A Novel, Reproducible, and Objective Method for Volumetric Magnetic Resonance Imaging Assessment of Enhancing Glioblastoma

Charles W. Kanaly, M.D.[1], **Ankit I. Mehta, M.D.**[1], **Dale Ding, M.D.**[2], **Jenny Hoang, M.D.**[3], **Peter G. Kranz, M.D.**[3], **James E. Herndon II, Ph.D.**[4], **April Coan, M.P.H.**[4], **Ian Crocker, M.D.**[5], **Anthony F. Waller**[5], **Allan H. Friedman, M.D.**[1], **David A. Reardon, M.D.**[6], and **John H. Sampson, M.D., Ph.D.**[1,7]

[1]Division of Neurosurgery, Department of Surgery, Duke University Medical Center, Durham, NC, USA

[2]Department of Neurosurgery, University of Virginia, Charlottesville, VA, USA

[3]Department of Radiology, Duke University Medical Center, Durham, NC, USA

[4]Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC, USA

[5]Department of Radiation Oncology, Emory University, Atlanta, GA, USA

[6]Dana-Farber Cancer Institute, Boston, MA, USA

[7]Department of Pathology, Duke University Medical Center, Durham, NC, USA

Abstract

Objective—Robust methodology that allows objective, automated, and observer-independent measurements of brain tumor volume, especially postresection, is lacking; hence, determination of tumor response and progression in neuro-oncology is unreliable. Our objective was to determine if a semi-automated, volumetric method for quantifying enhancing tissue would perform with high reproducibility and low interobserver variability.

Methods—Fifty-seven scans from 13 patients with glioblastoma (GBM) were assessed, using our method, by two neuroradiologists, one neurosurgeon, one neurosurgery resident, one nurse practitioner, and one medical student. The two neuroradiologists also performed traditional one-dimensional and two-dimensional measurements. Intraclass correlation coefficients (ICC) assessed interobserver variability between measurements. Radiographic response was determined using *Response Evaluation Criteria In Solid Tumors* (RECIST) guidelines and Macdonald criteria.

Corresponding Author: John H. Sampson, MD, PhD, MHSc, Division of Neurosurgery, Post Office Box 3050, Room 220, Sands Building, Research Drive, Duke University Medical Center, Durham, NC 27710; phone: (919) 684-9041; fax: (919) 684-9045; john.sampson@duke.edu.

Kappa statistics described interobserver variability of volumetric radiographic response determinations.

Results—There was strong agreement for one-dimensional (RECIST) and two-dimensional (Macdonald) measurements between neuroradiologists (ICC=0.42 and 0.61, respectively), but the agreement using our novel automated approach was significantly stronger (ICC=0.97). Our volumetric method had the strongest agreement with regard to radiographic response (kappa=0.96) when compared with two-dimensional (0.54) or one-dimensional (0.46) methods. Despite diverse levels of experience, measurements using our volumetric program by all users remained remarkably high (0.94).

Conclusion—Interobserver variability with our semiautomated method is less than the variability with traditional methods of tumor measurement. It is objective, quick, and highly reproducible among operators with varying expertise. This approach should be further evaluated as a potential standard for response assessment based on contrast enhancement in brain tumors.

Keywords

Brain Neoplasms; Disease Progression; Gliomas; Magnetic Resonance Imaging; Recurrence

Introduction

Progression-free survival time based on contrast enhancement has become an accepted end point in glioma trials.[6,13,14] However, current methods to assess radiographic response in brain tumors are highly subjective and have high interobserver variability. To both accurately detect small changes in complex tumor configurations, especially after resection, and to strengthen the results from clinical trials that assess progression-free survival, there is a dire need for reproducible determination of radiographic change in gliomas.[9,18]

Malignant gliomas are often very challenging to measure since they are commonly irregularly shaped, have cystic or hemorrhagic regions, have satellite lesions, are too small to be classified as measurable, or demonstrate a thin rim of enhancement around resection cavities. The Macdonald criteria are commonly used and apply two-dimensional (2D) measurements.[11] Many authors, however, have noted that the Macdonald criteria are still not comprehensive enough to describe subtle changes in complex tumor volumes, particularly when present in surgically created resection cavities, and a number of studies have questioned the validity of the criteria because they are subject to considerable interobserver variability. [18,20,21]

