FUNCTIONAL NEURORADIOLOGY

Single-subject analysis of regional brain volumetric measures can be strongly influenced by the method for head size adjustment

Roland Opfer[1] · Julia Krüger[1] · Lothar Spies[1] · Hagen H. Kitzler[2] · Sven Schippling[3] · Ralph Buchert[4]

Abstract

Purpose Total intracranial volume (TIV) is often a nuisance covariate in MRI-based brain volumetry. This study compared two TIV adjustment methods with respect to their impact on *z*-scores in single subject analyses of regional brain volume estimates.

Methods Brain parenchyma, hippocampus, thalamus, and TIV were segmented in a normal database comprising 5059 T1w images. Regional volume estimates were adjusted for TIV using the residual method or the proportion method. Age was taken into account by regression with both methods. TIV- and age-adjusted regional volumes were transformed to *z*-scores and then compared between the two adjustment methods. Their impact on the detection of thalamus atrophy was tested in 127 patients with multiple sclerosis.

Results The residual method removed the association with TIV in all regions. The proportion method resulted in a switch of the direction without relevant change of the strength of the association. The reduction of physiological between-subject variability was larger with the residual method than with the proportion method. The difference between *z*-scores obtained with the residual method versus the proportion method was strongly correlated with TIV. It was larger than one *z*-score point in 5% of the subjects. The area under the ROC curve of the TIV- and age-adjusted thalamus volume for identification of multiple sclerosis patients was larger with the residual method than with the proportion method (0.84 versus 0.79).

Conclusion The residual method should be preferred for TIV and age adjustments of T1w-MRI-based brain volume estimates in single subject analyses.

Keywords MRI · Total intracranial volume · Brain parenchyma · Thalamus · Hippocampus · Age

Abbreviations

BP/BPV/BPF	Brain parenchyma/BP volume (ml)/fraction of BPV relative to the TIV (%)
CoV	Coefficient of variance
HIPP/HIPPV/HIPPF	Hippocampus/bilateral HIPP volume (ml)/fraction of HIPPV relative to the TIV (%)
THAL/THALV/THALF	Thalamus/bilateral THAL volume (ml)/fraction of THALV relative to the TIV (%)
ROC	Receiver operating characteristic
ROI/ROIV/ROIF	Region-of-interest/ROI volume (ml)/fraction of ROIV relative to the TIV (%)
SD	Standard deviation
TIV	Total intracranial volume

✉ Ralph Buchert
r.buchert@uke.de

1 Jung Diagnostics GmbH, Hamburg, Germany

2 Institute of Diagnostic and Interventional Neuroradiology, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

3 Center for Neuroscience Zurich (ZNZ), Federal Institute of Technology (ETH), Multimodal Imaging in Neuroimmunological Diseases (MINDS), University of Zurich, Zurich, Switzerland

4 Department of Diagnostic and Interventional Radiology and Nuclear Medicine, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany

Introduction

MRI-based regional brain volume estimates are increasingly used to support diagnosis and monitoring of neurological and psychiatric diseases. For example, MRI-based hippocampus volume is used as a topographical biomarker to detect or exclude hippocampal degeneration in patients with suspected Alzheimer's disease [1]. In multiple sclerosis (MS), thalamus atrophy detected by MRI is amongst the earliest signs of neurodegeneration and a strong predictor of disability and cognitive impairment [2].

For most regions-of-interest (ROI) in the brain, MRI-based volume estimates (ROIV) are associated with head size [3, 4]. Thus, head size can be a confounder of no interest (nuisance covariate) in between-subject comparisons [3]. Careful adjustment for head size is required in these cases not only to improve the statistical power to detect effects of interest, but also to avoid spurious effects and misinterpretation of actual effects [3, 5, 6]. When considering ROIV estimates across the entire adult age range, the total intracranial volume (TIV) derived from the same T1w-MRI as ROIV is often used to account for head size [3, 7].

Commonly used methods for TIV adjustment of ROIV estimates include the residual method and the proportion method [5, 7–9]. In the residual method, ROIV is regressed on TIV in a sufficiently large, independent group of normal control subjects (normal database). The resulting regression model is used to compute the residuals of ROIV relative to the model in the subjects of interest (usually not included in the normal database). The residuals quantitatively characterize the extent to which each individual subject's ROIV deviates from that predicted for normal subjects with the same TIV. With the proportion method, the ROIV is scaled to the individual TIV to obtain the ROIV-to-TIV ratio, denoted as ROIF in the following ("F" for fraction).

The major advantage of the proportion method is that it is intuitive and straightforward to implement. In particular, it does not require an independent database of normal control subjects. A disadvantage of the proportion method is that it often does not remove the association of ROIV with TIV, whereas the residual method does [5, 7, 9–11].

A further advantage of the residual method is that it easily allows adjustment for other nuisance covariates such as age simultaneously with TIV, simply by adding further terms to the regression model. When using the proportion method for TIV adjustment, adjustment for other nuisance covariates can be achieved by computing residuals of ROIF with respect to a regression model of ROIF in the normal database with the other nuisance variables as explanatory variables. However, often, there is no clear rationale for treating TIV separately and differently from the other nuisance variables. Furthermore, curvilinear (allometric) relationships [4] between regional brain volumes and TIV are easily modelled with the residual method by including logarithmic, quadratic, or higher order polynomial terms in the regression model. The same is true for interaction between different nuisance covariates, e.g., TIV and age, on the regional brain volume of interest, by adding an interaction term to the regression model. Finally, one of the major assumptions of the residual method to be valid, namely that in the normal database the residuals are normally distributed, is easily checked for quality control. There is no analogous quality check of proportions.

