviving the Bonferroni correction for multiple tests with a Pearson correlation coefficient $|r| > 0.30$ (**Table 2** and Appendix D in **Supplementary Material**) and are therefore potential univariate predictors. As BV, GM, and WM volumes can be altered by neurodegenerative processes, these variables were not considered in further analyses.

Pearson correlation coefficients of all variables and brainstem substructure volumes are shown in Appendix D in **Supplementary Material**.

Comparison of Different Normalization Models

The backward selection procedure resulted in two models based on TICV and age (Model 1a) and v-scale (Model 2). The model based on TICV and age was further simplified to TICV alone (Model 1b) (**Table 3**).

Results of the linear regression analysis with brainstem substructure volumes as outcome are shown in Appendix E in **Supplementary Material**.

The model with TICV and age consistently yielded the highest adjusted r^2. However, eliminating the variable age from the Model 1a did not substantially reduce the variance explained. Brainstem volume normalization by TICV (Model 1b) or v-scale (Model 2) yielded comparably high r^2.

Efficiency of the normalization based on Models 1a, 1b, and 2 is reported as the %RSD, the relative %RSD reduction, and the mean % inter-individual variability achieved by the respective models with respect to measured brainstem volumes (**Table 4**). Normalization by Models 1a, 1b, 2 yielded similarly efficient relative %RSD reductions, with a reduction of 33–36% for total brainstem volumes and up to 46% for mesencephalon volumes. Normalization effects on mesencephalon, pons, and

Supplementary Material) with all other comparisons being insignificant after Bonferroni correction. Adjustment for field strength or protocol did not alter these observations.

With adjustment for sex, there was no significant association between age and total brainstem ($p = 0.4131$) volumes. In line with this observation, total brainstem volumes did not differ significantly between older subjects (aged above the group mean of 35 years; $n = 44$) and younger subjects (<35 years; $n = 66$) ($p = 0.3068$) with adjustment for sex. Results were comparable for mesencephalon, pons, and medulla oblongata volumes (Appendix B in **Supplementary Material**).

TABLE 4 | % relative standard deviation (RSD, standard deviation divided by the mean volume), relative %RSD reduction, mean % inter-individual variability [(measured volume – mean volume)/mean volume], and SD of the % inter-individual variability with respect to the measured total brainstem volumes for normalizations based on Models 1a, 1b, and 2.

	Non-normalized	Model 1a	Model 1b	Model 2
Brainstem				
%RSD	12.26	7.82	8.13	8.16
Relative %RSD reduction [%]		36.22	33.69	33.44
Mean % inter-individual variability	9.76	6.26	6.55	6.43
SD	7.36	4.64	4.78	4.98

medulla oblongata volumes are reported in Appendix F in **Supplementary Material**.

The total brainstem volume normalization based on each of the three models significantly reduced the inter-individual variability compared to the non-normalized brainstem volumes (p/ 95% CI of the variability difference for Model 1a $p < 0.0001$/[0.022; 0.048], Model 1b $p < 0.0001$/[0.019; 0.046], Model 2 $p < 0.0001$/[0.007; 0.020]) surviving the Bonferroni correction ($<0.05/3$), respectively. One-way ANOVA comparing the inter-individual variability across all models showed no significant difference between different models for total brainstem ($F_{2, 327} = 0.099$, $p = 0.905$), mesencephalon ($F_{2, 327} = 0.631$, $p = 0.533$), pons ($F_{2, 327} = 0.115$, $p = 0.892$), and medulla oblongata ($F_{2, 327} = 0.187$, $p = 0.829$) volumes.

DISCUSSION

Using a novel, accurate, fully automated, and rapid brainstem segmentation method (Sander et al., 2019) we explored sources of extrinsic (field strength, protocol) as well as intrinsic anatomical variability, investigated age and sex influences on brainstem volumes on high-resolution MPRAGE images in HS and developed potential normalization strategies for variability reduction in brainstem volumetry.

The extrinsic variability of our brainstem volumetry assessment method with respect to different acquisition protocols, hardware, and magnetic field strength was low; the comparisons of brainstem segmentations obtained in the same individuals assessed by different scanners as well as different protocols and both different scanners and protocols yielded very high Dice scores (≥ 0.94). These results confirm the robustness of the applied brainstem segmentation algorithm with respect to different image acquisition settings, i.e., different scanners with 1.5T and 3T field strength and different acquisition protocols.