Volumetric assessment of tumors intuitively should be superior to one-dimensional (1D) or 2D measurements in determining changes in tumor size, as it has the ability to overcome difficulties with cystic spaces, irregular shapes, small amounts of nodular rim enhancement, satellite lesions, and small volumes considered to be unmeasurable by conventional methods. [18] Precise and reproducible volumetric assessment is expected to be able to detect smaller changes in tumors and, therefore, could detect tumor response or progression sooner than 1D or 2D techniques. Prior studies have shown up to 26.1% discrepancy between response assessments made based on 2D or volumetric techniques[19]. By taking the entire

tumor volume into account, volumetric measurements should be more accurate and augment the power of clinical trials by reducing the variance in the measurements. [18] Furthermore, results from previous studies demonstrate radiographic assessment by volumetric techniques exhibit decreased measurement variability when compared with 1D and 2D methods.[5,12,19] Unfortunately, most previously published methods of volumetric assessment have required a considerable level of operator skill and experience because the operator must manually outline the specific region of tumor to calculate a tumor volume. As a result, volumetric assessments are prone to a significant amount of subjectivity and produce a high degree of interobserver variability. Additionally, these methods are hampered by the time required to perform the analysis and the difficulties specific to the central nervous system (CNS) location, including frequently necrotic or cystic tumors, blood or other obscuring lesions intrinsically bright on T1-weighted imaging, trouble determining where tumor margins end, and differences between scans in the slice acquisition and the timing of contrast boluses. Since the relationship between magnetic resonance imaging (MRI) contrast enhancement and gadolinium (Gd) concentration is not linear, [22] the degree of brightness on standard MRI scans cannot be easily correlated with quantitative numbers.

In this study, we evaluate the performance of a new method of assessing CNS tumor volume that is highly reproducible and has low interobserver variability even when used by examiners with vastly different levels of experience. The determination of radiographic progression also had high reproducibility and low interobserver variability.

Methods

Patients

A total of 57 scans were analyzed from 13 patients with recurrent grade IV GBM, as measured by the World Health Organization, on four different experimental chemotherapy protocols. The 13 patients with recurrent glioblastoma used in our study all received standard of care of therapy, including surgery (n=11) or biopsy (n=2), external beam radiation therapy, and temozolomide chemotherapy prior to recurrence. In addition, in all cases recurrence was documented by biopsy (n=4) or unequivocal radiographic progression according to standard protocols (n=9). They were then followed with serial scans, and placed on these chemotherapy protocols after radiographic, clinical or tissue diagnosis of progression had been made by the treating neurooncologist. These experimental protocols included topotecan with temozolomide, erlotinib with rapamycin, erlotinib with dasatinib, and an experimental chemotherapeutic agent BIBW 2992. A total of 13 baseline scans and 44 follow-up scans were performed on these patients between August 2004 and December 2008. Baseline scans were determined to be the first scan after resection. The median number of follow-up scans was three. Patient ages ranged from 27 to 71 years, with a median age of 48 years. The majority of the patients (85%) had a Karnofsky performance status 80 and some degree of surgical resection (93%). A total of 38% of the patients were alive at the time of analysis, with a median follow-up from study enrollment of 14.6 months. All 57 scans were assessed with our novel volumetric method by two neuroradiologists, one neurosurgeon, one neurosurgery resident, one nurse practitioner, and one medical student; the two neuroradiologists assessed the scans with the 1D and 2D criteria. The study was

conducted in accordance with the Declaration of Helsinki and the guidelines on Good Clinical Practice.

