

# Comparison of Diameter and Perimeter Methods for Tumor Volume Calculation

By A. Gregory Sorensen, Shveta Patel, Carla Harmath, Sarah Bridges, Jennifer Synnott, Amy Sievers, Young-Ho Yoon, E. John Lee, Michael C. Yang, Robert F. Lewis, Gordon J. Harris, Michael Lev, Pamela W. Schaefer, Bradley R. Buchbinder, Glenn Barest, Kei Yamada, John Ponzio, H. Young Kwon, Joseph Gemmete, Jeff Farkas, Andrew L. Tievsky, Richard B. Ziegler, Megan R.C. Sallus, and Robert Weisskoff

**Purpose:** Lesion volume is often used as an end point in clinical trials of oncology therapy. We sought to compare the common method of using orthogonal diameters to estimate lesion volume (the diameter method) with a computer-assisted planimetric technique (the perimeter method).

**Methods:** Radiologists reviewed 825 magnetic resonance imaging studies from 219 patients with glioblastoma multiforme. Each study had lesion volume independently estimated via the diameter and perimeter methods. Cystic areas were subtracted out or excluded from the outlined lesion. Inter- and intrareader variability was measured by using multiple readings on 48 cases. Where serial studies were available in noncystic cases, a mock response analysis was used.

**Results:** The perimeter method had a reduced interreader and intrareader variability compared with the

diameter method (using SD of differences): intrareader, 1.76 mL v 7.38 mL ( $P < .001$ ); interreader, 2.51 mL v 9.07 mL ( $P < .001$ ) for perimeter and diameter results, respectively. Of the 121 noncystic cases, 23 had serial data. In six (26.1%) of those 23, a classification difference occurred when the perimeter method was used versus the diameter method.

**Conclusion:** Variability of measurements was reduced with the computer-assisted perimeter method compared with the diameter method, which suggests that changes in volume can be detected more accurately with the perimeter method. The differences between these techniques seem large enough to have an impact on grading the response to therapy.

*J Clin Oncol* 19:551-557. © 2001 by American Society of Clinical Oncology.

IMAGING HAS BEEN used to monitor therapeutic end points since its inception, often informally. Recently, the United States Food and Drug Administration, in Section 112 of the Food and Drug Administration Modernization Act of 1997, indicated that, in certain instances, end points other than survival might be used to formally judge therapeutic efficacy. This legislation has spurred an interest in determining which types of end points might be reproducible and justifiable.

A common oncologic evaluation approach for lesions visible on cross-sectional imaging, such as magnetic resonance imaging (MRI) or x-ray computed tomography, is to choose the image or images on which the lesion seems to have the maximum size and then measure the orthogonal diameters of the lesion. One, two, or three diameters are then used for comparison. This approach typically makes some simplifying assumptions, such as that the lesion is solid and/or that an ellipsoid is a reasonable representation of the lesion. This model can be extended for cystic lesions by estimating the volume of the cystic component(s) in the same way but subtracting the cystic ellipsoid volume(s) from the overall lesion volume.<sup>1</sup> Although this method may suffer from inaccuracies due to its assumptions, it is widely used because of its simplicity.

Planimetry has also been used to estimate lesion volume. Various methods have been used, including manual thresholding,<sup>2</sup> semiautomated thresholding based on the average

of seed regions from two tissue types (such as gray matter and white matter or CSF and brain),<sup>3</sup> histogram-based best-fit semiautomated methods based on discriminant function analysis of seed regions,<sup>4</sup> and semiautomated thresholding based on lookup values.<sup>5</sup> Segmentation has also been based on edge detection between tissue types rather than thresholding.<sup>6,7</sup> In the planimetry approach, the lesion is outlined on each image in which it appears.<sup>8-10</sup> The area of the lesion is then multiplied by the slice thickness to compute a volume, and the volumes of each image are summed to compute a total volume.<sup>11,12</sup> In this approach, each voxel in which the lesion is judged to appear is included. This has the advantage that each voxel can be either included or excluded according to whatever criteria might be established, making (for example) the exclusion of cystic or necrotic areas straightforward.

