JNS

CLINICAL ARTICLE

Automatic estimation of extent of resection and residual tumor volume of patients with glioblastoma

***Raphael Meier, MSc,[1] Nicole Porz, MD,[2,3] Urspeter Knecht, MD,[2] Tina Loosli, MSc,[2] Philippe Schucht, MD,[3] Jürgen Beck, MD,[3] Johannes Slotboom, PhD,[2] Roland Wiest, MD,[2] and Mauricio Reyes, PhD[1]**

[1]Institute for Surgical Technology & Biomechanics, University of Bern; [2]Support Center for Advanced Neuroimaging, Institute for Diagnostic and Interventional Neuroradiology, University Hospital Inselspital and University of Bern; and [3]Department of Neurosurgery, University Hospital Inselspital and University of Bern, Switzerland

OBJECTIVE In the treatment of glioblastoma, residual tumor burden is the only prognostic factor that can be actively influenced by therapy. Therefore, an accurate, reproducible, and objective measurement of residual tumor burden is necessary. This study aimed to evaluate the use of a fully automatic segmentation method—brain tumor image analysis (BraTumIA)—for estimating the extent of resection (EOR) and residual tumor volume (RTV) of contrast-enhancing tumor after surgery.

METHODS The imaging data of 19 patients who underwent primary resection of histologically confirmed supratentorial glioblastoma were retrospectively reviewed. Contrast-enhancing tumors apparent on structural preoperative and immediate postoperative MR imaging in this patient cohort were segmented by 4 different raters and the automatic segmentation BraTumIA software. The manual and automatic results were quantitatively compared.

RESULTS First, the interrater variabilities in the estimates of EOR and RTV were assessed for all human raters. Interrater agreement in terms of the coefficient of concordance (W) was higher for RTV ($W = 0.812$; $p < 0.001$) than for EOR ($W = 0.775$; $p < 0.001$). Second, the volumetric estimates of BraTumIA for all 19 patients were compared with the estimates of the human raters, which showed that for both EOR ($W = 0.713$; $p < 0.001$) and RTV ($W = 0.693$; $p < 0.001$) the estimates of BraTumIA were generally located close to or between the estimates of the human raters. No statistically significant differences were detected between the manual and automatic estimates. BraTumIA showed a tendency to overestimate contrast-enhancing tumors, leading to moderate agreement with expert raters with respect to the literature-based, survival-relevant threshold values for EOR.

CONCLUSIONS BraTumIA can generate volumetric estimates of EOR and RTV, in a fully automatic fashion, which are comparable to the estimates of human experts. However, automated analysis showed a tendency to overestimate the volume of a contrast-enhancing tumor, whereas manual analysis is prone to subjectivity, thereby causing considerable interrater variability.

https://thejns.org/doi/abs/10.3171/2016.9.JNS16146

KEY WORDS glioblastoma; extent of resection; residual tumor volume; automatic tumor volumetry; BraTumIA; oncology

Glioblastoma (WHO Grade IV) is the most aggressive and most common type of primary brain tumor. The current treatment options depend on patient-specific factors such as the location and size of the glioma, patient age, symptoms, and neurological status. Treatment includes surgery, radiation therapy, and/or chemotherapy. The role of surgery has been debated for decades; however, modern guidelines[41] recommend primary surgical removal provided that neurological function is preserved.[37]

ABBREVIATIONS BraTumIA = brain tumor image analysis; CET = contrast-enhancing tumor; CRET = complete resection of the enhancing tumor; EOR = extent of resection; MPR = multiplanar reconstruction; PRET = partial resection of the enhancing tumor; RTV = residual tumor volume; T1w = T1-weighted; T2w = T2-weighted; W = Kendall's coefficient of concordance.

SUBMITTED January 18, 2016. **ACCEPTED** September 15, 2016.

INCLUDE WHEN CITING Published online January 6, 2017; DOI: 10.3171/2016.9.JNS16146.

* Mr. Meier and Dr. Porz contributed equally to this work. Drs. Wiest and Reyes share senior authorship of this work.

A growing body of evidence indicates a significant overall survival benefit for the radical resection of contrast-enhancing tumor (CET) compared with subtotal resection.[5,17] Resection of CET is usually quantified by reporting the extent of resection (EOR) or residual tumor volume (RTV). Consequently, both EOR[6,16,24–26,31,36,40] and RTV[6,16] were found to be associated with patient survival, suggesting their roles as prognostic biomarkers.[10,11,25] This has likewise motivated the use of intraoperative 5-aminolevulinic acid fluorescence and electrophysiological mapping and/or intraoperative MRI–assisted surgery in many neurosurgical units in order to reach most radical resections.[8,9,33,35,38]

The volumetric measurement of EOR and RTV can be obtained via the manual segmentation of the CET and the postoperative residual, respectively. Kubben et al.[23] studied the intrarater and interrater variability of EOR and RTV measurements in 8 patients across 3 different expert raters. They found high intrarater but low interrater agreement and suggested that computer-assisted methods may increase interrater agreement. Furthermore, manual segmentation is a time-consuming procedure; it can take up to 20 minutes per patient,[14,39] and this considerably limits its usage in clinics. Kanaly et al.[20] proposed a semiautomatic threshold-based method that requires the user to outline the tumor region and a region of normal brain parenchyma in coregistered precontrast and postcontrast T1-weighted images. However, their study lacked a quantitative comparison between the estimates of the computer-assisted method and manual segmentation, which makes it difficult to judge the performance of the proposed method. Chow et al.[7] developed a semiautomatic segmentation method for the quantification of residual tumor burden and evaluated it in a cohort of 29 glioblastoma patients. They reported a 10-fold decrease in the segmentation time (from 9.7 minutes to < 1 minute). Moreover, they found volumetric estimates provide higher interrater agreement than unidimensional (Response Evaluation Criteria in Solid Tumors [RECIST]) and bidimensional (Macdonald criteria) measures when used for tumor response assessment.[15] Cordova et al.[12] evaluated a semiautomatic segmentation method based on manual region of interest selection and fuzzy C-means clustering in 37 different patients. Similar to the study of Chow et al., Cordova et al. found good agreement between the volumes estimated by semiautomatic segmentation and the ground truth estimated by manual segmentation. Furthermore, they also observed a decrease in the segmentation time. Recently, a clinically oriented, fully automated segmentation tool for brain tumor image analysis (BraTumIA) was proposed.[3,27,29,32] BraTumIA performs compartmentalization of the glioma into necrosis, edema, and nonenhancing and enhancing tumor sections and estimates the respective volumes within an average computation time of 5 minutes. The software was evaluated in 25 patients within a prospective clinical trial by Porz et al.,[32] which showed good agreement with the volumetric ground truth estimated by manual segmentation.

