Comparison of Radiographic Approaches to Assess Treatment Response in Pituitary Adenomas: Is RECIST or RANO Good Enough?

Brandon S. Imber,[1]* Andrew L. Lin,[2]* Zhigang Zhang,[3]
Krishna Nand Keshavamurthy,[4] Amy Robin Deipolyi,[4] Kathryn Beal,[1]
Marc A. Cohen,[5] Viviane Tabar,[6] Lisa M. DeAngelis,[2] Eliza B. Geer,[7]
T. Jonathan Yang,[1]† and Robert J. Young[4]†

[1]Department of Radiation Oncology, Multidisciplinary Skull Base and Pituitary Center at Memorial Sloan-Kettering Cancer Center, New York, New York 10065; [2]Department of Neurology, Multidisciplinary Skull Base and Pituitary Center at Memorial Sloan-Kettering Cancer Center, New York, New York 10065; [3]Department of Epidemiology & Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York 10065; [4]Department of Radiology, Multidisciplinary Skull Base and Pituitary Center at Memorial Sloan-Kettering Cancer Center, New York, New York 10065; [5]Department of Surgery, Head & Neck Service, Multidisciplinary Skull Base and Pituitary Center at Memorial Sloan-Kettering Cancer Center, New York, New York 10065; [6]Department of Neurosurgery, Multidisciplinary Skull Base and Pituitary Center at Memorial Sloan-Kettering Cancer Center, New York, New York 10065; and [7]Department of Endocrinology, Multidisciplinary Skull Base and Pituitary Center at Memorial Sloan-Kettering Cancer Center, New York, New York 10065

ORCiD numbers: 0000-0002-1281-5915 (B. S. Imber); 0000-0003-0659-261X (A. L. Lin); 0000-0003-0989-6723 (K. Beal); 0000-0002-1284-1670 (V. Tabar); 0000-0002-6290-7818 (L. M. DeAngelis); 0000-0003-3722-1889 (E. B. Geer); 0000-0001-9375-7384 (T. J. Yang).

*B.S.I and A.L.L contributed equally as co-first authors.
†T.J.Y and R.J.Y contributed equally as co-senior authors.

Context: Pituitary adenomas (PA) are often irregularly shaped, particularly posttreatment. There are no standardized radiographic criteria for assessing treatment response, substantially complicating interpretation of prospective outcome data. Existing imaging frameworks for intracranial tumors assume perfectly spherical targets and may be suboptimal.

Objective: To compare a three-dimensional (3D) volumetric approach against accepted surrogate measurements to assess PA posttreatment response (PTR).

Design: Retrospective review of patients with available pre- and postradiotherapy (RT) imaging. A neuroradiologist determined tumor sizes in one dimensional (1D) per Response Evaluation in Solid Tumors (RECIST) criteria, two dimensional (2D) per Response Assessment in Neuro-Oncology (RANO) criteria, and 3D estimates assuming a perfect sphere or perfect ellipsoid. Each tumor was manually segmented for 3D volumetric measurements. The Hakon Wadell method was used to calculate sphericity.

Setting: Tertiary cancer center.

Patients or Other Participants: Patients (n = 34, median age = 50 years; 50% male) with PA and MRI scans before and after sellar RT.

Interventions: Patients received sellar RT for intact or surgically resected lesions.

Main Outcome Measure(s): Radiographic PTR, defined as percent tumor size change.

Results: Using 3D volumetrics, mean sphericity = 0.63 pre-RT and 0.60 post-RT. With all approaches, most patients had stable disease on post-RT scan. PTR for 1D, 2D, and 3D spherical measurements were moderately well correlated with 3D volumetrics (*e.g.,* for 1D: 0.66, $P < 0.0001$) and were superior to 3D ellipsoid. Intraclass correlation coefficient demonstrated moderate to good reliability for 1D, 2D, and 3D sphere ($P < 0.001$); 3D ellipsoid was inferior ($P = 0.009$). 3D volumetrics identified more potential partially responding and progressive lesions.

Conclusions: Although PAs are irregularly shaped, 1D and 2D approaches are adequately correlated with volumetric assessment.

Pituitary tumors represent the third most common primary tumor in the adult head after gliomas and meningiomas, accounting for ~15% of cases [1]. Although pituitary adenomas (PAs) are typically benign and slow-growing, their clinical course can vary widely; those that are hormonally active or located in the proximity of sensitive parasellar structures such as the cavernous sinuses, optic chiasm, and hypothalamus often cause the greatest morbidity [2, 3]. With the exception of prolactinomas, surgery remains the preferred initial treatment, and subtotally resected and recurrent tumors often require multimodality treatment including medication and radiotherapy (RT) [4]. Despite the current treatment efforts, a subset of these tumors progress.

Clinical trials investigating the role of cytotoxic chemotherapy, targeted therapy, and immunotherapy are needed to improve the treatment of PAs [5, 6]. Emerging retrospective data support the use of the alkylator chemotherapy, temozolomide, in the treatment of aggressive PAs. One challenge is that the field has not agreed on a standardized response criterion; for this reason, authors have defined meaningful responses to temozolomide as reductions of anywhere from 20% to 80%. Further complicating interpretation is hetero-geneity in how a reduction in tumor size is defined. For example, sometimes regression refers to a decrease in tumor volume and sometimes a decrease in longest diameter (or the product of diameters) [7]. For reference, a 65% decrease in volume, assuming tumor sphericity, translates to a 50% decrease in the product of the diameters and a 30% change in the longest diameter [8]. Thus, differences in how regression is defined can be meaningful in terms of overall interpretation.

Going forward, the ability to compare temozolomide and other therapies reliably necessitates confidence that overall response rates are generated comparably. As the number of experimental protocols for both medical therapies and RT continues to increase [9, 10], there is an increasing need to establish a standardized approach to imaging interpretation.

