AI-Guided RANO 2.0 Assessment and Pseudoprogression Incidence in Glioblastoma: A Comparative Analysis of Bidimensional and Volumetric Measurements

Michael J. Ciesielski, PhD[1,3], Danielle M. Casucci BA, BS, GN, CCRP, CCIP[1], Gene Kim, PharmD[2], Jonathan Loi, PhD[2], Shinji Yue, MS, MBA[2], Alyssa Bartholomew, RT (R)(CT)(MR)[2], Fanny Wang, MD[2], Ahmed Belal, MD[1,3]; Robert Fenstermaker, MD[1,3]

[1]MimiVax, Inc., Buffalo, NY, USA; [2]MERIT CRO, Inc., Madison, WI, USA; [3]Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA

IMG-143

Presenting Authors:
Gene Kim, gkim@meritcro.com; Jonathan Loi, jloi@meritcro.com;
Alyssa Bartholomew, abartholomew@meritcro.com

MIMIVAX

MERIT
Your Clinical Endpoint Expert

BACKGROUND

Accurate measurement of glioma burden is essential for assessing treatment response in glioblastoma (GBM). Traditional 2D RANO criteria rely on bidimensional measurements that poorly capture the irregular 3D morphology of high-grade gliomas, leading to inter-reader variability. While updated RANO 2.0 guidelines now include volumetric measurements, manual 3D segmentation remains time-intensive, limiting practical implementation in clinical trials. Artificial intelligence offers a scalable solution for rapid, reproducible volumetric quantification. However, rigorous validation against expert standards is needed to ensure reliability in regulatory and clinical trial contexts.

OBJECTIVES

- Develop and validate an AI-assisted volumetric segmentation pipeline for GBM
- Compare AI-assisted volumetric RANO 2.0 assessments with conventional 2D bidimensional RANO 2.0 criteria
- Assess variation of progression and psudoprogression in subjects assessed with AI volumetric RANO 2.0 and 2D bidimensional RANO 2.0

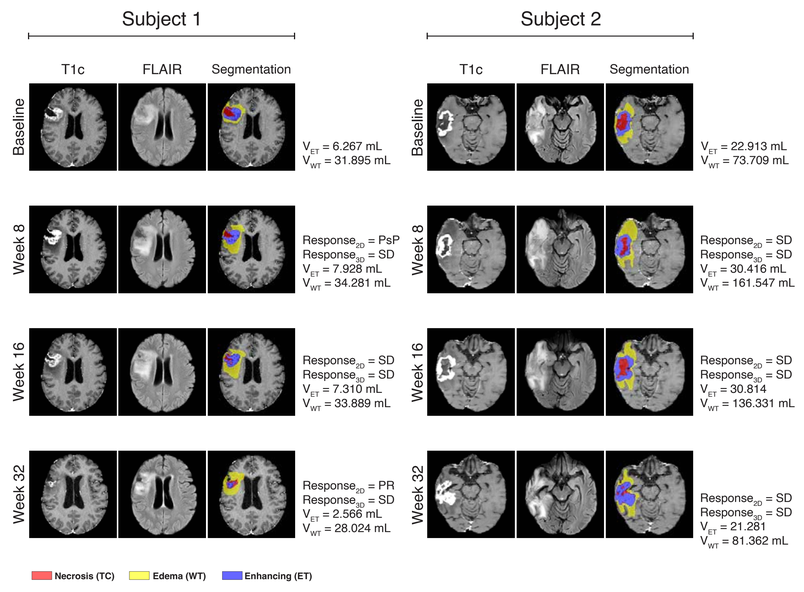

Figure 1. Representative T1c and FLAIR MRI images of two subjects with GBM at baseline, week 8, week 16, and week 32 with automated deep learning 3D volumetric segmentation of necrosis (TC) - red, edema (WT) - yellow, and enhancing tumor (ET) – blue, subregions. The RANO 2.0 response classification and ET and WT subregion volumes are shown.