Consistent with previous studies, our results showed no relevant age dependent volume reduction of the brainstem and its substructures in this cohort aged between 18 and 72 years. With a mean age of 34.9 years and a median age of 30.9 years, this cohort might be, however, more representative for middle-aged and younger adults. Based on the result we cannot

fully exclude a decline in brainstem volume in healthy persons of advanced age.

The lack of an age dependent volume reduction observed in this cohort is consistent with previous studies: Several cross-sectional brainstem segmentation studies based on manual brainstem segmentation reported no association of ventral pons volumes with age (Raz et al., 2001; Sullivan et al., 2004). Likewise, no age effects were found in total brainstem and medulla oblongata volumes (Luft et al., 1999; Lee et al., 2009). Lambert et al. (2013) found isolated midbrain atrophy in HS of age older than 60 years, predominantly due to a volume loss of the superior cerebellar fiber bundles which are not taken into account in our mesencephalon volumetry definition.

In our study, men showed significantly larger unadjusted volumes of the brainstem and its substructures compared to women, which is in line with findings by Raz et al. (2001) and Sullivan et al. (2004). Lee et al. (2009) also reported larger medulla oblongata volumes in men. However, after adjustment for TICV (to account for head size differences) and age, the differences observed between men and women remained only significant for medulla oblongata volumes.

Anatomical variations between HS are an important source of brainstem volume variability with this cohort showing an inter-individual variability of about 10% for brainstem volumes. Therefore, normalization of brainstem volumes is crucial to reduce measurement variation to facilitate the applicability of brainstem volumetrics as a surrogate marker for prognosis, disease course monitoring and therapeutic monitoring in neurodegenerative diseases as e.g., amyotrophic lateral sclerosis, Alzheimer's and Parkinson's disease.

Intracerebral metrics like GM, WM, and BV are expected to be altered by neurodegenerative pathologies, and their potential use as covariates of brainstem volumes might therefore only be adequate in studies involving HS. Hence these parameters were not considered as adequate brainstem normalization parameters.

Models based on FreeSurfer-derived TICV and SIENAX-derived v-scale, two commonly used normalization parameters, as well as TICV and age scored highest adjusted r^2 in linear regression analyses with brainstem and brainstem substructures as outcomes and were therefore further tested as normalization variables.

Normalization for anatomic variation of head size by TICV and age reduced the %RSD of total brainstem volumes by 36%, of mesencephalon volumes up to 46%. Normalization with TICV or v-scale alone showed comparable results.

Brainstem volume normalization based on each of the three normalization models significantly reduced the inter-individual variability compared to the non-normalized volumes. Comparison between the three normalization models showed no significant differences in inter-individual variability of brainstem and brainstem substructure volumes, indicating an equal efficiency of normalization by these models.

TICV and v-scale are frequently applied normalization parameters for brain volumes because in general not affected by neurological/neurodegenerative diseases. By normalization with TICV, inter-individual variation of brain volumes was previously reduced about 4% (Whitwell et al., 2001). Using a similar

methodological approach normalization with v-scale reduced variation in spinal cord volumetry by up to 10.24% (Papinutto et al., 2019).

By reducing measurement variability, we expect the proposed normalization methods to improve the sensitivity in detecting subtle brainstem volume differences between patients with diseases affecting the brainstem and/or its substructures and healthy controls or between patients' subgroups. Thus, previous studies showed improved detection of spinal cord volume differences between multiple sclerosis patients and controls after cervical volume normalization (Oh et al., 2014). Brainstem volume normalization, by reducing anatomical variability, might allow to reveal and strengthen clinical-radiological correlations in neurodegenerative diseases such as multiple sclerosis or Alzheimer's disease (Zhou et al., 2014).

The absence of brainstem volume reductions with increasing age observed in our cross-sectional study is in line with findings in other cross-sectional studies of Raz et al. (2001), Sullivan et al. (2004), and Walhovd et al. (2011). Walhovd et al. reported age-related volume differences in all examined brain structures except the brainstem based on FreeSurfer assessments in a large cohort of HS. To disentangle the exact mechanisms underlying the relative volume preservation of the brainstem with increasing age is beyond the scope of this descriptive study. As a phylogenetically relatively old structure the brainstem is crucial for survival. The reasons for its relative resilience to atrophy compared to other phylogenetically old structures like the hippocampus (Jack et al., 1998; Schröder and Pantel, 2016), amygdala (Kurth et al., 2019) and entorhinal cortex (Hasan et al., 2016) remain unknown. Potential limitations of this study include the underrepresentation of very advanced age and the cross-sectional design that does not allow intra-individual comparisons. Longitudinal studies covering a sufficiently long time-span are difficult to perform, but are certainly necessary to confirm our cross-sectional results in this regard.