There are imaging response criteria that have been validated for other tumor types. Response Evaluation Criteria in Solid Tumors (RECIST) [11] measures tumors in one di-mension and is used in clinical trials to define radiographic response for most malignancies outside of the central nervous system. For central nervous system malignancies, a two-dimensional criteria, Response Assessment in Neuro-Oncology (RANO) [12], is used most commonly. Two different volumetric approaches have been proposed: (i) an approach that uses geometric formulas that assume that intracranial tumors are either perfectly spherical or ellipsoid and (ii) computer-based approaches which allow tumor volume measurements based on three-dimensional (3D) voxel-based segmentation of relevant cross-sectional imaging

sequences. Segmentation can be performed either manually by a radiologist or using semi- or fully automated approaches. The 3D volumetric approach has been adopted in several contexts within neuro-oncology, *e.g.,* in glioblastoma, with promising results [13, 14]. Although 3D volumetric measurement is believed to be the most accurate reflection of tumor response, its correct application requires operator expertise and more neuroradiology time/resources per scan.

Importantly, there have not been formal validations of any of these imaging frameworks for pituitary tumors. We hypothesized that because of the irregular morphology of pituitary tumors, particularly after surgical resection, one dimensional (1D) and two dimensional (2D) as well as 3D spherical or ellipsoid surrogate measurement strategies are inadequate. In this study, we generated 3D volumetric measurements in patients receiving conventionally fractionated RT for PAs and compared the post-RT response with the other more standard neuro-oncological imaging response criteria.

1. Materials and Methods

A. Patient Cohort

We retrospectively identified patients from a departmental database treated between 2000 and 2017 who had pathologically confirmed pituitary macroadenomas and available MRI imaging before and after pituitary RT. The inclusion criteria included either functional or nonfunctional adenoma. A total of 34 consecutive patients were identified and included in our study (Table 1). Patients had a median of 1 (range, 0 to 4) resection at a median of 6.1 months before RT. One patient did not have a resection because of locally advanced disease but the diagnosis was confirmed pathologically via endoscopic biopsy. For uniformity, the MRI immediately preceding and following RT were analyzed. This imaging end point was selected given differing surveillance schedules as per the standard of care. The median duration between scans was 4.9 months (range, 2.9 to 9.8). All patients received full-dose RT, between 45 and 54 Gy, determined at the discretion of the treating radiation oncologist.

Table 1. Cohort Description

Patient Characteristics	Detail
Male sex	17 (50%)
Age at the time of RT, y	
Median	50.4
Range	14.2–72.4
Histological classification	
Nonfunctional adenoma	23 (68%)
Prolactinoma	4 (12%)
Growth hormone staining	4 (12%)
Adrenocorticotropic hormone staining	3 (9%)
Surgical history prior to RT	
No prior surgeries	1 (3%)
1 prior surgery	18 (53%)
2 prior surgeries	11 (32%)
3 or more prior surgeries	4 (12%)
Duration between most recent surgery and RT, mo	
Median	6.1
Range	1.1-169.1
RT dose, Gy	
45	6 (18%)
50.4	22 (65%)
54	6 (18%)
Duration between pre-RT and post-RT MRI, mo	
Median	4.9
Range	2.9-9.8

Because the goal of this project was to evaluate response criteria for prospective clinical trials, we decided to evaluate the first posttreatment MRI. One of the biggest criticisms of the available literature is that there is no standardized reporting of treatment response whether it is to RT or systemic therapy. Our review of recently reported or currently accruing protocols for PAs suggest that radiographic response at three or six months is often a primary or secondary endpoint (*e.g.,* NCT03930771, NCT00939523, NCT03309319). Even within these three listed protocols, there are three different imaging end points that have been described (*e.g.,* RECIST criteria, 40% reduction in tumor size, tumor volume). To our knowledge, neither RECIST nor RANO have been validated formally for PAs, therefore, the utility of these approaches for short posttreatment time points is already a salient clinical question. In other words, the assumption that RECIST and other response criteria are applicable to this clinical situation is already dictating protocol success or failure. We designed our pilot study to emulate a prospective trial and for these reasons, we selected the first posttreatment scan, which corresponds to approximately six months of follow-up. Selection of this end point also increased the homogeneity of our patient sample.

B. MRI Scans and Measurements

MRI scans were acquired according to the standard of care during the study period, including the administration of intravenous contrast usually consisting of gadobutrol at 0.1 mmol/kg (Gadavist, Bayer, Whippany, NJ). Pituitary tumors were usually measured on coronal contrast T1-weighted images targeted to the sella with a median slice thickness of 5.0 mm (range, 1.0 to 7.5).

A board-certified neuroradiologist with 18 years of experience measured the pre- and post-RT pituitary lesions according to the following accepted approaches: 1D measurements (as governed by RECIST 1.1) [11], 2D measurements (as governed by RANO) [12], and 3D spherical and 3D ellipsoid measurements (as governed by New Approaches to Brain Tumor Therapy and volumetric RECIST, respectively) [15–18]. For a 1D estimation, the longest diameter was measured. The product of the two largest perpendicular diameters was calculated for a 2D estimation.

3D volumetric measurements of the tumor were performed as follows. The neuroradiologist manually segmented each tumor using Food and Drug Administration-approved commercially available software (Aquarius iNtuition Edition 4.4.12, TeraRecon, Foster City, CA). After segmentation, tumor volumes were measured and recorded in cubic centimeters. The measured volumes were exported as 3D mesh model files and examined using 3D Slicer 4.8.0 (http://www.slicer.org) [19] to extract tumor volumes and surface areas, which were then used to calculate Hakon Wadell sphericity indices [20]. This unitless sphericity index describes how closely the shape of a lesion approximates a perfect sphere by measuring the ratio of the surface area of a sphere (equal in volume to the lesion) to the surface area of the lesion, whereby a ratio of 1 represents a perfect sphere and a ratio <1 represents less spherical shapes.

Finally, two different calculated 3D surrogates were obtained by utilizing standard geometric formulas for a 3D sphere or 3D ellipsoid. The maximal orthogonal diameters were automatically calculated from the segmented volume and recorded in centimeters. The radii were recorded and used to calculate surrogate volumes, assuming spherical shapes (volume = $4/3\pi r^3$) and alternatively ellipsoid shapes (volume = $4/3\pi r_1 r_2 r_3$) where $r_1 > r_2 = r_3$ [17, 21].