---

From the MGH NMR Center and Neuroradiology Division, Department of Radiology, Massachusetts General Hospital, Charlestown, MA.  
Submitted March 13, 2000; accepted August 14, 2000.

Supported in part by Novartis.

Address reprint requests to Gregory Sorensen, MD, MGH NMR Center, 149 13th St, Charlestown, MA 02129; email sorensen@nmr.mgh.harvard.edu.

© 2001 by American Society of Clinical Oncology.  
0732-183X/01/1902-551

Because of the recent increased interest in the establishment of verifiable end points, we sought to compare a planimetric approach with the diameter approach in the clinical trial context. Because no true gold standard is available in this setting, we chose to measure the internal consistency of the two methods as well as the difference between them on individual lesions. We further sought to determine whether any differences that we might detect were meaningful in a clinical setting by performing a mock response analysis, in which we compared the volume changes over time measured by one method with those measured by the other method. We chose glioblastoma cases because the lesions tend to be irregular in shape and often include cystic areas, which would stress the assumptions of the routine diameter approach. Additionally, a glioblastoma clinical trial was chosen because it was available and offered a sufficiently large database for study.

## METHODS

### Patient Database

Eight hundred twenty-five MRI studies in 219 patients formed the database for study. These MRI studies were part of a multicenter, randomized, controlled, phase III clinical trial of a novel biologic therapy in glioblastoma multiforme. Of these, 207 were preoperative studies and 618 were postoperative. The diagnosis was confirmed pathologically in each patient. All participants in this study remained blinded as to the arms to which any patients had been assigned. Each MRI study was performed according to a standardized protocol that included pre- and postcontrast T1-weighted imaging in multiple planes, as well as a high-resolution (3- or 1.5-mm slice thickness) gradient echo acquisition. Clinical data, such as patient survival, treatment parameters, and so on, are not described in this report because they are not relevant to the main focus of this work, the reproducibility of the measurement techniques.

### Data Acquisition

**General.** All studies were reviewed by board-certified radiologists with fellowship training and a minimum of 2 years' experience in neuroradiology. Fourteen different radiologists interpreted the 825 studies; each radiologist read an average of 60 studies (range, one to 247). Each study underwent two types of volume calculation, described below as the diameter method and the perimeter method. The readers were asked to include the volume of abnormal contrast enhancement seen on the MRI study and were aware that a comparison would be made between the two techniques. Two different radiologists performed readings at different reading sessions to ensure that the interpretations were independent. In no case did the same reader perform both analysis techniques on the same case, and in no case did the reader have available any data or volume information from the other evaluation method.

Interreader studies were performed on 48 studies. Interreader studies were performed by having a second reader analyze each of the 48 studies without knowledge of the first reader's results. Intra-reader studies were performed on 48 studies. This was accomplished by having the same reader return and reread the same study but with

