Neuro-Oncology Advances

5(1), 1–10, 2022 | https://doi.org/10.1093/noajnl/vdac184 | Advance Access date 20 December 2022

NS-HGlio: A generalizable and repeatable HGG segmentation and volumetric measurement AI algorithm for the longitudinal MRI assessment to inform RANO in trials and clinics

Aly H. Abayazeed, Ahmed Abbassy, Michael Müeller, Michael Hill, Mohamed Qayati, Shady Mohamed, Mahmoud Mekhaimar, Catalina Raymond, Prachi Dubey, Kambiz Nael, Saurabh Rohatgi, Vaishali Kapare, Ashwini Kulkarni, Tina Shiang, Atul Kumar, Nicolaus Andratschke, Jonas Willmann, Alexander Brawanski, Reordan De Jesus, Ibrahim Tuna, Steve H. Fung, Joseph C. Landolfi, Benjamin M. Ellingson, and Mauricio Reyes

Biomedical Engineering group, Neosoma Inc., Groton, Massachusetts, USA (Originating Institution address:44 Farmers Row, Groton, Massachusetts, USA 01450) (A.H.A, A.A., M.H., M.Q., S.M., M.Me.); ARTORG Biomedical Engineering group, University of Bern, Switzerland (M.Me., M.R.); Radiology Department, University of Cairo School of Medicine, Egypt (A.B., M.Q., S.M., M.M.); Radiology Department, Houston Methodist Hospital, Houston, Texas, USA (P.D., S.H.F.); Radiology Department, University of California Los Angeles, Los Angeles, California, USA (K.N.); Radiology Department, Massachusetts General Hospital, Boston, Massachusetts, USA (S.R.); Radiology Department, University of Massachusetts, Worcester, Massachusetts, USA (V.K., A.K., T.S.); Radiology Department, Yale School of Medicine, New Haven, Connecticut, USA (A.Ku.); Radiation Oncology Department, University of Zurich, Switzerland (N.A., J.W.); Radiation Oncology Department, University Hospital Regensburg, Cairo Egypt and Regensburg, Germany (A.B.); Radiology Department, University of Florida, Gainesville, Florida, USA (R.D.J., I.T.); Neurology/Neuro-oncology Department, Hackensack Meridian Health JFK Medical Center, Edison, New Jersey, USA (J.C.L.); Brain Tumor Imaging Laboratory, University of California Los Angeles, Los Angeles, California, USA (C.R., B.M.E.)

Corresponding Author: Aly H. Abayazeed, Chief Medical Officer Neosoma Inc. and Associate Neuroradiologist, 44 Farmers Row, Groton, MA 01450 (aly.abayazeed@neosomainc.com)

Abstract

Background. Accurate and repeatable measurement of high-grade glioma (HGG) enhancing (Enh.) and T2/FLAIR hyperintensity/edema (Ed.) is required for monitoring treatment response. 3D measurements can be used to inform the modified Response Assessment in Neuro-oncology criteria. We aim to develop an HGG volumetric measurement and visualization AI algorithm that is generalizable and repeatable.

Methods. A single 3D-Convoluted Neural Network, NS-HGlio, to analyze HGG on MRIs using 5-fold cross validation was developed using retrospective (557 MRIs), multicentre (38 sites) and multivendor (32 scanners) dataset divided into training (70%), validation (20%), and testing (10%). Six neuroradiologists created the ground truth (GT). Additional Internal validation (IV, three institutions) using 70 MRIs, and External validation (EV, single institution) using 40 MRIs through measuring the Dice Similarity Coefficient (DSC) of Enh., Ed. ,and Enh. + Ed. (WholeLesion/WL) tumor tissue and repeatability testing on 14 subjects from the TCIA MGH-QIN-GBM dataset using volume correlations between timepoints were performed.

Results. IV Preoperative median DSC Enh. 0.89 (SD 0.11), Ed. 0.88 (0.28), WL 0.88 (0.11). EV Preoperative median DSC Enh. 0.82 (0.09), Ed. 0.83 (0.11), WL 0.86 (0.06). IV Postoperative median DSC Enh. 0.77 (SD 0.20), Ed 0.78. (SD 0.09), WL 0.78 (SD 0.11). EV Postoperative median DSC Enh. 0.75 (0.21), Ed 0.74 (0.12), WL 0.79 (0.07). Repeatability testing; Intraclass Correlation Coefficient of 0.95 Enh. and 0.92 Ed.

Conclusion. NS-HGlio is accurate, repeatable, and generalizable. The output can be used for visualization, documentation, treatment response monitoring, radiation planning, intra-operative targeting, and estimation of Residual Tumor Volume among others.

© The Author(s) 2022. Published by Oxford University Press, the Society for Neuro-Oncology and the European Association of Neuro-Oncology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

source: https://doi.org/10.48350/177813 | downloaded: 12.10.2024

Downloaded from https://academic.oup.com/noa/article/5/1/vdac184/6948060 by Universitaetsbibliothek Bern user on 25 January 2023

Key Points

1. Deep Learning (DL) based glioma segmentation is needed for automated longitudinal RANO assessment in clinical trials and clinical practice.
2. Ground Truth quality and training dataset heterogeneity are crucial for glioma segmentation algorithms generalization and clinical translation.
3. NS-HGlio is accurate and repeatable in quantifying HGG Enh. volumes for mRANO, eliminating inter-rater and intra-rater variability, objectifying Ed. measurement for RANO, and accelerating the workflow.

Importance of the Study

The success of clinical trials and clinical management of gliomas is dependent on many factors. The accurate, repeatable, and rapid quantification of HGG Enh. volumes on MRIs are important to inform mRANO and the accurate and repeatable quantification of HGG Ed. volume is of value in objectifying FLAIR assessment for RANO. Volumetric measurement of HGG Enh. and Ed. volumes likely correlate better with Progression Free Survival (PFS) and Overall Survival (OS) when compared to standard bi-dimensional measures. However, the clinical translation of an automated AI technology to perform this task has not been successful given the complexity of the postoperative and post-treatment imaging in gliomas and the poor quality of the ground truth (GT) of the publicly available datasets. We describe the development and validation of a Deep Learning (DL) device to perform segmentation, volumetric measurement, and visualization of the Enh. and Ed. components of HGG at accuracy levels comparable to those of experts while eliminating inter-rater and intra-rater variability and potentially the need for adjudications in clinical trials. This would additionally allow for standardized tumor growth and regression analysis and quick bulk re-analysis of historical clinical trials data in need of re-categorization based on the 2021 WHO classification of CNS tumors fifth edition ($WHOCNS_5$). Furthermore, given that this AI is currently FDA cleared, this would allow for routine implementation of response assessment in clinical care allowing for better adherence of institutions to RANO and mRANO criteria.

