How reliable are volumetric techniques for high-grade gliomas? A comparison study of different available tools

How Reliable Are Volumetric Techniques for High-Grade Gliomas? A Comparison Study of Different Available Tools

Pietro Zeppa 1, Luca Neitzert 1, Marco Mammi 1, Matteo Monticelli 1, Roberto Altieri 2, Margherita Castaldo 1, Fabio Cofano 1, Alda Borrè 3, Francesco Zenga 1, Antonio Melcarne 1, Diego Garbossa 1

1Dipartimento di Neuroscienze, Università degli Studi di Torino, Turin, Italy.

2Dipartimento di Neurochirurgia, Policlinico G. Rodolico, Catania, Italy.

3Neuroradiologic Unit, Diagnostic Imaging and Interventistic Radiology Department, AOU Città della Salute e della Scienza of Turin, Turin, Italy.

Abstract

BACKGROUND

Gliomas are the most common malignant primary brain tumors. Assessment of the tumor volume represents a crucial point in preoperative and postoperative evaluation.

OBJECTIVE

To compare pre- and postoperative tumor volumes obtained with an automated, semi-automatic, and manual segmentation tool. Mean processing time of each segmentation techniques was measured.

METHODS

Manual segmentation was performed on preoperative and postoperative magnetic resonance images with the open-source software Horos (Horos Project). “SmartBrush,” a tool of the IPlan Cranial software (Brainlab, Feldkirchen, Germany), was used to carry out the semi-automatic segmentation. The open-source BraTumIA software (NeuroImaging Tools and Resources Collaboratory) was employed for the automated segmentation. Pearson correlation coefficient was used to assess volumetric comparison. Subsequently deviation/range and average discrepancy were determined. The Wilcoxon signed-rank test was used to assess statistical significance.

RESULTS

A total of 58 patients with a newly diagnosed high-grade glioma were enrolled. The comparison of the volumes calculated with Horos and IPlan showed a strong agreement both on preoperative and postoperative images (respectively: “enhancing” ρ = 0.99-0.78, “fluid-attenuated inversion recovery” ρ = 0.97-0.92, and “total tumor volume” ρ = 0.98-0.95). Agreement between BraTumIA and the other 2 techniques appeared to be strong for preoperative images, but showed a higher disagreement on postoperative images. Mean time expenditure for tumor segmentation was 27 min with manual segmentation, 17 min with semi-automated, and 8 min with automated software.

CONCLUSION

The considered segmentation tools showed high agreement in preoperative volumetric assessment. Both manual and semi-automated software appear adequate for the postoperative quantification of residual volume. The evaluated automated software is not yet reliable. Automated software considerably reduces the time expenditure.

Gliomas are the most common malignant primary brain tumors in adults and among them glioblastoma multiforme represents the most frequent and lethal glial tumor, with a 5 yr survival rate of approximately 5%.1,2

The main goal in glioma surgery is to maximize the extent of resection (EOR) and at the same time reduce the occurrence of neurological deficits in order to improve the patient's prognosis and quality of life. Indeed, several studies have demonstrated that a radical surgical resection positively affects prognosis.3-5 Volumetric assessment of the tumor represents a crucial point in preoperative evaluation and a precise determination of the EOR plays an important prognostic role. RANO (Response Assessment in Neuro-Oncology) criteria represent a standardized tool for the response assessment in neuro-oncology.6 The current response criteria recommend both the use of volumetric measurement of the enhancing volume, which is considered less operator dependent than the product of the maximum cross-sectional diameters (Macdonald criteria), and the signal abnormalities on fluid-attenuated inversion recovery (FLAIR) images as markers for tumor progression.6

It is, therefore, essential to identify reliable segmentation tools that allow to measure pre- and postoperative tumor volume. Furthermore, 3-dimensional volumetry might play a complementary role to the postoperative determination of response to treatment and tumor progression, overcoming the current limitations of 2-dimensional assessment.7

Manual segmentation tools represent the gold standard techniques for volumetric assessment to date; however, their use in clinical practice is limited by the intra- and inter-

rater variability. In addition, the use of manual segmentation tools in clinical practice is limited by the high expenditure of time that these techniques require.7 In recent years, several software for semi-automatic and automatic segmentation have been developed.

Semi-automated software is currently applied in clinical practice because of the reliability and saving of time, even though they still are user dependent. Recently, fully automated software has been developed in order to shorten segmentation time and improve accuracy of the volumetric assessment by reducing the intra- and inter-rater variability. Fully automated segmentation tools do not have a clinical application yet.8

The aim of this study was to compare pre- and postoperative tumor volumes in 58 patients with high-grade glioma, obtained respectively with a fully automated, a semi-automatic, and a manual segmentation tool. Such comparison was intended to provide information about the clinical applicability and utility of the different available segmentation tools. Furthermore, we measured the mean processing time of each segmentation tool, in order to provide data on their efficiency and time expenditure.