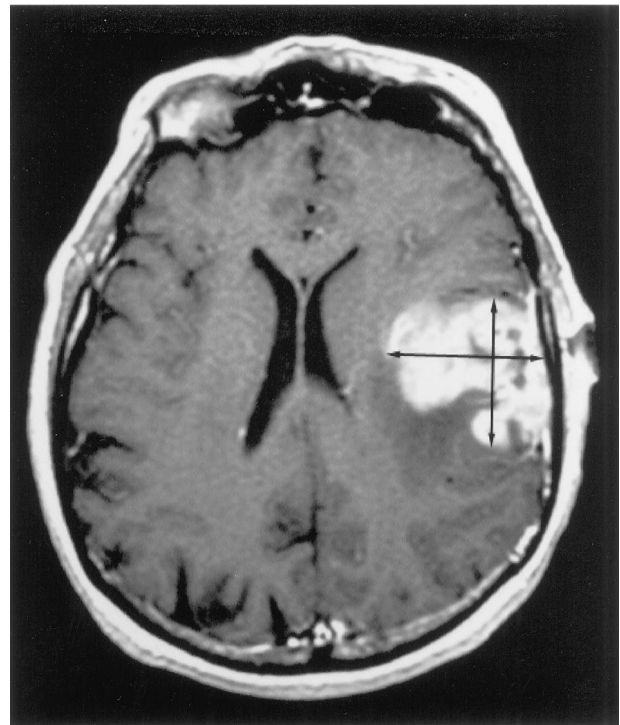


Fig 1. Diameter method uses calipers (dividers) for orthogonal diameter lesion measurements.

sufficient time between readings to decrease the likelihood of remembering the first analysis (average time, 13 weeks; minimum time, 11 days). In both approaches, the studies were randomly selected and the readers were assisted by a technologist who prepared the MRI studies for presentation and recorded the measurements. Nonenhancing hemorrhagic components were treated as cystic components and were excluded from the calculations. Scale bars included on the films were used for calibration, and image thickness and interimage gaps were recorded. All data were stored in a relational database management system (FileMaker Pro; Claris, Cupertino, CA).

**Diameter method.** Readers were instructed to identify the three largest orthogonal diameters of the lesion. Diameters were manually measured on MRI films with calipers (dividers) (Fig 1). In each case where a second lesion was present, or the shape of the lesion was best characterized by two ellipsoids, a second set of three diameters was also recorded, and the volumes were summed. If one or more necrotic or cystic areas were thought to be present, additional diameters for the cystic component were recorded, and the computed cystic volume was subtracted from the overall volume. Readers were not aware that an end point of this study was the determination of how many sets of diameters (one v two v three diameter measurements) were thought to be required to accurately characterize the lesion volume. The formula used to compute volumes was the standard volume of an ellipsoid, as follows:

$$V = 4/3 \pi (a * b * c) \quad (1)$$

where  $a$ ,  $b$ , and  $c$  are the three radii (half the diameters).

In addition to the total volume, the individual diameters were also recorded to allow analysis on a single- or dual-diameter basis, ie, diameter or area rather than a volume estimate.

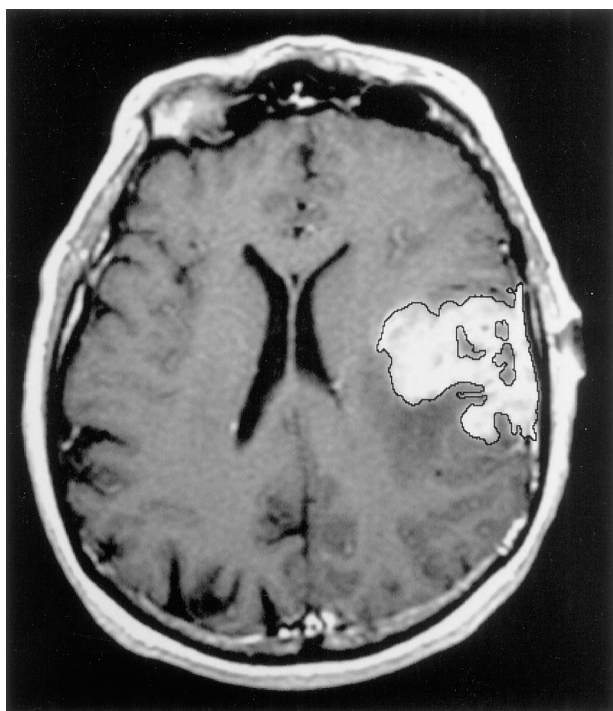


Fig 2. Perimeter method with lesion outlined by using Alice software program.