C. Imaging Response Assessment

Posttreatment response (PTR) was defined as the percent change in post-RT tumor size vs pre-RT, and a negative change was defined as a reduction in tumor size after treatment. PTR was then categorized as complete response (CR) if there was complete disappearance of the lesion, partial response (PR), progressive disease (PD), or otherwise stable disease (SD) based on accepted criteria. For RECIST, patients were classified as PR if there was a 30% decrease

in maximum diameter or PD if there was at least a 20% increase in maximum diameter [11]. For RANO, patients were categorized as PR or PD if there was decrease by at least 50% or increase by 25%, respectively [12]. For the 3D spherical and 3D ellipsoid approaches, the classification of imaging response was based on simple mathematical extrapolation of RECIST to spherical or ellipsoid volumes, as described previously [17, 21–23]. Table 2 summarizes our response assessment criteria.

For the 3D volumetric approach, there is no defined consensus for imaging response thresholds for pituitary (or any solid) tumors using volumetric analysis. For PD, we selected a minimum threshold of >20% because this cutoff has precedence in neuro-oncology [24] and has been used in other volumetric studies as a minimum to indicate clinically meaningful change [25]. As small changes in the size of sellar tumors can have anatomic implications, we purposefully selected PD criteria to maximize sensitivity. Given the assumption that pituitary tumors are not spherical, we did not believe that a simple spherical extrapolation of RECIST (which would have required >70% increase) was suitable. For PR, a 30% reduction in volume was required.

D. Statistical Approach

Mean pre- and post-RT sphericities were compared using the Wilcoxon signed rank test with continuity correction. 3D volumetrics were assumed to be the gold standard for PA measurement. We used several statistical approaches to assess the accuracy of the various surrogate estimations to predict PTR compared with the 3D volumetric gold standard. The associations were assessed using Pearson correlation and intraclass correlation coefficient (ICC) [26] and were visualized using the Bland-Altman plot [27]. ICC analysis quantitatively tests the agreement or repeatability between two different methods to measure the same quantity; values close to 1 indicate strong agreement or repeatability. Bland-Altman plots are often used to graphically illustrate the agreement or repeatability between two different methods on measuring the same quantity. In this study, the quantity under comparison was the PTR, which was measured in several different ways (1D, 2D, 3D volumetrics, etc.). Graphically, the points on the Bland-Altman plots were plotted on a horizontal coordinate representing the average of the two measurements and a vertical coordinate representing the difference between them. If the two measurements agree with each other to a reasonable degree, then most of the points should lie within roughly two SDs of the difference (*i.e.,* 95% CI) and without depicting any obvious pattern.

Statistical calculations were performed using R v. 3.4.3 (R Foundation for Statistical Computing, Vienna, Austria), SPSS V. 25 (IBM, Armonk, NY), or GraphPad Prism 7 (GraphPad Software, La Jolla, CA) and a *P* value of less than 0.05 was considered significant.

Table 2. Radiographic Response Assessment Criteria

Measurement Strategy	Standardized Response Criteria	PR	PD	SD
[1D] Longest diameter, cm	RECIST	Decrease by 30%	Increase by 20%	Neither PR nor PD criteria met
[2D] Product of perpendicular diameters, cm^2	RANO	Decrease by 50%	Increase by 25%	Neither PR nor PD criteria met
[3D spherical] Surrogate volume using geometric formula for perfect sphere, cm^3		Decrease by 65%	Increase by 73%	Neither PR nor PD criteria met
[3D ellipsoid] Surrogate volume using geometric formula of perfect ellipsoid, cm^3		Decrease by 30%	Increase by 20%	Neither PR nor PD criteria met
[3D volumetric] Measured volume using segmentation, cm^3		Decrease by 30%	Increase by 20%	Neither PR nor PD criteria met

2. Results

A. Patient Characteristics

There were 34 patients with a median age of 50.4 years; 17 (50%) were men (Table 1). All except 1 patient were treated with intensity-modulated radiotherapy; the latter was treated with a 3D conformal RT plan. All but 1 patient (97%) received RT as either adjuvant or salvage therapy. Thus, in our cohort, tumor measurements were performed principally on postsurgical and/or recurrent tumors. One patient was deemed as nonoperable at diagnosis and received definitive intent RT after endoscopic biopsy. For this patient, tumor measurements were performed on an intact tumor.

B. Summary of Different Measurement Approaches

Regardless of 1D, 2D, or 3D techniques, there was little change in the pre- and post-RT measurements when examining the mean or the median (Table 3). In general, the mean measurement was reduced post-RT whereas the median measurement was stable or slightly larger with all surrogate measurement techniques. Using 3D volumetric segmentation, the median size of the pre-RT and post-RT tumors was 6.5 cm^3 (range $\pm$ SD, 0.5 to 51.6 $\pm$ 10.5) and 6.1 cm^3 (range $\pm$ SD, 0.4 to 44.8 $\pm$ 9.8), respectively. This difference did not achieve statistical significance ($P = 0.11$) using the Wilcoxon signed rank test.

C. Sphericity Estimation

Mean pre- and post-RT sphericity were 0.63 and 0.60, respectively, indicating variable irregular shapes (Fig. 1). Post-RT tumors tended to be less spherical and the difference trended toward significance ($P = 0.054$). The pre-RT sphericity of the nonoperable tumor was 0.39, indicating the possibility of highly irregular shapes even in large intact lesions involving the skull base.

D. Assessment of Posttreatment Imaging Responses

Figure 2 summarizes the percent changes in tumor measurements pre- and post-RT by imaging convention. In general, the tumors remained stable after RT as reflected by all five measurement approaches. The 3D spherical approach had the widest PTR range. Overall imaging response at the first post-RT assessment was then determined for each of the five measurement approaches and is summarized in Table 4. No patient achieved CR. Using the 3D volumetric approach, 24% and 21% of patients had PR or PD, respectively, and the remainder had SD. A greater proportion of patients were classified as having SD using either the 1D, 2D, or 3D spherical surrogate approaches. In contrast, fewer cases were classified as SD under the 3D ellipsoid approach. Response classification was significantly different ($P = 0.007$, χ^2) among the five approaches.