METHODS

In this study, pre- and postoperative tumor volumes were derived from preoperative and postoperative magnetic resonance imaging (MRI) using segmentation methods that integrate the information derived from 2-dimensional images to recreate a 3-dimensional volume. We considered 3 volumes: “enhancing volume” (cystic-necrotic and enhancing part of the tumor), “FLAIR volume” (T2/FLAIR hyperintense areas) and “total tumor volume,” represented by the sum of the enhancing volume and FLAIR volume.

Study population: patients who received surgical treatment at our institution between April 2014 and May 2019 were enrolled in this single center retrospective study. The exclusion criteria were: (i) recurrent tumor, (ii) low-grade gliomas, (iii) unavailable or low-quality magnetic resonance (MR) images (motion artefacts and insufficient slices), and (iv) previous brain tumor surgery or radiotherapy. For each patient, manual, semi-automated, and fully automated segmentation was performed, both on preoperative and postoperative MRI images. Patient records were de-identified and analyzed anonymously. Written informed consent was obtained from all the patients enrolled in this study.

All neuroradiological and histological data were collected and retrospectively analyzed. This study does not require any variations in patient's treatment and no formal ethics committee approval is required. All procedures performed for this study were in accordance with the ethical standards of our institution and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

MR imaging protocol: preoperative and postoperative MRI scans were obtained from different MRI scanners with varying field strengths (1.5 or 3 Tesla). Each patient underwent the same MR protocol that included: (i) precontrast T1w in axial or sagittal acquisition; (ii) postcontrast T1w in axial or sagittal acquisition; (iii) T2w in axial or sagittal acquisition; and (iv) FLAIR in axial or sagittal acquisition. The protocol was applied for the acquisition of both preoperative and postoperative images.

Manual segmentation: manual segmentation was performed by a medical student (NL) previously trained on medical images by a neurosurgery resident (ZP) and the segmentations were subsequently double-checked by a fully trained neurosurgeon (MA) to exclude errors in the delineation of the lesion and minimize the inter-rater variability. The manual segmentation was performed with the open-source software Horos (www.horosproject.org; Horos Project) for MacOS. Horos allows to manually delineate, slice-by-slice, the contour of the tumor by using a virtual drawing tool. The tool "region of interest (ROI) volume calculation" was used to obtain a volume starting from a ROI dataset. Manual segmentation of the abnormal signal detected in the postcontrast T1w sequences, which represented the necrotic (hypointense) and enhancing (hyperintense) areas, was carried out and the resulting volume was labeled as "enhancing volume." The same method was applied to obtain the total tumor volume (TTV) on FLAIR images, including all the hyper intense areas indicative for edema and abnormal tissue. The "FLAIR volume" was defined as the difference of TTV and enhancing volume. Similarly, the 3 volumes were also calculated on postoperative images. The resection cavity and the hyperintense boundary were not included, respectively, in the enhancing and FLAIR volume.

Semi-automated segmentation: "SmartBrush," a tool of the IPlan Cranial v3.0 (Brainlab, Feldkirchen, Germany) software, was used to carry out the semi-automated segmentation. The tool is based on a region-growing algorithm that permits to automatically expand the segmentation to adjacent areas characterized by signal alterations that are similar to the ones in the selected area.9 Manual changes to the segmentation were then performed by using the "brush" and "eraser" tools. SmartBrush has previously shown to be a reliable tool for volumetric segmentation of high-grade glioma.10,11 This software is routinely used for volumetric assessment of glial tumors in clinical practice at our institution. Semi-automated segmentation was performed by a medical student (NL) under the supervision of a resident neurosurgeon (ZP); segmentations were subsequently double-checked by a fully trained neurosurgeon (MA). To obtain preoperative and postoperative enhancing volume, FLAIR volume, and TTV, semi-automated segmentation was carried out both on postcontrast T1w and FLAIR images.

Automated segmentation: the open-source Brain Tumor Image Analysis (BraTumIA) v.1.2. (www.nitrc.org/projects/bratumia; NeuroImaging Tools and Resources Collaboratory) software was employed for the fully automated tumor segmentation. The software requires

the user to upload the MRI sequences (T1w, postcontrast T1w, T2w, and FLAIR) to the program interface. The four different MRI modalities are used to delineate the tumor and its subregions. BraTumIA allows not only to segment the brain into healthy and pathological tissue, but also subdivides healthy tissue into white matter, grey matter and cerebrospinal fluid and the tumor tissue into four sub-compartments (edema, necrosis, enhancing, and nonenhancing). The images initially undergo a preprocessing pipeline that include the alignment of the images, the extraction of the brain tissue from the images and the removal of the noise from the signal. Subsequently, the preprocessed images undergo feature extraction, that includes for every voxel the delineation of different features for distinguishing pathologic and healthy tissue. Classification is done using a Support Vector Machine classifier, which determines, based on the features of each voxel, to allocate it in one of the sub-compartments considered through a probability distribution. Finally, spatial regularization using a Conditional Random Fields method on the generated label map allows to enforce the spatial consistency of classified voxels with respect to the neighboring voxels. The functioning of the software has previously been described elsewhere.12-14 The edema and nonenhancing sub-compartments were considered together as "FLAIR volume," whereas the "enhancing volume" was given by the sum of the enhancing and the necrotic sub-compartments. The processing of the tumor volume is user independent.