*Perimeter method.* All MRI films were digitized at 150 dots per inch using a film scanner (model VXR-12; Vidar Systems, Herndon, VA). Commercial software (Photoshop 3.0.5; Adobe Systems, Inc., Mountain View, CA) was used to acquire data from the film scanner, and images were stored in 8-bit TIFF (tag image file format) on a personal computer (Power Macintosh 7,300/132; Apple Computer, Cupertino, CA). Planimetry was carried out in a multistep process using additional commercial software (Alice; Hayden Image Processing Group/Parexel International Corp, Waltham, MA). A technologist using a variety of outlining tools in the Alice software program outlined the lesion. The software's built-in edge-detecting algorithm was supplemented with a custom-developed plug-in module. The technologist could use either or both of these techniques, followed by manual editing (Fig 2). This technologist-produced outline was then reviewed with a physician reader who modified the outline as necessary. When the physician reader was satisfied with the outlines, the software program calculated the volume by counting the number of voxels and multiplying the count by the volume of a single voxel. The same staff were used in each arm of the study. The physician readers for the perimeter method were from the same group as the readers for the diameter method, but they did not read both perimeter and diameter methods for the same patient. For inter- and intrareader lesion outlining, the repeat outlining was performed either by a separate technologist without knowledge of the previous outline or by the same technologist but delayed by 13 weeks. It took the average technician approximately 1 day to learn to use the outlining software and 3 days of shadowing by a more experienced technician to consolidate the knowledge gained. The technicians were computer literate and had a basic exposure to human anatomy before training.

### Data Analysis

The results of the diameter and perimeter methods were compared by linear correlation analysis. Intrareader variability and interreader variability were analyzed identically for both diameter and perimeter methods. For each of the four data cohorts (interreader diameter, interreader perimeter, intrareader diameter, and intrareader perimeter), the first volume was plotted against the second volume. The mean difference between the two volumes for each MRI study and the SD of the difference were also computed. The similarities for SDs were compared using an *f* test. We compared SDs rather than means because SDs capture the variability of the techniques better. For example, in the case in which one technique was markedly more variable than another, the average difference between the two methods would regress toward zero as the number of cases increased. However, the SD would provide an estimate of how frequently this mean was missed and by how much.

An estimate of how the perimeter and diameter methods might affect clinical response outcomes was derived by performing a mock response analysis as follows. Patients for whom it was thought a single set of diameters was adequate to measure lesion volume (necrotic, single lesion) and for whom more than one MRI study was available were selected for further analysis. For each patient, each study was compared with the other available serial studies to determine the clinical response. The decision algorithm compares two studies and classifies the difference between them as indicating a complete response (100% regression or absence of the lesion), partial response (50% to 99% decrease), or nonresponse (less than 50% decrease or an increase in lesion volume). The clinical response was determined for both the diameter and the perimeter methods. Patients were scored as indicating a discrepancy between the diameter method and the perimeter method if any pairing of a patient's studies could produce a discrepancy. This same mock response analysis was repeated with substitution of a single largest diameter and the calculation of a cross-sectional area by using two diameters and the formula for the area of an ellipse,