**Table 3. Descriptive Summary of Tumor Sizes Using Five Different Measurement Approaches**

	Pre-RT (n = 34)			Post-RT (n = 34)		
	Mean	Median	SD	Mean	Median	SD
[1D] Longest diameter, cm	3.1	2.7	1.2	2.9	2.8	1.3
[2D] Product of perpendicular diameters, cm^2	6.5	4.9	5.3	6.0	5.1	5.4
[3D spherical] Surrogate volume using geometric formula for perfect sphere, cm^3	23.2	10.2	28.9	20.7	11.2	27.3
[3D ellipsoid] Surrogate volume using geometric formula of ellipsoid, cm^3	8.4	4.3	9.9	7.7	4.7	10.9
[3D volumetric] Measured volume using segmentation, cm^3	9.2	6.5	10.5	8.5	6.1	9.8
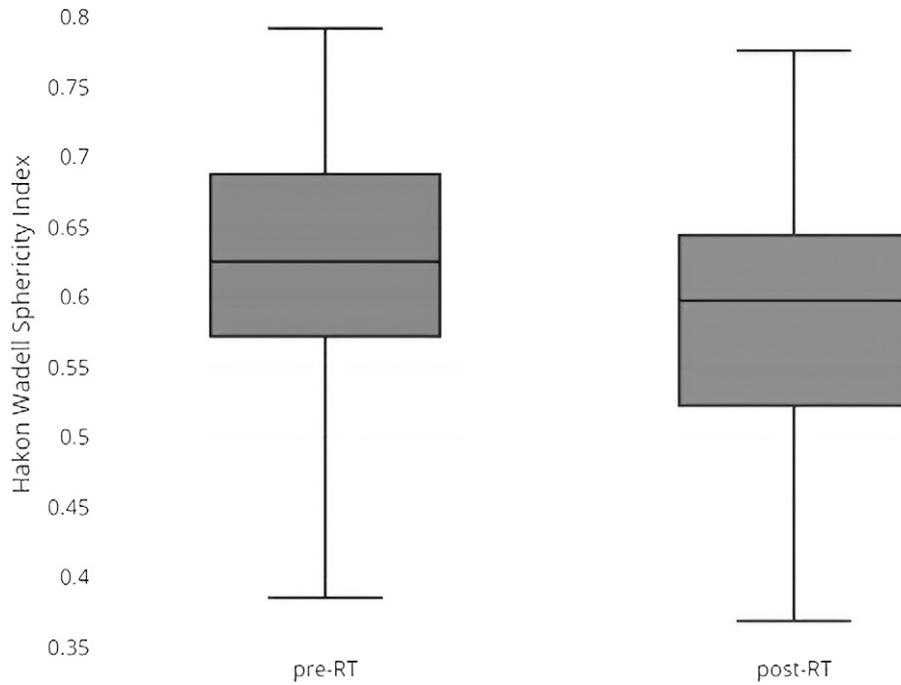
Figure 1. Box and whisker distribution of pre- and post-RT sphericities in the cohort.

E. Associations Between Volumetric and Surrogate PTR

Under the assumption that the 3D volumetric approach is the gold standard we then assessed the correlation between the PTR calculated using the various surrogate approaches and the
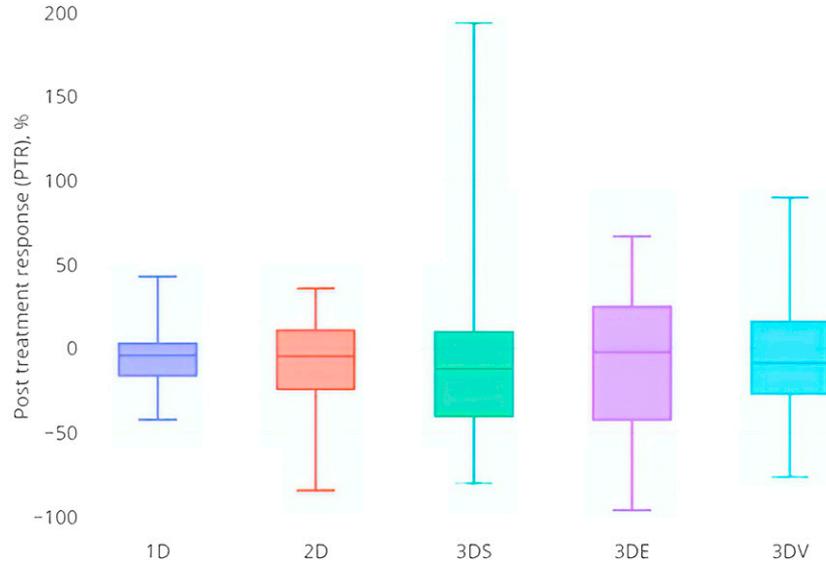
|	1D	2D	3DS	3DE	3DV
Median	-4%	-4%	-12%	-2%	-8%
Mean	-5%	-9%	-5%	-8%	-8%
Minimum	-42%	-84%	-80%	-96%	-76%
Maximum	43%	36%	194%	67%	90%

Figure 2. PTR distribution for the five different measurement approaches. The table below the box and whisker plots summarizes the distribution.

Table 4.   Overall Post-RT MRI Response Assessment Using the Different Measurement Approaches

	1D	2D	3D Spherical	3D Ellipsoid	3D Volumetric
CR	0	0	0	0	0
PR	3	3	3	10	8
SD	28	28	28	15	19
PD	3	3	3	9	7

PTR calculated using 3D volumetrics. The scatter plots and associated Pearson product moment correlation coefficients suggested that moderate correlation exists between each of the four surrogate measurement approaches and the 3D volumetric approach (Fig. 3). Of these associations, the strongest correlation was found between the 1D/RECIST framework and 3D volumetric whereas the weakest correlation was found between the estimates derived from the 3D ellipsoid approach and 3D volumetrics (Fig. 3F). Not surprisingly, given the mathematical relation between the two estimates, the Pearson coefficient was found to the strongest between the 1D and 2D surrogate estimates (coefficient = 0.71; 95% CI: 0.48, 0.84; $P < 0.00001$).