Volumetric Method

The volumetric method used in this study to calculate changes in enhancing tumor volume was a semiautomated, atlas-based segmentation program called VelocityAI[TM], (Velocity Medical Solutions, Atlanta, GA). The detailed methodology behind this new software has been previously reported[10] (Fig. 1). In brief, each user was provided with a simple typed outline of the steps required to use the program and given a brief sample demonstration of the program that lasted less than 5 minutes. Afterwards, they were left to perform the volumetric analysis independently. Briefly, each user loaded the Digital Imaging and Communications in Medicine files of both precontrast and postcontrast MRI T1-weighted axial sequences of the brain into the computer program. The program automatically adjusted for motion and aligned the images between the two sequences. After alignment, the program automatically subtracted the precontrast images from the postcontrast images, so that only contrast-enhancing objects (and not intrinsically bright T1 objects, such as hemorrhage in a resection cavity) were used in later analyses. An atlas was loaded that automatically located and conformed to the nasal mucosa to normalize for Gd signal between scans. Next, the user manually used a drawing tool to grossly outline a region of interest of any size that contained the tumor and a separate small region of normal brain parenchyma on the same axial slices as the previously outlined tumor. This rough outline of the area of interest consistently took only a few minutes per scan by every user to complete, regardless of the user's prior level of expertise. The computer program then automatically compared the normal region of the precontrast scans with that of the postcontrast scans to determine a correction factor that was applied to all subsequent analyses to normalize between the two scans. Subsequently, the normalized, subtracted values in the enhancing nasal mucosa structure were used by the program to determine a quantitative cutoff for enhancement by excluding the top 5% of enhancement values and then calculating the 25% value of this remaining maximum value. Finally, the volume of pixels in the outlined tumor structure that met or exceeded this normalized enhancement threshold was automatically calculated by the program to be the enhancing tumor volume.

The two neuroradiologists also performed traditional 1D (RECIST) [9] and 2D (Macdonald) [18] measurements (Fig. 2). The radiographic response assessment criteria used percent changes from baseline in 1D and 2D measurements. The RECIST guidelines state that a decrease of at least 30% from baseline value is considered regression and an increase of 20% of the baseline is progression. Macdonald criterion for regression was defined as a decrease of at least 50% from baseline values, and progression was defined as an increase of at least 25%. Our volumetric method defined regression as a decrease of at least 65% of the volume and progression as an increase in volume of at least 40% based on previously published studies. [7]

Statistics

To determine the interobserver variability of radiographic scan measurements (actual, percent change from baseline, and percent change from smallest) for each measurement

method (1D, 2D, and novel volumetric), intraclass correlation coefficients (ICC) were estimated from variance components of a hierarchical linear model analysis.

The interobserver variability of radiographic calls (complete response/ regression, stable disease, and progression) was examined by computing Cohen's kappa coefficient for the 1D and 2D methods (two examiners) and Fleiss' kappa coefficient for the novel volumetric method (more than two examiners). Analyses were conducted under the assumption that each individual patient scan was independent.

Statistical analyses were conducted using SAS® 9.2 (SAS Institute Inc., Cary, NC) and an SAS macro MKAPPA developed by Westat Inc.

Results

Interobserver Variability Between Experts Is Less with Volumetric Software Compared with 1D and 2D Methods

Two neuroradiologists, who were blinded to the study objectives, used our volumetric software or standard 1D (RECIST) and 2D (Macdonald) methods to calculate the tumor size for each patient. The traditional RECIST guidelines and Macdonald criteria determine progression based on the percent change from baseline tumor size. Therefore, comparisons were made between the methods for percent change from baseline tumor size. Calculated ICC is shown as an assessment of the interobserver variability of radiographic measurements for each measurement method (Table 1). For the three methods used for assessing tumor size—1D, 2D, and our novel method—the actual volume measurement showed the percent change from baseline to be 0.42, 0.61, and 0.97, respectively. Our novel method demonstrated the highest ICC amongst the two neuroradiologists in comparison to the other 1D and 2D methods.

ICC High with Examiners of Varying Experience

We had six examiners with various experience levels use our volumetric software to assess the change in tumor volume using the volumetric, 1D, and 2D criteria. In Table 2, the actual volume measurements between the six examiners had a high ICC at 0.94. The percent change from baseline of tumor volume was also consistent with a high ICC at 0.93. Even in the hands of examiners without clinical or radiographic experience, there was a strong ICC using our novel volumetric software method.