In contrast to semiautomatic segmentation methods, fully automatic methods offer the advantage of reproducible and objective estimates of tumor volume. This is of great importance for longitudinal studies in glioma patients.[29] Consequently, we employ BraTumIA to estimate the EOR and RTV of CETs. We hypothesize that BraTumIA is able to generate estimates of EOR and RTV that are comparable to the estimates of human raters. Specifically, we aimed to evaluate 1) interrater variability between 4 raters for the available patient data in order to highlight the subjectivity of the task at hand, 2) if the EOR estimated by BraTumIA is comparable to the estimates of human raters, and 3) if the RTV estimated by BraTumIA is comparable to the estimates of human raters.

Methods

Study Population

Data on patients with newly diagnosed and histologically confirmed glioblastoma who were preoperatively admitted to our institution between October 2012 and July 2013 were extracted for the study at hand. Patients were included if the image acquisition was complete (i.e., all required MR sequences were obtained) and no previous cranial neurosurgery or biopsy was performed. The study was approved by the local research ethics commission (Kantonale Ethikkommission Bern). All patients provided written informed consent. From a total of 19 patients, 9 patients underwent subtotal extirpations or partial resection of the enhancing tumor (PRET), whereas 10 patients underwent complete resections of the enhancing tumor (CRET). All diagnoses were confirmed by histopathological analysis. A standardized MR protocol was performed on all patients. Manual and automatic segmentation were performed on the preoperative and immediately postoperative MRI studies obtained in all 19 patients. The raters were blinded to the outcome of surgery (i.e., the radiological reports) and performed segmentation of the immediate postoperative image subsequent to the preoperative image.

MRI Protocol

The MR images were acquired preoperatively and postoperatively (no later than 72 hours after resection) on two 1.5-T MR scanners from 1 vendor (Siemens Avanto and Siemens Aera). Every patient underwent a standardized MRI protocol, including: 1) precontrast 3D T1-weighted (T1w) multiplanar reconstruction (MPR) for sagittal acquisition with 1-mm isotropic resolution; 2) postcontrast 3D T1w MPR for sagittal acquisition with 1-mm isotropic resolution; 3) 3D T2-weighted (T2w) SPACE for sagittal acquisition with 1-mm isotropic resolution; and 4) FLAIR (2D turbo inversion recovery) for axial acquisition. The sequence parameters were: 1) for the precontrast 3D T1-weighted MPR sequences, TE 2.67 msec, TR 1580 msec, FOV 256 × 256 mm^2, and FA 8° with an isotropic voxel resolution of 1 × 1 × 1 mm; 2) for postcontrast T1-weighted imaging, TE 4.57 msec, TR 2070 msec, FOV 256 × 256 mm^2, and FA 15° using isotropic 1 × 1 × 1–mm voxels; 3) for 3D-T2-weighted SPACE for sagittal acquisition, TE 380 msec, TR 3000 msec, FOV 256 × 256 mm^2, and FA 120° using isotropic 1 × 1 × 1–mm voxels; and 4) for 2D FLAIR sequencing, TE 80 msec, TR 8000 msec, FOV 256 × 256 mm^2, and FA 120° using a nonisotropic voxel size of 1 × 1 × 3 mm.

Manual Segmentation

Prior to manual segmentation, the different MR images were skull stripped[2] and coregistered on the postcontrast T1-weighted image using a rigid transformation. This procedure is part of the BraTumIA software and was performed to facilitate the comparison of automatically generated segmentations, which were obtained from the same coregistered images. The corresponding preoperative and postoperative MRI sequences obtained in all 19 patients were segmented by 4 human raters. Rater 1 is a neurosurgeon experienced (> 5 years) in brain tumor imaging, Rater 2 is a neuroradiologist with several years of experience (> 5 years) in brain tumor diagnostics, Rater 3 is an experienced (3 years) researcher in brain tumor image analysis, and Rater 4 is an medical master student who was previously trained in neuroimaging with more than 1 year of experience in the field (the numbering of the raters reflects their experience in descending order). All raters were counseled by a neuroradiologist with more than 15 years of experience in brain tumor imaging. The raters performed segmentation of the tumor into necrosis, edema, and contrast-enhancing and nonenhancing tumor sections and adhered to a predefined segmentation protocol[19] using a 3D slicer. The protocol was adapted for segmenting the immediate postoperative images in the following manner: 1) the necrotic core was not segmented as a subcompartment because it is removed during surgery; and 2) the coregistered T1-weighted and postcontrast T1-weighted images were overlaid and used to differentiate the blood products from the enhancing tumor. The average time required to manually segment all tumor compartments (preoperatively and postoperatively) in a study patient was approximately 1 hour.

Automated Segmentation

The automatic segmentation was performed using BraTumIA software (https://www.nitrc.org/projects/bratumia/). This software offers a completely integrated segmentation pipeline, where the user loads the original DICOM images of the 4 relevant MRI modalities (T1-weighted, postcontrast T1-weighted, T2-weighted, and FLAIR images). Subsequently, the images are fully automatically processed, including skull stripping[2] and subsequent rigid coregistration to ensure voxel-to-voxel correspondence between the different MRI sequences. Based on the registered images, segmentation into unaffected tissue and tumor tissue, which encompass 4 different compartments (necrosis, edema, and enhancing and nonenhancing tumor sections), is performed using combined supervised classification and regularization.[3] The machine learning–based methodology relies on a voxel-wise feature extraction[27] followed by classification via a decision forest[4,13] and final spatial regularization through conditional random field–based optimization. In contrast to the study of Porz et al.,[32] BraTumIA was trained on an enlarged patient image set containing 36 preoperative images, 9 immediate postoperative images, and 9 follow-up images (acquired within 1–6 months after surgery). The intention was to make BraTumIA applicable to preoperative, immediate postoperative, and follow-up images, thus enabling the software to segment longitudinal imaging studies.[29]