Statistical analysis: volumetric comparison between manual, semi-automated, and fully automated segmentation was determined with the Pearson correlation coefficient (ρ), which was calculated both on preoperative and postoperative images for all the considered tumor sub-compartments (enhancing, FLAIR, and TTV). The techniques were compared 2 by 2. It is important to emphasize that the main limit of the Pearson coefficient is the strong sensibility to outliers. In order to quantify and analyze the different trend of the 3 segmentation techniques, the deviation/range for each considered volume was calculated. Subsequently, we determined the average discrepancy to determine a potential under- or overestimation of one of the 3 techniques. The Wilcoxon signed-rank test was used to determine the statistical significance due to the non-normal distribution of the data. Statistical significance level was set at $\alpha = 0.05$.

Finally, we calculated the mean processing time for each technique in order to highlight a potential discrepancy between the considered techniques. The processing time was measured on a randomly chosen sample that included 10 preoperative images and 10 postoperative images for each segmentation software.

RESULTS

Study population: 58 patients (mean age: 62 y) with a newly diagnosed high-grade glioma (54 WHO grade IV; 4 WHO grade III) were enrolled. The diagnosis of high-grade glioma was histologically confirmed at the local neuropathology department.

Mean volumes and standard deviations are shown in Table 1.

Scatter plots that illustrate the volume correlation out of preoperative and postoperative images are shown respectively in Figures 1 and 2.

Volumetric comparison on preoperative MR images: Pearson correlation coefficient (ρ) and Wilcoxon signed-rank test (T) for each comparison are shown in Table 2.

The comparison of the volumes calculated with Horos (Horos Project) and IPlan (Brainlab, Feldkirchen, Germany) showed a very strong agreement between the 2 segmentation techniques. Indeed, the comparison of the preoperative volumes showed a very high Pearson correlation coefficient for the enhancing ($\rho = 0.99$), FLAIR ($\rho = 0.97$), and TTV volume ($\rho = 0.98$). Scatter plots that illustrate the volume correlation are shown in Figure 1A-1C. After analyzing the average discrepancies, we found a small underestimation of the volumes measured with semi-automated technique compared to the manual volumes, but it did not turn out to be statistically significant. Results are shown in Figure 3A-3C.

The volumetric agreement between BraTumIA (NeuroImaging Tools and Resources Collaboratory) and Horos appeared to be strong. The comparison of the preoperative enhancing, FLAIR, and TTV volume showed a high Pearson correlation coefficient, respectively $\rho = 0.88$, $\rho = 0.86$, and $\rho = 0.89$. BraTumIA, however, showed a statistically significant overestimation of the volumetric measurements of the tumor components, as can be seen from Figure 1D-1F.

The Pearson correlation coefficient between BraTumIA and IPlan was $\rho = 0.87$ for the enhancing volume, $\rho = 0.86$ for the FLAIR volume, and $\rho = 0.89$ for the TTV volume. However, again, we noticed a statistically significant overestimation of the volumes measured with BraTumIA (Figure 1G-1I).

Volumetric comparison on postoperative MR images: Pearson correlation coefficient (ρ) and Wilcoxon signed-rank test (T) for each comparison are shown in Table 2.

The volumetric comparison between the three segmentation techniques showed a greater disagreement on postoperative MR images, particularly regarding the enhancing volume.

The correlation between Horos and IPlan appeared to be strong for all the considered tumor sub-components. The Pearson correlation was ρ = 0.78 for the enhancing volume, ρ = 0.92 for the FLAIR volume, and ρ = 0.95 for the TTV. No statistically significant trend discrepancies were found (Figure 2A-2C).

The comparison between BraTumIA and the other two segmentation techniques showed a higher disagreement.

The comparison between the fully automatic and the manual technique showed a Pearson correlation coefficient of ρ = 0.62 for the enhancing volume, ρ = 0.72 for the FLAIR volume, and ρ = 0.75 for the TTV.

When we compared the fully automatic with the semi-automatic technique, the Pearson correlation coefficient was ρ = 0.42 for the enhancing volume, ρ = 0.70 for the FLAIR volume, and ρ = 0.76 for the TTV (Figure 2G-2I).

The comparison between BraTumIA and the other segmentation software (IPlan and Horos), highlighted a systematic overestimation of the tumor volumes by the automatic segmentation software. The overestimation appeared to be statistically significant for all the considered volumes, as shown in Figure 3D-3F.

Processing time analysis: the mean time expenditure for tumor segmentation with the manual segmentation was 27 min (range: 24-29 min). The user-dependent IPlan required a mean time expenditure of 17 min (range: 15-22 min) to complete the segmentation. BraTumIA, being fully automated, required only 2 min to upload the images on the program interface and it took a mean time of 6 min (range: 5-8 min) to complete the segmentation.