Improved Concordance on Radiographic Response Assessment Between Neuroradiologists Using Volumetric Method

Radiographic response assessment was determined by neuroradiologists to be either complete response, stable disease, or progression based on the 1D method, 2D method, or our volumetric method. Cohen's and Fleiss' kappa coefficients are presented in Table 3 as an assessment of the interobserver variability of determination of radiographic progression or response for each of the measure methods. The novel volumetric method had the strongest agreement among the two neuroradiologists (Cohen's kappa=0.96) and also had strong agreement (Fleiss' kappa=0.79) among all examiners. Cohen's kappa statistic for

agreement among the two neuroradiologists findings for 2D and 1D were 0.54 and 0.46, respectively (see Table 3).

Progression Called Earlier Between Neuroradiologists Using Volumetric Method

Of the 13 study patients, eight were determined to have progressed during the study period per an assessment by the treating neurooncologist. A comparison was performed between the times at which the initial call of progression was made by each neuroradiologist to the times determined to be progression by the treating neurooncologist, using the three methods. If simultaneous calls of progression occurred with different methods, it was counted as the initial call of progression for all methods that made that simultaneous call. Comparing across the three methods, the neurooncologist detected progression first in one of eight patients, while one neuroradiologist using the novel volumetric method detected progression first in six of the eight patients. The same neuroradiologist detected progression first in three of eight patients using the 1D method and detected progression first in four of eight patients with the 2D method. A similar comparison with the calls of progression made by the second neuroradiologist to the neurooncologist's calls yielded similar results: two of eight patients for the neurooncologist first, five of eight patients using the novel volumetric method, two of eight patients using the 1D method, and four of eight patients using the 2D method.

Discussion

Our semiautomated method appears to represent a significant improvement over previous measurement techniques for numerous reasons. Since it measures tumor volume, it incorporates more detail than 1D and 2D methods, whereby it is also able to calculate changes in tumors even in complicated, postresection shapes or in the presence of residual blood or other intrinsically bright T1 entities. It is able to automatically correct for motion between the two sequences and automatically subtracts the precontrast images from the postcontrast images, which eliminates intrinsically bright T1 objects from the calculation of enhancement. [10] This subtraction enables our method to detect enhancing tumor in the midst of any intrinsically bright T1 signals. This is a particularly important advantage for postoperative scans in which residual tumor may be hidden within blood in the resection cavity. Other studies have shown that similar subtraction techniques resulted in improved contrast detection. [3] Additionally, since the relationship between MRI contrast enhancement and Gd concentration is not linear, [22] the degree of brightness on MRI scans cannot be easily correlated with quantitative numbers. Our program overcomes this difficulty by normalizing the scans with comparisons of normal brain parenchyma between precontrast and postcontrast sequences, and it determines an individual enhancement threshold for each scan based on the automated nasal mucosa atlas. The top 5% of enhancement values in the nasal mucosa were excluded to avoid error because of the nonlinear relationship between enhancement and Gd concentration. The enhancement threshold was selected to be 25% based on expert radiologist opinion as to which threshold value most closely corresponded with their subjective judgement of tissue enhancement. Our prior publication on this method presents data showing that as long as the threshold value is consistently maintained, the results are highly correlated[10]. This suggests that response assessment conclusions are largely insensitive to specific threshold level. This threshold method compensates for

differences between scans in contrast boluses and timing and is similar to one that is commonly employed and has been validated with positron emission tomography scans. [2,4]

One possible concern about our method is that although the volumes generated by our method are highly precise, this does not necessarily mean that the measurements are accurate. This is unfortunately a limitation of all radiographic measurements, as there is no definite gold standard volumetric measurement technique. The only way to absolutely know the volume of tumor is to resect it entirely and physically measure the volume. Comparisons between our method and any other volumetric method would simply be measuring agreement between two different estimations, neither of which has been proven to be accurate. However, it does appear that increasing automation of radiographic measurements improves accuracy and reproducibility when compared with manual techniques. [16] Computer-assisted methods of volume calculation have been shown to have less variability than manual calculations[19]. Furthermore, our method does not require the time or expertise that manual volumetric methods require. The increased automation of our approach significantly improves the reproducibility of tumor volumes and determination of progression obtained across different users. Our results suggest that reliable tumor volumes can be determined even by users with minimal expertise in the field of neuroradiology. Therefore, our method does not rely on the subjective interpretation of scans that is inherent in many other volumetric methods. The program calculates the amount of enhancing tissue within the grossly outlined tumor region. The processing steps are highly automated, which minimizes both the time required for analysis and the expertise needed by the user. When comparing neuroradiologic response assessments in gliomas, Shah *et al.* demonstrated an ICC for 1D, 2D, and 3D criteria as 0.874, 0.822, and 0.889, respectively. [17] Our volumetric method demonstrates a superior interobserver ICC at 0.97, while 1D and 2D are comparable at 0.42 and 0.61, respectively. Vos *et al.* showed a 2D ICC of 0.64 (0.61 in our study) and a kappa of response classification of 0.51 (0.54 in our study). [21]