Statistical Analysis

The statistical analysis was performed on the volumetric estimates of CET as defined by the different raters. Multiple differences between manual and automatic estimates of EOR and RTV were analyzed using the Kruskal-Wallis test. To assess interrater variability, Kendall's coefficient of concordance (W) was computed due to the possible nonnormality of the data at hand (a normal assumption of the data was rejected based on the results of the Shapiro-Wilk test; $p < 0.001$). The significance level was defined to be $\alpha = 0.05$. To compare the difference in EOR and RTV between the estimates of 2 raters, we computed the approximate 95% confidence intervals (CIs) of the median using a paired, exact Wilcoxon signed-rank test (R package "exactRankTests"; a Bonferroni correction was applied for multiple comparisons). The confidence intervals are reported as a tuple (median [95% CI range]). Based on the study of Grabowski et al.,[16] we defined thresholds for EOR (0.98) and RTV (2000 mm^3) that are relevant for patient survival. Estimates of the raters on different sides of the threshold are considered disagreements between the raters. The agreement between couples of raters with respect to these thresholds is reported as a percentage. The thresholds of Grabowski et al. were chosen since they measured a median RTV (1.2 cm^3) that was substantially closer to the median RTV (1.1 cm^3) of the study at hand than the alternative values of other studies.[6]

The results of Patient 9 were excluded from the statistical analysis because BraTumIA yielded a negative estimation of EOR, which is not plausible (resulting in the data of 18 patients being analyzed). The cause of this misestimation is discussed in detail.

Results

Interrater Variability in EOR and RTV

The measurements of EOR and RTV that were estimated by the 4 different raters are shown in Figs. 1 and 2, respectively. Postoperative segmentations of the different raters for 2 exemplary slices of Patient 5 are shown in Fig. 3. The Kruskal-Wallis test did not detect a statistically significant difference among the measurements of the 4 raters for either EOR ($p = 0.841$) or RTV ($p = 0.861$). For all 4 raters, $W = 0.775$ ($p < 0.001$) for EOR, whereas $W = 0.812$ ($p < 0.001$) for RTV. For both EOR and RTV, the average agreement with respect to the survival-relevant threshold among all 4 raters was evaluated (i.e., the mean agreement for all 6 possible pairings of raters) and corresponds to 84.3% and 90.7%, respectively. In 5 of 18 patients, disagreement among the human raters occurred for EOR, but for RTV a disagreement occurred in 3 of 18 patients.

Comparison Between Manual and Automatic EOR

Measurements of the preoperative CETs by BraTumIA were plotted and correlated against the estimates of all human raters, as shown in Fig. 4. The median preoperative CET volumes were 12.68 cm^3, 12.06 cm^3, 14 cm^3, 17.75 cm^3, and 23.04 cm^3 for Raters 1 to 4 and BraTumIA. The resulting EOR measurements for BraTumIA, as well as

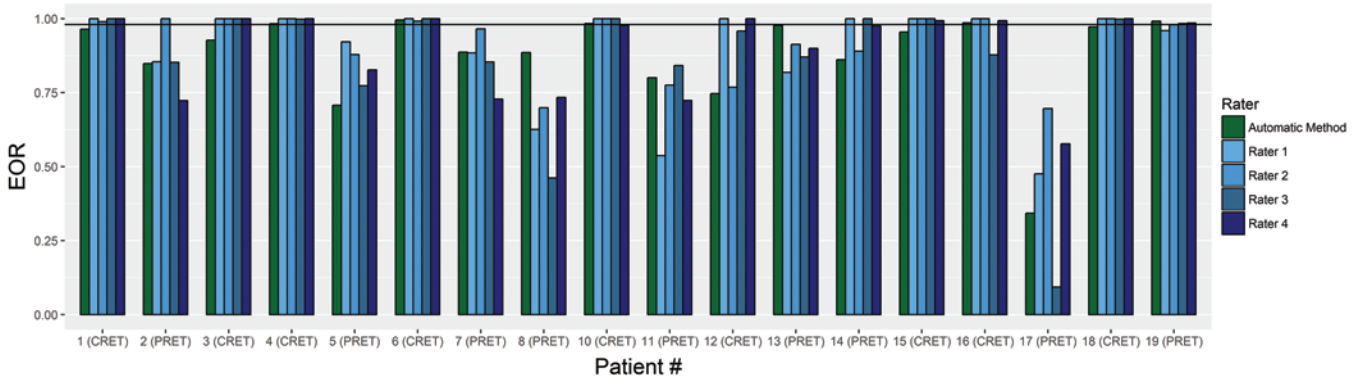

FIG. 1. EOR measurements of BraTumIA (*green*) and different human raters (*blue*) for 18 patients with PRET and CRET. The survival-relevant threshold value (0.98) is shown as a *solid line*. Figure is available in color online only.

for the human raters, for all 18 patients are shown in Fig. 1. The Kruskal-Wallis test did not detect a statistically significant difference between the measurements of the human raters and BraTumIA (5 groups; $p = 0.384$). The agreement between all 5 raters in terms of Kendall's coefficient of concordance was $W = 0.713$ ($p < 0.001$). The approximate (Bonferroni-corrected) 95% CIs (median [95% CI range]) shown in Fig. 5 were computed between BraTumIA and Rater 1 (–0.02 [–0.12 to 0.11]), BraTumIA and Rater 2 (–0.03 [–0.11 to 0.02]), BraTumIA and Rater 3 (–0.01 [–0.07 to 0.12]), and BraTumIA and Rater 4 (–0.02 [–0.12 to 0.07]). With respect to the survival-relevant threshold, agreement between the estimates of Rater 1 and BraTumIA (55.6%; 10 of 18 patients), Rater 2 and BraTumIA (66.7%; 12 of 18 patients), Rater 3 and BraTumIA (61.1%; 11 of 18 patients), and Rater 4 and BraTumIA (66.7%; 12 of 18 patients) were assessed. This resulted in an average agreement of 62.5%. In 5 of 18 patients, BraTumIA disagreed with all 4 human raters with respect to the survival-relevant threshold (0.98).