DISCUSSION

In the context of glioma segmentation and volumetry, it is possible to employ three different segmentation techniques that differ from one another for the user dependency and the time expenditure: (i) manual segmentation; (ii) semi-automated segmentation; and (iii) fully automated segmentation.

In clinical practice, only manual and semi-automatic software packages have so far found an application, even though they are limited in particular by inter- and intraobserver variability and time expenditure. Previous studies pointed out the importance of minimizing user

interaction to reduce possible errors due to the interobserver and intraobserver variability, which appeared to be especially relevant on postoperative images.15,16

The emerging involvement of computational science in daily clinical practice has led to the development of new software that aid and supplements the user in the interpretation of medical images.

The MICCAI-Brain Tumor Segmentation Challenge (BRATS) has been organized since 2012 in order to yearly update the state-of-the-art in automated brain tumor segmentation and compare the performances and accuracy of the newly developed segmentation algorithms.8

The decision to employ BraTumIA (NeuroImaging Tools and Resources Collaboratory) as fully automated segmentation program was taken because of its encouraging performances in terms of agreement between fully automatic and manual tumor volume assessment that has led the software to be awarded as one of the top performing at the MICCAI-BRATS 2012 and 2013.8,12

This study aimed to compare three different segmentation techniques, including a manual, a semi-automatic, and a fully automatic software, in terms of volumetric agreement and time expenditure. The volumetric measurement was carried out both on preoperative and postoperative images.

The results obtained by the comparison of the segmentation techniques on the preoperative images showed a high agreement between the 3 software, particularly between Horos (Horos Project) and IPlan (Brainlab, Feldkirchen, Germany). Accordingly, it is possible to claim that both the manual and semi-automated segmentation techniques appear reliable and accurate for the determination of the preoperative tumor volume. On the other hand, BraTumIA is characterized by a slight, statistically significant overestimation of the tumor sub-compartments.

It is, however, important to consider that fully automated tools are not user dependent, therefore are not characterized by interobserver variability. In addition, BraTumIA allows to obtain a segmentation in a short amount of time. Therefore, it can be considered as a complementary technology in clinical practice.

Porz et al12 have previously highlighted a possible application of BraTumIA for complex data analysis, tumor growth modelling, and radiomic/radiogenomic analyses due to the rater independence of the segmentation tool.

The volumetric agreement on postoperative images resulted not as strong as for preoperative ones. Despite the fact that the Pearson correlation coefficient appeared to be high, when we compared the volumes obtained with IPlan and Horos, we noticed differences in the detection of the residual enhancing volume. A possible reason that could explain the difference is the inter-rater variability in the discrimination between pathological alteration and postoperative blood-brain barrier alteration. Further studies on the role of inter-rater and intrarater variability are therefore necessary. BraTumIA, on the other hand, showed systematic errors in the detection of residual tumor. This implied that volumetric agreement between the automatic segmentation and the other 2 segmentation techniques was not strong, particularly for the enhancing volume.

The automatic segmentation showed an evident overestimation of the 3 volumes considered as well as gross accuracy errors in the delineation of the tumor. The main explanation for these errors is that BraTumIA has been developed only for preoperative images. A new version of the fully automated software has been presented recently. Further comparative analysis that investigate the reliability of BraTumIA v.2.0. for the postoperative volumetric assessment should be carried out.

In terms of time expenditure, the manual tool resulted significantly more time consuming compared to the other segmentation tools. BraTumIA, on the other hand, considerably reduces the segmentation time expenditure.

Limitations

A limitation to our study is that the volumetric comparison between the 3 technique was not carried out by using the Dice similarity coefficient (DSC), a standard evaluation metric that is routinely used to evaluate the performance, reproducibility, and spatial overlap accuracy of MRI image segmentation algorithms.17-19 DSC was not calculated as image segmentation was not always recordable; tumor volume was therefore chosen as a surrogate measure of segmentation agreement.

A further limitation is represented by the impossibility to determine the inter-rater agreement of manual and semi-automated volume assessment because only one rater carried out the segmentation. A previous study has showed that the inter-rater agreement on manual segmentation of high-grade glioma appears to be comparable between experts and novices, whereas it is poor when postoperative images are considered.15 A further study that aimed to investigate the inter-rater agreement on semi-automated segmentation of glioblastoma showed an excellent concordance between experts and novices.11

Fully automated segmentation algorithms are not user dependent, therefore are not subject to inter-rater variability and lack of reproducibility.

CONCLUSION

In this study, the 3 considered segmentation tools showed high agreement in terms of preoperative volumetric assessment. Despite the slight overestimation of the segmented volume, BraTumIA (NeuroImaging Tools and Resources Collaboratory) considerably reduced the time expenditure and was not characterized by inter-rater and intrarater variability. Both IPlan (Brainlab, Feldkirchen, Germany) and Horos (Horos Project) appear adequate for the postoperative quantification of the EOR of high-grade gliomas, whereas BraTumIA is not yet reliable.