Although our method represents a considerable advance in radiographic assessment for high-grade gliomas, there are a number of remaining difficulties that it does not yet solve. Most significantly, this method was developed to measure the volume of enhancing tumor and, therefore, will not detect nonenhancing tumor. This drawback will be of increasing importance as more patients begin antiangiogenic therapy with agents like bevacizumab that normalize vasculature and dramatically decrease enhancement. [8,15] For patients in this situation, the measurement of enhancing tumor is no longer an adequate assessment of tumor burden, although even in these studies tumor response burden measurements of enhancing tumor predicts survival. [1] The measurement of nonenhancing tumor is also extremely important for the assessment of low grade tumors. The Response Assessment in Neuro-Oncology Working Group (RANO) developed new standardized response criteria for high grade gliomas in clinical trials that accounts for nonenhancing tumor. [23] The RANO criteria use T2/FLAIR imaging to incorporate an increase in nonenhancing tumor burden to determine progression. Adaptation of our method for use with T2/FLAIR imaging sequences is currently being attempted to assess nonenhancing tumor volumes. Although technology will need to be developed to accurately measure nonenhancing tumor volume, no other currently accepted method is able to accurately address this issue. This method is able to

quantify enhancing volume, however this calculated volume is not necessarily tumor and the measured enhancement could be secondary to radiation necrosis or pseudoprogression. This is an inherent limitation of any method that measures enhancement, as a measurement change does not necessarily imply a specific diagnosis of tumor recurrence. Clinical judgment or even possibly tissue biopsy may be required to determine the true meaning of these changes in enhancement. It is possible that knowledge of the time course of these enhancement changes or serial imaging to determine trends in these changes may be able to distinguish between treatment effect and recurrence. Future investigations are planned to determine if recurrence can be distinguished from other enhancement changes based on the serial measurements of enhancement over time.

Conclusion

We show that our semi-automated method of quantifying enhancing intracranial tumor volume has very low variability among users with widely divergent levels of expertise. Even when comparing the measurements made by two neuroradiologists, this method had higher agreement than traditional one dimension and two dimensional methods. This technique has the potential to gain wide acceptance in clinical practice since users with all levels of expertise appear to obtain highly similar results, it only requires a few minutes to perform the analysis, and there is no need for special MRI sequences or computers. Such an approach may also then be applicable to the analysis of benign tumors, such as meningioma, and might also be useful to follow radiosurgery responses over time as well. Future studies with larger numbers of patients will need to be performed to validate this technique.

Acknowledgments

References

1. Ananthnarayan S, Bahng J, Roring J, Nghiemphu P, Lai A, Cloughesy T, et al. Time course of imaging changes of GBM during extended bevacizumab treatment. J Neurooncol. 2008; 88:339–347. [PubMed: 18389177]

2. Burri RJ, Rangaswamy B, Kostakoglu L, Hoch B, Genden EM, Som PM, et al. Correlation of positron emission tomography standard uptake value and pathologic specimen size in cancer of the head and neck. Int J Radiat Oncol Biol Phys. 2008; 71:682–688. [PubMed: 18258379]

3. Curati WL, Williams EJ, Oatridge A, Hajnal JV, Saeed N, Bydder GM. Use of subvoxel registration and subtraction to improve demonstration of contrast enhancement in MRI of the brain. Neuroradiology. 1996; 38:717–723. [PubMed: 8957793]