Comparison Between Manual and Automatic RTV

The median RTVs of the patients with PRET were 1.106 cm^3, 0.733 cm^3, 1.071 cm^3, 1.552 cm^3, and 0.834 cm^3 for Raters 1 to 4 and BraTumIA. The median RTV of the patients with CRET was 0 cm^3 for the human raters. BraTumIA estimated a median RTV of 0.564 cm^3. The measurements of RTV by BraTumIA, as well as by the human raters, for all 18 patients are shown in Figure 2. The Kruskal-Wallis test did not detect a statistically significant difference among the measurements of the human raters and BraTumIA (5 groups; $p = 0.16$). The agreement between all 5 raters in terms of Kendall's coefficient of concordance was $W = 0.693$ ($p < 0.001$). The approximate (Bonferroni-corrected) 95% CIs (median [95% CI range]) in cubic millimeters, as shown in Fig. 6, were computed between BraTumIA and Rater 1 (642 mm^3 [–174 to 1953] mm^3), BraTumIA and Rater 2 (471 mm^3 [–86 to 2599] mm^3), BraTumIA and Rater 3 (351 mm^3 [–2174 to 1275] mm^3), and BraTumIA and Rater 4 (377 mm^3 [–308 to 2130] mm^3). With respect to the survival-relevant threshold, agreement between the estimates of Rater 1 and BraTumIA (83.3%; 15 of 18 patients), Rater 2 and BraTumIA (88.9%; 16 of 18 patients), Rater 3 and BraTumIA (88.9%; 16 of 18 patients), and Rater 4 and BraTumIA (94.4%; 17 of 18 patients) were assessed. This resulted in an average agreement of 88.9%. In 1 of 18 patients (Patient 3), BraTumIA estimated an RTV that disagreed with the estimates of all human raters with respect to the survival-relevant threshold (2000 mm^3). Representative segmentation results (Patient 17) of the RTV for BraTumIA and all 4 raters are shown in Fig. 7A. In addition, an example of incor-

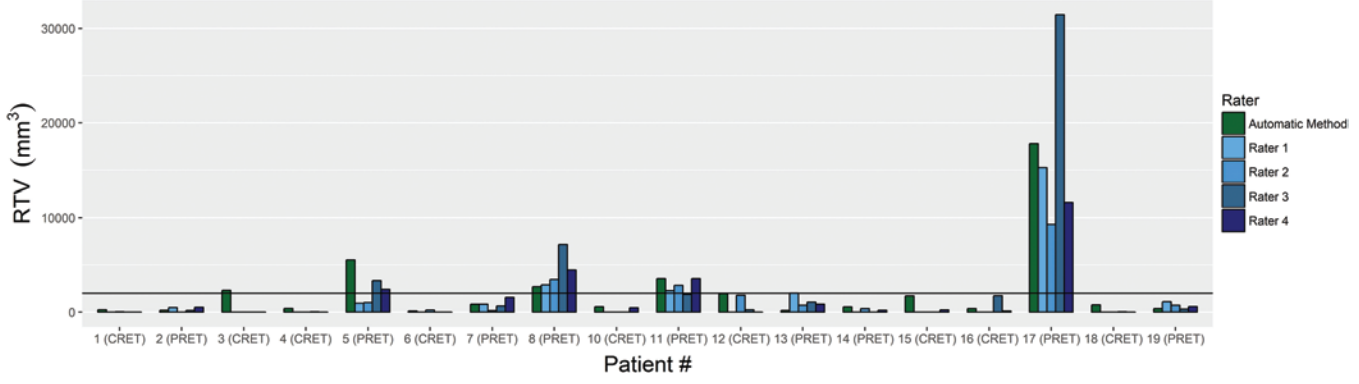

FIG. 2. RTV measurements of BraTumIA (*green*) and different human raters (*blue*) for 18 patients with PRET and CRET. The survival-relevant threshold value (2000 mm^3) is shown as a *solid line*. Figure is available in color online only.

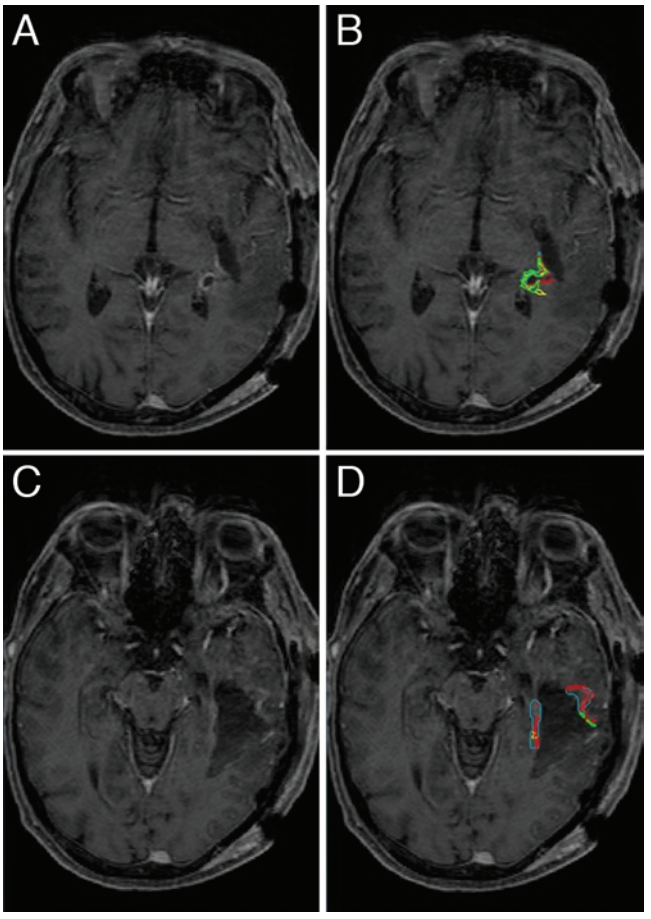

FIG. 3. Patient 5. Postoperative gadolinium-enhanced T1-weighted MR images overlaid with the manual segmentations of the different human raters. The segmentations for Rater 1 (*yellow*), Rater 2 (*blue*), Rater 3 (*red*), and Rater 4 (*green*) are visualized. **A and B:** Image slices showing high agreement among the different raters. **C and D:** Image slices showing low agreement between the segmentations of the different raters. Figure is available in color online only.

rectly labeled blood products (Patient 9) by BraTumIA is shown in Fig. 7B.