The impact of the TIV adjustment method on regional brain volumetry has been studied before. However, previous studies entirely focused on group level analyses to characterize disease-specific effects [10, 12, 13], or the impact of age and gender on brain volumetric measures [4, 8, 14–19], or associations between brain volumetric measures and cognitive performance [11]. The aim of the present study was to evaluate the impact of the TIV adjustment method on z-scores used to support the interpretation of regional brain volume estimates in single subject analyses of individual patients. T1w-MRI-based volume estimates of brain parenchyma (large volume), thalamus (intermediate volume), and hippocampus (small volume) were considered as clinically relevant examples. The regional brain volumes were adjusted for age in addition to TIV, because longitudinal and cross-sectional studies clearly demonstrated age-related shrinkage of the brain independent of the TIV [20, 21].

Materials and methods

Normal database of T1w-MRI scans

The normal database comprised 5059 clinical 3D gradient echo T1w-MR images of the brain from 5059 different patients referred to MR imaging for unspecific symptoms (headache, dizziness). The age of the patients ranged between 20 and 90 years (mean 49.6 ± 16.7 years, the age distribution is shown in Online Supplementary Fig. 1); 57.6% were females (sex was not known in 540 patients). None of the patients had a history of or currently ongoing neurological or psychiatric disease at the time of MR imaging. All images were free of abnormalities beyond those expected for the patients' age based on visual inspection by the local radiologists. Clinical follow-up after MRI was not available. The images had been acquired with 160 different MRI scanners from three different manufacturers (Siemens/Philips/GE: $n = 109/37/14$). Most of the scanners operated at 1.5 T field strength ($n = 116$), the remaining scanners at 3 T ($n = 44$). The MR images had been acquired with scanner-specific sequences as recommended for brain volumetric

analyses by the scanner manufacturers. The slice thickness ranged between 0.9 and 1.3 mm.

The MR images of the normal database had been acquired in clinical routine and were transferred to jung diagnostics GmbH under the terms and conditions of the European general data protection regulation for remote image analysis. Subsequently, the data had been anonymized. The need for written informed consent for the retrospective use of the anonymized data in the present study was waived by the ethics review board of the general medical council of the state of Hamburg, Germany (reference number 2021–300,047-WF).

MS patient dataset

T1w-MR images of 127 MS patients were included retrospectively to compare the two TIV adjustment methods with respect to their impact on the sensitivity to detect thalamus atrophy in MS.

Thirty-three of the MS patients had been enrolled in a clinical study at the Institute of Diagnostic and Interventional Neuroradiology of the University Hospital Carl Gustav Carus, Dresden, Germany (age 42.2 ± 10.1 years, Expanded Disability Status Scale 2.7 ± 1.6, disease duration 5.2 ± 4.8 years). T1w-MR images had been acquired with a 3.0 T Siemens Verio scanner. The remaining 94 MS patients had participated in an observational study at the University Hospital of Zurich, Switzerland (age 37.3 ± 8.9 years, Expanded Disability Status Scale 1.3 ± 1.3, disease duration 2.7 ± 4.5 years). The T1w-MR images had been acquired with a 3.0 T Philips Ingenia scanner. Both studies used a 3D gradient echo T1w-MR sequence.

Both studies had been approved by the local ethics committee. All patients had given written informed consent. This included the retrospective use of the data for the present study. All procedures were in accordance with the 2013 Helsinki declaration.

The MRI data of the MS subsamples have been used previously in a study on age-dependent cutoffs for the detection of pathological deep gray matter volume loss using Jacobian integration [22].

Regional volumetry

A custom three-dimensional convolutional neural network with U-net architecture [23] recently introduced for the segmentation of pathological lesions in T2w-MRI [24, 25] was used to segment the whole brain parenchyma (BP, = gray and white matter), the total intracranial volume (TIV, = BP + inner and outer cerebrospinal fluid), the hippocampi (HIPP), and the thalamus (THAL) in each T1w-MRI. Training and validation of the 3D-CNN for segmentation of these brain structures with respect to stability and accuracy will be described elsewhere. The segmentation

results were used to compute the corresponding volumes in ml in each T1w-MRI (TIV, BPV, THALV, HIPPV). THALV and HIPPV comprised thalamus and hippocampus in both hemispheres.

TIV and age adjustments with the residual method

For the residual method [17–19, 26], the physiological relationship of the regional volume ROIV with TIV and age was modelled in the normal database by

$$
\begin{aligned}
ROIV = &\ a \times TIV^2 \\
&+ b \times age^2 + c \\
&\times TIV \times age + d \\
&\times TIV + e \times age + f.
\end{aligned} \tag{1}
$$

More precisely, the regression coefficients a, b, c, d, e, and f were determined by minimizing the sum of squared differences between the individual ROIV estimates in the normal database and the model according to Eq. 1. Then, the residual of an individual ROIV estimate with respect to the model, denoted resROIV, was computed as

$$
\begin{aligned}
resROIV = &\ ROIV \\
&- \left(a \times TIV^2 + b \times age^2 + c \times TIV \times age + d \times TIV + e \times age + f \right).
\end{aligned} \tag{2}
$$

TIV and age adjustments with the proportion method

First, the individual ROIV was divided by the individual TIV to obtain ROIF, the ROIV-to-TIV fraction [27–29]. Then, adjustment for age was performed by computing residuals of ROIF with respect to the following model

$$
ROIF = r \times age^2 + s \times age + t \tag{3}
$$

The regression coefficients r, s, and t were determined by minimizing the sum of squared differences between the individual ROIF estimates in the normal database and this model. The residual of individual ROIF estimates with respect to this model was computed as

$$
resROIF = ROIF - \left(r \times age^2 + s \times age + t \right). \tag{4}
$$

Outliers

Regression can be affected by outliers. In order to account for this, an iterative two-step approach was used for the regression modelling. After the first regression, scans with

residual < lower quartile − 1.5 × inter-quartile-range or residual > upper quartile + 1.5 × inter-quartile range were considered outliers and were excluded from the second (and final) regression. The identification of outliers was performed separately for the residual method and for the proportion method and separately for each ROI.