Glioblastoma (GBM) is the most aggressive malignant primary brain tumor with very poor survival of approximately 15–18months from time of diagnosis and approximately 6% at 5 years. The incidence rate is 3.19 per 100,000 persons in the United States with a median age of 64 years, it is uncommon in children. The incidence is 1.6 times higher in males compared to females and 2.0 times higher in Caucasians compared to Africans and Afro-Americans, with lower incidence in Asians and American Indians.[1] The accurate and repeatable longitudinal MRI measurement of the enhancing (Enh.) and T2/FLAIR edema (Ed.) components of High-Grade Glioma (HGG) is of crucial importance for clinical trials, treatment planning, and monitoring response to therapy. 3D measurements are preferred over 2D measurements and can be used to monitor treatment response according to the modified Response Assessment in Neuro-oncology criteria (mRANO),[2,3] and for surgical and radiation planning. Currently these measurements are done manually which is time consuming, inconsistent and suffers from high inter-rater and intra-rater variability specially in the post-treatment phase.[4] Our aim is to evaluate the accuracy, generalizability, and repeatability of a DL HGG segmentation and volumetric analysis AI algorithm (NS-HGlio) for clinical translation.

Materials and Methods

The Data Sets

Starting in 2012, the BraTS (Brain Tumor Segmentation Challenge) group has sponsored a yearly competition through MICCAI for the development and testing of ML algorithms for HGG segmentation on MRIs with good success. However, most of their efforts were focused on the analysis of preoperative imaging, which is limited if used to inform mRANO, tumor growth (*g*) or tumor regression (*d*) analysis. RANO and mRANO criteria are the cornerstones of imaging analysis in clinical trials and are gaining traction in clinical practice in part due to the increasing availability of AI technology. These criteria depend on the ability to perform accurate and repeatable postoperative analysis of the whole tumor as well as the sub-components mainly the Enh. and T2/FLAIR hyperintensity/Ed. tumor segments.

NS-HGlio AI was developed as a Software as a Medical Device that is generalizable at analyzing HGG Enh. volumes on both preoperative and postoperative MRIs to inform mRANO, Residual Tumor Volumes (RTV) post-surgically, tumor growth (*g*), and tumor regression (*d*) analysis.

Downloaded from https://academic.oup.com/noa/article/5/1/vdac184/6948060 by Universitaetsbibliothek Bern user on 25 January 2023

NS-HGlio also provides contouring and volumetry for the Ed. subcompartment to allow for objective assessment of the RANO[5] criteria given that the assessment of this subcompartment is currently based on subjective assessment. Additionally, the contouring and volumetry of the surgical cavity would be of great value for radiation planning. It is expected that RANO 2.0, which is soon to be released, will allow investigators and clinicians the option to use 3D volumetry if available and is well validated.

The training, internal validation, and external validation datasets for this version of the software were sourced retrospectively between 1996 and 2021 from public (TCIA; training/loss validation/testing dataset) and private institutions (UF, HM, JFK; internal validation dataset and UCLA; external validation dataset) as shown in Figure 1. The training dataset consisted of consecutive patients with age range of 24–81 (average age 52.3; M:F, 3:2), Table 1, with pathology proven 2016 WHO grade 3 or 4 astrocytoma or 2016 WHO grade 4 glioblastoma who had a brain MRI containing the four required sequences; T1WI without contrast (T1c), T1WI with contrast (T1c+), T2WI (T2) and Fluid Attenuated Inversion Recovery (T2-FLAIR). The dataset consisted of 557 public, multicentre (38 sites), and multivendor (32 scanners) MRIs (300 preoperative and 257 postoperative) used for training (70%), validation (20%), and testing (10%) of NS-HGlio. Approximately 50% of the training dataset parameters were meeting the parameters set in the standardized Brain Tumor Imaging Protocol (BTIP).[6] The other 50% of the dataset included less stringent parameters to aid generalizability of the performance beyond the parameters of BTIP in institutions that do not adhere to BTIP routinely and to aid in the analysis of data acquired prior to BTIP adoption. The minimum parameters of the MRI sequences included in the datasets are summarized in Table 2.

The internal validation (IV) dataset consisted of 70 consecutive MRIs (46 preoperative and 24 postoperative) adult patients who are 18 years or older with pathology proven 2016 WHO grade 3 and 4 astrocytoma or 2016 WHO grade 4 glioblastoma (average age 64.5; M:F, 1.6:2) from 3 institutions; University of Florida, Gainesville FL, Houston Methodist Hospital, Houston TX and JFK Medical Center, Edison NJ. The external validation (EV) dataset consisted of 40 retrospective brain MRIs (20 preoperative and 20 postoperative) from IDH wild type, high-grade (2016 WHO grade 3 and 4) glioma patients (average age 58.6; M:F, 3:1) that were selected from the UCLA Neuro-Oncology Database (UCLA IRB #10-000655) by the Brain Tumor Imaging Laboratory (BTIL) group in a single blinded design where the authors were blinded to the imaging data. The IV and EV Subjects' characteristics are summarized in Table 1.

The Ground Truth

For the training and internal validation datasets, six neuroradiologists each with >6 years of neuroradiology practice experience created the ground truth (GT) with 2 rounds of overreads for consensus. Disagreements were handled by in-person agreement (see Supplementary Table 1). Analyze 14.0 (Mayo Clinic, AnalyzeDirect, Overland Park, KS), was used to create the GT. The GT protocol followed the standards as defined by the National Cancer Institute Cancer Imaging Achieve VASARI[7] (Visually AcceSAble Rembrandt Images) features set for HGG segmentation.