ICC values to assess the overall association of each surrogate measurement approach with the 3D volumetric gold standard at the individual patient level are shown in Table 5. ICC demonstrated moderate to good reliability of response for 1D, 2D, and 3D spherical (ICC = 0.54, 0.61, 0.52, respectively; $P < 0.001$); 3D ellipsoid was again inferior (ICC = 0.47, $P = 0.002$). Bland-Altman visualization confirmed similar moderate to good concordance as few points fell outside the 95% CI of the difference between the two estimates.
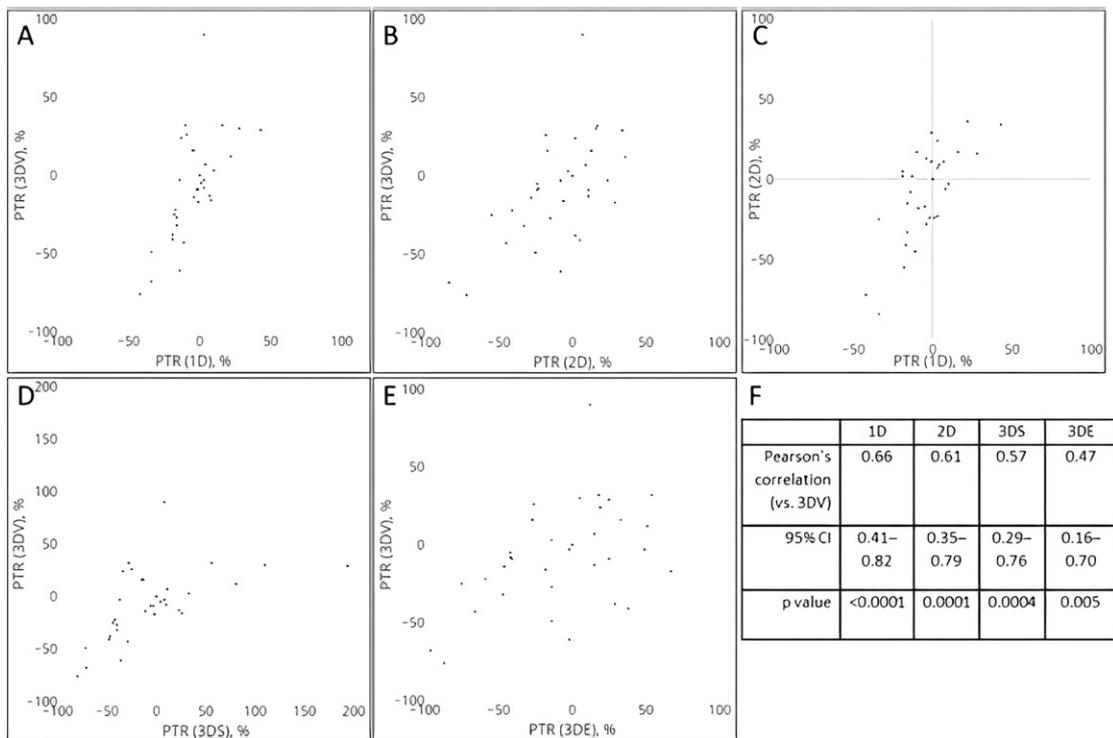
Figure 3. Scatter plot distributions showing the PTR estimates as determined by the (A) 1D, (B) 2D, (D) 3D spherical, and (E) 3D volumetric approaches, all vs the presumed gold standard 3D volumetric approach. (C) PTR scatter plot association for the 2D vs 1D approaches. (F) Pearson product-moment correlation coefficients and 95% CIs for the individual surrogate measurements vs the 3D volumetric approach.

Table 5. ICCs for PTR Determined by Different Measurement Approaches

	ICC	95% CI	P
1D vs 3D volumetric	0.54	0.30–0.78	0.0004
2D vs 3D volumetric	0.61	0.40–0.83	<0.00004
3D spherical vs 3D volumetric	0.52	0.28–0.77	0.0006
3D ellipsoid vs 3D volumetric	0.47	0.21–0.74	0.002
1D vs 2D	0.62	0.41–0.83	<0.00003

F. Characterizing Discordance Between 3D Volumetric and Surrogate Measurement Approaches

False-positive assignments of PD as well as false-negative assignments of SD/PR/CR could have important implications for clinical management. Therefore, we sought to better understand patients in whom the surrogate and volumetric classifications were discordant. From the correlative analysis, the estimated 3D surrogate approaches, 3D spherical and 3D ellipsoid, did not appear to be more accurate than the simpler 1D and 2D measurements. Therefore, we focused on comparing the more widely used 1D/2D and 3D volumetric approaches. Figure 4 shows the patient-level PTR measurements in order of decreasing post-RT tumor sphericity as well as potential classification discrepancies.

Assuming 3D volumetrics as the gold standard, there were few false-positive classifications of tumor progression; using 1D, only one of three patients classified as PD was discordant with the 3D volumetric interpretation (which was SD). Using 2D, two of three patients classified as PD were discordant with the 3D volumetric interpretation (both SD). Six of the seven patients classified as PD using 3D volumetrics were classified as SD by either 1D and/or 2D measurements. Of these potentially false-negative patients, the majority (5/6) had PTR (as measured by 3D volumetrics) in the range of 25% to 35%. There was no indication that less spherical tumors were more likely to have discordant imaging classifications.
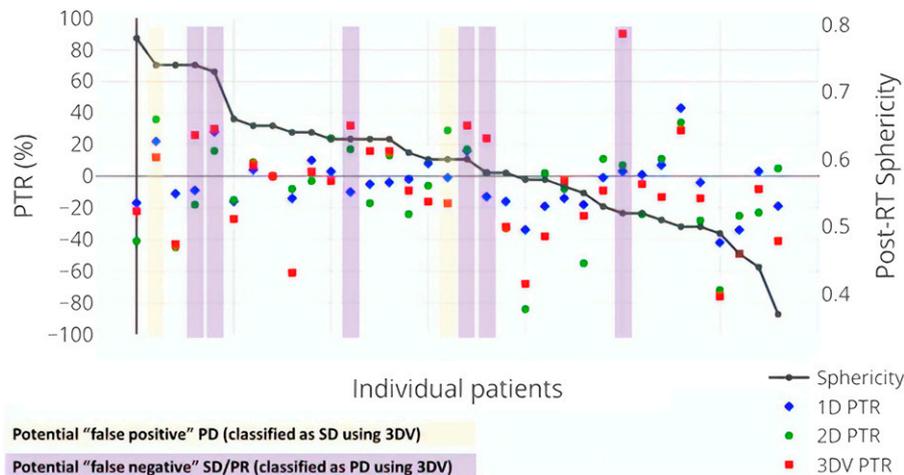
Figure 4. Per-patient distribution of the PTR as calculated by the 1D, 2D, and 3D volumetric measurement approaches. Each column reflects an individual patient in the cohort and the PTR is listed as measured by the left-sided *y*-axis. The individual patients are ordered according to post-RT sphericity, *i.e.,* patients with less spherical tumors are further to the right. Patients whose imaging classifications using 1D or 2D surrogates are discordant from 3D volumetric interpretations are highlighted in yellow (for potential false-positive PD) or purple (for potential false-negative PR or SD).