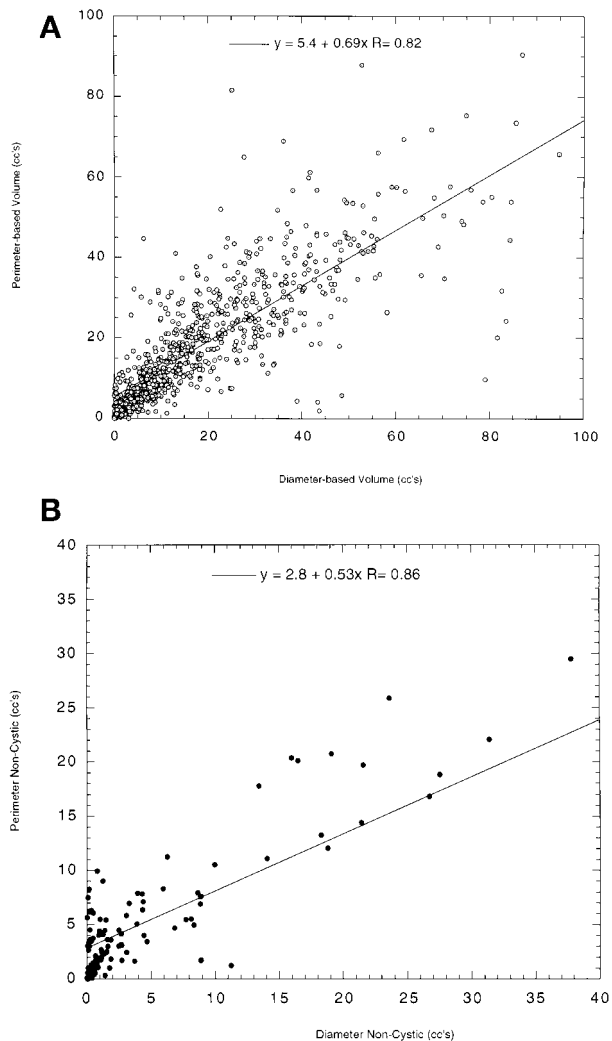
$$\text{Area} = \pi * a * b \quad (2)$$

where *a* and *b* are the two radii.

### RESULTS

Seven hundred four of the 825 MRI studies evaluated were judged to require more than one set of three diameters to adequately characterize lesion volume. Four hundred ninety-two of the 704 were judged to have a cystic component, 47 were judged to require a second set of diameters to characterize an additional solid component, and 165 were judged to have both a cystic component and an additional solid component.

Figure 3A compares the volumes determined by the diameter method with the volumes determined by the perimeter method. Fitting a curve to this data indicates that the slope is 0.69, with a *y* intercept of 5.4 mL (*r* = .82). This indicates that when the diameter method indicates a near-zero volume, the perimeter method still finds a substantial volume. The slope of less than 1 indicates a trend for the diameter-based volume to be smaller than the perimeter-based volume at larger volumes. Figure 3B compares volumes determined by the diameter method with the



**Fig 3.** (A) Plot of diameter-based volumes versus perimeter-based volumes for all 825 MRI studies. (B) Plot of diameter-based volumes versus perimeter-based volumes for the 121 solid, noncystic lesions.

volumes determined by the perimeter method using only solid, noncystic lesions. The same results are apparent, with a slope of 0.53 and a y intercept of 2.8 mL ( $r = .86$ ).

Figures 4A through 4D plot the inter- and intrareader results for the diameter and perimeter methods. The mean and SDs for each data cohort are listed in Table 1. Both the intrareader variability and interreader variability were lower with the perimeter method than with the diameter method, and the differences for both were statistically significant by *f* test. Figures 4E and 4F show the agreement between the methods for both the inter- and the intrareader variability, by plotting the difference of the mean against the mean for each technique.

Among the 121 MRI studies in which a single set of diameters was judged to be adequate to characterize lesion volumes, 23 patients had more than one MRI study. Results of the clinical responses in the mock response analysis are listed in Table 2. There was a discrepancy in whether disease progression was present as determined by the perimeter method compared with the diameter method in six (26%) of 23 patients.

## DISCUSSION

Our results indicate that planimetry provides an improved measure of lesion volume compared with simple diameter measurements. Logically, planimetry should be a more accurate approach than diameter measurements because simplifying assumptions do not need to be made. However, because there is no way to assess the true volume of enhancing tissue and to use that volume as a gold standard, in a clinical trial setting it is impossible to say with certainty that the lesion volume with one technique is closer to the true volume than with another technique. This lack of a verifiable standard can be challenging and can lead to uncertainty about optimizing approaches to reduce measurement variance. Nevertheless, the reduced variability evident in the perimeter method, as well as the inherent attractiveness of being able to include and independently choose whether each voxel belongs in the lesion, makes computer-assisted planimetry likely to be one of the more accurate techniques available.