Discussion

Radiological assessment of the borders of resection and possible tumor residuals remains a challenge. The surrounding tissue is subjected to large deformations, and the presence of residual CET can be easily confounded with benign enhancements (e.g., the choroid plexus, early blood-brain barrier disruption along the rim of resection, and the presence of deoxyhemoglobin). Consequently, volumetric measurements of residual tumor burden are subject to large interrater variability.[23] The study at hand provides evidence for the potential use of a fully automatic segmentation method to perform volumetric analysis of a CET on immediate postoperative images.

To reduce confounders between the CET and nonspecific T1-weighted hyperintensity caused by blood products, we employed an MR acquisition protocol that required the postoperative images be obtained within 72 hours after surgery.[1] Differentiation between the residual tumor and

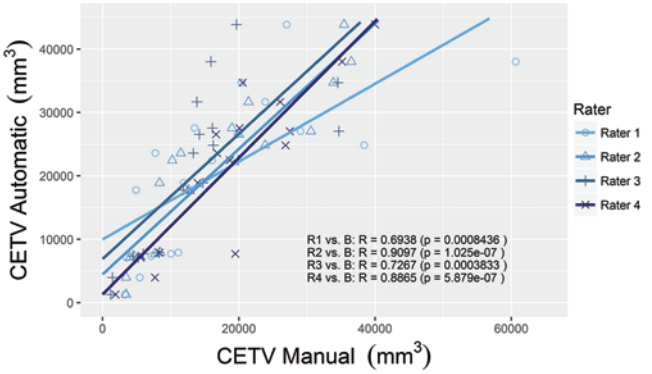

FIG. 4. Estimates of preoperative CET volume (CETV) by the human raters (R1–R4) plotted against the estimates of BraTumIA (B). Figure is available in color online only.

other benign enhancements (e.g., the choroid plexus) was facilitated by the coregistration of the unenhanced T1-weighted and gadolinium-enhanced T1-weighted MR sequences via the BraTumIA software.

BraTumIA is a fully automatic, machine learning–based segmentation method capable of performing compartmentalization of a glioblastoma into necrosis, edema, and enhancing and nonenhancing tumor sections. We decided to employ BraTumIA for the following reasons. First, considerable evidence about the capability of BraTumIA to segment preoperative high-grade glioma was generated recently. BraTumIA's performance was compared against other fully automated methods in the Medical Image Computing and Computer-Assisted Intervention Brain Tumor Segmentation (MICCAI BRATS) challenges,[30] where it showed competitive performance as well as superiority in terms of computational running time (average runtime 5 minutes). Moreover, BraTumIA was evaluated prospectively for the purpose of performing preoperative segmentation on a clinical data set[32] that included 25 patients. The automatically generated segmentations of CET showed good agreement (in terms of the Dice coefficient) with the corresponding manual ground truth. In a recent study by Rios Velazquez et al.,[34] the results were confirmed on an independent data set. Moreover, the preoperative volumetric estimates of BraTumIA for CET were found to be significantly associated with overall and 1-year survival. Second, BraTumIA has been made publicly available (https://www.nitrc.org/projects/bratumia/) and is equipped with a graphical user interface. This facilitates its use by other researchers and thus allows for independent reevaluation of the tool on different data sets.

The first aim of our study was to illustrate the subjectivity of manually assessing residual tumor burden. In Figs. 1 and 2, considerable interrater variability can be observed in the measurements of partial resections. This variability is even more emphasized for EOR, which also incorporates the estimate of the preoperative volume of CET. The additional preoperative measurement required for the calculation of EOR is also an additional source of measurement error and could explain the reduced interrater agreement for EOR (W = 0.775) when compared with RTV (W = 0.812). Figure 3 shows that the amount of intervariability is also visually apparent and can vary from slice to slice.

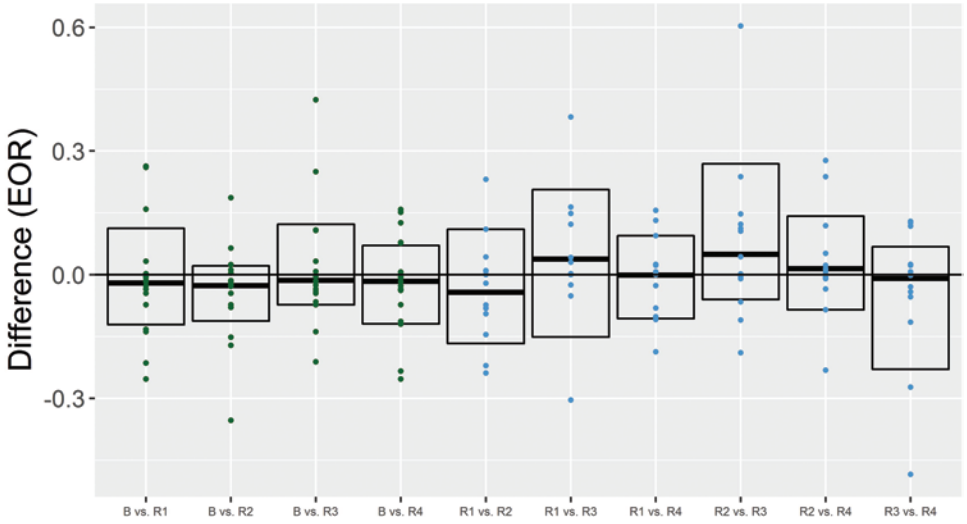

FIG. 5. Approximate 95% CIs for EOR determined by different pairs of human raters (R1–R4) and BraTumIA (B). The *solid line* indicates the line of zero difference. The *center line* of the CI indicates the estimated median value. The *dots* indicate the values of the paired differences (*green*, results involving BraTumIA; *blue*, results involving only human raters). Figure is available in color online only.