The vital function of the brainstem, its clinical involvement in neurodegenerative and neuroinflammatory diseases, and the absence of volume reductions observed in HS aged from 18 to 72 years in this study render atrophy assessments of the brainstem and its substructures an interesting imaging surrogate candidate for the study of neurodegeneration as e.g., in progressive multiple sclerosis. This study analyzed different sources of both extrinsic and intrinsic variability of brainstem volumetry assessments and evaluated normalization models for variability reduction in healthy controls. The inter-individual anatomical variability of total brainstem volumes is relatively high but can be efficiently reduced by 36% using a normalization based on both TICV and age, and by about 34% based on TICV or v-scale alone.

This study's automated segmentation approach proved to be robust across different scanners, field strengths and imaging protocols and allows very fast, efficient, anatomically accurate, and reliable automated brainstem segmentation.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: Upon reasonable request, we will render the detailed results derived from the reported analyses available. Requests to access these datasets should be directed to regina.schlaeger@usb.ch.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethikkommission Nordwest- und Zentralschweiz. The participants provided their written informed consent to participate in these studies.

AUTHOR CONTRIBUTIONS

LS: conceptualization, methodology, analysis, and writing. AH, SP, and SA: methodology and analysis. MA: data collection. TS: analysis. MW and EK: proof reading and analysis. ÖY: data collection and proof reading. LK and CG: proof reading and methodology. JW and PC: supervision, proof reading, and acquiring funding. RS: conceptualization, analysis, writing, supervision, and acquiring funding. All authors contributed to the article and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2020.609422/full#supplementary-material

REFERENCES

Andermatt, S., Pezold, S., and Cattin, P. C. (2016). "Multi-dimensional gated recurrent units for the segmentation of biomedical 3D-data," in *Deep Learning and Data Labeling for Medical Applications. DLMIA 2016, LABELS 2016. Lecture Notes in Computer Science (Vol 10008)*, ed G. Carneiro (Cham: Springer). doi: 10.1007/978–3–319–46976–8_15

Andermatt, S., Pezold, S., and Cattin, P. C. (2018). "Automated segmentation of multiple sclerosis lesions using multi-dimensional gated recurrent units,"

in *Brain Lesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Brain Les 2017. Lecture Notes in Computer Science,* eds A. Crimi, S. Bakas, H. Kuijf, B. Menze, and M. Reyes (Cham: Springer), 10670. doi: 10.1007/978–3–319–75238–9_3

Anson, E. R., Bigelow, R. T., Carey, J. P., Xue, Q. L., Studenski, S., Schubert, M. C., et al. (2016). Aging increases compensatory saccade amplitude in the video head impulse test. *Front Neurol.* 7:113. doi: 10.3389/fneur.2016.00113

Basinger, H., and Hogg, J. P. (eds.). (2020). "Neuroanatomy, brainstem," in *StatPearls* (Treasure Island, FL: StatPearls Publishing).

Bendfeldt, K., Kuster, P., Traud, S., Egger, H., Winklhofer, S., Mueller-Lenke, N., et al. (2009). Association of regional gray matter volume loss and progression of white matter lesions in multiple sclerosis - a longitudinal voxel-based morphometry study. *Neuroimage* 45, 60–67. doi: 10.1016/j.neuroimage.2008.10.006

Bolser, D. C., Pitts, T. E., Davenport, P. W., and Morris, K. F. (2015). Role of the dorsal medulla in the neurogenesis of airway protection. *Pulm. Pharmacol. Ther.* 35, 105–110. doi: 10.1016/j.pupt.2015.10.012

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, 2nd Edn.* Hillsdale, NJ: L. Erlbaum Associates.