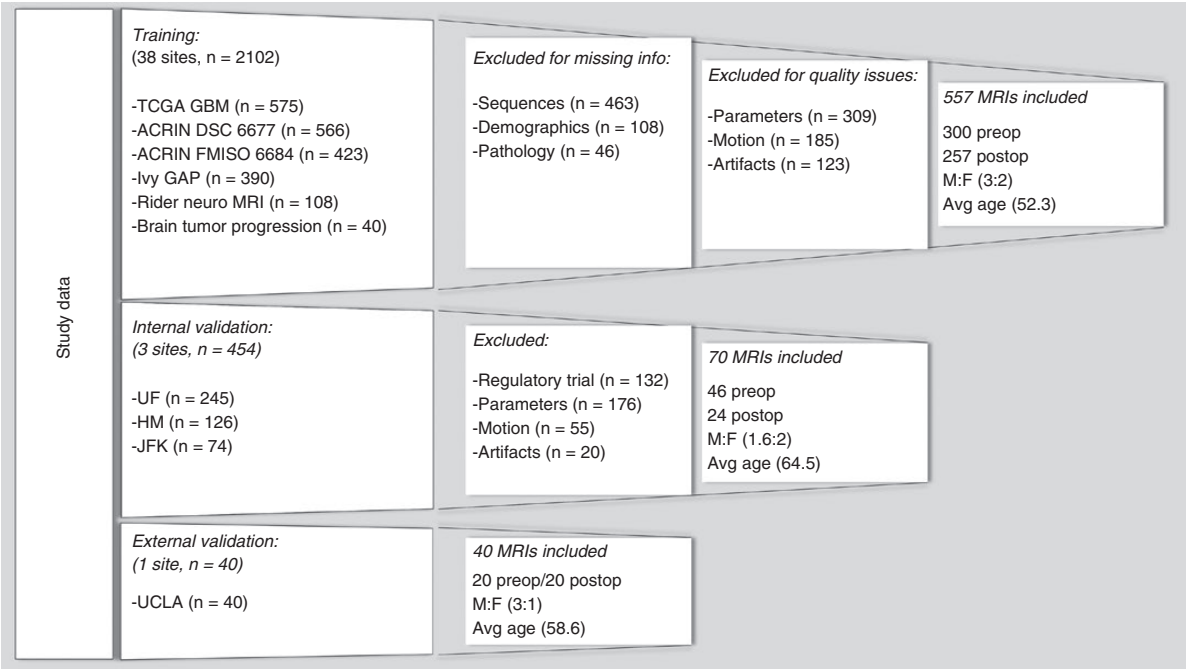

Figure 1. Training, Internal Validation and External Validation dataset flow chart.

Downloaded from https://academic.oup.com/noa/article/5/1/vdac184/6948060 by Universitaetsbibliothek Bern user on 25 January 2023

Table 1. Overall datasets subjects characteristics

	Public dataset	Internal validation dataset	External validation dataset
Mean age (yr, range)	52.3 (24–81)	64.5 (18–86)	58.6 (32–94)
HGG 2016 WHO grade 3 % vs grade 4 %	12%/88%	50%/50%	3%/87%
Preoperative: Postoperative	1:0.8	1:0.5	1:1
Gender M:F	3:2	1.6:2	3:1

Table 2. Minimum training dataset MRI sequences parameters for both 1.5T and 3T scanners

T1c-	Pre-contrast 2DT1-weighted (T1w) axial acquisition with ≤1 × 1 × 5 mm anisotropic voxel size *OR* 3DT1w axial, coronal, or sagittal acquisition with ≤2 mm. No interslice gap.
T1c+	Post-contrast 3DT1w axial, sagittal or coronal with ≤2 mm.
T2	Pre-contrast or post-contrast; 2DT2-weighted (T2w) axial acquisition with ≤1 × 1 × 5 mm non-isotropic voxel size *OR* 3DT2w axial, coronal, or sagittal acquisition with ≤2 mm. No interslice gap.
T2-FLAIR	Pre-contrast or post-contrast; 2DT2-FLAIR axial acquisition with ≤1 × 1 × 5 mm non-isotropic voxel size *OR* 3D FLAIR axial, sagittal or coronal acquisition with ≤2 mm. No interslice gap.

The quality of the GT for the training and internal validation datasets was validated using a double blinded web-based voting system where the GT designed for NS-HGlio development by the authors was compared against the publicly available BraTS GT for a set of randomly selected 20 preoperative MRIs form the first 100 MRIs of the publicly available BraTS 2020 dataset[8–11] as shown in Figure 2. The results of this validation demonstrated that the GT designed by the authors was preferred in 70% of cases by the 12 clinicians who participated in the validation (8 neuroradiologists, 2 radiation oncologists, and 2 neuroradiology fellows).

Two neuroradiologists from UCLA created the GT for the external validation dataset using the institution standards of practice and no over-read.

AI Architecture and Training

NS-HGlio was developed between January 2019 and October 2021 as a single algorithm that can segment HGG on preoperative and postoperative MRIs. NS-HGlio is formed of a preprocessing pipeline, the AI core algorithm, and a post-processing pipeline.

The preprocessing pipeline performs image registration, normalization, and skull stripping. The AI core algorithm is a 3D U-Net.[12,13] The 3D U-net consists of 5 sub-models, which are trained by the concept of 5-fold cross-validation to create an ensemble model using majority voting. The majority voting design is where every individual sub-model votes for a class, and the majority wins. In statistical terms, the predicted target label of the ensemble is the mode of the distribution of individually predicted labels. Dropout based regularization[14] was used with random dropout rate of 50% across all layers of the model. Adam optimization with learning rate of 3e-4 was used. Data augmentation was employed as elastic deformation with alpha [0–200] and sigma [9–13], gamma transformation [0.7–1.5] which changes the mean and standard derivation of the patch, random rotations in range [–30, +30] degrees, random scale range [0.85, 1.25] and random image cropping and padding with a crop size of 128 were used. The result of the deep learning model is a label mask comprising four labels: Label "0" for background/healthy tissue, label "1" for T2/FLAIR hyperintensity/Ed; label "2" for enhancing tissue/Enh; label "3" for necrosis+ resection cavity/Nec. NS-HGlio uses 4 standard MRI sequences as input; T1c–, T1c+, T2, and FLAIR sequences. After the segmentation is created using the pre-processed T1c–, T1c+, T2, and FLAIR sequences, the label maps are then resampled and re-oriented to the original sequences. This step allows the user to view the segmentations as an overlay to the original un-processed sequences in NS-HGlio clinical decision support platform or any other viewer modality/PACS system (Figure 2).

