Title

Volumetric measurements are preferred in the evaluation of mutant IDH inhibition in non-enhancing diffuse gliomas: Evidence from a phase I trial of ivosidenib

Journal

Neuro-Oncology, 24(5)

ISSN

1522-8517

Authors

Ellingson, Benjamin M
Kim, Grace Hyun J
Brown, Matt
et al.

Publication Date

2022-05-04

DOI

10.1093/neuonc/noab256

Peer reviewed

Neuro-Oncology
24(5), 770–778, 2022 | https://doi.org/10.1093/neuonc/noab256 | Advance Access date 9 November 2021

Volumetric measurements are preferred in the evaluation of mutant IDH inhibition in non-enhancing diffuse gliomas: Evidence from a phase I trial of ivosidenib

Benjamin M. Ellingson, Grace Hyun J. Kim, Matt Brown, Jihey Lee, Noriko Salamon, Lori Steelman, Islam Hassan, Shuchi S. Pandya, Saewon Chun, Michael Linetsky, Bryan Yoo, Patrick Y. Wen, Ingo K. Mellinghoff, Jonathan Goldin, and Timothy F. Cloughesy

UCLA Brain Tumor Imaging Laboratory, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA (B.M.E., S.C.); UCLA Center for Computer Vision and Imaging Biomarkers, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA (B.M.E., G.H.J.K., M.B., J.L., J.G.); Department of Radiological Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA (B.M.E., G.H.J.K., M.B., J.L., N.S., M.L., B.Y., J.G.); Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA (B.M.E.); Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA (S.C., T.F.C.); Servier Pharmaceuticals, LLC, Boston, Massachusetts, USA (L.S., I.H., S.S.P.); Center for Neuro-Oncology, Dana-Farber Cancer Institute & Harvard Medical School, Boston, Massachusetts, USA (P.Y.W.); Department of Neurology, Memorial Sloan Kettering Cancer Center, New York, New York, USA (I.K.M.)

Corresponding Author: Benjamin M. Ellingson, PhD, Director, UCLA Brain Tumor Imaging Laboratory (BTIL), Professor of Radiology and Psychiatry, Departments of Radiological Sciences, Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, 924 Westwood Blvd., Suite 615, Los Angeles, CA 90024, USA (bellingson@mednet.ucla.edu).

Abstract

Background. Since IDH-mutant (mIDH) low-grade gliomas (LGGs) progress slowly and have a relatively long survival, there is a significant need for earlier measurements of clinical benefit. Guidance using the LGG RANO criteria recommends serial bidirectional (2D) measurements on a single slice; however, questions remain as to whether volumetric (3D) measurements are better, since they would allow for more accurate measurements in irregular shaped lesions and allow readers to better assess areas of subtle change.

Methods. Twenty-one (out of 24) non-enhancing, recurrent mIDH1 LGGs were enrolled in a phase I, multicenter, open-label study of oral ivosidenib (NCT02073994), and with imaging pre- and post-treatment as part of this exploratory ad hoc analysis. 2D and 3D measurements on T2-weighted FLAIR images were centrally evaluated at an imaging contract research organization using a paired read and forced adjudication paradigm. The effects of 2D vs 3D measurements on progression-free survival (PFS), growth rate measurement variability, and reader concordance and adjudication rates were quantified.

Results. 3D volumetric measurements showed significantly longer estimated PFS (P = .0181), more stable (P = .0063) and considerably slower measures of tumor growth rate (P = .0037), the highest inter-reader agreement (weighted kappa = 0.7057), and significantly lower reader discordance rates (P = .0002) with 2D LGG RANO.

Conclusion. 3D volumetric measurements are better for determining response assessment in LGGs due to more stable measures of tumor growth rates (ie, less "yo-yo-ing" of measurements over time), highest inter-reader agreement, and lowest reader discordance rates. Continued evaluation in future studies is warranted to determine whether these measurements reflect clinical benefit.

Key Points

- 3D is better than 2D lesion measurements in non-enhancing IDH-mutant LGGs due to higher inter-reader agreement and lower reader discordance rates.
- 3D measurements lead to a longer PFS and lower growth rate variability compared with 2D measures.

Importance of the Study

Therapeutic advances in the treatment of IDH-mutant low-grade glioma (LGG) are needed due to the non-curative effect of radiation and chemotherapy for IDH-mutant astrocytoma or as a substitution for radiation and chemotherapy in IDH-mutant oligodendroglioma where long-term survival is more common and accompanied by progressive cognitive decline. Since IDH-mutant LGGs grow slowly and patients may have a relatively long survival, testing of therapies typically requires large, randomized trials using survival as the primary endpoint and can take longer than a decade to report. Thus, there is a significant need for the use of imaging measurements and serial surveillance to provide early evidence of therapeutic benefit of new agents in clinical trials. In the current study, we utilize imaging and clinical data from independent radiological facility measurements from a prospective phase I trial of ivosidenib (AG-120) to determine the best approach for evaluation of IDH-mutant LGG response to therapy. By looking at discordance between, and variation within, volumetric (3D) and bidirectional (2D) measurements, we show that volumetric measurements are better than 2D bidirectional measurements using the traditional blinded locked sequential read paradigm, resulting in highest inter-reader agreement and lowest reader discordance rates. Additionally, data suggests 3D measurements result in significantly longer estimate of progression-free survival (PFS) and more stable measures of tumor growth rates compared with 2D measurements, presumably due to less "yo-yo-ing" of measurements over time causing fewer early and erroneous calls of progression as well as more accurate estimates of growth rate.