Consequently, such interrater variability likely renders any association to clinical end points (e.g., the response to therapy or overall survival) problematic (as previously reported by Kubben et al.[23]). Interrater variability can be caused by differing educational backgrounds among raters. However, in this study, measures for normalizing the different backgrounds of the raters were taken into account in advance; 3 of the 4 raters were educated at the same clinical education program and all 4 raters were counseled by the same senior neuroradiologist (who has > 15 years of experience). A further cause of the interrater variability in the study at hand could also be the different levels of experience of the raters. Grabowski et al.[16] observed considerable variability between the measurements of residual contrast-enhancing volumes when raters have different levels of experience. They employed a quantitative, semiautomatic segmentation method (Brainlab's prototype iPlan software) to generate the volumetric information. The remaining variability between the different raters is due to the manual interaction, and thus Grabowski et al. suggest using fully automated methods in future studies in order to further standardize measurements. In a recent study by Huber et al.,[18] the same semiautomatic segmentation was used by raters with varying levels of experience to segment the preoperative and postoperative follow-up MRI data of 5 patients. Despite the different levels of expertise, they found high interrater agreement. However, their study did not include the segmentation of immediate postoperative images, and

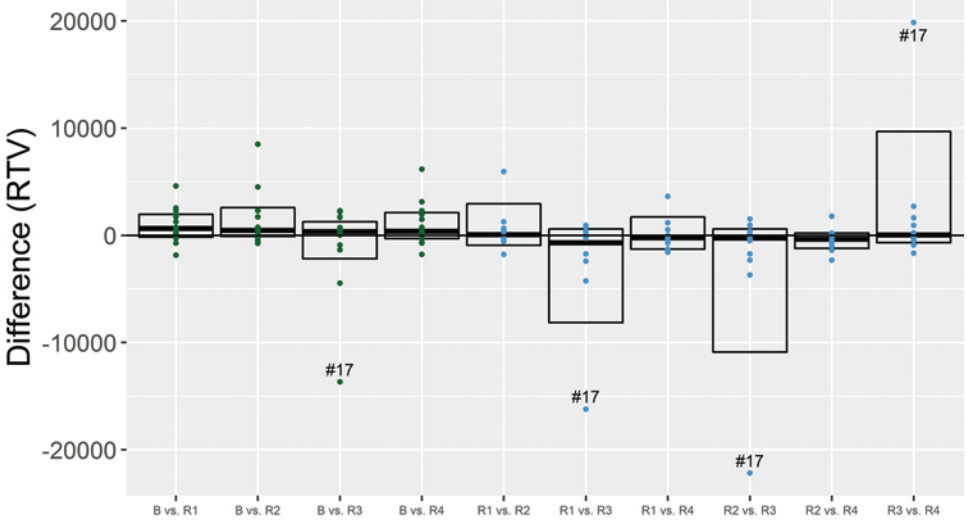

FIG. 6. Approximate 95% CIs in cubic millimeters for RTV determined by different pairs of human raters and BraTumIA. The *solid line* indicates the line of zero difference. The *center line* of the CI indicates the estimated median value. The *dots* indicate the values of the paired differences (*green*, results involving BraTumIA; *blue*, results involving only human raters). The outlier is Patient 17 *(#17)*. Figure is available in color online only.

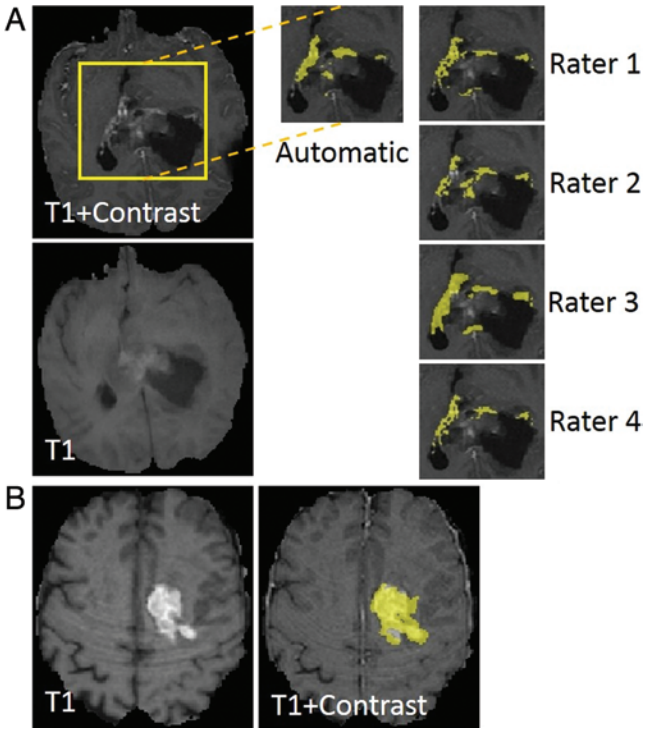

FIG. 7. Qualitative results of 2 patients. **A:** Patient 17. Segmentation result (*yellow*) of BraTumIA and the different human raters. **B:** Patient 9. Segmentation result of BraTumIA (CRET) showing blood products that were incorrectly identified by the algorithm as an enhancing tumor. Figure is available in color online only.

thus their findings cannot be compared with our findings or with the observations by Grabowski et al.

The 2 remaining aims of the study at hand were to perform a comparison of automatic and manual estimates of EOR and RTV. In Fig. 4, one can see that BraTumIA showed a general tendency toward the overestimation of preoperative volume. For a given RTV, this results in the overestimation of the EOR (e.g., Patients 8 and 11). In general, the estimates of BraTumIA tended to underestimate EOR when compared with the estimates of the human raters. This tendency can be seen in Fig. 5, where all estimated median differences in EOR between BraTumIA and manual measurements are negative. In the context of the survival-relevant threshold, this led to more disagreements with the human raters and explains the lower percentage of agreement between the (binarized) estimates of BraTumIA and the estimates of the human raters when compared with agreement among the human raters themselves. The analysis of the paired differences showed that the confidence interval between BraTumIA and each of the human raters is about equal or smaller in size than the confidence intervals for human raters. This suggests that the volumetric estimates of EOR by BraTumIA were, in general, located either close to or between the estimates of the human raters. The tendency of BraTumIA to overestimate the volume of a CET also applies to the postoperative situation. For RTV, an overestimation by BraTumIA compared with all human raters occurred in 9 of 18 patients. Seven of these 9 patients had CRET. For these cases (e.g., Patient 3), BraTumIA incorrectly identified benign enhancements as CET and thus shows that the software is overly sensitive in predicting the presence of a residual tumor (thereby causing the previously mentioned underestimation of EOR). However, the observed disagreement with respect to the survival-relevant threshold is less severe than in the case of EOR. This is also reflected in the percentage agreement between BraTumIA and the human raters that is closer to the agreement between the human raters themselves. In Fig. 6, the confidence intervals of the paired differences in RTV show a similar situation as EOR. Although no statistically significant differences were detected, Raters 1, 2, and 4 clearly segmented the residual enhancing tumor more conservatively than BraTumIA and Rater 3, leading to confidence intervals that barely include the zero line. This is also visible in Fig. 7A , where Patient 17's tumor infiltration of the choroid plexus led to differing segmentation results. For Patient 9, BraTumIA strongly overestimated RTV, which led to a negative EOR. Since a negative EOR is not plausible, we excluded this patient from the statistical analysis. After visual inspection (Fig. 7B), we can conclude that Patient 9 exhibited a large presence of blood products, which was wrongly identified by BraTumIA as a CET. In fact, the hemorrhage in Patient 9 was significantly larger than in any other patient and altered the image intensity of the resection cavity and confounded the appearance of the tumoral tissue. In general, we identified the tendency to overestimate the volume of a CET to be the main weakness of BraTumIA. This tendency can be linked to the algorithmic core of BraTumIA. The final segmentation of BraTumIA is obtained by optimization of the energy function of a pairwise conditional random field. Segmentations obtained from pairwise conditional random fields suffer from short-boundary bias, leading to overestimation of the true object size. Modifications[21,22] that would neutralize this bias have been proposed and are currently being investigated by us (along with complementary approaches[28]) because they would likely increase the agreement of BraTumIA's estimates with the estimates of the human raters.