Cosottini, M., Ceravolo, R., Faggioni, L., Lazzarotti, G., Michelassi, M. C., Bonuccelli, U., et al. (2007). Assessment of midbrain atrophy in patients with progressive supranuclear palsy with routine magnetic resonance imaging. *Acta Neurol. Scand.* 116, 37–42. doi: 10.1111/j.1600-0404.2006.00767.x

Fein, G., Di Sclafani, V., Taylor, C., Moon, K., Barakos, J., Tran, H., et al. (2004). Controlling for premorbid brain size in imaging studies: T1-derived cranium scaling factor vs. T2-derived intracranial vault volume. *Psychiatr. Res.* 131, 169–176. doi: 10.1016/j.pscychresns.2003.10.003

Ghorayeb, I., Yekhlef, F., Chrysostome, V., Balestre, E., Bioulac, B., and Tison, F. (2002). Sleep disorders and their determinants in multiple system atrophy. *J. Neurol. Neurosurg. Psychiatr.* 72, 798–800. doi: 10.1136/jnnp.72.6.798

Grinberg, L. T., Rüb, U., Ferretti, R. E., Nitrini, R., Farfel, J. M., Polichiso, L., et al. (2009). The dorsal raphe nucleus shows phospho-tau neurofibrillary changes before the transentorhinal region in Alzheimer's disease. A precocious onset? *Neuropathol. Appl. Neurobiol.* 35, 406–416. doi: 10.1111/j.1365-2990.2008.00997.x

Grinberg, L. T., Rueb, U., Alho, A. T., and Heinsen, H. (2010). Brainstem pathology and non-motor symptoms in PD. *J. Neurol Sci.* 289, 81–88. doi: 10.1016/j.jns.2009.08.021

Hasan, K. M., Mwangi, B., Cao, B., Keser, Z., Tustison, N. J., Kochunov, P., et al. (2016). Entorhinal cortex thickness across the human lifespan. *J. Neuroimaging* 26, 278–282. doi: 10.1111/jon.12297

Ikeda, K., Kawakami, K., Onimaru, H., Okada, Y., Yokota, S., Koshiya, N., et al. (2017). The respiratory control mechanisms in the brainstem and spinal cord: integrative views of the neuroanatomy and neurophysiology. *J. Physiol. Sci.* 67, 45–62. doi: 10.1007/s12576-016-0475-y

Jack, C. R. Jr, Petersen, R. C., Xu, Y., O'Brien, P. C., Smith, G. E., Ivnik, R. J., Tangalos, E. G., et al. (1998). Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. *Neurology* 51, 993–999. doi: 10.1212/WNL.51.4.993

Kim, Y., Kim, Y. E., Park, E. O., Shin, C. W., Kim, H. J., and Jeon, B. (2018). REM sleep behavior disorder portends poor prognosis in Parkinson's disease: a systematic review. *J. Clin. Neurosci.* 47, 6–13. doi: 10.1016/j.jocn.2017.09.019

Kurth, F., Cherbuin, N., and Luders, E. (2019). Age but no sex effects on subareas of the amygdala. *Hum. Brain Mapp.* 40, 1697–1704. doi: 10.1002/hbm.24481

Lambert, C., Chowdhury, R., Fitzgerald, T. H., Fleming, S. M., Lutti, A., Hutton, C., et al. (2013). Characterizing aging in the human brainstem using quantitative multimodal MRI analysis. *Front. Hum. Neurosci.* 7:462. doi: 10.3389/fnhum.2013.00462

Lee, N. J., Park, I. S., Koh, I., Jung, T. W., and Rhyu, I. J. (2009). No volume difference of medulla oblongata between young and old Korean people. *Brain Res.* 1276, 77–82. doi: 10.1016/j.brainres.2009.04.027

Luft, A. R., Skalej, M., Schulz, J. B., Welte, D., Kolb, R., Bürk, K., et al. (1999). Patterns of age-related shrinkage in cerebellum and brainstem observed *in vivo* using three-dimensional MRI volumetry. *Cereb Cortex.* 9, 712–721. doi: 10.1093/cercor/9.7.712

Naidich, T. P., Duvernoy, H. M., Delman, B. N., Sorensen, A. G., Kollias, S. S., and Haacke, E. M. (2009). *Duvernoy's Atlas of the Human Brain Stem and Cerebellum.* Wien New York, NY: Springer.

Nieuwenhuys, R. (1985). *Chemoarchitecture of the Brain.* New York, NY: Springer. doi: 10.1007/978-3-642-70426-0

Noseworthy, J. H., Lucchinetti, C., Rodriguez, M., and Weinshenker, B. G. (2000). Multiple sclerosis. *N. Engl. J. Med.* 343, 938–952. doi: 10.1056/NEJM200009283431307

Oh, J., Seigo, M., Saidha, S., Sotirchos, E., Zackowski, K., Chen, M., et al. (2014). Spinal cord normalization in multiple sclerosis. *J. Neuroimaging* 24, 577–584. doi: 10.1111/jon.12097

Papinutto, N., Asteggiano, C., Bischof, A., Gundel, T. J., Caverzasi, E., Stern, W. A., et al. (2019). Intersubject variability and normalization strategies for spinal cord total cross-sectional and gray matter areas. *J. Neuroimaging* 30, 110–118. doi: 10.1111/jon.12666

Raz, N., Gunning-Dixon, F., Head, D., Williamson, A., and Acker, J. D. (2001). Age and sex differences in the cerebellum and the ventral pons: a prospective MR study of healthy adults. *AJNR Am. J. Neuroradiol.* 22, 1161–1167.