Statistical analysis

The coefficient of variance (CoV[%] = 100 × standard deviation/mean) in the normal database was used to characterize physiological between-subject variability of raw and adjusted volume measures. The CoV of residuals was computed using the mean value of the corresponding raw measure as reference (residuals have zero mean).

Adjusted volumes were tested for (residual) correlation with TIV and age by bivariate correlation analysis using Pearson's method.

Z-scores of the residuals were used for standardized characterization of the deviation of TIV- and age-adjusted ROIV from the norm, independent from the adjustment method [30]. Residuals were transformed to z-scores by scaling them to the standard deviation of the residuals in the normal database, that is,

$$z - \text{resROIV} = \text{resROIV}/\text{standard deviation of resROIV} \tag{5}$$

and

$$z - \text{resROIF} = \text{resROIF}/\text{standard deviation of resROIF.} \tag{6}$$

The individual difference between z-resROIF (obtained with the proportion method) and z-resROIV (obtained with the residual method), that is,

$$z - \text{diff} = z - \text{resROIF} - z - \text{resROIV} \tag{7}$$

was used to assess the impact of the adjustment method on the interpretation of regional brain volume estimates in individual subjects. The z-score difference (z-diff) was tested for possible association with TIV and age.

The impact of TIV adjustment on the power to detect MS-associated thalamus atrophy was assessed by comparing the receiver operating characteristic (ROC) curves to separate the MS patients from the control subjects in the normal database between z-resTHALV and z-resTHALF.

Results

The number of outliers of the residual after the first iteration of regression modelling in the normal database was 79 (BP), 71 (THAL), and 104 (HIPP) for the residual method and 64, 52, and 73 for the proportion method. Figure 1 shows the final regression models obtained with the residual method. Scatter plots of ROIF versus TIV are shown in Fig. 2, together with the final regression models of ROIF versus age. All terms except the TIV^2 term (coefficient a) of the final regression model of the residual method contributed significantly.

The distribution of the residuals in the normal database (without the outliers) was Gaussian to good approximation in all cases (Online Supplementary Fig. 2).