G. Radiographic Response and Correlation With Clinical Outcomes

Although the primary objective was to compare treatment response, we sought to explore the predictive power of the various imaging strategies. Clinical outcomes after the analyzed post-RT MRI were reviewed (median follow-up = 40 months; range, 2 to 196). Of the 34 patients, there were only 2 documented failures (as defined by a change in treatment) following the referenced course of RT. Given the limited number of events, we did not have adequate power to meaningfully prognosticate based on PTR. However, all 7 patients found to have PD by 3D volumetrics had no subsequent documented clinical evidence of progression, as defined by the need for salvage treatment, over the period of chart review (median, 39.8 months; range, 2 to 196).

3. Discussion

Prior studies principally focusing on primary intra-axial brain tumors have advocated for the usage of linear 1D and 2D measurements because they demonstrate high correlation and concordance for determining tumor progression [28, 29] and may be better than estimated 3D measurements [30]. Across oncology, because these 1D/2D measurements are easier to implement than 3D volumetrics, most modern response criteria propose them as reasonable surrogates for determining tumor size changes. These criteria have become the standard for response assessment in prospective neuro-oncology trials. To our knowledge, there has been no formal comparison or validation of these imaging conventions specifically for pituitary lesions. Our main concern was that the 1D/2D approaches might be suboptimal for pituitary lesions because the tumors are often highly irregularly shaped.

This study aimed to evaluate four accepted surrogate measurements against 3D volumetric measurements to assess the PTR of PAs. We first demonstrated that PAs are highly nonspherical. Whereas other groups have argued empirically that pituitary tumors, particularly postsurgical lesions, are irregularly shaped, we quantified this low sphericity using the Hakon Wadell approximation. In our samples, mean sphericity was roughly 0.6 but was found to be as low as 0.36 (with 1 representing a perfect sphere).

Given this geometric irregularity, we expected that 1D and 2D approaches as well as calculated 3D approaches would correlate poorly with volumetric measurements. To the contrary, our data suggest that the 1D and 2D approaches as well as the 3D spherical approach are reasonably well correlated with volumetric prediction, as evidenced by similar moderate to good Pearson correlation and ICC values. We observed that the PTRs predicted for the various imaging conventions were often strongly clustered, even for tumors with the lowest sphericity (Fig. 4). This suggests that RECIST and/or RANO can be applied to this patient population.

The volumetric approach is substantially more labor intensive and requires a highly experienced neuroradiology operator for maximal validity. Recommendation of the volumetric approach as the gold standard for pituitary tumor assessment is nontrivial and would require widespread availability of the software and consensus recommendations for performing this analysis. Although the 3D volumetric approach declared more treatment responses and failures, its clinically meaningful effects remained uncertain because none of the patients identified as having PD by 3D volumetrics on the first post-RT scan went on to have sustained progression requiring additional salvage treatment.

With that caveat, there may be situations in which 3D volumetric measurements may add value above RECIST or RANO. The 3D volumetric approach appears to have improved the capacity to discriminate between subtle PR and otherwise SD. Figure 5 shows a representative patient example in which the 1D interpretation was SD but volumetric measurement suggested PR that persisted on the subsequent surveillance scan. Potential clinical scenarios in which 3D volumetrics might be more accurate include multiloculated or cystic adenomas, small recurrences or areas of residual disease, and multifocal and bony invasive adenomas.
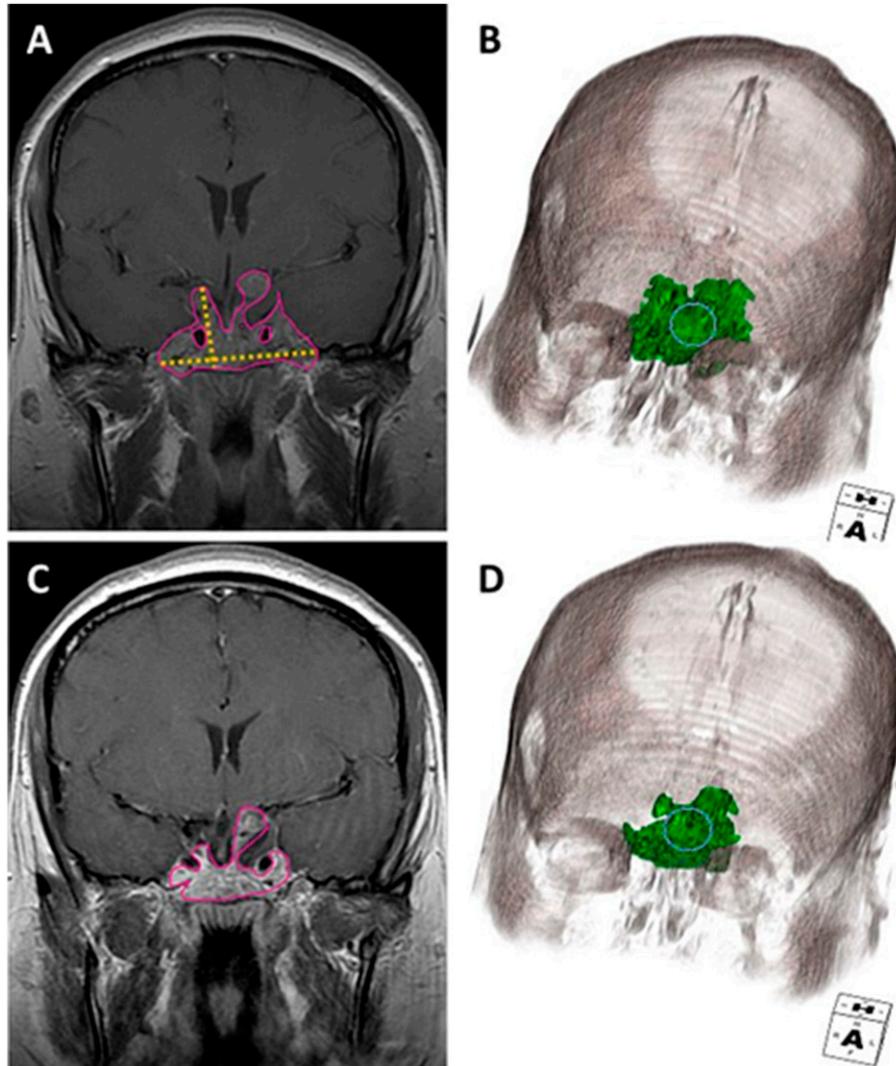
Figure 5. Coronal contrast T1-weighted and 3D volumetric images (A, B) before and (C, D) after radiation therapy illustrate discordant response assessment in a highly irregular shaped tumor. The tumor volume outlined in pink (A, C) and coded in green (B, D) demonstrated PR, whereas the orthogonal yellow lines in (A) demarcate the maximal perpendicular diameters used for RANO demonstrated SD.