There is no standard of care for adult or pediatric low-grade isocitrate dehydrogenase (IDH)-1 or 2 mutant gliomas,[1] the most common type of low-grade gliomas (LGGs),[2–4] and management of LGGs remains one of the most controversial areas in clinical neuro-oncology.[5] Patients with IDH-mutant 1p19q codeleted LGGs tend to be young[6] and can have a relatively long survival (5-13 years from diagnosis[6–9]). Aggressive chemoradiation can result in significant treatment-related morbidity (~98% of patients in RTOG-9802),[9] leading some oncologists to adopt a "watch and wait" strategy[10,11] prior to chemoradiation.[10,12] Thus, novel treatment strategies that are effective, have low toxicity, and prolong the time before using chemoradiation, such as direct targeting of the IDH mutation are highly desired.

Since IDH-mutant LGGs grow very slowly and patients may have a relatively long survival, testing of new (often non-beneficial) therapies in large clinical trials based solely on survival outcomes can take more than 20 years (eg, RTOG-9802).[7] Hence, there is a significant need for noninvasive imaging biomarkers that can quickly and reliably determine antitumor activity in patients with low-grade IDH-mutant tumors. To address this, guidance on radiographic evaluation of LGGs (ie, non-enhancing IDH-mutant gliomas) was provided by the response assessment in neuro-oncology (RANO) working group in 2011.[13] In this guidance, bidirectional measurement of T2 hyperintense lesion size on a single slice was recommended for serial monitoring and response assessment, despite more than 30 years of documented limitations of planar measurements in gliomas.[14–18] While the use of bidirectional measurements of T2 hyperintense diffuse gliomas is the recommended approach for evaluation of IDH-mutant glioma response assessment in part due to their purported reproducibility and perceived difficulty in performing 3D measurements, significant questions remain in terms of whether alternative approaches, including volumetric measurements, are potentially more reliable for determining radiographic response in LGGs.

Ivosidenib (AG-120) represents a first-in-class IDH inhibitor[19] that has been tested in 2 phase 1 studies in adult patients with IDH1 mutant gliomas.[20,21] In the current study, we utilize imaging data from independent radiological facility measurements, along with clinical data collected as part of one of a phase I trial of ivosidenib, to determine the best approach for radiographic evaluation of IDH-mutant (or low-grade) gliomas. By looking at discordance between and variation within volumetric and bidirectional measurements, as well as stability of measurements used to estimate growth rates, we demonstrate that volumetric measurements are better than bidirectional measurements for the radiographic assessment of IDH-mutant gliomas.

Methods

Patients

Men and women aged ≥18 years of age, with an Eastern Cooperative Oncology Group performance status of 0 to 1, and expected survival of ≥3 months, were eligible. All patients had an established diagnosis of m*IDH1* glioma that had recurred after—or not responded to—initial surgery, radiation, or chemotherapy. *IDH1* mutation status was based on local laboratory testing with retrospective central confirmation. Because this study was initiated before the most recent revision of the WHO Classification of Tumors of the Central Nervous System,[22] tumor subtypes were classified using the 2007 classification.[23]

Transformation of LGGs to a higher tumor grade is frequently associated with the appearance of tumor contrast enhancement on T1-weighted brain MRI. For the dose-expansion phase, patients were therefore separated into 2 cohorts based on the presence or absence of tumor contrast enhancement at time of enrollment according to the investigator. For the current study, a total of 21 patients (out of 24 originally reported as part of the expansion cohort[20,21]) *without* contrast-enhancing tumor at the time of enrollment and both pre- and post-treatment MRI scans available for radiographic assessment were evaluated at a total of 162 follow-up time points after treatment. All institutions involved in the trial received appropriate institutional review board and/or ethics committee approval. All patients provided informed written consent to participate in this clinical trial at the respective institutions. Table 1 describes patient demographic details for the study.