Results

The Dice Similarity Coefficient (DSC) which measures the degree of spatial overlap of NS-HGlio output (Enh., FLAIR hyperintense/Ed. and Enh + Ed/Whole Lesion,WL) and the consensus GT was used to measure the device performance. Repeatability was tested using 14 subjects with newly diagnosed GBM from the MGH QIN GBM[15] treatment response dataset that contained the needed 4 sequences by analyzing the Intraclass correlation coefficient (ICC) of the Enh. and Ed. Labels across all timepoints as well as the mean, standard deviation, and Coefficient of Variance (CV) of the log10volumes of the whole lesion. The 14 subject dataset contained 2 postoperative MRIs per subject performed after surgery but prior to radio/chemotherapy, typically 2–5 days apart.

Internal and External Validation (Table 3)

The internal validation (IV) dataset consisted of 70 MRIs (46 preoperative and 24 postoperative) from 3 institutions (UF,

Downloaded from https://academic.oup.com/noa/article/5/1/vdac184/6948060 by Universitaetsbibliothek Bern user on 25 January 2023

Neuro-Oncology Advances

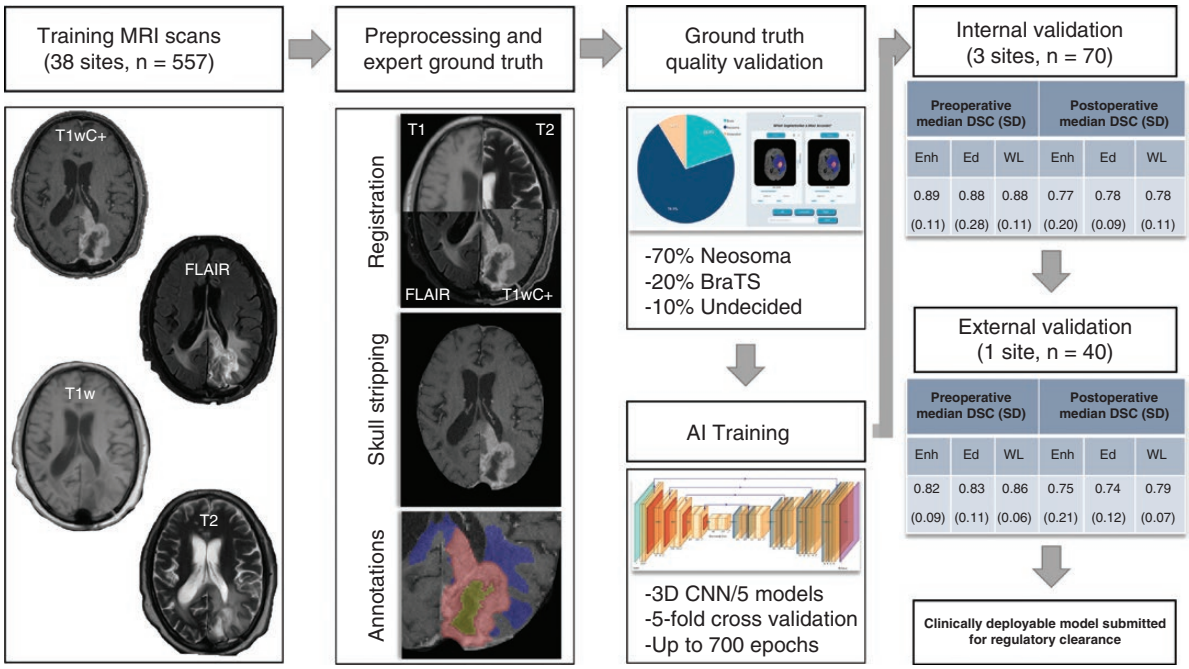

Figure 2. Algorithm inputs, processing pipeline, Ground Truth (GT) example, GT validation, and performance validation results. Final segmentation output is color coded; blue (T2-FLAIR hyperintensity), red (enhancing), and yellow (necrosis and surgical cavity).

Table 3. Internal and external validation DSC results

	Preoperative median DSC (SD)			Postoperative median DSC (SD)		
	Enh	Ed	WL	Enh	Ed	WL
IV (Internal Validation)	0.89 (0.11)	0.88 (0.28)	0.88 (0.11)	0.77 (0.20)	0.78 (0.09)	0.78 (0.11)
EV (External Validation)	0.82 (0.09)	0.83 (0.11)	0.86 (0.06)	0.75 (0.21)	0.74 (0.12)	0.79 (0.07)

HM, JFK) and the external Validation (EV) dataset consisted of 40 MRIs (20 preoperative and 20 postoperative) from a single institution (UCLA) as discussed previously.

In the preoperative setting (Figure 3), NS-HGlio was able to achieve median DSC Enh. 0.89 (SD 0.11), Ed. 0.88 (SD 0.28), Enh+Ed/WL 0.88 (SD 0.11) in the internal validation dataset and showed good generalizability in the external validation dataset where it achieved median DSC Enh. 0.82 (0.09), Ed. 0.83 (0.11), Enh+Ed/WL 0.86 (0.06). This corresponds to an 8% drop in performance for the preoperative DSC mean for Enh. label which is suggestive of good generalizability and reproducibility to unseen data in the preoperative setting.

In the postoperative setting (Figure 3), NS-HGlio was able to achieve median DSC Enh. 0.77 (SD 0.20), Ed 0.78. (SD 0.09), Enh+Ed/WL 0.78 (SD 0.11) and showed good generalizability in the external validation dataset where it achieved median DSC Enh. 0.75 (0.21), Ed 0.74 (0.12), Enh+Ed/WL 0.79 (0.07). This corresponds to a 3.5% drop in performance for the postoperative DSC mean for Enh. label which is suggestive of excellent generalizability and reproducibility to unseen data in the preoperative setting.

These results show that the device performance is better than the reported inter-rater performance of experts when compared to results from Visser et al.[16] evaluating the inter-rater agreement in glioma segmentations on longitudinal MRIs between 4 experts and 4 novices.

When evaluating the relationship between the volume of the Enh. lesions and the device DSC performance for the internal and external validation datasets both in the preoperative and the postoperative setting a general trend was observed were higher Enh. volumes were associated with better DSC performance. However, statistical significance was only observed in the postoperative setting of the external validation dataset (see Supplementary Figure 1).