This study has 2 limitations. First, only segmentation of CETs was analyzed. Compared with manual segmentation of a CET, segmentation of a nonenhancing tumor on immediate postoperative images is even more challenging. Furthermore, we lacked histologically confirmed ground truth data for the postoperative images. Therefore, we limited our analysis to the morphologically most discriminative tumor compartment (i.e., the CET). Second, the moderate number of samples (n = 19) prevented us from studying the association of the volumetric estimates of BraTumIA with patient survival. Future studies using larger patient cohorts should certainly investigate this aspect.

Conclusions

In the light of the fact that the residual tumor burden appears to be the only prognostic factor that can be actively influenced by the clinician, an objective and reproducible way of quantifying it is of the outmost importance. Our results suggest that BraTumIA, though in general being overly sensitive, can automatically yield estimates of EOR and RTV that are comparable to the estimates of expert raters.

Acknowledgments

We thank the MR technicians of our department for their excellent support. This study was supported by the Swiss National Science Foundation (SNF grant no. 320030_140958), the Bernese Cancer League, the Swiss Cancer League, and the European Union Seventh Framework Programme for research, technological development, and demonstration under grant agreement no. 600841.

References

1. Albert FK, Forsting M, Sartor K, Adams HP, Kunze S: Early postoperative magnetic resonance imaging after resection of malignant glioma: objective evaluation of residual tumor and its influence on regrowth and prognosis. **Neurosurgery 34:**45–61, 1994
2. Bauer S, Fejes T, Reyes M: A skull-stripping filter for ITK. **Insight J 20:**1–7, 2012
3. 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, in Fichtinger G, Martel A, Peters T (eds): **Medical Image Computing and Computer-Assisted Intervention—MICCAI 2011.** Heidelberg: Springer, 2011, pp 354–361
4. Breiman L: Random forests. **Mach Learn 45:**5–32, 2001
5. Brown TJ, Brennan MC, Li M, Church EW, Brandmeir NJ, Rakszawski KL, et al: Association of the extent of resection with survival in glioblastoma: a systematic review and meta-analysis. **JAMA Oncol 352:**987–996, 2016
6. Chaichana KL, Jusue-Torres I, Navarro-Ramirez R, Raza SM, Pascual-Gallego M, Ibrahim A, et al: Establishing percent resection and residual volume thresholds affecting survival and recurrence for patients with newly diagnosed intracranial glioblastoma. **Neuro Oncol 16:**113–122, 2014
7. Chow DS, Qi J, Guo X, Miloushev VZ, Iwamoto FM, Bruce JN, et al: Semiautomated volumetric measurement on postcontrast MR imaging for analysis of recurrent and residual disease in glioblastoma multiforme. **AJNR Am J Neuroradiol 35:**498–503, 2014
8. Coburger J, Engelke J, Scheuerle A, Thal DR, Hlavac M, Wirtz CR, et al: Tumor detection with 5-aminolevulinic acid fluorescence and Gd-DTPA-enhanced intraoperative MRI at the border of contrast-enhancing lesions: a prospective study based on histopathological assessment. **Neurosurg Focus 36(2):**E3, 2014
9. Coburger J, Hagel V, Wirtz CR, König R: Surgery for glioblastoma: impact of the combined use of 5-aminolevulinic acid and intraoperative MRI on extent of resection and survival. **PLoS One 10:**e0131872, 2015
10. Coburger J, Wirtz CR, König RW: Impact of extent of resection and recurrent surgery on clinical outcome and overall survival in a consecutive series of 170 patients for glioblastoma in intraoperative high field iMRI. **J Neurosurg Sci** [epub ahead of print], 2015
11. Cordova JS, Gurbani SS, Holder CA, Olson JJ, Schreibmann E, Shi R, et al: Semi-automated volumetric and morphological assessment of glioblastoma resection with fluorescence-guided surgery. **Mol Imaging Biol 18:**454–462, 2016
12. Cordova JS, Schreibmann E, Hadjipanayis CG, Guo Y, Shu HKG, Shim H, et al: Quantitative tumor segmentation for evaluation of extent of glioblastoma resection to facilitate multisite clinical trials. **Transl Oncol 7:**40–47, 2014
13. Criminisi A, Shotton J: **Decision Forests for Computer Vision and Medical Image Analysis.** London: Springer, 2013, pp 25–45
14. Deeley MA, Chen A, Datteri R, Noble JH, Cmelak AJ, Donnelly EF, et al: Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study. **Phys Med Biol 56:**4557–4577, 2011
15. Gállego Pérez-Larraya J, Lahutte M, Petrirena G, Reyes-Botero G, González-Aguilar A, Houillier C, et al: Response assessment in recurrent glioblastoma treated with irinotecan-bevacizumab: comparative analysis of the Macdonald, RECIST, RANO, and RECIST + F criteria. **Neuro Oncol 14:**667–673, 2012
16. Grabowski MM, Recinos PF, Nowacki AS, Schroeder JL, Angelov L, Barnett GH, et al: Residual tumor volume versus extent of resection: predictors of survival after surgery for glioblastoma. **J Neurosurg 121:**1115–1123, 2014
17. Hardesty DA, Sanai N: The value of glioma extent of resection in the modern neurosurgical era. **Front Neurol 3:**140, 2012
18. Huber T, Alber G, Bette S, Boeckh-Behrens T, Gempt J, Ringel F, et al: Reliability of semi-automated segmentations in glioblastoma. **Clin Neuroradiol** [epub ahead of print], 2015
19. Jakab A: **Segmenting Brain Tumors with the Slicer 3D Software.** (http://www2.imm.dtu.dk/projects/BRATS2012/Jakab_TumorSegmentation_Manual.pdf.) [Accessed October 25, 2016]
20. Kanaly CW, Ding D, Mehta AI, Waller AF, Crocker I, Desjardins A, et al: A novel method for volumetric MRI response assessment of enhancing brain tumors. **PLoS One 6:**e16031, 2011
21. Kohli P, Osokin A, Jegelka S: A principled deep random field model for image segmentation, in Kellenberger P (ed): **Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2013.** Piscataway, NJ: IEEE Press, 2013, pp 1971–1978
22. Krähenbühl P, Koltun V: Efficient inference in fully connected CRFs with Gaussian edge potentials, in Shawe-Taylor J, Zemel RS, Bartlett PL, et al (eds): **Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011.** Cambridge, MA: MIT Press, 2011, pp 109–117
23. Kubben PL, Postma AA, Kessels AGH, van Overbeeke JJ, van Santbrink H: Intraobserver and interobserver agreement in volumetric assessment of glioblastoma multiforme resection. **Neurosurgery 67:**1329–1334, 2010
24. Lacroix M, Abi-Said D, Fourney DR, Gokaslan ZL, Shi W, DeMonte F, et al: A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. **J Neurosurg 95:**190–198, 2001
25. Marko NF, Weil RJ, Schroeder JL, Lang FF, Suki D, Sawaya RE: Extent of resection of glioblastoma revisited: personalized survival modeling facilitates more accurate survival prediction and supports a maximum-safe-resection approach to surgery. **J Clin Oncol 32:**774–782, 2014
26. McGirt MJ, Chaichana KL, Gathinji M, Attenello FJ, Than K, Olivi A, et al: Independent association of extent of resection with survival in patients with malignant brain astrocytoma. **J Neurosurg 110:**156–162, 2009
27. Meier R, Bauer S, Slotboom J, Wiest R, Reyes M: Appearance-and context-sensitive features for brain tumor segmentation, in **Proceedings of MICCAI BRATS Challenge**, 2014, pp 020–026 (http://people.csail.mit.edu/menze/papers/proceedings_miccai_brats_2014.pdf) [Accessed October 25, 2016]
28. Meier R, Bauer S, Slotboom J, Wiest R, Reyes M: Patient-specific semi-supervised learning for postoperative brain tumor segmentation, in Golland P, Nobuhiko H, Barillot C, et al (eds): **Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014.** Heidelberg: Springer, 2014, pp 714–721
29. Meier R, Knecht U, Loosli T, Bauer S, Slotboom J, Wiest R, et al: Clinical evaluation of a fully-automatic segmentation method for longitudinal brain tumor volumetry. **Sci Rep 6:**23376, 2016

30. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). **IEEE Trans Med Imaging 34:**1993–2024, 2015
31. Orringer D, Lau D, Khatri S, Zamora-Berridi GJ, Zhang K, Wu C, et al: Extent of resection in patients with glioblastoma: limiting factors, perception of resectability, and effect on survival. **J Neurosurg 117:**851–859, 2012
32. Porz N, Bauer S, Pica A, Schucht P, Beck J, Verma RK, et al: Multi-modal glioblastoma segmentation: man versus machine. **PLoS One 9:**e96873, 2014
33. Raabe A, Beck J, Schucht P, Seidel K: Continuous dynamic mapping of the corticospinal tract during surgery of motor eloquent brain tumors: evaluation of a new method. **J Neurosurg 120:**1015–1024, 2014
34. Rios Velazquez E, Meier R, Dunn WD Jr, Alexander B, Wiest R, Bauer S, et al: Fully automatic GBM segmentation in the TCGA-GBM dataset: Prognosis and correlation with VASARI features. **Sci Rep 5:**16822, 2015
35. Roder C, Bisdas S, Ebner FH, Honegger J, Naegele T, Ernemann U, et al: Maximizing the extent of resection and survival benefit of patients in glioblastoma surgery: high-field iMRI versus conventional and 5-ALA-assisted surgery. **Eur J Surg Oncol 40:**297–304, 2014
36. Sanai N, Polley MY, McDermott MW, Parsa AT, Berger MS: An extent of resection threshold for newly diagnosed glioblastomas. **J Neurosurg 115:**3–8, 2011
37. Schucht P, Beck J, Seidel K, Raabe A: Extending resection and preserving function: modern concepts of glioma surgery. **Swiss Med Wkly 145:**w14082, 2015
38. Schucht P, Seidel K, Beck J, Murek M, Jilch A, Wiest R, et al: Intraoperative monopolar mapping during 5-ALA-guided resections of glioblastomas adjacent to motor eloquent areas: evaluation of resection rates and neurological outcome. **Neurosurg Focus 37(6):**E16, 2014
39. Sorensen AG, Patel S, Harmath C, Bridges S, Synnott J, Sievers A, et al: Comparison of diameter and perimeter methods for tumor volume calculation. **J Clin Oncol 19:**551–557, 2001
40. Stummer W, Reulen HJ, Meinel T, Pichlmeier U, Schumacher W, Tonn JC, et al: Extent of resection and survival in glioblastoma multiforme: identification of and adjustment for bias. **Neurosurgery 62:**564–576, 2008
41. Stupp R, Brada M, van den Bent MJ, Tonn JC, Pentheroudakis G: High-grade glioma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. **Ann Oncol 25 (Suppl 3):**iii93–iii101, 2014

Disclosures

The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

Author Contributions

Conception and design: Meier, Porz, Wiest. Acquisition of data: Meier, Porz, Knecht, Loosli. Analysis and interpretation of data: Meier, Porz. Drafting the article: Meier, Porz. Critically revising the article: Knecht, Loosli, Wiest, Reyes. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Meier. Statistical analysis: Meier. Administrative/technical/material support: Slotboom. Study supervision: Wiest, Reyes.

Correspondence

Raphael Meier, Institute for Surgical Technology & Biomechanics, University of Bern, Bern 3014, Switzerland. email: raphael.meier@istb.unibe.ch.