Although the distinction between SD and PR may not often change routine clinical practice, accurate assessment of imaging response could have important implications for clinical trials and affect the approval of new treatments, where overall response rate is typically the proportion of complete and partial responders. Given that the overall response rate is often a primary outcome for early stage oncologic trials, insensitive radiographic response criteria could mean the difference between trial success and failure.

There is no established threshold for tumor progression when assessed using volumetric segmentation. Given the proximity to critical structures such as the optic chiasm, a 20% increase in PA volume could constitute a clinically important change that would warrant a change in treatment. We selected this threshold given precedents [24, 25] and to maximize sensitivity. The radiographic balance of appropriate thresholds and resultant test sensitivity is not unique to pituitary tumors; we would have had considerably better concordance between the 1D/2D and 3D volumetric response assessments had a volumetric progression threshold of >35% been used.

For pituitary tumors, defining a strict progression cutoff may be challenging because of the heterogeneity of tumor location. For example, if a tumor encroaches or contacts the optic

chiasm, a tiny change could have a detrimental effect, whereas inferior tumors extending into the sphenoid sinus could expand substantially without having any appreciable clinical impact. This would argue that response assessment criteria should incorporate data on the anatomic location of the tumor, as even minor changes in size in a high-risk location could support a change in treatment strategy.

There are several limitations to this study. The analyzed cohort is fairly small with few clinical events after RT. Given the natural history of PAs, a larger, and perhaps multi-institutional, cohort would likely be required to correlate radiographic or radiomic characteristics with clinical outcomes. Furthermore, the natural history of PAs is slow growth and our relatively short median follow-up may translate to lower power to detect subtle differences between the different methodologies. We selected this time point because the first posttreatment scan is often a clinically reported end point and is already being used as the response assessment time point for primary or secondary outcomes in prospective protocols. Selection of this time point enabled greater homogeneity for our pilot since the longer term follow-up imaging intervals tended to be more irregular given diverse practice patterns.

Because this was a retrospective study, there was heterogeneity in the MRI slice thicknesses drawn from actual clinical practices that may lead to overestimation of tumor volumes. It is also possible that the response on the first post-RT scan might include some residual local inflammatory change or tumor flare phenomenon because of the RT that does not reflect the inevitability of ongoing growth. This remains a commonly used and clinically meaningful end point; furthermore, our primary goal to study radiographic performance was enabled by using this homogeneous time point.

We do not feel that our study alone is sufficient to say definitively whether RECIST or RANO is suitable to capture the full degree of nuance required for all clinical situations. Specifically, longer term follow-up is critical to further validate these surrogate measurement approaches. Additional work is ongoing with larger cohorts and more longitudinal and longer-term imaging assessments to categorize typical posttreatment patterns and to better prognosticate which clinical scenarios might benefit from more detailed volumetric assessment. We hope that this pilot study is a first step to homogenize the reporting criteria and structure for PAs which hopefully will empower clinical trials and facilitate the comparison of cross-institutional outcome data.

4. Conclusion

Although pituitary tumors are inherently nonspherical, unidimensional or bidimensional imaging measurement appear to be suitable surrogates for routine clinical surveillance and response assessment. These methods, as governed by RECIST or RANO, have acceptable correlation with more sophisticated 3D volumetric approaches. 3D volumetrics may be a more nuanced measurement tool for prospective trials that require accurate response assessment.

Acknowledgments

Additional Information

Correspondence: Andrew L. Lin, MD, Department of Neurology, Memorial Sloan-Kettering Cancer Center, New York, New York 10065. E-mail: lina1@mskcc.org.

References and Notes

1. Ostrom QT, Gittleman H, Liao P, Vecchione-Koval T, Wolinsky Y, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: primary brain and other central nervous system tumors diagnosed in the United States in 2010–2014. *Neuro Oncol.* 2017;**19**(suppl_5):v1–v88.

2. Gittleman H, Ostrom QT, Farah PD, Ondracek A, Chen Y, Wolinsky Y, Kruchko C, Singer J, Kshettry VR, Laws ER, Sloan AE, Selman WR, Barnholtz-Sloan JS. Descriptive epidemiology of pituitary tumors in the United States, 2004-2009. *J Neurosurg.* 2014;**121**(3):527–535.

3. Ntali G, Wass JA. Epidemiology, clinical presentation and diagnosis of non-functioning pituitary adenomas. *Pituitary.* 2018;**21**(2):111–118.

4. Raverot G, Burman P, McCormack A, Heaney A, Petersenn S, Popovic V, Trouillas J, Dekkers OM; European Society of Endocrinology. European Society of Endocrinology Clinical Practice Guidelines for the management of aggressive pituitary tumours and carcinomas. *Eur J Endocrinol.* 2018;**178**(1): G1–G24.

5. Reddy R, Cudlip S, Byrne JV, Karavitaki N, Wass JAH. Can we ever stop imaging in surgically treated and radiotherapy-naive patients with non-functioning pituitary adenoma? *Eur J Endocrinol.* 2011; **165**(5):739–744.