Repeatability

Repeatability was tested using the 14 subjects with newly diagnosed GBM from the MGH QIN GBM treatment response dataset that contained the needed 4 sequences by analyzing the Intraclass correlation coefficient (ICC) of the Enh. and Ed. Labels across all timepoints as well as the mean, standard deviation, and Coefficient of Variance

Downloaded from https://academic.oup.com/noa/article/5/1/vdac184/6948060 by Universitaetsbibliothek Bern user on 25 January 2023

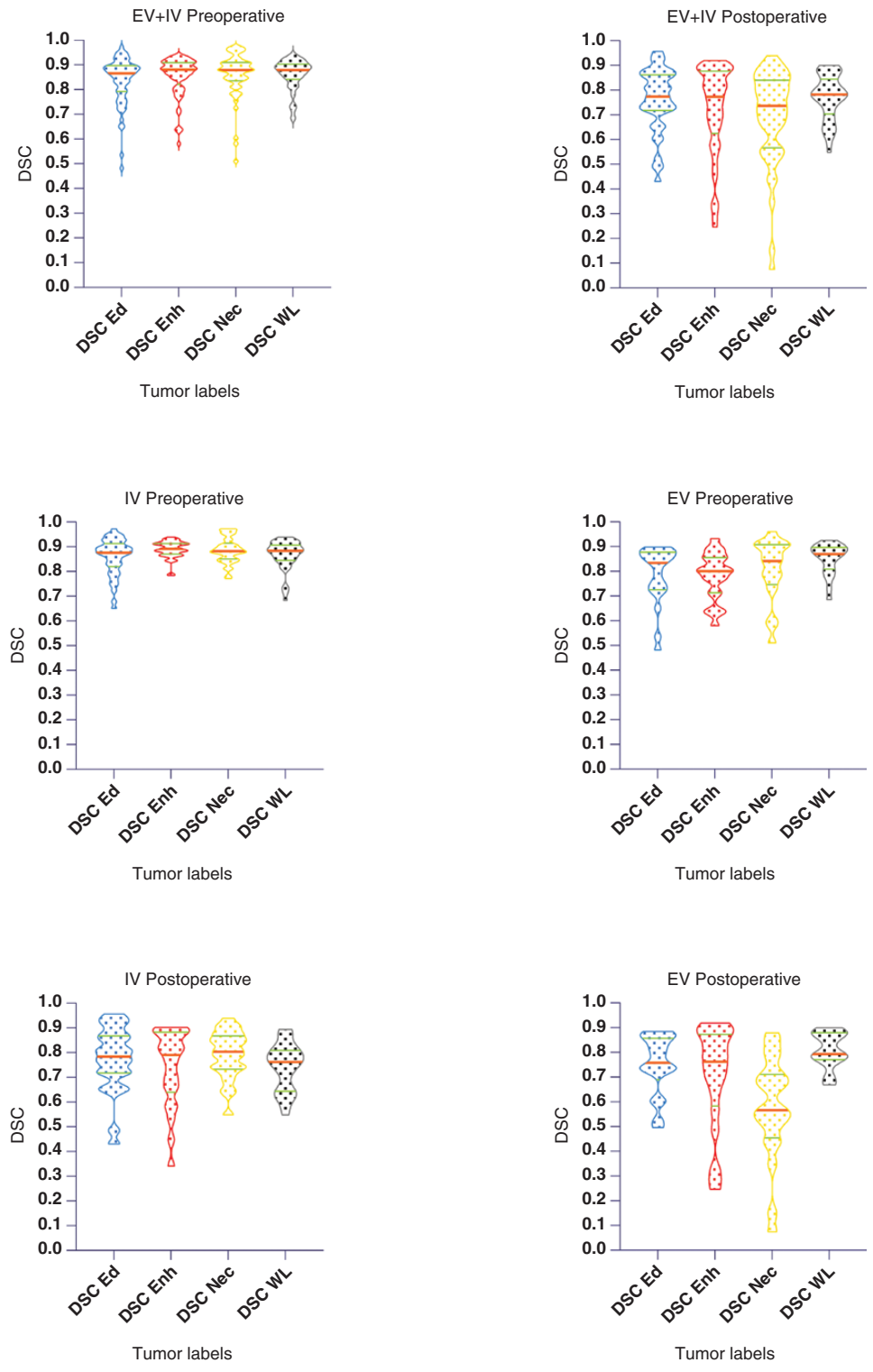

Figure 3. Violin-plot graphs illustrating the DSC performance for the preoperative and postoperative MRIs from the internal and external validation datasets of NS-HGlio. As expected, generalizability results in very modest drop in performance (8% for preoperative and 3.5% for postoperative DSC mean for Enh. label). For the postoperative DSC, Nec label included necrotic tissue and surgical cavity combined in one label to allow for a single algorithm training for both preoperative and postoperative imaging.

Downloaded from https://academic.oup.com/noa/article/5/1/vdac184/6948060 by Universitaetsbibliothek Bern user on 25 January 2023

(CV) of the log10volumes of the whole lesion. The 14 subject dataset contained 2 postoperative MRIs per subject performed after surgery but prior to radio/chemotherapy, typically 2–5 days apart. NS-HGlio showed excellent repeatability achieving ICC of 0.95 for Enh. and 0.92 for Ed. labels. The WL log10 volumes demonstrated a mean of 4.87, SD of 0.21 and CV of 0.53 (see Supplementary Table 2).

Discussion

Background

High-grade glioma is the most common malignant primary brain cancer, with more than 13,000 Americans expected to be diagnosed this year and 10,000 expected deaths. The prognosis for HGG has been virtually unchanged for decades with an average length of survival of 15–18 months and a 5-year survival rate of only 6%.[17–19]The poor survival rate of 40% in the first year and 6% at 5 years as well as the lack of effective therapeutics in the last 2 decades highlights the importance of standardizing and objectifying imaging assessment of these tumors to enhance the chances of clinical trials success and facilitate inter-institutional and international collaborations.

The Response Assessment in Neuro-Oncology (RANO) is the standard for HGG assessment in clinical trials and is increasingly being adopted in clinical care. RANO uses conventional 2D measurements (carried over from the 1981 WHO and the 1990 Macdonald criteria) that has proven largely inadequate for accurately characterizing the growth of certain complex geometric shapes that are not uncommon in HGG.[20] Schmitt P. et al[21] concluded that bi-dimensional measurements are inadequate in evaluating tumor growth rates when the slice thickness is ≥ 3 mm (regardless of head positioning) and to properly evaluate small gliomas manually (20 mm in diameter or less), it is necessary to exactly replicate the MRI scanning conditions longitudinally. Furthermore, it is known that manual measurements suffer from significant inter-rater and intra-rater variability leading to high rate of adjudications in clinical trials and time delays.