Disclosures

The authors have no personal, financial, or institutional interest in any of the drugs, materials, or devices described in this article.

Notes

Part of the results of this study were presented as a poster at the 2019 AINO (Associazione Italiana di Neuro Oncologia) Annual Meeting on November 10-12, 2019, in Udine, Italy.

REFERENCES

1. Ostrom QT, Bauchet L, Davis FG, et al. The epidemiology of glioma in adults: a state of the science review. Neuro Oncol. 2014;16(7):896-913.

2. Monticelli M, Zeppa P, Zenga F, et al. The post-surgical era of GBM: how molecular biology has impacted on our clinical management. A review. Clin Neurol Neurosurg. 2018;170:120-126.

3. Lacroix M, Abi-Said D, Fourney DR, et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. J Neurosurg. 2001;95(2):190-198.

4. Li YM, Suki D, Hess K, Sawaya R. The influence of maximum safe resection of glioblastoma on survival in 1229 patients: can we do better than gross-total resection? J Neurosurg. 2016;124(4):977-988.

5. Altieri R, Melcarne A, Soffietti R, et al. Supratotal resection of glioblastoma: is less

more? Surg Technol Int. 2019;35:432-440.
6. Chukwueke UN, Wen PY. Use of the response assessment in neuro-oncology (RANO) criteria in clinical trials and clinical practice. CNS Oncol. 2019;8(1):CNS28.
7. Gordillo N, Montseny E, Sobrevilla P. State of the art survey on MRI brain tumor segmentation. Magn Reson Imaging. 2013;31(8):1426-1438.
8. Menze BH, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging. 2015;34(10):1993-2024.
9. Shilpa Kamdi R. Image segmentation and region growing algorithm. Int J Comput Technol Electron Eng. 2012;2(1):103-107.
10. Huber T, Alber G, Bette S, et al. Progressive disease in glioblastoma: benefits and limitations of semi-automated volumetry. PLoS One. 2017;12(2):e0173112.
11. Huber T, Alber G, Bette S, et al. Reliability of semi-automated segmentations in glioblastoma. Clin Neuroradiol. 2017;27(2):153-161.
12. Porz N, Habegger S, Meier R, et al. Fully automated enhanced tumor compartmentalization: man vs Machine reloaded. PLoS One. 2016;2(11):e0165302.
13. Bauer S, Nolte LP, Reyes M. Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. Med Image Comput Comput Assist Interv. 2011;14(Pt 3):354-361.
14. Rios Velazquez E, Meier R, Dunn WD, et al. Fully automatic GBM segmentation in the TCGA-GBM dataset: prognosis and correlation with VASARI features. Sci Rep. 2015;5:16822.
15. Visser M, Müller DMJ, van Duijn RJM, et al. Inter-rater agreement in glioma segmentations on longitudinal MRI. NeuroImage Clin. 2019;22:101727.
16. Vos MJ, Uitdehaag BMJ, Barkhof F, et al. Interobserver variability in the radiological assessment of response to chemotherapy in glioma. Neurology. 2003;11;60(5):826-830.
17. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmen-

tation quality based on a spatial overlap index. Acad Radiol. 2004;11(2):178-189.

18. Cui S, Mao L, Jiang J, Liu C, Xiong S. Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network. J Healthc Eng. 2018;2018:4940593.

19. Chang K, Beers AL, Bai HX, et al. Automatic assessment of glioma burden: a deep learning algorithm for fully automated volumetric and bidimensional measurement. Neuro Oncol. 2019;21(11):1412-1422.

Preoperative Images	Horos	SmartBrush	BraTumIA
FLAIR volume	μ: 67.3 cm^3 (0.3-162.5); δ: 46,6 cm^3	μ: 65 cm^3 (0.4-164.1); δ: 46.2 cm^3	μ: 72 cm^3 (2.5-157.7); δ: 42.6 cm^3
Enhancing volume	μ: 29.8 cm^3 (0-93.1); δ: 26.7 cm^3	μ: 29.7 cm^3 (0-98.9); δ: 27.2 cm^3	μ: 31.9 cm^3 (0.1-84.7); δ: 28.2 cm^3
Total volume	μ: 97.1 cm^3 (0.7-226.9); δ: 63.5 cm^3	μ: 94.7 cm^3 (0.8-203.7); δ: 62 cm^3	μ: 103.9 cm^3 (2.5-199.9); δ: 60.6 cm^3
Postoperative Images	**Horos**	**SmartBrush**	**BraTumIA**
FLAIR volume	μ: 29.6 cm^3 (0-100.1); δ: 26.4 cm^3	μ: 28.9 cm^3 (0.4-124.7); δ: 30 cm^3	μ: 58 cm^3 (1.7-164.6); δ: 40.8 cm^3
Enhancing volume	μ: 2.7 cm^3 (0-29.3); δ: 6.5 cm^3	μ: 1.7 cm^3 (0-37.7); δ: 6 cm^3	μ: 8.8 cm^3 (0-54.2); δ: 11.3 cm^3
Total volume	μ: 27.9 cm^3 (0-139.4); δ: 30 cm^3	μ: 30.8 cm^3 (0.4-162.4); δ: 33.1 cm^3	μ: 67.4 cm^3 (1.8-193.5); δ: 46 cm^3