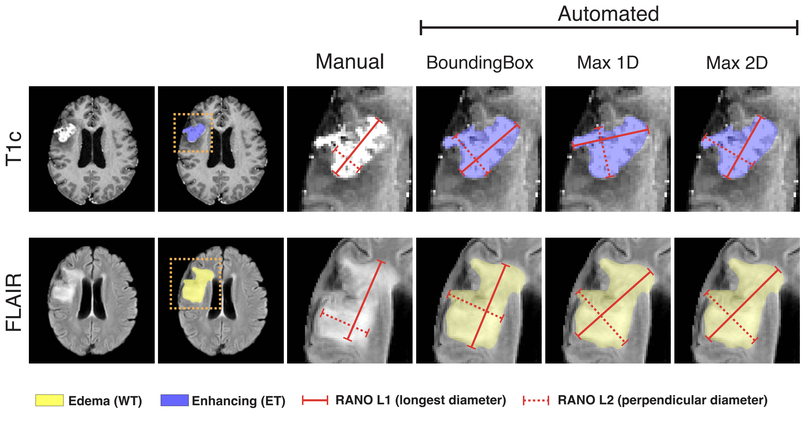

Figure 2. Representative T1c and FLAIR MRI images of Subject 1 from Fig. 1 comparing the RANO L1 (longest diameter) and L2 (perpendicular diameter) bidimensional measurements assessed using the manual, BoundingBox, Max1D, and Max2D methods using the WT and ET subregions.

METHODS

Patient Cohort

Eight patients with high-grade gliomas (GBM IDH wild-type or IDH-mutant grade 4 astrocytoma) were included.

Deep Learning Model

A 3D U-Net trained on publicly available glioma MRI datasets performed automated volumetric segmentation using multi-parametric input (T1, T1c, T2, FLAIR) to delineate enhancing tumor (ET), tumor core (TC), and whole tumor (WT) subregions.

Preprocessing

Standardized preprocessing included skull stripping, bias field correction, affine registration, and intensity normalization.

RANO 2.0 Assessment

Response assessment followed RANO 2.0 criteria for enhancing lesions with confirmation of progression beyond 12 weeks to account for pseudoprogression.

3D volumetric segmentation computed tumor volumes and automated bidimensional measurements on the slice with largest cross-sectional area, compared to manual measurements using four approaches:

1. BoundingBox - minimum area bounding box guiding measurement lines
2. Max1D - maximum length line followed by perpendicular line
3. Max2D - two maximum length lines placed simultaneously
4. 3D volumetric - AI-derived volumes

Measurement correlation and response agreement were analyzed across all five methods (manual + 4 automated).