Fig. 1 Final regression models of the relationship of ROIV with TIV and age obtained for the residual method (top left: BPV, top right: THALV, bottom: HIPPV)
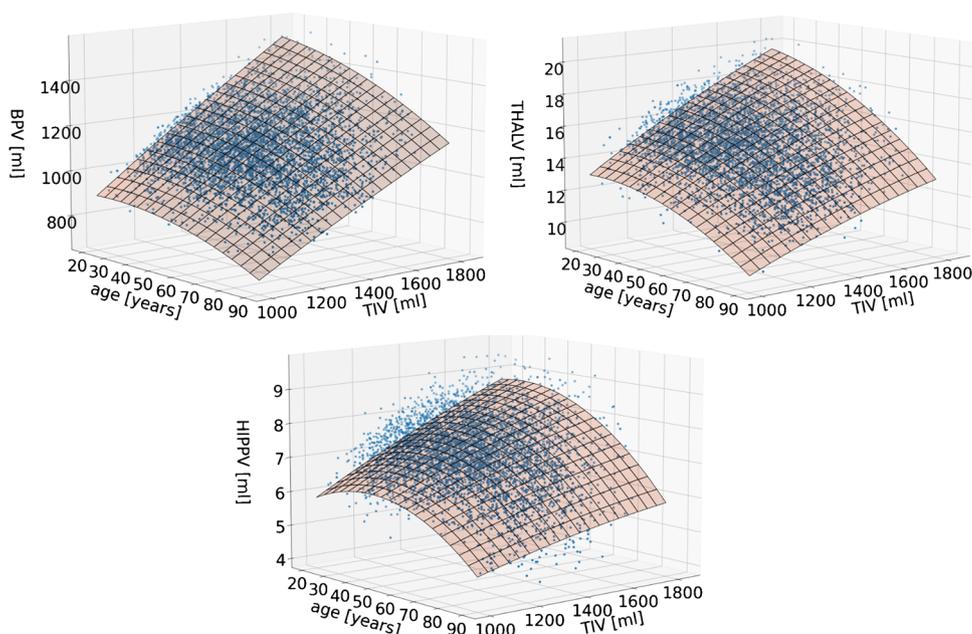
Fig. 2 Scatter plots of ROIF versus TIV (left) and the final regression models of the relationship of ROIF with age for the proportion method (right) (top: BP, middle: THAL, bottom: HIPP)
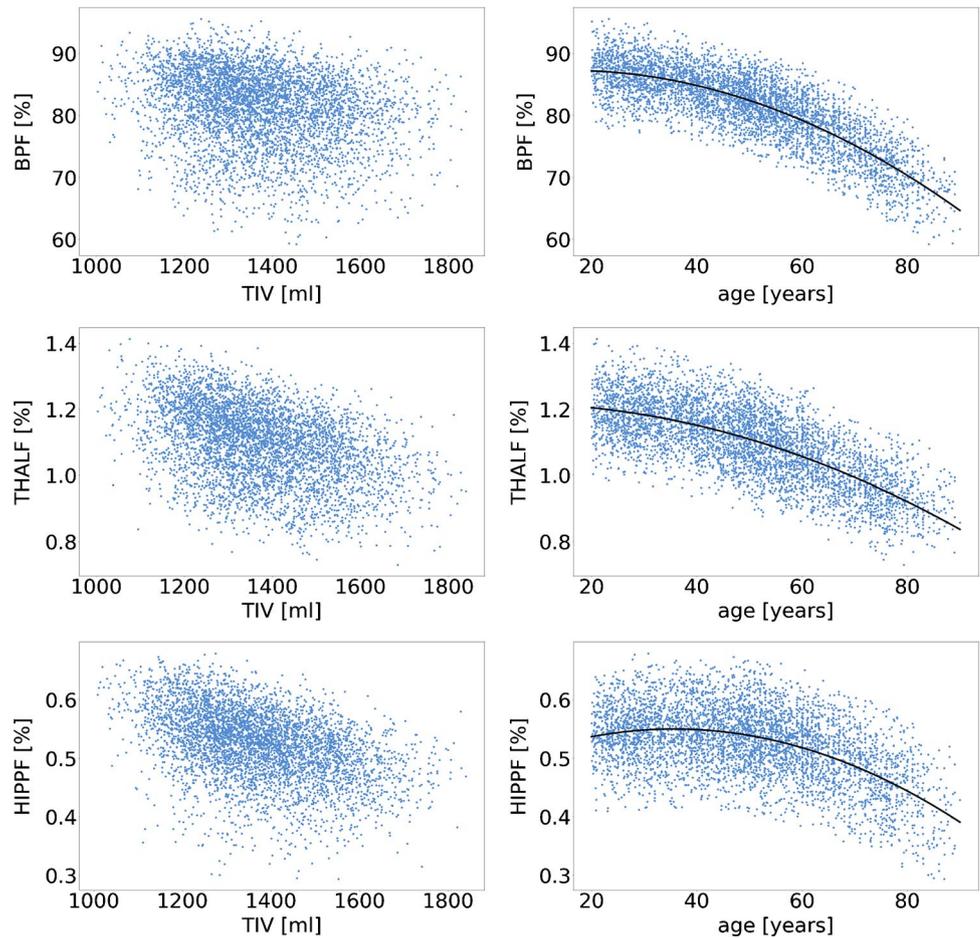
Table 1 Mean value, standard deviation (*SD*), and coefficient of variance (CoV[%] = 100 × *SD*/mean) of adjusted and non-adjusted volume estimates of brain parenchyma (BP), bilateral thalamus (THAL), and bilateral hippocampus (HIPP) in the normal database. The last two columns give the Pearson correlation coefficients with TIV and age. Statistically significant (*P* < .0001) correlations are marked with an asterisk

Residual method						
ROI	Adjustment	Volume (ROIV)			Correlation with	
		Mean [ml]	SD [ml]	CoV [%]	TIV	age
BP (*n* = 4980)	No adjustment	1112.66	117.28	10.54	0.79*	− 0.40*
	After adjustment	0.00	40.96	3.68	0.00	− 0.00
THAL (*n* = 4988)	No adjustment	14.97	1.67	11.17	0.52*	− 0.57*
	After adjustment	0.00	0.89	5.95	0.00	0.00
HIPP (*n* = 4955)	No adjustment	7.16	0.81	11.26	0.42*	− 0.38*
	After adjustment	0.00	0.58	8.16	− 0.00	− 0.00
Proportion method						
ROI	Adjustment	Proportion of TIV (ROIF)			Correlation with	
		Mean [%]	SD [%]	CoV [%]	TIV	Age
BP (*n* = 4995)	No adjustment	81.3	5.45	6.70	− 0.31*	− 0.76*
	After adjustment	0.00	3.28	4.04	− 0.39*	− 0.00
THAL (*n* = 5007)	No adjustment	1.10	0.11	10.41	− 0.42*	− 0.70*
	After adjustment	0.00	0.08	7.14	− 0.52*	0.00
HIPP (*n* = 4986)	No adjustment	0.52	0.06	11.41	− 0.45*	− 0.45*
	After adjustment	0.00	0.05	9.65	− 0.52*	− 0.00

Scaling the regional volume (ROIV) to the TIV to obtain the ROIV-to-TIV fraction (ROIF) for the proportion method resulted in a considerable decrease of the CoV for BP (from 10.54 to 6.70%; Table 1), a small decrease for THAL (from 11.17 to 10.41%), and a small *in*crease of the CoV for HIPP (from 11.26 to 11.41%). Adjustment for age in the proportion method resulted in a decrease of the CoV for all ROIs (BP: from 6.70 to 4.04%, THAL: from 10.41 to 7.14%, HIPP: from 11.41 to 9.65%; Table 1). When comparing TIV- and age-adjusted measures between the residual method and the proportion method, the CoV was lower with the residual method for all ROIs (BP: 3.68% versus 4.04%, THAL: 5.95% versus 7.14%, HIPP: 8.16% versus 9.65%; Table 1).

Raw ROIV was positively correlated with TIV and negatively with age in all ROIs (Table 1). The residual method removed the correlation with TIV and age in all ROIs, whereas the proportion method removed only the correlation

with age (Table 1). The sign of the correlation with TIV was switched by the proportion method in all ROIs (Table 1).