The assumption that a solid ellipsoid shape will accurately characterize lesions such as glioblastoma multiforme seems to be contradicted by our data, in which more than 85% of the lesions were judged to require more than a single set of diameters. Typically, this was due to a cystic component within the lesion. Brain tumors are sometimes classified as solid lesions by some tumor classification schemes,<sup>13</sup> but particularly after surgical therapy this is not the case. Even when such cystic lesions were excluded, there was still marked variation between the perimeter and diameter methods. This suggests that even in solid lesions, using the assumption that an ellipsoid accurately characterizes the lesion increases the variance of the measurement.

The cause for the increased variance in the ellipsoid technique most likely originates from the complex shape of the lesions. When the diameter method is used, the complexity of the lesion shape allows each reader an opportunity to vary where the diameters are chosen and measured. In the perimeter method, decisions about the location of the lesion's boundaries seem to be more straightforward. Although a boundary decision may vary by a few voxels in any given location from one reader to another using the perimeter method, most of the time the readers seemed to be

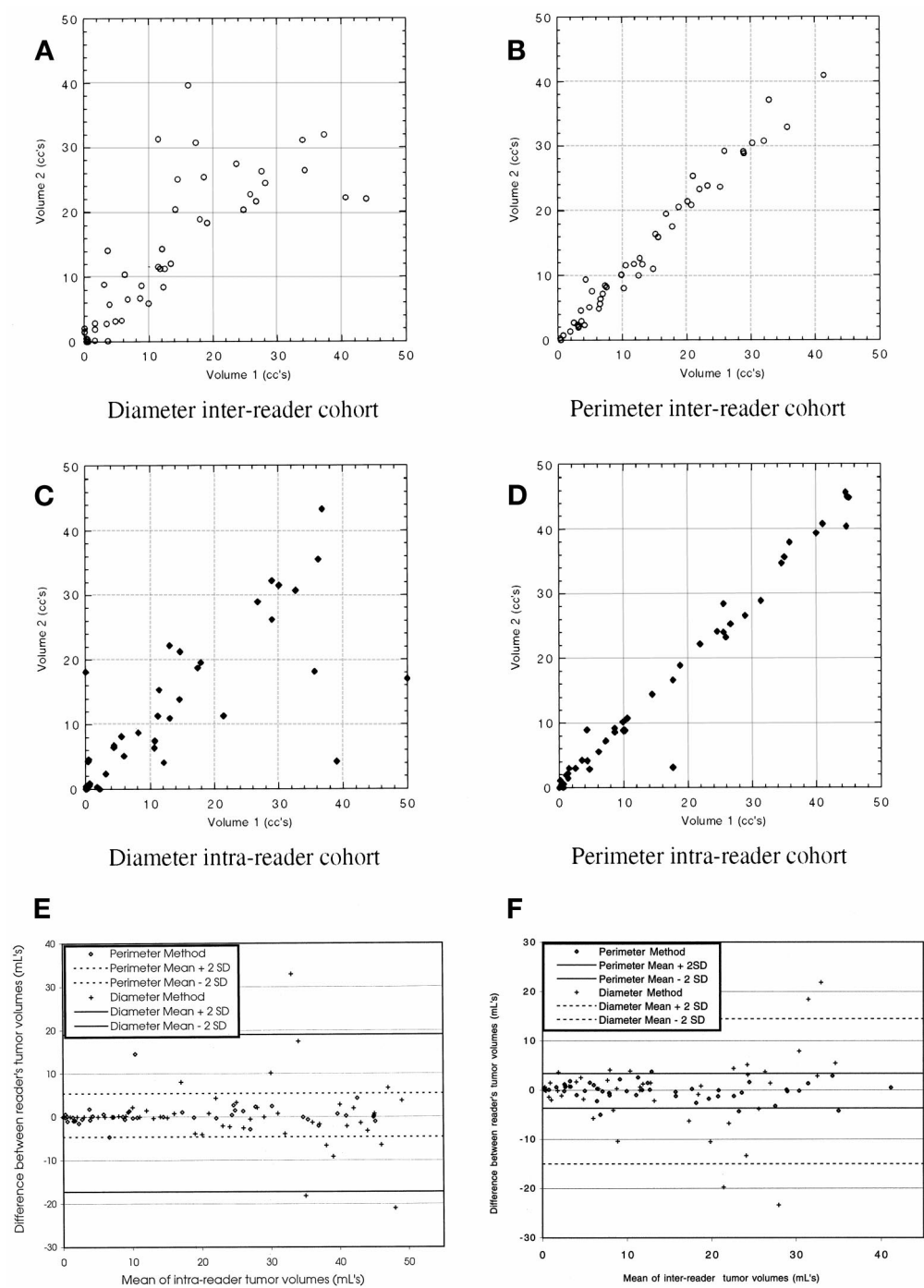


Fig 4. (A-D) Plot of the inter- and intrareader results in 48 cases. (E) Plots of the difference against the mean for intrareader tumor volumes in 48 cases. (F) Plots of the difference against the mean for interreader tumor volumes using the perimeter method in 48 cases.

able to agree on the boundaries, whereas the different volumes with the diameter method indicate that the readers often did not agree on the same diameters. Reader variability might be reduced if the diameter directions were standardized (eg, right-left, superior-inferior, anterior-posterior)

rather than allowing the largest three orthogonal diameters, but this may well introduce another source of variability (the rotation of the axes of the ellipsoid).

Comparing the diameter method with the perimeter method indicates that there is not a simple relationship to

**Table 1. Comparison of Mean and SD of Inter- and IntraReader Variability for both Diameter and Perimeter Methods**

	Mean	SD
InterDiameter	-0.31	7.38
InterPerimeter	-0.19	1.76
IntraDiameter	0.91	9.07
IntraPerimeter	0.39	2.51

account for the different volumes produced by the two methods. The relatively large  $y$  intercept and the slope of less than 1 suggest that at small values, the estimation given by the diameter method (ellipsoid) is smaller than that produced by the perimeter method. If we assume that the planimetry results are closer to truth, perhaps this is due to lesions with shapes that are highly irregular.