Numerous studies have demonstrated that 3D volume measurements are more reliable and less variable than 2D measurements.[22,23] In three studies 3D tumor measurements have been found to be predictive of survival and progression free survival in recurrent HGG using Gd-enhanced T1-weighted images,[24] T1 subtraction maps[19] and DWI[25] sequences. Another study, semiautomated postoperative tumor volume estimations were found to be more reliable, reproducible, and faster than manual assessments in diagnosing tumor recurrence and progression.[26] In another study it was shown that quantifying residual and recurrent glioblastomas alike, not only was semi-automatic volumetric segmentation faster than the manual technique, but it was also more reliable and reproducible compared with 1D and 2D measurements.[27] In 2017, the modified volumetric mRANO criteria were introduced and included recommendations for volumetric change for the enhancing tumor volume to allow categorization of the imaging changes according to the original RANO categories of Complete Response (CR), Partial Response (PR), Stable Disease (SD), and Progressed Disease (PD). A strong correlation was observed between mRANO progression free survival (PFS) and overall survival (OS) while no correlation was observed between radiographic PFS and OS for standard RANO or iRANO in a cohort of 47 patients with recurrent glioblastoma (rGBM, WHO grade 4 HGG) enrolled in a prospective phase II convection-enhanced delivery of an IL4R-targeted immunotoxin (MDNA55-05, NCT02858895).[28] Recently, 3D volumetric measurements showed significantly longer estimated PFS, more stable and considerably slower measures of tumor growth rate, the highest inter-reader agreement and significantly lower reader discordance rates vs 2D RANO for low grade glioma (LGG).[29] Finally, there is an increasing interest in using tumor growth (*g*) and regression (*d*) metrics to measure response to treatment[30] using 3D measures.

For this, NS-HGlio was developed between January 2019 and October 2021 as a single algorithm that can segment high-grade glioma (HGG) on preoperative and postoperative MRIs.

Impact

We expect this technology to have an impact on HGG clinical trials, clinical care, adherence to RANO and mRANO utilization and in the future the expanding utilization of tumor growth (*g*) and tumor regression (*d*) metrics.

Historically clinical trials have suffered from high failure rate due to many reasons; this is compounded with the uncertainty in manual longitudinal tumor measurements on brain MRIs causing high rate of adjudications due to inter-rater and intra-rater variability of the 2D measurements. NS-HGlio has the potential to lower the need for adjudications while standardizing and objectifying longitudinal tumor volumes measurements which will potentially make HGG clinical trials more cost effective. The turnaround time per MRI segmentation is on average reduced by 97% versus manual volumetric segmentation (2 min vs 60 min) and 90% versus semiautomated volumetric segmentation (2 min vs 20 min). Additionally, the technology can easily and quickly be used to analyze historical clinical trials data to evaluate for possible new insights given the recent changes in CNS tumors classification based on the 2021 WHO Classification of CNS Tumors, fifth edition ($WHOCNS_5$) that would require re-classification of historical trials data.

Additionally, NS-HGlio is expected to have an impact on the current neuro-oncology practice. The use of RANO and mRANO criteria in clinical practice is significantly lagging clinical trials which is primarily due to the time and financial requirements of having the neuroradiologist perform those measurements on every MRI and the absence of an FDA cleared technology that can replace the tedious, time consuming, and error prone manual process. The technology can also be used in centers that do not have the neuro-oncology expertise of the tertiary care centers allowing for better streamlining of resources and lower financial burden on the patients.

As for future technologies, NS-HGlio will provide accurate and reproducible Regions of Interest (ROI) for

Downloaded from https://academic.oup.com/noa/article/5/1/vdac184/6948060 by Universitaetsbibliothek Bern user on 25 January 2023

radiomics and radiogenomics based analysis and predictions.

Limitations

NS-HGlio performance was not without pitfalls. For example, we noticed in a small percentage of cases erroneous segmentation of Ed. in the cerebellar hemispheres due to exaggerated pulsation artifacts from the sigmoid sinuses inducing artifactually increased cerebellar FLAIR signal. Another example would be missed small Enh. lesions below 5 mm in diameter in cases where the original T1c+ sequence voxel size was more than 1mm isovolumetric and the T1c– sequence parameters were not matching those of the T1c+ sequence. Very low resolution of any of the 4 sequences (ie, slice thickness above 5 mm or presence of interslice gaps) resulted in false positive Enh. and Ed. segmentations in the cortical and subcortical tissue specially in the anterior aspect of the frontal lobes and in the occipital lobes. Similarly, errors in detecting postoperative enhancing thickened dura at the operative site as Enh. segmentation or over- or under-segmentation of the resection cavity specially if containing hemorrhage were observed if the T1c– acquisition parameters were not equivalent to the T1c+ acquisition parameters as defined by BTIP. The cases with lower DSC values as shown in Figure 3 suffered from some of the data quality issues described here. Additionally, the DSC performance drop in the postoperative setting compared to the preoperative setting could be in part related to the known higher GT variability between raters in the postoperative phase. These observations required the development of a protocol guidance for the optimum MRI acquisitions for NS-HGlio which is consistent with the BTIP guidance. Further validation of the technology after FDA clearance through a multi-institutional external validation study using prospective longitudinal datasets is desired.

Workflow Integration

To ensure adoption of the technology by clinicians, researchers, and clinical trials it is important to facilitate the integration of any AI technology with the already existing hospital systems. For this, NS-HGlio would seamlessly integrate with the already existing PACS, EMR and research systems through a technology that we developed and refer to as Edge System. Additionally, NS-HGlio can also be accessed and used through a standalone client dashboard that would allow for visualization, editing and structured reporting of the results. The dashboard also allows virtual Multidisciplinary Team meetings (Tumor Board meetings) with audit trailing, decision support, and zoom-like videoconferencing capabilities among its many capabilities (Figure 4). Additionally, the segmentations can be used as an ROI for further analysis of additional MRI sequences such as diffusion and perfusion imaging.