The distribution of z-diff, the individual difference of the regional brain volume z-scores with the proportion method minus the residual method, in the normal database is shown in Online Supplementary Fig. 3. Mean, 95th percentile, and maximum of the absolute value of z-diff are given in Table 2. The maximum z-score difference was larger or equal to two z-value points in all ROIs. Scatter plots of z-diff versus TIV and versus age are shown in Fig. 3. There was a strong correlation between z-diff and TIV in all ROIs ($r < -0.92$, $p < 0.001$). Relative to the residual method, the proportion method resulted in overestimation of individual z-scores at small TIV and in underestimation of z-scores at large TIV. z-diff was not associated with age (Fig. 3).

The ROC curves of z-resTHALV and z-resTHALF for the discrimination of MS patients from the control subjects in the normal database are shown in Online Supplementary Fig. 4. The area under the ROC curve was 0.84 for z-resTHALV and 0.79 for z-resTHALF.

Table 2 Absolute difference of regional brain volume z-scores with the proportion method minus the residual method (abs(z-diff)) in the normal database

ROI	Mean	95th percentile	Max
BPV	0.34	0.86	2.00
THALV	0.46	1.11	2.05
HIPPV	0.44	1.09	2.20

Discussion

The residual method was superior to the proportion method with respect to reduction of physiological between-subject variability and with respect to removing the association with
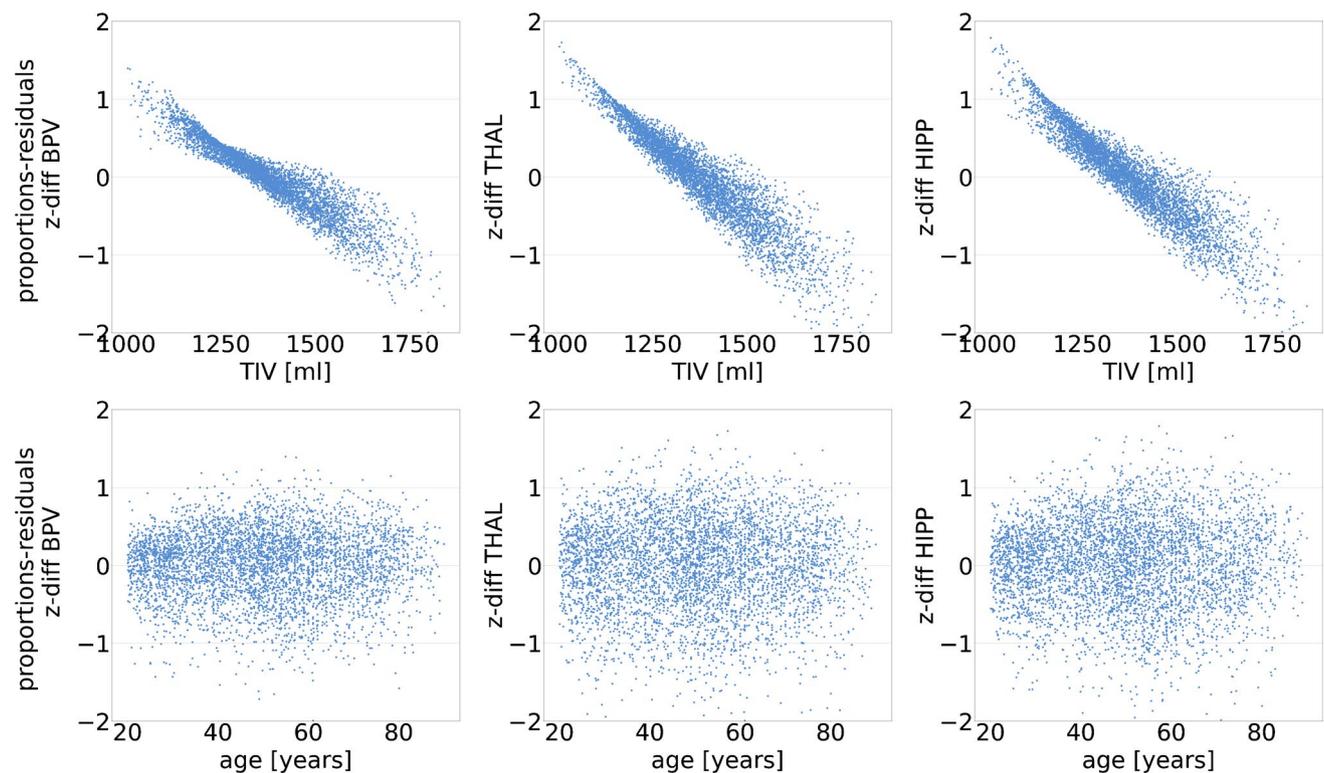

Fig. 3 Scatter plots of z-diff versus TIV (top) and versus age (bottom) (left: BP, middle: THAL, right: HIPP)

TIV of the considered regional brain volumes. The first suggests improved statistical power with the residual method, which was supported by a higher area under the ROC curve with the residual method when testing the thalamus volume for the discrimination of MS patients from control subjects (Online Supplementary Fig. 4). The latter suggests an increased risk of systematic misinterpretation of MRI-based volume estimates with the proportion method, depending on the subject's TIV.