Study Design

The current study used prospective data from the dose-expansion cohort of a phase I, multicenter, open-label study (ClinicalTrials.gov NCT02073994). The primary objectives of the clinical study were to assess the safety and tolerability of oral ivosidenib as a single agent and determine the maximum tolerated dose or recommended phase 2 dose of ivosidenib in patients with solid tumors. Patients underwent baseline screening evaluations within 28 days prior to study day 1. Treatment with oral ivosidenib was continuous; one cycle was defined as 28 days. Images used for analyses included 2D or 3D T2-weighted images and/or 2D or 3D fluid-attenuated inversion recovery (FLAIR) images collected in compliance with the international standardized brain tumor imaging protocol.[24] (Supplementary Table 1) When possible, respective patients were scanned

Table 1 Patient Demographics

Subject ID	Age	Sex	ECOG @ Baseline	IDH Genotype	1p19q Status	Mutated ATRX	WHO Grade @ Screening	Prior Radiotherapy	Prior Systemic Therapy
1	46	M	0	Unknown	Not co-del	Unknown	2	Y	Y
2	39	M	0	R132H	Co-del	Unknown	2	Y	Y
3	55	M	0	R132H	Not co-del	Y	3	Y	Y
4	38	M	1	R132H	Co-del	Y	3	Y	Y
5	47	M	0	R132H	Co-del	Unknown	2	N	N
6	54	M	1	R132H	Unknown	Unknown	2	Y	Y
7	55	F	1	Unknown	Unknown	Unknown	2	Y	Y
8	58	F	0	R132H	Co-del	Unknown	2	N	Y
9	34	M	1	R132H	Not co-del	Unknown	2	N	N
10	41	F	1	R132H	Not co-del	Unknown	2	N	N
11	71	M	0	R132H	Co-del	Unknown	2	N	N
12	35	F	1	R132H	Not co-del	Y	2	N	N
13	24	M	1	R132H	Not co-del	Y	2	N	N
14	36	M	0	R132H	Unknown	Y	2	Y	N
15	64	M	0	Unknown	Not co-del	Unknown	2	N	N
16	34	F	1	R132H	Not co-del	Y	2	Y	Y
17	21	M	1	R132H	Not co-del	Unknown	2	Y	N
18	44	M	0	R132H	Not co-del	Unknown	Unknown (3 at diagnosis)	N	Y
19	38	M	0	R132H	Not co-del	Y	2	Y	Y
20	25	M	0	R132C	Unknown	Y	3	Y	Y
21	30	F	1	R132H	Not co-del	Y	Unknown (2 at diagnosis)	Y	Y

M = male; F = female; Y = yes; N = no.

on the same MRI system at the respective institutions throughout the entire trial. Tumor volumes were first segmented using a semi-automated approach on FLAIR images, then manually edited and signed off by the respective neuroradiologist, through an independent imaging contract research organization (MedQIA, Los Angeles, CA, USA) according to FDA guidelines. A total of 3 board-certified neuroradiologists with at least 10 years of experience performed the 2D and 3D imaging assessment in the current trial. Bidirectional measurements were extracted from volumetric measurements, and these measurements were explicitly reviewed and approved by the respective neuroradiologist. A standard, blinded locked sequential paired read with forced adjudication radiographic read paradigm[25] was used for the current study. In this paradigm, 2 readers were blinded to the number of follow-up studies and made measurements at each time point. This was followed by adjudication by a third reader for outcome discordance and final measurements. Response assessment was evaluated using MRI every 2 cycles (56 ± 3 days) according to the response assessment in neuro-oncology for LGGs (LGG RANO) guidelines[13] for bidirectional measurements (2D) and volumetric adaptations using spherical equivalent approximations (3D), as outlined by Chappell et al.[26] Table 2 compares 2D and 3D LGG RANO response categories.

Hypothesis Testing

The weighted kappa (κ_w) statistic[27] was used to evaluate inter-observer reliability in terms of response categorization for 2D compared with 3D measurements. We hypothesized that 3D measurements would have a higher kappa compared with 2D measurements. In addition, to test discordance, a logistic linear mixed-effects model was used with the binary variable as an outcome variable and fixed and random intercept as covariates, to integrate the inherent correlation among repeated measures per subject for reader. The discordance between LGG RANO categorization of "best response" for each patient and all measurement time points were evaluated between 2D and 3D of the adjudicated measurements. The McNemar test and bootstrapped McNemar test with 10 000 replicates were used to test discordance for each patient and all measurement time points, respectively. Next, we tested whether 2D measurements resulted in a significantly shorter progression-free survival (PFS) compared with 3D measurements using log-rank analysis applied to Kaplan-Meir data along with a 1-sided paired *t*-test to test, since patients were paired between the different measurements. Lastly, we quantified the growth rates and variability in measurements between 2D and 3D measurements, examining the percentage change with respect to baseline at each time point along with the residuals after application of a linear mixed-effects model.[28] For this model, 3D changes were normalized by a factor of $x^{2/3}$ for direct comparison with 2D changes (as outlined in[29]).