When the diameter methodology is assessed by a clinical response model, the differences between the perimeter and diameter methods are perhaps even more striking. The change in outcome classification in 26% of patients may imply that calculations of treatment efficacy based on lesion shrinkage may need to include this additional source of variance. Alternatively, by using a perimeter method rather than a diameter method, measurement variance might be reduced enough to detect changes with a smaller patient sample size. Outcome classification is based on a change in measurements; because all measurements include errors or noise, a misclassification may be due to the observed change being a noise effect rather than a true effect. Because the perimeter method has less variance (or noise) than the diameter method, it logically follows that the rate of misclassification due to measurement noise would also be lower. This is not simply a matter of the cystic or postoperative nature of the lesions invalidating the assumption of ellipsoid shape; even when restricted to the solid lesions, enough asymmetry and variability in lesion geometry is present to allow changes in classification of disease progression (or nonprogression) when using the diameter method as compared with the perimeter method. Presumably, cases with cystic and/or multiple lesions would have even more variability with the diameter method and thus more misclassification of treatment outcomes.

**Table 2. Results of the Clinical Response in the Mock Response Analysis**

Method of Comparison	Discrepancy (%)	No. of Patients Showing Difference (of 23)
Diameter v perimeter	26	6
Single diameter v perimeter	22	5
Diameter area v perimeter	13	3

Interestingly, in our experience the effort required to actually measure these lesions with the perimeter method is less than the effort required to collect, digitize, and archive the scans in preparation for either form of measurement. This suggests that if central image archiving or review is to be performed as part of a clinical trial, the additional costs associated with lesion quantification with the perimeter method might be well worth bearing. The amount of time the radiologist actually spent reading when using either the perimeter method or the diameter method was approximately the same at 7 minutes per study. The perimeter method required approximately 20 minutes of technician time to digitize the films and prepare the outlines for the physician reader.