4. Erdi YE, Mawlawi O, Larson SM, Imbriaco M, Yeung H, Finn R, et al. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. Cancer. 1997; 80:2505–2509. [PubMed: 9406703]

5. Fraioli F, Bertoletti L, Napoli A, Calabrese FA, Masciangelo R, Cortesi E, et al. Volumetric evaluation of therapy response in patients with lung metastases. Preliminary results with a computer system (CAD) and comparison with unidimensional measurements. Radiol Med. 2006; 111:365–375. [PubMed: 16683084]

6. Friedman HS, Prados MD, Wen PY, Mikkelsen T, Schiff D, Abrey LE, et al. Bevacizumab alone and in combination with irinotecan in recurrent glioblastoma. J Clin Oncol. 2009; 27:4733–4740. [PubMed: 19720927]

7. Galanis E, Buckner JC, Maurer MJ, Sykora R, Castillo R, Ballman KV, et al. Validation of neuroradiologic response assessment in gliomas: measurement by RECIST, two-dimensional, computer-assisted tumor area, and computer-assisted tumor volume methods. Neuro Oncol. 2006; 8:156–165. [PubMed: 16533757]

8. Gonzalez J, Kumar AJ, Conrad CA, Levin VA. Effect of bevacizumab on radiation necrosis of the brain. Int J Radiat Oncol Biol Phys. 2007; 67:323–326. [PubMed: 17236958]

9. Jaffe CC. Measures of response: RECIST, WHO, and new alternatives. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2006; 24:3245–3251. [PubMed: 16829648]

10. 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. 2011; 6:e16031. [PubMed: 21298088]

11. Macdonald DR, Cascino TL, Schold SC Jr. Cairncross JG. Response criteria for phase II studies of supratentorial malignant glioma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1990; 8:1277–1280. [PubMed: 2358840]

12. Marten K, Auer F, Schmidt S, Kohl G, Rummeny EJ, Engelke C. Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria. Eur Radiol. 2006; 16:781–790. [PubMed: 16331462]

13. Oh JK, Sahu D, Ahn YH, Lee SJ, Tsutsumi S, Hwang JH, et al. Effect of fracture gap on stability of compression plate fixation: a finite element study. J Orthop Res. 2010; 28:462–467. [PubMed: 19862799]

14. Polley MY, Lamborn KR, Chang SM, Butowski N, Clarke JL, Prados M. Six-month progression-free survival as an alternative primary efficacy endpoint to overall survival in newly diagnosed glioblastoma patients receiving temozolomide. Neuro Oncol. 2010; 12:274–282. [PubMed: 20167815]

15. Pope WB, Lai A, Nghiemphu P, Mischel P, Cloughesy TF. MRI in patients with high-grade gliomas treated with bevacizumab and chemotherapy. Neurology. 2006; 66:1258–1260. [PubMed: 16636248]

16. Schwartz LH, Ginsberg MS, DeCorato D, Rothenberg LN, Einstein S, Kijewski P, et al. Evaluation of tumor measurements in oncology: use of film-based and electronic techniques. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2000; 18:2179–2184. [PubMed: 10811683]

17. Shah GD, Kesari S, Xu R, Batchelor TT, O'Neill AM, Hochberg FH, et al. Comparison of linear and volumetric criteria in assessing tumor response in adult high-grade gliomas. Neuro Oncol. 2006; 8:38–46. [PubMed: 16443946]

18. Sorensen AG, Batchelor TT, Wen PY, Zhang WT, Jain RK. Response criteria for glioma. Nat Clin Pract Oncol. 2008; 5:634–644. [PubMed: 18711427]

19. Sorensen AG, Patel S, Harmath C, Bridges S, Synnott J, Sievers A, et al. Comparison of diameter and perimeter methods for tumor volume calculation. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2001; 19:551–557. [PubMed: 11208850]

20. van den Bent MJ, Vogelbaum MA, Wen PY, Macdonald DR, Chang SM. End point assessment in gliomas: novel treatments limit usefulness of classical Macdonald's Criteria. J Clin Oncol. 2009; 27:2905–2908. [PubMed: 19451418]