Table 2 Bidimensional (2D) and Volumetric (3D) Definitions of Radiographic Response and Progression

State of Disease	Change in Bidimensional Product (2D LGG RANO)	Estimated Volumetric Change (3D Spherical Equivalent)[26]
Complete response (CR)	100% Decrease	100% Decrease
Partial response (PR)	≥50% Decrease	≥65% Decrease
Minor response (MR)	Included as stable disease (SD) for the current study	
Stable disease (SD)	<50% Decrease to <25% Increase	<65% Decrease to <40% Increase
Progressive disease (PD)	≥25% Increase	≥40% Increase

Results

Inter-Observer Agreement for Response at Each Time Point

Bidirectional measurements (2D) were more impacted by subtle differences in measurements compared with volumetric (3D) measurements, which affected the overall response determination. Figure 1 shows an example of a patient measured by both 2D bidimensional and 3D volumetric measurements, where 2D measurements were indicative of a minor response (>33% reduction), but volumetric measurements of the entire lesion size showed a modest 4% reduction in volume. Consistent with these observations, 3D measurement trended toward higher inter-reader agreement compared with 2D measurements (κ_w =0.7057 vs κ_w = 0.5665; P = .0672), but this difference was not statistically significant.

Discordance in 2D vs 3D Response

After adjudication, differences remained between both the best response for each patient and the response at each follow-up time point due to the measurement technique used (2D vs 3D) (Figure 2). Stable disease (SD) was the best response in 64% of patients (16 of 21) and progressive disease (PD) was the best response for 14% of patients (3 of 21) for both 2D and 3D measurements (Figure 2A and B; McNemar test, P = .1573). 2D and 3D measurements were discordant in 2 patients, with 2D measurements showing PD as the best response while 3D measurements suggesting the best response was SD (Figure 2C).

When examining the response at each follow-up time point for each patient, a similar trend emerged (Figure 2D). A total of 56.8% of follow-up time points demonstrated SD (92 of 162) using 2D measurements, whereas 3D measurements suggested SD on 69% (112 of 162) of follow-up time points. Approximately 11% of time points (18 of 162) showed PD on both 2D and 3D measurements, while 18.5% of exams showed discordance between 2D and 3D measurements (30 of 162 exams). Specifically, 2D measurements called PD and 3D measurements called SD 15.4% of the time (25 of 162 exams), while 3D measurements only called PD when 2D measurements suggested SD in 3.1% of exams (5 of 162). The difference between 2D and

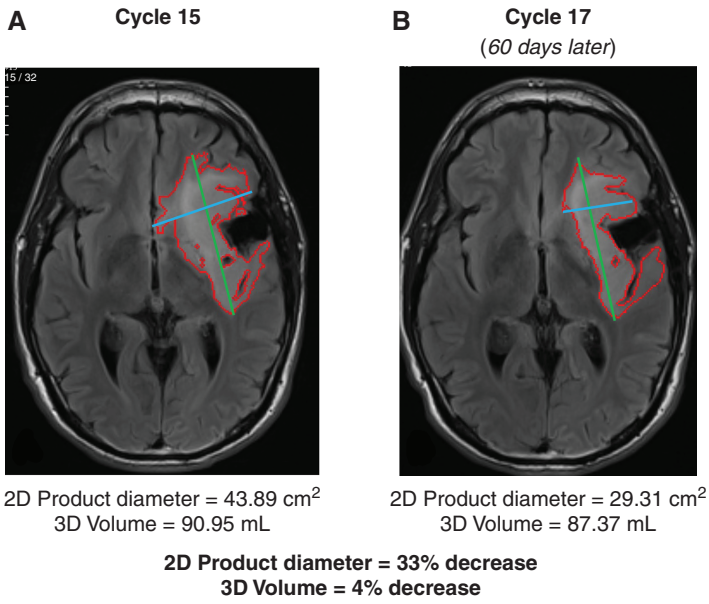

Fig. 1 Example of a IDH-mutant LGG patient measured by both 2D bidimensional (green and blue lines) and 3D volumetric measurements (red outline) at 2 different time points, a baseline time point (A) and the subsequent time point (B) 60 days (eg, cycle 3 day 1, first response assessment per protocol) later. While 2D measurements show a 33% reduction in bidimensional product, volumetric measurements only showed a 4% reduction in total volume. Abbreviation: LGG, low-grade gliomas.

3D response at each time point was significantly different (Bootstrapped McNemar test, $P = .0002$).

Comparison of Growth Rates and Measurement Variation

In order to compare the sensitivity of tumor growth rate measurements to the particular measurement technique used, the percentage change every 2 cycles (cycle = 28 days, response assessment performed every 2 cycles or 56 days) was quantified at each time point for 2D and 3D measurements after adjudication, using a logistic linear mixed-effects model[28] described previously. Examination of changes in tumor size using 2D measurements clearly showed more variability in measurements (Figure 3A), or "yo-yo-ing," compared with 3D measurements (Figure 3B). Significant differences in growth rates were observed between 2D and 3D reads, with 2D and 3D measurements showing a mean percentage change in tumor size every 58 days (2 cycles) with respect to nadir of 53.13% and 6.56%, respectively ($P = .0037$). When comparing 2D and 3D measurements and growth rate estimates after adjudication, 2D measurements had significantly higher measurement variability compared to 3D measurements (Figure 3C; $P = .0063$) as measured by the residual values after the linear mixed-effects model fit.