With the trend toward radiology PACS (picture archiving and communication systems) and digital image transfer, the process necessary before actual lesion measurement will become easier and less time-consuming. Continuing development in computer-assisted edge-detection software, and the possibility of having such software usable within the PACS systems, will decrease or eliminate the difference between the overall time necessary for the perimeter method and the diameter method, which will make the perimeter method even more attractive.

One possible limitation of our study is that the volumes used were measured from enhancing tissue. This is known in some cases to include tissue that is not truly malignant or even neoplastic but might be inflammatory in nature. This measurement of all enhancing tissue, however, should not affect the intra- or interreader variability. Indeed, it is common in imaging that one is uncertain as to the true pathologic correlate of imaging findings such as breakdown of the blood-brain barrier or tissue enhancement, but this does not preclude measurement and quantification of these findings.

In conclusion, our data suggest that using a planimetry-based method for lesion quantification reduces reader variability and improves measurement reproducibility. We speculate that this planimetry method therefore produces a more accurate representation of lesion volume than traditional diameter methods. Furthermore, the difference in volumes produced by the perimeter method and the diameter method is of sufficient magnitude to potentially lead to different outcomes in clinical situations. This implies that greater power for detecting lesion changes and therapeutic efficacy may be available by using the perimeter method for tumor volume calculation.

#### ACKNOWLEDGMENT

We thank Tara Ullrich for her assistance.

## REFERENCES

1. Clarke LP, Velthuizen RP, Clark M, et al: MRI measurement of brain tumor response: Comparison of visual metric and automatic segmentation. *Magn Reson Imaging* 16:271-279, 1998
2. Lim KO, Pfefferbaum A: Segmentation of MR brain images into cerebrospinal fluid spaces, white and gray matter. *J Comput Assist Tomogr* 13:588-593, 1989
3. Jernigan TL, Press GA, Hesselink JR: Methods for measuring brain morphologic features on magnetic resonance images: Validation and normal aging. *Arch Neurol* 47:27-32, 1990
4. Cohen G, Andreasen NC, Alliger R, et al: Segmentation techniques for the classification of brain tissue using magnetic resonance imaging. *Psychiatry Res* 45:33-51, 1992
5. Shenton ME, Kikinis R, McCarley RW, et al: Application of automated MRI volumetric measurement techniques to the ventricular system in schizophrenics and normal controls. *Schizophr Res* 5:103-113, 1991
6. Filipek PA, Kennedy DN, Caviness VS, et al: Magnetic resonance imaging-based brain morphometry: Development and application to normal subjects. *Ann Neurol* 25:61-67, 1989
7. Kennedy DN, Filipek PA, Caviness VS: Automatic segmentation and volumetric calculations in nuclear magnetic resonance imaging. *IEEE Trans Med Imaging* 8:1-7, 1989
8. Pass HI, Temeck BK, Kranda K, et al: Preoperative tumor volume is associated with outcome in malignant pleural mesothelioma. *J Thorac Cardiovasc Surg* 115:310-318, 1998
9. Keller JM, Edwards FM, Rundle R: Automatic outlining of regions on CT scans. *J Comput Assist Tomogr* 5:240-245, 1981
10. Lai MTY, Suen CY: Contour tracing and encodings of binary pattern, in Simon JC, Haralick RM (eds): *Digital Image Processing*. Hingham, PA, D. Reidel, 1981, pp 149-153
11. Harris GJ, Rhew EH, Noga T, et al: User-friendly method for rapid brain and CSF volume calculation using transaxial MRI images. *Psychiatry Res* 40:61-68, 1991
12. Harris GJ, Barta PE, Peng LW, et al: MR volume segmentation of gray matter and white matter using manual thresholding: Dependence on image brightness. *AJNR Am J Neuroradiol* 15:225-230, 1994
13. World Health Organization: *Handbook for reporting results of cancer treatment*. Geneva, Switzerland, World Health Organization, 1979