Sander, L., Pezold, S., Andermatt, S., Amann, M., Meier, D., Wendebourg, M. J., et al. (2019). Accurate, rapid and reliable, fully automated MRI brainstem segmentation for application in multiple sclerosis and neurodegenerative diseases. *Hum. Brain Mapp.* 40, 4091–4104. doi: 10.1002/hbm.24687

Sanfilipo, M. P., Benedict, R. H., Zivadinov, R., and Bakshi, R. (2004). Correction for intracranial volume in analysis of whole brain atrophy in multiple sclerosis: the proportion vs. residual method. *Neuroimage* 22, 1732–1743. doi: 10.1016/j.neuroimage.2004.03.037

Schröder, J., and Pantel, J. (2016). Neuroimaging of hippocampal atrophy in early recognition of Alzheimer's disease–a critical appraisal after two decades of research. *Psychiatr. Res. Neuroimaging* 247, 71–78. doi: 10.1016/j.pscychresns.2015.08.014

Skoe, E., Krizman, J., Anderson, S., and Kraus, N. (2015). Stability and plasticity of auditorybrainstem function across the lifespan. *Cereb. Cortex* 25, 1415–1426. doi: 10.1093/cercor/bht311

Sullivan, E. V., Rosenbloom, M., Serventi, K. L., and Pfefferbaum, A. (2004). Effects of age and sex on volumes of the thalamus, pons, and cortex. *Neurobiol. Aging.* 25, 185–192. doi: 10.1016/S0197-4580(03)00044-7

Sura, L., Madhavan, A., Carnaby, G., and Crary, M. A. (2012). Dysphagia in the elderly: management and nutritional considerations. *Clin. Interv. Aging.* 7, 287–298. doi: 10.2147/CIA.S23404

Vita, G., Princi, P., Calabro, R., Toscano, A., Manna, L., and Messina, C. (1986). Cardiovascular reflex tests. Assessment of age-adjusted normal range. *J. Neurol. Sci.* 75, 263–274. doi: 10.1016/0022-510X(86)90074-2

Walhovd, K. B., Westlye, L. T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., et al. (2011). Consistent neuroanatomical age-related volume differences across multiple samples. *Neurobiol. Aging* 32, 916–932. doi: 10.1016/j.neurobiolaging.2009.05.013

Warabi, Y., Hayashi, K., Nagao, M., and Shimizu, T. (2017). Marked widespread atrophy of the cerebral cortex and brainstem in sporadic amyotrophic lateral sclerosis in a totally locked-in state. *BMJ Case Rep.* 2017:218952. doi: 10.1136/bcr-2016-218952

Weier, K., Penner, I. K., Magon, S., Amann, M., Naegelin, Y., Andelova, M., et al. (2014). Cerebellar abnormalities contribute to disability including cognitive impairment in multiple sclerosis. *PLoS ONE* 9:e86916. doi: 10.1371/journal.pone.0086916

Whitwell, J. L., Crum, W. R., Watt, H. C., and Fox, N. C. (2001). Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging. *AJNR Am. J. Neuroradiol.* 22, 1483–1489.

Williams, D. R., and Lees, A. J. (2009). Progressive supranuclear palsy: clinicopathological concepts and diagnostic challenges. *Lancet Neurol.* 8, 270–279. doi: 10.1016/S1474-4422(09)70042-0

Woodruff-Pak, D. S., Foy, M. R., Akopian, G. G., Lee, K. H., Zach, J., Nguyen, K. P., et al. (2010). Differential effects and rates of normal aging in cerebellum and hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1624–1629. doi: 10.1073/pnas.0914207107

Zhou, Q., Goryawala, M., Cabrerizo, M., Barker, W., Duara, R., and Adjouadi, M. (2014). Significance of normalization on anatomical mri measures in predicting Alzheimer's disease. *Sci. World J.* 2014:541802. doi: 10.1155/2014/541802