Discordance in PFS Between 2D and 3D Measurements

We theorized 2D measurements would result in a shorter PFS compared with 3D measurements, based at least in part on the observation that 2D measurements have higher measurement variation and therefore may result in earlier calls of PD. Consistent with this hypothesis, results from the current study suggest 2D measurements had a significantly shorter PFS compared with 3D measurements (Figure 3D; median PFS = 6.4 vs 12.9 months, hazard ratio [HR] = 2.322; log-rank, $P = .0181$). Since each of the PFS measurements were paired between the 2D and 3D measurements, we also used a paired *t*-test to look for differences in PFS. Results confirmed that 2D measurements resulted in a significantly shorter PFS compared to 3D measurements (paired *t*-test, $P = .0035$).

Discussion

Results from the current exploratory ad hoc analysis of data from a prospective phase 1 study (NCT02073994) with ivosidenib suggest 3D volumetric measurements are preferred as the best approach for radiographic response assessment in LGGs. This is based on the data suggesting 3D measurements had the highest inter-reader agreement, were significantly more stable and slower tumor growth rate, demonstrated a statistically significant discordance in best response and time-point responses, and exhibited significantly different estimates of PFS compared with RANO recommended 2D approaches.

These results have implications in recommendations used for prospective LGG trials, but require further study to validate our findings. High-quality data collected from prospective phase 3 randomized and controlled studies are needed to establish the necessary modifications to the

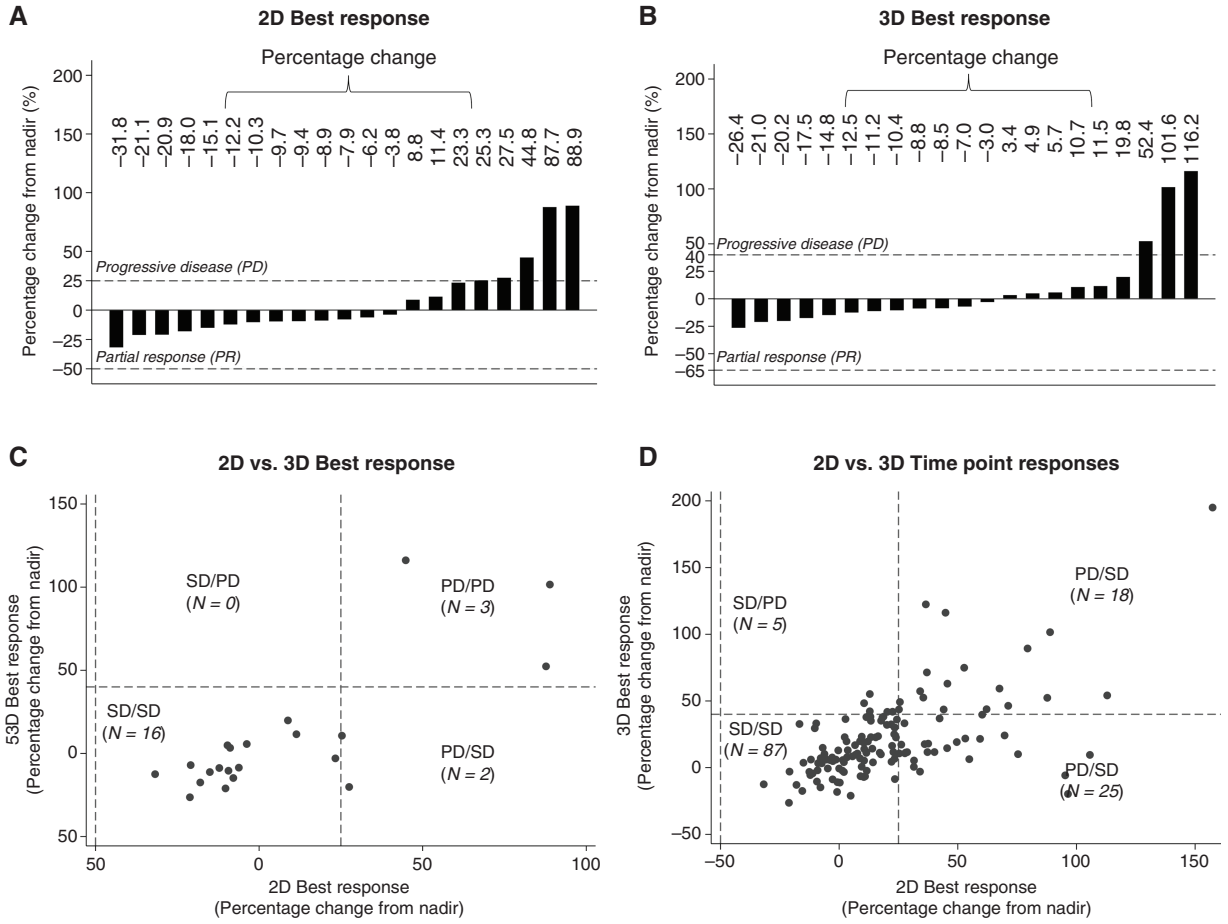

Fig. 2 Waterfall plots showing (A) "best response" using 2D bidirectional measurements and (B) 3D volumetric measurements after implementing a retrospective, consensus review where all readers agree on the measurements. (C) Concordance in "best response" for individual patients and (D) concordance in all time-point responses using 2D and 3D measurements after retrospective, consensus review by all radiologists.