21. Vos MJ, Uitdehaag BM, Barkhof F, Heimans JJ, Baayen HC, Boogerd W, et al. Interobserver variability in the radiological assessment of response to chemotherapy in glioma. Neurology. 2003; 60:826–830. [PubMed: 12629241]

22. Wang SC, White DL, Pope JM, Brasch RC. Magnetic resonance imaging contrast enhancement versus tissue gadolinium concentration. Invest Radiol. 1990; 25(Suppl 1):S44–45. [PubMed: 2283253]

23. Wen PY, Macdonald DR, Reardon DA, Cloughesy TF, Sorensen AG, Galanis E, et al. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010; 28:1963–1972. [PubMed: 20231676]
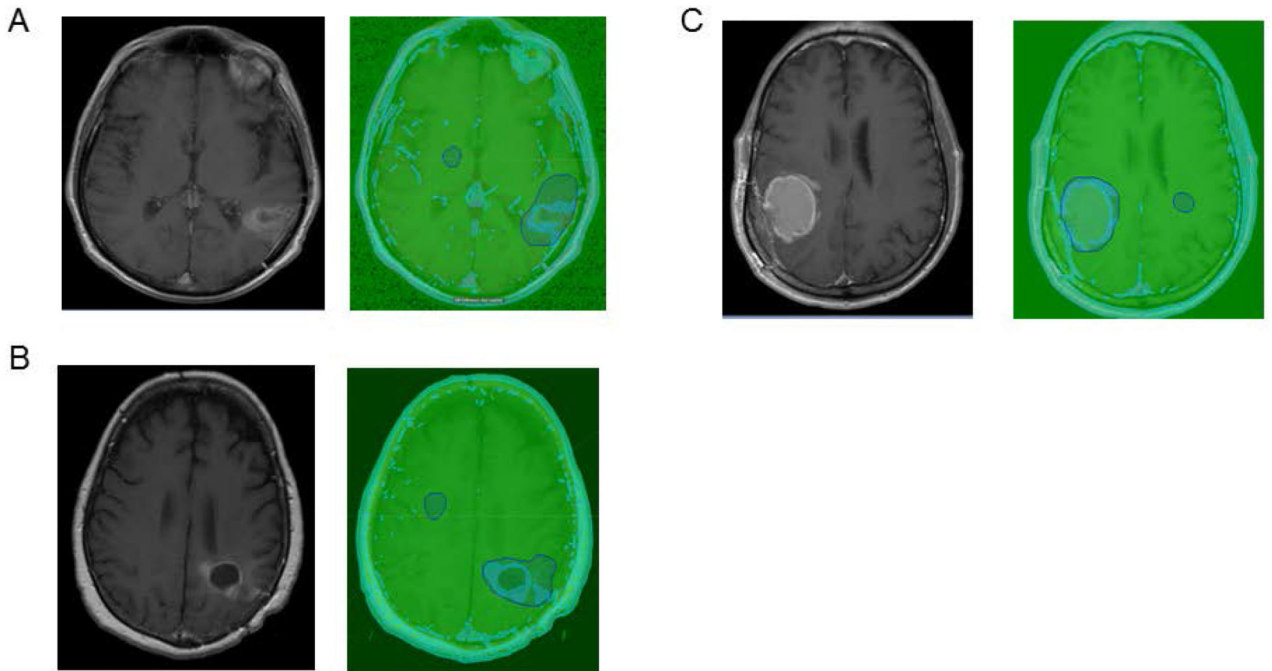
Figure 1.
A) Demonstrates a T1 with contrast image on the left and on the right with our software the blue in the cavity demonstrates the enhancing tumor volume. The enhancing tumor is limited to the region of interest. B) The tumor enhancement is picked up from the T1 with contrast to our velocity method without including the non-enhancing mass. C) A cystic mass with a rim of enhancement is picked up by our method as a rim.
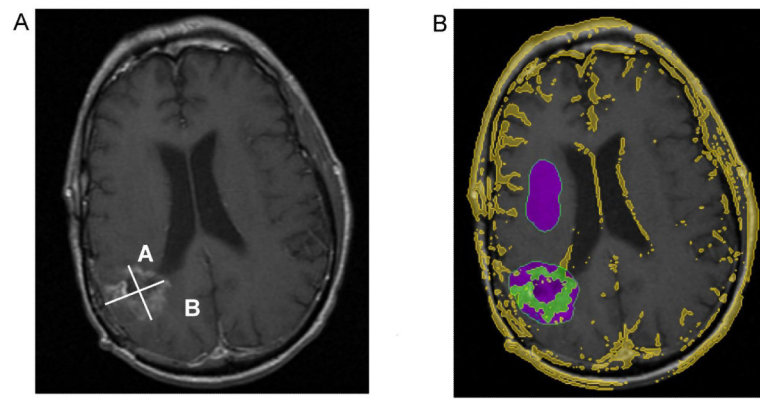
Figure 2.
A) T1-weighted post-contrast axial image demonstrate the 1-D RECIST measurement would be "A" and 2-D Macdonald measurement would be "A * B" B). Our volumetric analysis demonstrates the tumor tissue (shown in green).