TABLE 1. Mean Volumes and Standard Deviation for FLAIR, Enhancing, and TTV Obtained by the Different Segmentation Techniques out of Preoperative Images and Postoperative Images

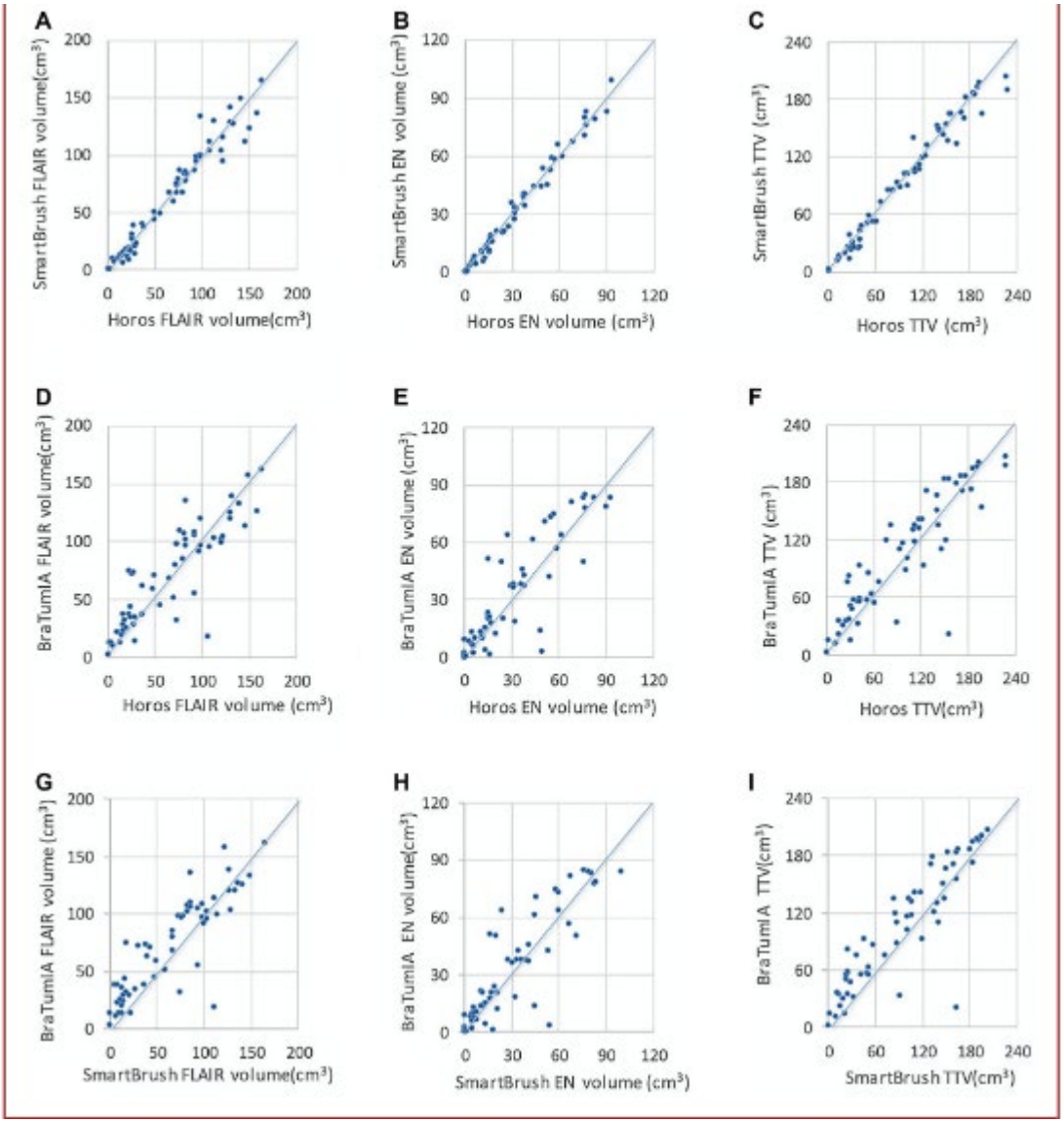

FIGURE 1. Scatter plots representing the correlation between volumes obtained by different segmentation software out of preoperative images. Horos (Horos Project) and SmartBrush (Brainlab, Feldkirchen, Germany) show a very strong correlation for the FLAIR volume A, Enhancing volume B, and TTV C. The correlation between BraTumIA (NeuroImaging Tools and Resources Collaboratory) and the other two segmentation techniques D-I shows a larger dispersion of the data compared to A-C, despite the strong Pearson correlation coefficient.