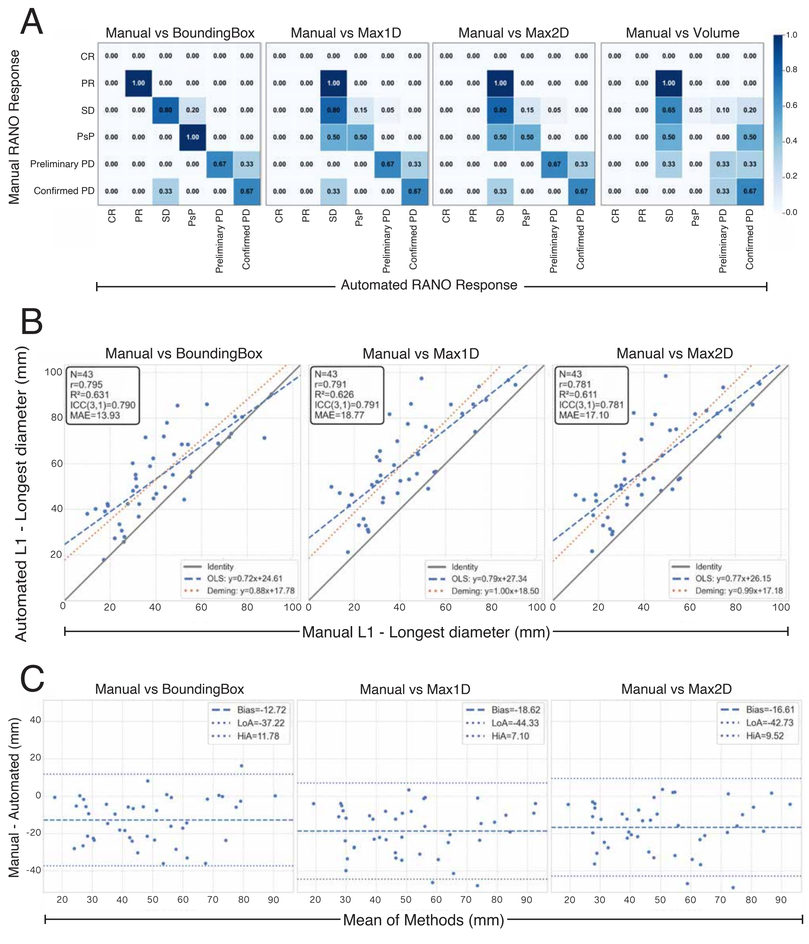

Figure 3. A. Confusion matrices showing concordance between RANO 2.0 response classifications using 2D manual versus automated 2D BoundingBox, Max1D, Max2D, and 3D volumetric approaches. Values represent proportions normalized per manual (true) label. **B.** Scatter plots comparing manual and automated longest diameters (L1) for enhancing lesions across all evaluable timepoints. Each plot includes the identity line (gray), ordinary least squares (ONL) regression (orange), and Deming regression (blue dashed) with correlation coefficient (r), intraclass correlation coefficient ICC (3,1), and mean absolute error (MAE). **C.** Bland-Altman plots showing mean bias and 95% limits of agreement (LoA) between manual and automated L1 measurements for the same methods as in (B). All plots share identical axes and scales for direct comparison.

RESULTS

- Three out of eight patients had different timing of confirmed radiographic pseudoprogression based upon manual 2D RANO 2.0 assessment to AI automated 3D RANO 2.0 assessment
- Deep learning-based volumetric segmentation provides consistent subregion delineation (Figure 1)
- Deep learning-based volumetric segmentation provides consistent subregion delineation with visual overlays confirming alignment to enhancing and FLAIR-defined tumor boundaries (Figures 1-2)
- Minor measurement deviations arise from tumor boundary determination differences between reader and AI model (Figure 2)
- Automated methods show 70-80% categorical concordance with manual RANO 2.0 labels for enhancing lesions, with BoundingBox method performing better (Figure 3A)
- Automated bidimensional measurements demonstrate strong correlation for L1 longest diameter (r = 0.78-0.80, ICC = 0.78-0.79), moderate for L2 (r = 0.54-0.57, ICC = 0.53-0.57), and good correlation for SPD (r = 0.66-0.72, ICC = 0.60-0.70) (Figure 3B)

CONCLUSIONS

This pilot study demonstrates feasibility of AI-assisted volumetric segmentation for high-grade gliomas with reasonable alignment between manual and automated assessments (70-80% categorical concordance). Despite the limited sample size (n=8), five subjects (62.5%) showed differences in pseudoprogression or progressive disease determination between 2D and 3D RANO 2.0 methods, highlighting the potential impact of measurement methodology on response classification.

Our findings reveal measurement variability in both approaches stemming from tumor boundary differences between readers and AI segmentation, consistent with prior GBM measurement studies. GBM's morphological heterogeneity and irregular geometry may be better captured through volumetric assessment.

A collaborative AI-human workflow offers a pragmatic approach for clinical translation, balancing automation efficiency with the clinical accuracy and expert oversight required for regulatory submissions. Future studies with larger cohorts should correlate volumetric and 2D RANO 2.0 assessments with clinical endpoints (PFS, OS) to determine which method provides superior prognostic value. This work represents an important step toward more reproducible and potentially more sensitive tumor burden assessment in GBM trials.

REFERENCES

1. Wen PY, van den Bent M, Vogelbaum MA, Chang SM. RANO 2.0: The revised Response Assessment in Neuro-Oncology (RANO) criteria for high- and low-grade glial tumors in adults designed for the future. Neuro Oncol. 2024 Jan 5;26(1):2-4. doi: 10.1093/neuonc/noad189. PMID: 37774741; PMCID: PMC10768981. **2.** 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. doi:10.1200/JCO.2009.26.3541. **3.** Villanueva-Meyer JE, Bakas S, Tiwari P, et al. Artificial Intelligence for Response Assessment in Neuro Oncology (AI-RANO), part 1: review of current advancements [published correction appears in Lancet Oncol. 2024 Dec;25(12):e626. doi: 10.1016/S1470-2045(24)00693-4.]. *Lancet Oncol*. 2024;25(11):e581-e588. doi:10.1016/S1470-2045(24)00316-4. **4.** Yang D. Standardized MRI assessment of high-grade glioma response: a review of the essential elements and pitfalls of the RANO criteria. *Neurooncol Pract*. 2016;3(1):59-67. doi:10.1093/nop/npv023.

DISCLOSURES

The authors have no financial conflicts of interest to disclose.

CONTACT INFORMATION

Gene Kim, MERIT CRO, Inc., gkim@meritcro.com

Michael Ciesielski, PhD, MimiVax, Inc., mciesielski@mimivax.com