Table 1

Correlation of Standard and Volumetric Measurements between Two Neuroradiologists. Correlation of measurement between two neuroradiologists is demonstrated. The calculated ICC is shown as an assessment of the interobserver variability of radiographic measurements for each measurement method.[a]

		ICC
Percent change from baseline	1D	0.42
	2D	0.61
	Volumetric	**0.97**

Abbreviations: 1D, one-dimensional; 2D, two-dimensional; ICC, intraclass correlation coefficients.

[a]N=57 scans, 44 comparisons were made to 13 baseline scans.

Table 2

Correlation of Volumetric Measurements. The volume measurements between the six examiners had a high ICC at 0.94; the percent change from baseline of tumor volume was also consistent with a high ICC at 0.93. Even in the hands of examiners without clinical or radiographic experience, there was a strong ICC using our novel volumetric software method.

Radiographic Method and Measurement[a]		N	Examiner 1		Examiner 2		Examiner 3		Examiner 4		Examiner 5		Examiner 6		ICC
			Median	Range	Median	Range	Median	Range	Median	Range	Median	Range	Median	Range	
Novel Volume Method	Actual Measurement	57	6.75	0.4 to 53.22	13.35	1.6 to 73.56	11.17	0.16 to 65.59	8.50	1.3 to 59.37	11.61	1.29 to 65.28	10.35	1.11 to 89.85	0.94
	Percent Change From Baseline	44	−12.91	97.25 to 1212.98	5.67	−85.06 to 1189.66	2.11	−79.53 to 1691.55	−15.01	−93.43 to 1006.39	5.16	−86.22 to 1749.38	−4.06	−91.94 to 1495.39	0.93

Abbreviations: ICC, intraclass correlation coefficients.

[a] Interobserver agreement on radiographic scan measurements of both expert and nonexpert examiners.

Table 3

Assessment of Interobserver Variability. Cohen's and Fleiss' Kappa Coefficients are presented in Table 3 as an assessment of the interobserver variability of determining radiographic progression or response for each of the methods. The novel volumetric method had the strongest agreement among the two neuroradiologists (Cohen's Kappa=0.96) and also had strong agreement (Fleiss' Kappa=0.79) among all examiners. Cohen's Kappa Statistic for Agreement among the two neuroradiologists for 2D and 1D were 0.54 and 0.46, respectively.

		Neuroradiologist 1														
		1D method				2D method				Volumetric method				
		Complete Response/ Regression	Stable	Progression	Total	Complete Response/ Regression	Stable	Progression	Total	Complete Response/ Regression	Stable	Progression	Total
Neuroradiologist 2	Complete Response/ Regression	4	3	4	11	5	2	4	11	5	0	0	5
	Stable	2	17	1	20	1	15	2	18	0	18	0	18
	Progression	2	3	8	13	2	2	11	15	0	1	20	21
	Concordance	65.9% (29 of 44)				70.5% (31 of 44)				97.7% (43 of 44)			
	Cohen's Kappa[a]	0.46				0.54				0.96			

Abbreviations: 1D, one-dimensional; 2D, two-dimensional.

[a]Interobserver agreement on radiographic calls of examiners for the 1D, 2D, and volumetric methods.