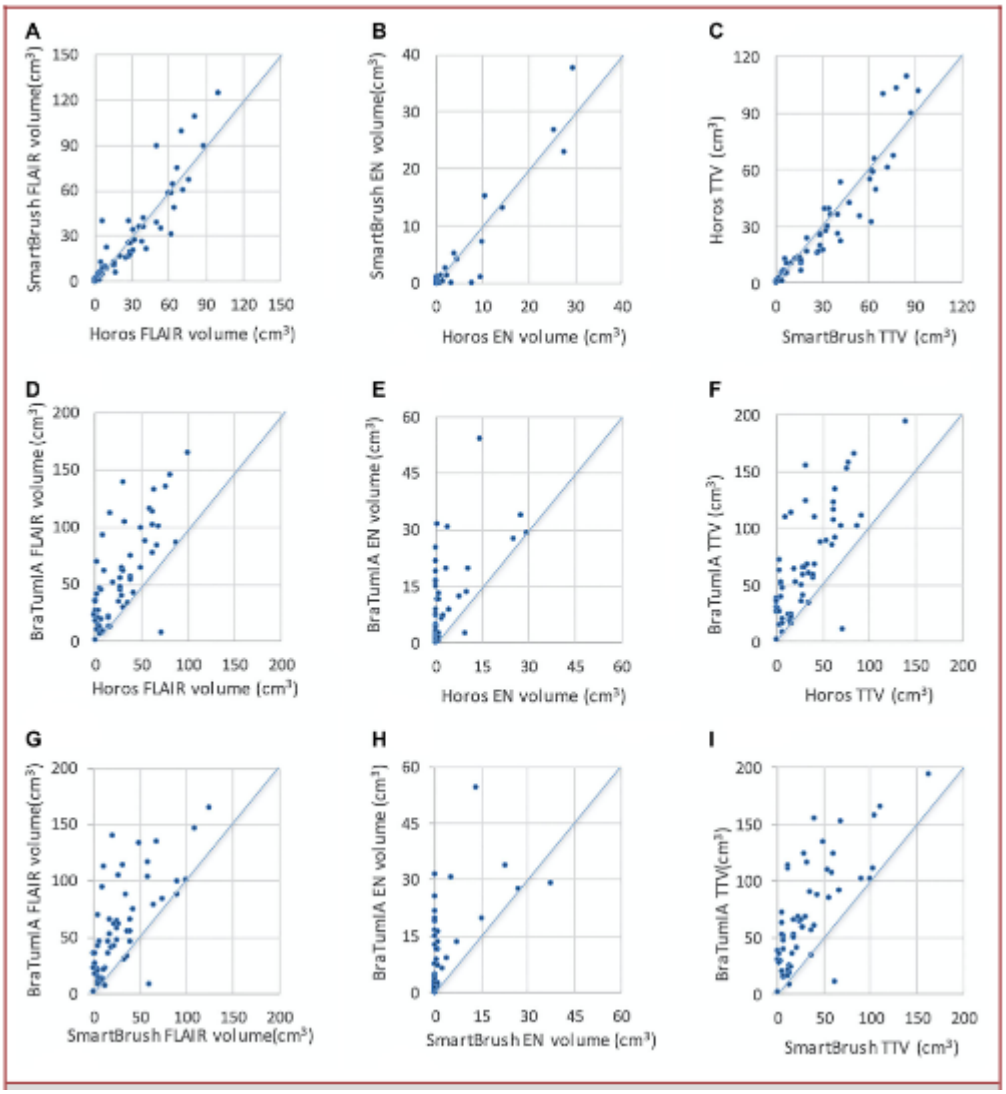

FIGURE 2. Scatter plots representing the correlation between volumes obtained by the three segmentation software out of postoperative images. The charts show a strong correlation between the manual and the semi-automated software for the FLAIR volume A, enhancing volume B, and the TTV C. The fully automated software tends to overestimate the three tumor sub-compartments D-I, particularly the enhancing volume E-H.