The mean difference of the z-scores from the two adjustment methods ranged between 0.34 and 0.46 z-score points, depending on the ROI (Table 2), which might appear rather small. However, the z-score difference was larger than one z-score point in about 5% of the 5059 scans in each ROI (Table 2). The maximum z-score difference observed in this sample was about two z-score points in each ROI (Table 2). Very similar findings were obtained for hippocampus volumetry using two alternative segmentation methods suggesting that the findings are not specific for segmentation with the custom convolutional neuronal network used in this study but hold more generally (section "Impact of the segmentation method" in the Online Supplementary material). Given that typical cutoffs used for the detection of regional brain atrophy range between -1.5 and -2.5 on the z-score, a systematic error of one z-score point or more in about 5% of the subjects questions the use of the proportion method in single subject analyses of MRI-based brain volumetry in clinical routine, although it is still widely used [31, 32].

The findings of the present study are in line with previous studies demonstrating that the proportion method often does not remove the association of ROIV estimates with TIV [5, 7, 9–11]. The proportion method implicitly assumes that in healthy subjects, ROIV is proportional to the TIV, that is, ROIV = slope × TIV. In this case, the fraction ROIF of ROIV relative to the TIV is constant (independent of TIV): ROIF = ROIV/TIV = slope. The fact that the proportion method did not remove the association between ROIV and TIV in the present study suggests that the assumption of proportionality between ROIV and TIV was violated in the patient data by a non-zero intercept: ROIV = slope × TIV + intercept [4]. In this case, it is ROIF = ROIV/TIV = slope + intercept/TIV. Thus, a positive intercept results in an inverse association between ROIF and TIV (smaller ROIF at larger TIV) as observed in this study (Fig. 2) [11, 12]. The sign flip of the correlation with TIV was observed for all ROIs. For the HIPP, the proportion method even caused a slight increase of the correlation strength (correlation coefficient from 0.42 to -0.52; Table 1). Thus, if the goal is to eliminate the impact of variable TIV on regional brain volume measures, the residual method is strongly preferred over the proportion method [11].

A secondary finding of the present study was the statistical significance of the TIV × age interaction term in the regression model used for the residual method (Eq. 1), indicating that the relationship between ROIV and TIV depends on age in all considered ROIs. Evaluating the final regression model of the residual method at different fixed ages demonstrated a steeper relationship between ROIV and TIV in younger subjects compared to older subjects (Online Supplementary Fig. 5). TIV × age interaction cannot be taken into account with the proportion method. The same is true for the residual method when applied stepwise, that is, when TIV is regressed out first and the resulting residuals are subsequently regressed over age [21, 33]. This stepwise approach is expected to result in overestimation of z-scores in young subjects (reduced sensitivity to detect volume loss) and in underestimation of z-scores in old subjects (increased rate of false positive findings).

Another secondary finding was the statistical significance of the quadratic age term in the regression model used for the residual method (Eq. 1). This indicates that a simple linear model is not sufficient to adequately describe the age dependency of regional brain volume over the entire range of adult age, in line with previous studies [22, 33]. Non-linear age effects can be taken into account with both the residuals and the proportion method. In this study, a quadratic age term was included in the regression model of both methods in order to avoid bias that might have been caused by including this term in only one method (Eqs. 1 and 3).

Analysis of covariance is also widely used for TIV adjustment with ROIV as dependent variable and TIV and other nuisance covariates as predictors together with the factors of primary interest such as disease status for example [7, 34]. However, analysis of covariance is designed for the comparison of two or more groups and, therefore, was not considered in this study on the impact of TIV adjustment on single-subject analyses.

Matching subjects with respect to all relevant nuisance variables might be considered the standard-of-truth to account for them. This is feasible in group comparisons [15, 18]. However, in single-subject analysis, this would require a huge normal database that can be stratified with respect to TIV and age such that each combination of TIV and age encountered in clinical routine is covered with adequate sample size. Such databases are currently not available at most sites.

Several parameters have been proposed to account for between-subject variability of head size in studies on regional brain volumes in addition to the TIV including body height, head circumference, and total brain volume (excluding ventricles) [7]. Which parameter is most appropriate depends on the question of interest [7]. When considering regional brain volumes across the entire adult age range and/or to support the diagnosis of neurodegenerative diseases, the TIV usually is preferred, because it is rather independent of age and atrophy and, therefore, can be considered

an index of maximum brain volume across the individual's entire adult life span [3, 7, 35, 36]. This was the rationale for using the TIV for head size adjustment in the present study. Furthermore, there are numerous software tools for automatic estimation of the TIV from the same high-resolution T1w-MRI of the brain (as used for regional volumetry) with high accuracy and precision. Thus, reliable TIV estimates are easily available to account for head size with only moderate amplification of noise (measurement error).

Hyperostosis frontalis interna, characterized by bony nodules situated on the inner lamina of the frontal (and parietal) bone, causes underestimation of the head size by the TIV. Hyperostosis frontalis interna is rather frequent in post-menopausal females but rarely found in males [37]. Repeat analyses performed separately in females and males suggested that the potential impact of limitations of TIV estimates associated with hyperostosis frontalis interna on the primary findings of the present study was small (section "Potential impact of hyperostosis frontalis interna" in the Online Supplementary material).

A limitation of the present study is that the interpretation of individual terms in the regression models might be limited by multicollinearity between the terms. However, this does not affect the utility of the full regression models to reduce between-subject variability associated with TIV and age.

Conclusion

The residual method should be preferred over the proportion method for TIV and age adjustments of T1w-MRI-based brain volume estimates for the detection or exclusion of regional brain atrophy in single subjects in clinical routine.