6. Tampourlou M, Ntali G, Ahmed S, Arlt W, Ayuk J, Byrne JV, Chavda S, Cudlip S, Gittoes N, Grossman A, Mitchell R, O'Reilly MW, Paluzzi A, Toogood A, Wass JAH, Karavitaki N. Outcome of nonfunctioning pituitary adenomas that regrow after primary treatment: a study from two large UK centers. *J Clin Endocrinol Metab.* 2017;**102**(6):1889–1897.

7. Lin AL, Sum MW, DeAngelis LM. Is there a role for early chemotherapy in the management of pituitary adenomas? *Neuro Oncol.* 2016;**18**(10):1350–1356.

8. James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Muldal A, Therasse P. Measuring response in solid tumors: unidimensional versus bidimensional measurement. *J Natl Cancer Inst.* 1999; **91**(6):523–528.

9. Ben-Shlomo A, Cooper O. Role of tyrosine kinase inhibitors in the treatment of pituitary tumours: from bench to bedside. *Curr Opin Endocrinol Diabetes Obes.* 2017;**24**(4):301–305.

10. Syro LV, Rotondo F, Camargo M, Ortiz LD, Serna CA, Kovacs K. Temozolomide and pituitary tumors: current understanding, unresolved issues, and future directions. *Front Endocrinol (Lausanne).* 2018;**9**: 318.

11. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer.* 2009;**45**(2): 228–247.

12. Wen PY, Chang SM, Van den Bent MJ, Vogelbaum MA, Macdonald DR, Lee EQ. Response assessment in neuro-oncology clinical trials. *J Clin Oncol.* 2017;**35**(21):2439–2449.

13. Henker C, Kriesen T, Glass Ä, Schneider B, Piek J. Volumetric quantification of glioblastoma: experiences with different measurement techniques and impact on survival. *J Neurooncol.* 2017;**135**(2): 391–402.

14. Huber T, Alber G, Bette S, Kaesmacher J, Boeckh-Behrens T, Gempt J, Ringel F, Specht HM, Meyer B, Zimmer C, Wiestler B, Kirschke JS. Progressive disease in glioblastoma: benefits and limitations of semi-automated volumetry. *PLoS One.* 2017;**12**(2):e0173112.

15. Batchelor TT, Gilbert MR, Supko JG, Carson KA, Nabors LB, Grossman SA, Lesser GJ, Mikkelsen T, Phuphanich S; NABTT CNS Consortium. Phase 2 study of weekly irinotecan in adults with recurrent malignant glioma: final report of NABTT 97-11. *Neuro Oncol.* 2004;**6**(1):21–27.

16. Sorensen AG, Batchelor TT, Wen PY, Zhang W-T, Jain RK. Response criteria for glioma. *Nat Clin Pract Oncol.* 2008;**5**(11):634–644.

17. Schiavon G, Ruggiero A, Schöffski P, van der Holt B, Bekers DJ, Eechoute K, Vandecaveye V, Krestin GP, Verweij J, Sleijfer S, Mathijssen RH. Tumor volume as an alternative response measurement for imatinib treated GIST patients [published correction appears in PLoS One. 2013;8(1)]. *PLoS One.* 2012; **7**(11):e48372.

18. Lin M, Pellerin O, Bhagat N, Rao PP, Loffroy R, Ardon R, Mory B, Reyes DK, Geschwind JF. Quantitative and volumetric European Association for the Study of the Liver and Response Evaluation Criteria in Solid Tumors measurements: feasibility of a semiautomated software method to assess tumor response after transcatheter arterial chemoembolization. *J Vasc Interv Radiol.* 2012;**23**(12): 1629–1637.

19. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller JV, Pieper S, Kikinis R. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging.* 2012;**30**(9): 1323–1341.

20. Wadell H. Volume, shape, and roundness of quartz particles. *J Geol.* 1935;**43**(3):250–280.

21. Hayes SA, Pietanza MC, O'Driscoll D, Zheng J, Moskowitz CS, Kris MG, Ginsberg MS. Comparison of CT volumetric measurement with RECIST response in patients with lung cancer. *Eur J Radiol.* 2016; **85**(3):524–533.

22. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, van Oosterom AT, Christian MC, Gwyther SG. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst.* 2000; **92**(3):205–216.

23. Welsh JL, Bodeker K, Fallon E, Bhatia SK, Buatti JM, Cullen JJ. Comparison of response evaluation criteria in solid tumors with volumetric measurements for estimation of tumor burden in pancreatic adenocarcinoma and hepatocellular carcinoma. *Am J Surg.* 2012;**204**(5):580–585.

24. Dombi E, Ardern-Holmes SL, Babovic-Vuksanovic D, Barker FG, Connor S, Evans DG, Fisher MJ, Goutagny S, Harris GJ, Jaramillo D, Karajannis MA, Korf BR, Mautner V, Plotkin SR, Poussaint TY, Robertson K, Shih CS, Widemann BC; REiNS International Collaboration. Recommendations for imaging tumor response in neurofibromatosis clinical trials. *Neurology.* 2013;**81**(21 Suppl 1):S33–S40.

25. Pietanza M, James LP, Schwartz LH, Ginsberg MS, Zhao B, Moskowitz CS, Zheng J, Rizvi NA, Kris MG. Assessing changes in tumor size with CT scans in lung cancer: are volumetric measurements better than unidimensional (1D) and bidimensional (2D) assessments? *J Clin Oncol.* 2008;**26**(15 suppl): 14562–14562.

26. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep.* 1966;**19**(1): 3–11.

27. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med.* 1990;**20**(5):337–340.

28. Galanis E, Buckner JC, Maurer MJ, Sykora R, Castillo R, Ballman KV, Erickson BJ. 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**(2):156–165.

29. Shah GD, Kesari S, Xu R, Batchelor TT, O'Neill AM, Hochberg FH, Levy B, Bradshaw J, Wen PY. Comparison of linear and volumetric criteria in assessing tumor response in adult high-grade gliomas. *Neuro-oncol.* 2006;**8**(1):38–46.

30. Warren KE, Patronas N, Aikin AA, Albert PS, Balis FM. Comparison of one-, two-, and three-dimensional measurements of childhood brain tumors. *J Natl Cancer Inst.* 2001;**93**(18):1401–1405.