existing LGG RANO criteria to enable the use of 3D measurements as a more reliable alternative to the conventional 2D measurements. While the combination of bidirectional measurements of T2 hyperintense diffuse gliomas is the currently recommended approach for evaluation of IDH-mutant glioma response assessment according to LGG RANO,[13] practical implementations in clinical trials have not been addressed. Volumetric measurements are better at capturing the gestalt tumor behavior compared with choosing a single set of axial planar measurements, whether the same slice over time or the slice with the largest cross-sectional tumor size. Results from the current study confirm that the 2D measurements may result in higher discordance in evaluation of response for diffuse gliomas with irregular margins and infiltrative tumor growth. Also, the slow-growing (and potentially slow responding) IDH-mutant diffuse gliomas require a more holistic approach when evaluating long-term tumor growth behavior. Thus, measurement techniques that are susceptible to significantly variable measurements (ie, 2D measurements) may further decrease confidence in the ability for radiographic response to reflect clinical benefit.

Study Limitations

It is important to point out a few limitations of the current study that may impact the interpretation and recommendations. First, the current study involves a relatively small patient cohort (N = 21), so some of the trends may not be generalizable for larger trials or to patients with more advanced disease or tumor grade. Another potential limitation to the current study is the restricted number of radiologists and paradigms tested in the current study. Although there are many possible radiographic read paradigms,[25] testing all these possible approaches was not possible. Instead, we chose a pragmatic strategy that included a traditional approach (ie, 2D bidimensional and 3D volumetric measurements and the blinded locked sequential read paradigm) and did not describe results from recommended hybrid approaches[25] that allow for radiologists to make changes to their measurements after retrospective review of all time points to get a sense of overall tumor growth behavior. Future studies are needed to determine whether hybrid approaches provide additional value beyond these more traditional read paradigms. While the

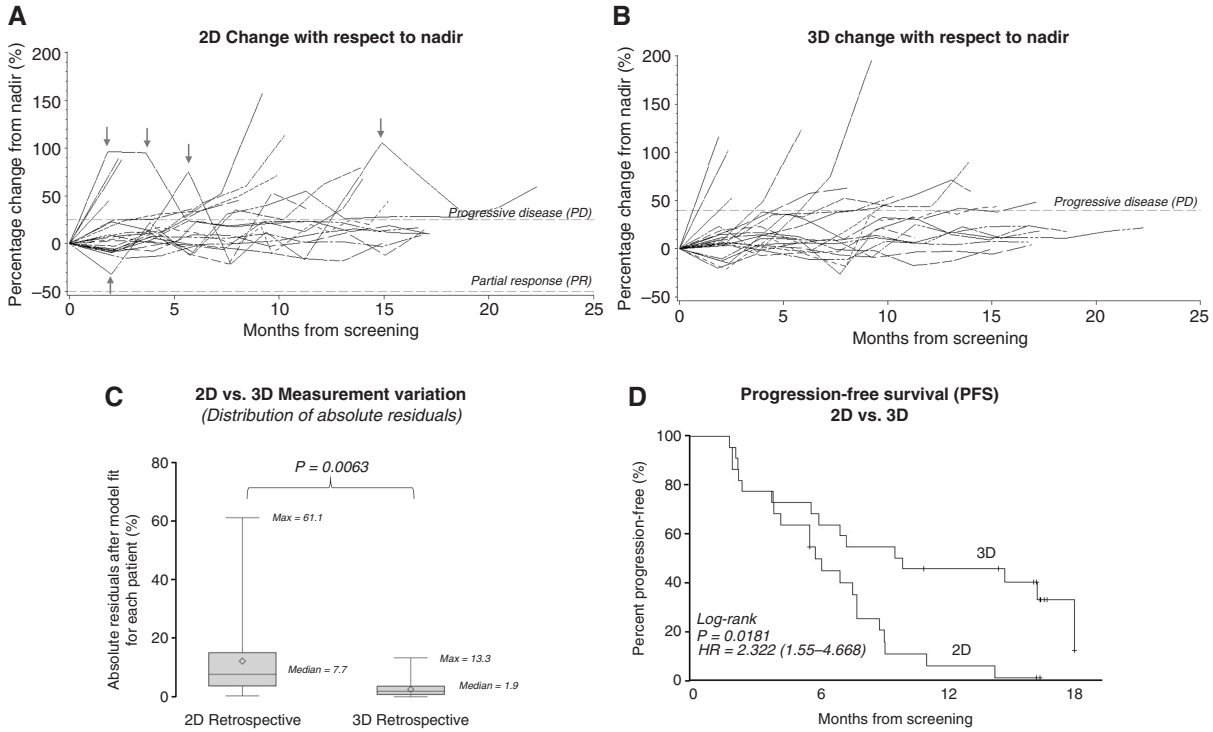

Fig. 3 (A) Spider plot showing percentage change in tumor size using 2D bidirectional measurements and (B) 3D volumetric measurements with respect to baseline or nadir for all patients. Arrows show examples of potential "yo-yo-ing" from erroneous 2D measurements. (C) Measurement variation estimates using the distribution of residuals (errors) after model fitting to 2D and 3D growth rate estimates. (D) Kaplan-Meier curves comparing progression-free survival (PFS) between 2D and 3D measurements.