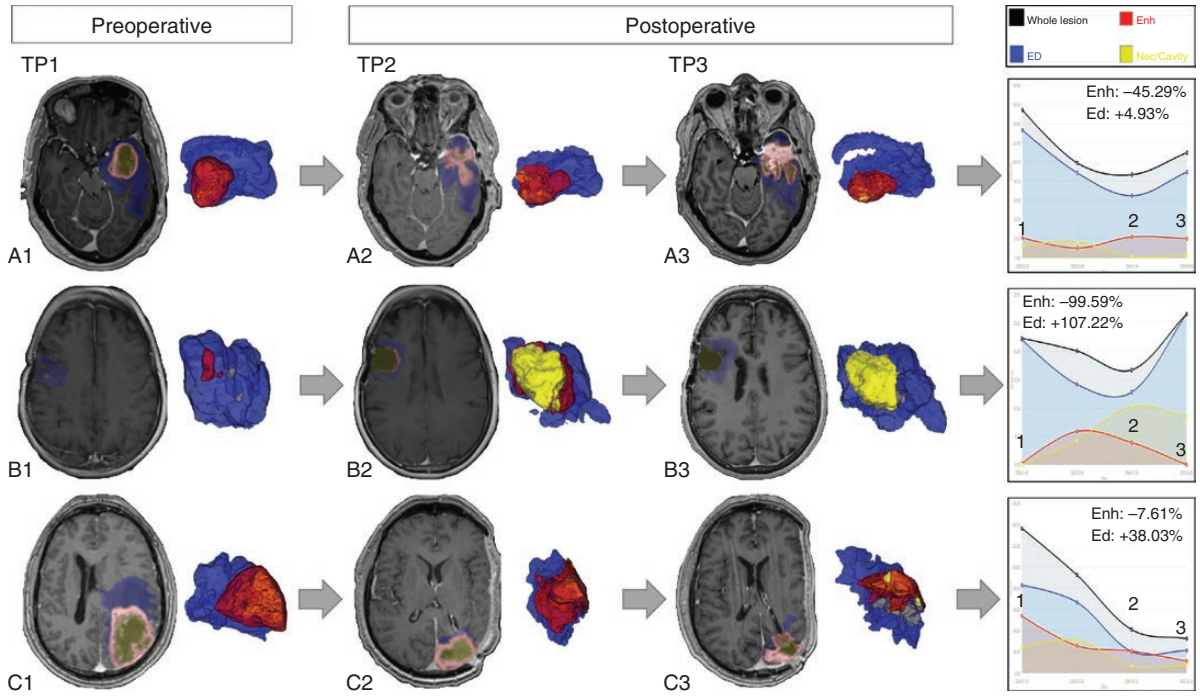

Figure 4. Three subjects (A–C) with representative MRI with segmentation color overlay and 3D reconstruction of the segmentation at 3 different timepoints (1) preoperative, (2) 3-months postoperative and (3) 9-months postoperative. (4) Graphical output from NS-HGlio demonstrating the different segments changes over the 3 time points (black = WL, Red = Enh, Blue = Ed, and yellow = Nec + Cavity). Volumetric measurements derived from the detected segmentations are used to calculate the rate of change over time and inform mRANO.

Downloaded from https://academic.oup.com/noa/article/5/1/vdac184/6948060 by Universitaetsbibliothek Bern user on 25 January 2023

Conclusion

NS-HGlio is accurate, repeatable, and generalizable. The output can be used for visualization, documentation, treatment response monitoring, radiation planning, intra-operative targeting, estimation of Extent of Resection (EOR) and RTV among others. Additional re-training with user-site specific data is likely to improve performance and generalizability. A larger multi-institutional external validation of the algorithm performance is desired after FDA clearance.

Supplementary material

Supplementary material is available online at *Neuro-Oncology Advances* online.

Keywords:

artificial intelligence | glioma | machine learning | RANO | segmentation.

Acknowledgments

We would like to acknowledge the patients and their families for their contribution to this work. We would also like to thank the Swiss Cancer League (grant KFS-3979-08-2016) and the Swiss Personalized Health Network (SPHN) initiative.

Unblinded acknowledgment

Data generated or analyzed during the study is available from the corresponding author by request.

Funding

Neosoma Inc., The Swiss Cancer League (grant KFS-3979-08-2016) and the Swiss Personalized Health Network (SPHN) initiative.

Conflict of interest statement: Dr. Abayazeed and Michael Hill own stocks in Neosoma Inc. Ben Ellingson has the following conflicts of interest; For profit: Medicenna—*Paid Consultant, Ad Board,* MedQIA, LLC—*Paid Consultant, Ad Board*, Servier Pharmaceuticals—*Paid Consultant, Ad Board*, Siemens—*Paid Consultant, Research Grant, Ad Board*, Janssen Pharmaceuticals—*Research Grant*, Imaging Endpoints—*Paid Consultant*, Kazia—*Paid Consultant, Ad Board*, Oncoceutics/Chimerix Inc—*Consultant*, Sumitomo Dainippon Pharma Oncology—*Consultant, Ad Board*, ImmunoGenesis—*Consultant, Ad Board*, Ellipses Pharma—*Consultant, Ad Board*, Monteris—*Paid Consultant, Ad Board*, Global Coalition for Adaptive Research (GCAR)—*Consultant, Ad Board*, Neosoma—*Paid Consultant, Ad Board*, Alpheus Medical, Inc.—*Paid Consultant, Ad Board*, Curtana Pharma—*Paid Consultant, Ad Board*, Sagimet Biosciences—*Paid Consultant, Ad Board*, Sapience Therapeutics—*Paid Consultant, Ad Board*. Non-profits: National Brain Tumor Society—*Research Grant*, American Cancer Society—*Research Grant*, NIH/NCI Cancer Imaging Steering Committee—*Paid Consultant.*

This paper is being submitted as one of a series that will include the results of the 501K FDA multi-institutional longitudinal MRMC validation for the technology in this publication, algorithms for RANO-BM (in R&D), RANO-LGG (in R&D), True Progression Prediction (in R&D), Tumor Infiltration Prediction (in R&D), IDH mutation prediction (in R&D), and MGMT promotor methylation prediction (in R&D).

Authorship statement: Data collection, Q.C. and G.T.: A.Abayazeed, M.Q., S.M., M.M., S.F., I.T., J.L., B.E., and A.Abassy. Algorithm training, validation, and testing: M.M., M.H., A.Abayazeed, and M.R. Ground Truth comparison: P.D., K.N., V.K., J.W., N.A., V.K., A.B., A.K., A.K., S.R., T.S., and R.D. Data analysis: M.M., M.H., C.R., and A.Abayazeed. Manuscript writing: A.Abayazeed and M.R. Manuscript review: A.Abayazeed, J.W., S.F., I.T., R.D., B.E., and M.R.