Declarations

The MR images of the normal database had been transferred to jung diagnostics GmbH under the terms and conditions of the European general data protection regulation for remote image analysis. Subsequently, the data had been anonymized. The need for written informed consent for the retrospective use of the anonymized data in the present study was waived by the ethics review board of the general medical council of the state of Hamburg, Germany (reference number 2021–300047-WF). The MS patients had been enrolled in clinical studies at the Institute of Diagnostic and Interventional Neuroradiology of the University Hospital Carl Gustav Carus, Dresden, Germany, and at the University Hospital of Zurich, Switzerland. Both studies had been approved by the local ethics committee. All patients had given written informed consent. This included the retrospective use of the data for the present study. All procedures were in accordance with the 2013 Helsinki declaration.

Competing interests Roland Opfer, Julia Krüger, and Lother Spies are full-time employees of jung diagnostics GmbH. The other authors have no competing interests to declare that are relevant to the content of this article.

References

1. Teipel S, Kilimann I, Thyrian JR, Kloppel S, Hoffmann W (2018) Potential role of neuroimaging markers for early diagnosis of dementia in primary care. Curr Alzheimer Res 15:18–27. https://doi.org/10.2174/1567205014666170908093846

2. Amin M, Ontaneda D (2020) Thalamic injury and cognition in multiple sclerosis. Front Neurol 11:623914. https://doi.org/10.3389/fneur.2020.623914

3. Barnes J, Ridgway GR, Bartlett J, Henley SM, Lehmann M, Hobbs N, Clarkson MJ, MacManus DG, Ourselin S, Fox NC (2010) Head size, age and gender adjustment in MRI studies: a necessary nuisance? Neuroimage 53:1244–1255. https://doi.org/10.1016/j.neuroimage.2010.06.025

4. Williams CM, Peyre H, Toro R, Ramus F (2021) Neuroanatomical norms in the UK Biobank: the impact of allometric scaling, sex, and age. Hum Brain Mapp 42:4623–4642. https://doi.org/10.1002/hbm.25572

5. Mathalon DH, Sullivan EV, Rawles JM, Pfefferbaum A (1993) Correction for head size in brain-imaging measurements. Psychiatry Res 50:121–139. https://doi.org/10.1016/0925-4927(93)90016-b

6. Whitwell JL, Crum WR, Watt HC, Fox NC (2001) Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging. AJNR Am J Neuroradiol 22:1483–1489

7. O'Brien LM, Ziegler DA, Deutsch CK, Kennedy DN, Goldstein JM, Seidman LJ, Hodge S, Makris N, Caviness V, Frazier JA, Herbert MR (2006) Adjustment for whole brain and cranial size in volumetric brain studies: a review of common adjustment factors and statistical methods. Harv Rev Psychiatry 14:141–151. https://doi.org/10.1080/10673220600784119

8. Sullivan EV, Rosenbloom MJ, Desmond JE, Pfefferbaum A (2001) Sex differences in corpus callosum size: relationship to age and intracranial size. Neurobiol Aging 22:603–611. https://doi.org/10.1016/s0197-4580(01)00232-9

9. Arndt S, Cohen G, Alliger RJ, Swayze VW 2nd, Andreasen NC (1991) Problems with ratio and proportion measures of imaged cerebral structures. Psychiatry Res 40:79–89. https://doi.org/10.1016/0925-4927(91)90031-k

10. Sanfilipo MP, Benedict RH, Zivadinov R, 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. https://doi.org/10.1016/j.neuroimage.2004.03.037

11. Van Petten C (2004) Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: review and meta-analysis. Neuropsychologia 42:1394–1413. https://doi.org/10.1016/j.neuropsychologia.2004.04.006

12. Voevodskaya O, Simmons A, Nordenskjold R, Kullberg J, Ahlstrom H, Lind L, Wahlund LO, Larsson EM, Westman E, Alzheimer's Disease Neuroimaging I (2014) The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. Front Aging Neurosci 6:264. https://doi.org/10.3389/fnagi.2014.00264

13. Ma D, Popuri K, Bhalla M, Sangha O, Lu D, Cao J, Jacova C, Wang L, Beg MF, Alzheimer's Disease Neuroimaging I (2019) Quantitative assessment of field strength, total intracranial volume, sex, and age effects on the goodness of harmonization for volumetric analysis on the ADNI database. Hum Brain Mapp 40:1507–1527. https://doi.org/10.1002/hbm.24463

14. Greenberg DL, Messer DF, Payne ME, Macfall JR, Provenzale JM, Steffens DC, Krishnan RR (2008) Aging, gender, and the elderly adult brain: an examination of analytical strategies. Neurobiol Aging 29:290–302. https://doi.org/10.1016/j.neurobiolaging.2006.09.016

15. Pintzka CW, Hansen TI, Evensmoen HR, Haberg AK (2015) Marked effects of intracranial volume correction methods on sex differences in neuroanatomical structures: a HUNT MRI study. Front Neurosci 9:238. https://doi.org/10.3389/fnins.2015.00238

16. Perlaki G, Orsi G, Plozer E, Altbacker A, Darnai G, Nagy SA, Horvath R, Toth A, Doczi T, Kovacs N, Bogner P, Schwarcz A, Janszky J (2014) Are there any gender differences in the hippocampus volume after head-size correction? A volumetric and voxel-based morphometric study. Neurosci Lett 570:119–123. https://doi.org/10.1016/j.neulet.2014.04.013

17. Nordenskjold R, Malmberg F, Larsson EM, Simmons A, Ahlstrom H, Johansson L, Kullberg J (2015) Intracranial volume normalization methods: considerations when investigating gender differences in regional brain volume. Psychiatry Res 231:227–235. https://doi.org/10.1016/j.pscychresns.2014.11.011