Preoperative Images	Horos	SmartBrush	BraTumIA
FLAIR volume	ρ: 0.97; T: 664.5 ($T_\alpha = 0.05 : 602$)	ρ: 0.86; T: 480 ($T_\alpha = 0.05 : 602$)	ρ: 0.86; T: 584 ($T_\alpha = 0.05 : 602$)
Enhancing volume	ρ: 0.99; T: 620.5 ($T_\alpha = 0.05 : 454$)	ρ: 0.87; T: 582.5 ($T_\alpha = 0.05 : 602$)	ρ: 0.88; T: 565 ($T_\alpha = 0.05 : 602$)
Total volume	ρ: 0.98; T: 650.5 ($T_\alpha = 0.05 : 558$)	ρ: 0.89; T: 357 ($T_\alpha = 0.05 : 602$)	ρ: 0.89; T: 490 ($T_\alpha = 0.05 : 602$)
Preoperative Images	**Horos**	**SmartBrush**	**BraTumIA**
FLAIR volume	ρ: 0.92; T: 686 ($T_\alpha = 0.05 : 80$)	ρ: 0.70; T: 80 ($T_\alpha = 0.05 : 602$)	ρ: 0.72; T: 72.5 ($T_\alpha = 0.05 : 580$)
Enhancing volume	ρ: 0.78; T: 289.5 ($T_\alpha = 0.05 : 223$)	ρ: 0.42; T: 40.5 ($T_\alpha = 0.05 : 494$)	ρ: 0.62; T: 55.5 ($T_\alpha = 0.05 : 474$)
Total volume	ρ: 0.95; T: 665 ($T_\alpha = 0.05 : 602$)	ρ: 0.76; T: 54 ($T_\alpha = 0.05 : 602$)	ρ: 0.75; T: 52 ($T_\alpha = 0.05 : 602$)

TABLE 2. Comparison of the Different Segmentation Techniques in Terms of Pearson Correlation Coefficient (P) and Wilcoxon Signed-Rank Test (T) on Preoperative Images and Postoperative Images

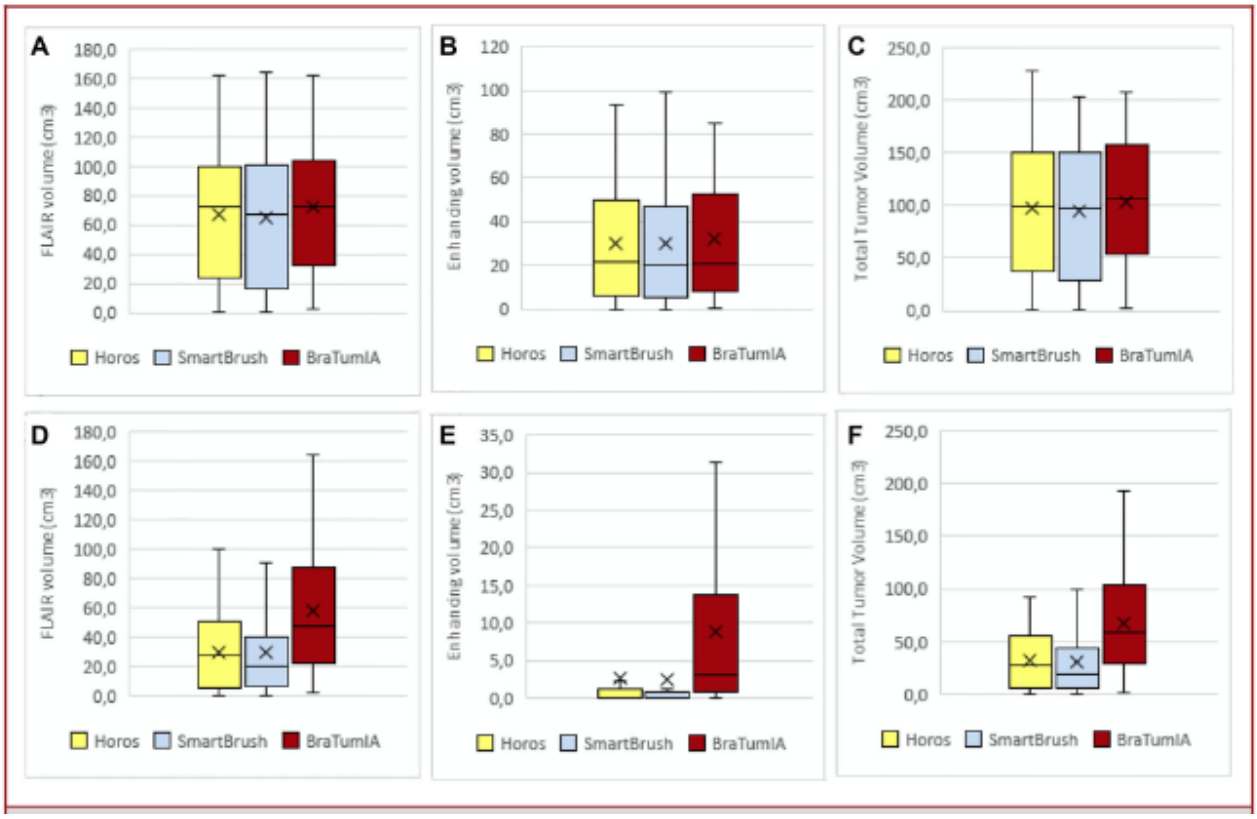

FIGURE 3. Box plot showing the distribution of the preoperative A-C and postoperative D-F FLAIR volume A and D, enhancing volume B and E, and TTV C and F. The Wilcoxon signed-rank test shows a statistically significant difference between volumes obtained with BratumIA and with the other 2 segmentation techniques for all of the tumor sub-compartments considered. Particularly, BraTumIA is characterized by a statistically significant overestimation of the postoperative tumor sub-compartments D-F.