References

1. Ostrom QT, Gittleman H, Farah P, et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2006–2010. *Neuro Oncol.* 2013;15(Suppl):2ii–256.
2. Ellingson BM, Wen PY, Cloughesy TF. Modified criteria for radiographic response assessment in glioblastoma clinical trials. *Neurotherapeutics.* 2017;14(2):307–320.
3. Kickingereder P, Isensee F, Tursunova I, et al. Automated quantitative tumor response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. *Lancet Oncol.* 2019;20(5):728–740. doi:10.1016/S1470-2045(19)30098-1
4. Meier R, Porz N, Knecht U, et al. Automatic estimation of extent of resection and residual tumor volume of patients with glioblastoma. *J Neurosurg.* 2017;127(4):798–806. doi:10.3171/2016.9.JNS16146
5. Wen PY, Macdonald DR, Reardon DA, et al. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *J Clin Oncol.* 2010;28(11):1963–1972.
6. Ellingson BM, Bendszus M, Boxerman J, et al; Jumpstarting Brain Tumor Drug Development Coalition Imaging Standardization Steering Committee. Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials. *Neuro Oncol.* 2015;17(9):1188–1198.
7. VASARI Research Project. *The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki*; TCIA. 2015. https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project

Downloaded from https://academic.oup.com/noa/article/5/1/vdac184/6948060 by Universitaetsbibliothek Bern user on 25 January 2023

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. Bakas S, Akbari H, Sotiras A, et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Nat Sci Data*. 2017;4:170117.
10. Bakas S, Reyes S, Jakab A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge, *arXiv*, 2018, arXiv:1811.02629, preprint: not peer reviewed.
11. Bakas S, Akbari H, Sotiras A, et al. Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. *The Cancer Imaging Archive* 2017.
12. Ronneberger O, Fischer P, Brox T et al. U-Net: convolutional networks for biomedical image wegmentation BT - medical image computing and computer-assisted intervention – MICCAI 2015. In Navab N, Hornegger J, Wells WM, Frangi AF, eds. Cham: Springer International Publishing; 2015:234–241.
13. Isensee, F., Petersen, J., Klein, A. et al nnu-net: self-adapting framework for u-net-based medical image segmentation. *arXiv*, 2018, arXiv:1809.10486, preprint: not peer reviewed.
14. Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014;15(1):1929–1958.
15. Prah MA, Stufflebeam SM, Paulson ES, et al. Repeatability of Standardized and Normalized Relative CBV in Patients with Newly Diagnosed Glioblastoma. *AJNR Am J Neuroradiol* 2015;36(9):1654–1661. doi:10.3174/ajnr.A4374
16. Visser M, Müller DMJ, van Duijn RJM, et al. Inter-rater agreement in glioma segmentations on longitudinal MRI. *NeuroImage Clin*. 2019;22(July 2018):101727.
17. Campos B, Olsen LR, Urup T, Poulsen HS. A comprehensive profile of recurrent glioblastoma. *Oncogene*. 201635(45):5819–5825.
18. Omuro A, DeAngelis LM. Glioblastoma and other malignant gliomas: a clinical review. *JAMA*. 2013;310(17):1842–50.
19. Eder K, Kalman B. Molecular heterogeneity of glioblastoma and its clinical relevance. *Pathol Oncol Res*. 2014;20:777–787.
20. Ellingson BM, Kim HJ, Woodworth DC, et al. Recurrent glioblastoma treated with bevacizumab: contrast-enhanced T1-weighted subtraction maps improve tumor delineation and aid prediction of survival in a multicenter clinical trial. *Radiology*. 2014;271(1):200–210.
21. Schmitt P, Mandonnet E, Perdreau A, Angelini ED. Effects of slice thickness and head rotation when measuring glioma sizes on MRI: in support of volume segmentation versus two largest diameters methods. *J Neurooncol*. 2013;112(2):165–172.
22. Sorensen AG, Patel S, Harmath C, et al. Comparison of diameter and perimeter methods for tumor volume calculation. *J Clin Oncol*. 2001;19(2):551–557.
23. Meier R, Knecht U, Loosli T, et al. Clinical Evaluation of a Fully-automatic Segmentation Method for Longitudinal Brain Tumor Volumetry. *Sci Rep*. 2016;6:23376. doi:10.1038/srep23376
24. Dempsey MF, Condon BR, Hadley DM. Measurement of tumor "size" in recurrent malignant glioma: 1D, 2D, or 3D? *AJNR Am J Neuroradiol*. 2005;26(4):770–776.
25. Hwang EJ, Cha Y, Lee AL, et al. Early response evaluation for recurrent high-grade gliomas treated with bevacizumab: a volumetric analysis using diffusion-weighted imaging. *J Neurooncol*. 2013;112(3):427–435.
26. Ertl-Wagner BB, Blume JD, Peck D, et al; Members of the American College of Radiology Imaging Network 6662 Study Group. Reliability of tumor volume estimation from MR images in patients with malignant glioma. Results from the American College of Radiology Imaging Network (ACRIN) 6662 Trial. *Eur Radiol*. 2009;19(3):599–609.
27. Chow DS, Qi J, Guo X, et al. Semiautomated volumetric measurement on postcontrast MR imaging for analysis of recurrent and residual disease in glioblastoma multiforme. *AJNR Am J Neuroradiol*. 2014;35(3):498–503.
28. Ellingson BM, Sampson J, Achrol AS, et al. Modified RANO, immunotherapy RANO, and standard RANO response to convection-enhanced delivery of IL4R-targeted immunotoxin MDNA55 in recurrent glioblastoma. *Clin Cancer Res*. 2021;27(14):3916–3925.
29. Ellingson BM, Kim GHJ, Brown M, et al. Volumetric measurements are preferred in the evaluation of mutant IDH inhibition in non-enhancing diffuse gliomas: evidence from a phase I trial of ivosidenib. *Neuro Oncol*. 2022;24(5):770–778.
30. Huang RY, Young RJ, Ellingson BM, et al. Volumetric analysis of IDH-mutant lower-grade glioma: a natural history study of tumor growth rates before and after treatment. *Neuro Oncol*. 2020;22(12):1822–1830.

Downloaded from https://academic.oup.com/noa/article/5/1/vdac184/6948060 by Universitaetsbibliothek Bern user on 25 January 2023