current study used the same software and the same radiologists to review each time point, consideration should be given into the potential heterogeneity or differences in segmentation software/algorithms used, tissue contrast differences from slightly altered acquisition parameters, and/or software operator experience in future studies to optimize the reproducibility of tumor measurements over time. Lastly, it remains unclear whether 2D or 3D measurements and/or response determination have any differential impact on overall survival. Given the long survival in IDH-mutant LGGs, it was not feasible to correlate measures of PFS or response with overall survival.

It is important to point out the potential implications of using 3D measurements for clinical trial response assessment for LGGs, and how this might impact trial design and conduct. For example, a prospective trial might consider a central lab providing 3D measurements and consultation to local sites in "real-time" so they can make clinical decisions. Since LGGs are relatively slow-growing and typically have follow-up MRI examinations every 3-6 months, real-time feedback to sites from the central imaging core lab is reasonable and is currently being done in some trials. If this is not feasible, local sites have many potential options available to them. A sponsor may choose to provide a software package to the local site for patient management. Additionally, most neurosurgery departments routinely utilize presurgical planning software that can easily create 3D contours of tumors. Many radiology departments have "3D Labs" with expert technologists that routinely segment vascular structures for cardiovascular or neurovascular procedures and tumors for solid cancer cases, so resources may already exist to aid in tumor segmentation. Yet another approach might be to allow local sites to use 2D measurements as a preliminary or surrogate measurement of the centrally determined 3D measurements, since there is reasonable agreement between 2D and 3D response categories for most time points.

Conclusion

Diffuse gliomas are infiltrative by nature and have highly irregular growth patterns, so measurements of tumor burden are inherently difficult to capture using a single image slice. Consistent with this concept, the current study suggests 3D volumetric measurements of non-enhancing tumors are more reliable for radiographic response assessment in LGGs compared to 2D bidirectional measurements due to higher inter-reader agreement, lower discordance rates, and more stable and conservative measurements of tumor growth rates. Additionally, 3D volumetric measurements demonstrated a longer PFS due to less "yo-yo-ing" of measurements over time causing fewer erroneous calls of progression. The differences between 2D and 3D evaluations in terms of both best response and estimates of PFS

should be investigated in larger prospective controlled studies to determine their clinical significance.

Supplementary Material

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

Keywords

IDH-mutant gliomas | ivosidenib | LGG RANO | low-grade gliomas

Funding

This study was supported by the American Cancer Society (ACS) Research Scholar Grant (RSG-15-003-01-CCE) (Ellingson) and UCLA SPORE in Brain Cancer (NIH/NCI 1P50CA211015-01A1) (Ellingson, Cloughesy).

Conflict of interest statement. B.M.E. is a Paid Consultant and Advisor for Siemens; Medicenna; MedQIA; Imaging Endpoints; Agios Pharmaceuticals; Neosoma; Janssen; Kazia; VBL; Oncoceutics; Boston Biomedical Inc.; ImmunoGenesis; and Ellipses Pharma. B.M.E. has received grant funding from Siemens, Agios, and Janssen. T.F.C. is on the advisory board for Roche/Genentech, Amgen, Tocagen, NewGen, LPath, Proximagen, Celgene, Vascular Biogenics Ltd, Insys, Agios, Cortice Biosciences, Pfizer, Human Longevity, BMS, Merck, Notable Lab, and MedQIA. I.K.M. reports serving as a consultant for Agios Pharmaceuticals, Inc., Black Diamond, DC Europa, Debiopharm, Puma Biotechnology, and Voyager and research grants from Amgen, Eli Lilly, and General Electric. L.S., I.H., and S.S.P. are paid employees of Servier Pharmaceuticals, LLC.

Authorship statement. The following authors contributed to experimental design: B.M.E., G.H.J.K., M.B., J.G., and T.F.C. The following authors contributed to study implementation: N.S., L.S., S.S.P., P.Y.W., I.K.M., and T.F.C. The following authors contributed to analysis and interpretation; B.M.E., G.H.J.K., M.B., J.L., S.C., M.L., B.Y., J.G., and T.F.C. All authors were involved in manuscript writing, review, and approval.