18. Sanchis-Segura C, Ibañez-Gual MV, Adrián-Ventura J, Aguirre N, Gómez-Cruz ÁJ, Avila C, Forn C (2019) Sex differences in gray matter volume: how many and how large are they really? Biol Sex Differ 10:32. https://doi.org/10.1186/s13293-019-0245-7

19. Sanchis-Segura C, Ibañez-Gual MV, Aguirre N, Cruz-Gómez ÁJ, Forn C (2020) Effects of different intracranial volume correction methods on univariate sex differences in grey matter volume and multivariate sex prediction. Sci Rep 10:12953. https://doi.org/10.1038/s41598-020-69361-9

20. Scahill RI, Frost C, Jenkins R, Whitwell JL, Rossor MN, Fox NC (2003) A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. Arch Neurol 60:989–994. https://doi.org/10.1001/archneur.60.7.989

21. Schippling S, Ostwaldt A-C, Suppa P, Spies L, Manogaran P, Gocke C, Huppertz H-J, Opfer R (2017) Global and regional annual brain volume loss rates in physiological aging. Journal of Neurology:1–9. https://doi.org/10.1007/s00415-016-8374-y

22. Opfer R, Kruger J, Spies L, Hamann M, Wicki CA, Kitzler HH, Gocke C, Silva D, Schippling S (2020) Age-dependent cut-offs for pathological deep gray matter and thalamic volume loss using Jacobian integration. Neuroimage Clin 28:102478. https://doi.org/10.1016/j.nicl.2020.102478

23. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp 234–241

24. Krüger J, Opfer R, Gessert N, Ostwaldt AC, Manogaran P, Kitzler HH, Schlaefer A, Schippling S (2020) Fully automated longitudinal segmentation of new or enlarged multiple sclerosis lesions using 3D convolutional neural networks. NeuroImage Clinical 28:102445. https://doi.org/10.1016/j.nicl.2020.102445

25. Krüger J, Ostwaldt AC, Spies L, Geisler B, Schlaefer A, Kitzler HH, Schippling S, Opfer R (2021) Infratentorial lesions in multiple sclerosis patients: intra- and inter-rater variability in comparison to a fully automated segmentation using 3D convolutional neural networks. Eur Radiol. https://doi.org/10.1007/s00330-021-08329-3

26. Kijonka M, Borys D, Psiuk-Maksymowicz K, Gorczewski K, Wojcieszek P, Kossowski B, Marchewka A, Swierniak A, Sokol M, Bobek-Billewicz B (2020) Whole brain and cranial size adjustments in volumetric brain analyses of sex- and age-related trends. Front Neurosci 14:278. https://doi.org/10.3389/fnins.2020.00278

27. Rudick RA, Fisher E, Lee JC, Simon J, Jacobs L (1999) Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing-remitting MS. Multiple Sclerosis Collaborative Research Group Neurology 53:1698–1704

28. Reite M, Reite E, Collins D, Teale P, Rojas DC, Sandberg E (2010) Brain size and brain/intracranial volume ratio in major mental illness. BMC Psychiatry 10:79. https://doi.org/10.1186/1471-244x-10-79

29. Fotenos AF, Snyder AZ, Girton LE, Morris JC, Buckner RL (2005) Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. Neurology 64:1032–1039. https://doi.org/10.1212/01.wnl.0000154530.72969.11

30. O'Brien PC, Dyck PJ (1995) Procedures for setting normal values. Neurology 45:17–23. https://doi.org/10.1212/wnl.45.1.17

31. Louis S, Morita-Sherman M, Jones S, Vegh D, Bingaman W, Blumcke I, Obuchowski N, Cendes F, Jehi L (2020) Hippocampal sclerosis detection with NeuroQuant compared with neuroradiologists. AJNR Am J Neuroradiol 41:591–597. https://doi.org/10.3174/ajnr.A6454

32. Tommasin S, Cocozza S, Taloni A, Gianni C, Petsas N, Pontillo G, Petracca M, Ruggieri S, De Giglio L, Pozzilli C, Brunetti A, Pantano P (2021) Machine learning classifier to identify clinical and radiological features relevant to disability progression in multiple sclerosis. J Neurol 268:4834–4845. https://doi.org/10.1007/s00415-021-10605-7

33. Fjell AM, Walhovd KB, Westlye LT, Ostby Y, Tamnes CK, Jernigan TL, Gamst A, Dale AM (2010) When does brain aging accelerate? Dangers of quadratic fits in cross-sectional studies. Neuroimage 50:1376–1383. https://doi.org/10.1016/j.neuroimage.2010.01.061

34. O'Brien LM, Ziegler DA, Deutsch CK, Frazier JA, Herbert MR, Locascio JJ (2011) Statistical adjustments for brain size in volumetric neuroimaging studies: some practical implications in methods. Psychiatry Res 193:113–122. https://doi.org/10.1016/j.pscychresns.2011.01.007

35. Buckner RL, Head D, Parker J, Fotenos AF, Marcus D, Morris JC, Snyder AZ (2004) A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage 23:724–738. https://doi.org/10.1016/j.neuroimage.2004.06.018

36. Jenkins R, Fox NC, Rossor AM, Harvey RJ, Rossor MN (2000) Intracranial volume and Alzheimer disease: evidence against the cerebral reserve hypothesis. Arch Neurol 57:220–224. https://doi.org/10.1001/archneur.57.2.220

37. Raikos A, Paraskevas GK, Yusuf F, Kordali P, Meditskou S, Al-Haj A, Brand-Saberi B (2011) Etiopathogenesis of hyperostosis frontalis interna: a mystery still. Ann Anat 193:453–458. https://doi.org/10.1016/j.aanat.2011.05.004