References

1. Parsons DW, Jones S, Zhang X, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science*. 2008;321(5897):1807–1812.
2. Balss J, Meyer J, Mueller W, Korshunov A, Hartmann C, von Deimling A. Analysis of the IDH1 codon 132 mutation in brain tumors. *Acta Neuropathol*. 2008;116(6):597–602.
3. Bleeker FE, Lamba S, Leenstra S, et al. *IDH1* mutations at residue p.R132 (*IDH1*^{R132}) occur frequently in high-grade gliomas but not in other solid tumors. *Hum Mutat*. 2009;30(1):7–11.
4. Yan H, Parsons DW, Jin G, et al. IDH1 and IDH2 mutations in gliomas. *N Engl J Med*. 2009;360(8):765–773.
5. Whittle IR. The dilemma of low grade glioma. *J Neurol Neurosurg Psychiatry*. 2004;75(Suppl 2):ii31–ii36.
6. Claus EB, Walsh KM, Wiencke JK, et al. Survival and low-grade glioma: the emergence of genetic information. *Neurosurg Focus*. 2015;38(1):E6.
7. Shaw EG, Wang M, Coons SW, et al. Randomized trial of radiation therapy plus procarbazine, lomustine, and vincristine chemotherapy for supratentorial adult low-grade glioma: initial results of RTOG 9802. *J Clin Oncol*. 2012;30(25):3065–3070.
8. Buckner JC, Chakravarti A, Curran WJ Jr. Radiation plus chemotherapy in low-grade glioma. *N Engl J Med*. 2016;375(5):490–491.
9. Buckner JC, Shaw EG, Pugh SL, et al. Radiation plus procarbazine, CCNU, and vincristine in low-grade glioma. *N Engl J Med*. 2016;374(14):1344–1355.
10. Wessels PH, Weber WE, Raven G, Ramaekers FC, Hopman AH, Twijnstra A. Supratentorial grade II astrocytoma: biological features and clinical course. *Lancet Neurol*. 2003;2(7):395–403.
11. Feigenberg SJ, Amdur RJ, Morris CG, Mendenhall WM, Marcus RB Jr, Friedman WA. Oligodendroglioma: does deferring treatment compromise outcome? *Am J Clin Oncol*. 2003;26(3):e60–e66.
12. Keles GE, Lamborn KR, Berger MS. Low-grade hemispheric gliomas in adults: a critical review of extent of resection as a factor influencing outcome. *J Neurosurg*. 2001;95(5):735–745.
13. van den Bent MJ, Wefel JS, Schiff D, et al. Response assessment in neuro-oncology (a report of the RANO group): assessment of outcome in trials of diffuse low-grade gliomas. *Lancet Oncol*. 2011;12(6):583–593.
14. Vos MJ, Uitdehaag BM, Barkhof F, et al. Interobserver variability in the radiological assessment of response to chemotherapy in glioma. *Neurology*. 2003;60(5):826–830.
15. Shah GD, Kesari S, Xu R, et al. Comparison of linear and volumetric criteria in assessing tumor response in adult high-grade gliomas. *Neuro Oncol*. 2006;8(1):38–46.
16. 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.
17. Provenzale JM, Ison C, Delong D. Bidimensional measurements in brain tumors: assessment of interobserver variability. *AJR Am J Roentgenol*. 2009;193(6):W515–W522.
18. Ellingson BM, Bendszus M, Sorensen AG, Pope WB. Emerging techniques and technologies in brain tumor imaging. *Neuro Oncol*. 2014;16(Suppl 7):vii12–vii23.
19. Popovici-Muller J, Saunders JO, Salituro FG, et al. Discovery of the first potent inhibitors of mutant IDH1 that lower tumor 2-HG in vivo. *ACS Med Chem Lett*. 2012;3(10):850–855.
20. Mellinghoff IK, Touat M, Maher E, et al. ACTR-46. AG120, a first-in-class mutant IDH1 inhibitor in patients with recurrent or progressive IDH1 mutant glioma: results from the phase 1 glioma expansion cohorts. *Neuro Oncol*. 2016;18(suppl_6):vi12.
21. Mellinghoff IK, Ellingson BM, Touat M, et al. Ivosidenib in isocitrate dehydrogenase 1-mutated advanced glioma. *J Clin Oncol*. 2020;38(29):3398–3406.
22. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol*. 2016;131(6):803–820.
23. Louis DN, Ohgaki H, Wiestler OD, et al. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol*. 2007;114(2):97–109.
24. 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.

25. Ellingson BM, Brown MS, Boxerman JL, et al. Radiographic read paradigms and the roles of the central imaging laboratory in neuro-oncology clinical trials. *Neuro Oncol*. 2021;23(2):189–198.
26. Chappell R, Miranpuri SS, Mehta MP. Dimension in defining tumor response. *J Clin Oncol*. 1998;16(3):1234.
27. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276–282.
28. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2011.
29. Petrick N, Kim HJ, Clunie D, et al. Comparison of 1D, 2D, and 3D nodule sizing methods by radiologists for spherical and complex nodules on thoracic CT phantom images. *Acad Radiol*. 2014;21(1):30–40.