Brain Tumor Imaging in Clinical Trials

J.W. Henson, S. Ulmer and G.J. Harris

AJNR Am J Neuroradiol 2008, 29 (3) 419-424
doi: https://doi.org/10.3174/ajnr.A0963
http://www.ajnr.org/content/29/3/419

This information is current as of September 16, 2023.
SUMMARY: There are substantial challenges in the radiologic evaluation of tumor size during clinical trials, and it is important for neuroradiologists to have a firm understanding of these issues. This review will examine measurement approaches, response criteria, selection of lesions for measurement, technical imaging considerations, interval between tumor measurements and response confirmation, and validity of imaging as a measure of efficacy.

From the Stephen E. and Catherine Pappas Brain Tumor Imaging Research Program (J.W.H., S.U.), Division of Neuroradiology (J.W.H.), 3D Imaging Service (J.W.H., G.J.H.), Department of Radiology (J.W.H., G.J.H.), and Neurology Service (J.W.H.), Massachusetts General Hospital, Boston, Mass; and Harvard Medical School (J.W.H.), Boston, Mass.

This work was supported by the Pappas Brain Tumor Imaging Research Program, Massachusetts General Hospital, and a grant form the National Cancer Institute to the Dana-Farber/Harvard Cancer Center, which supports Dr. Harris as Principal Investigator of the Tumor Imaging Metrics Core.

Please address correspondence to John W. Henson, MD, Stephen E. and Catherine Pappas Brain Tumor Imaging Research Program, Massachusetts General Hospital, 55 Fruit St, Yawkey Nine East, Boston, MA 02114; e-mail: henson@helix.mgh.harvard.edu

DOI 10.3174/ajnr.A0963

E xtraordinary effort has been invested in clinical trials for malignant gliomas and brain metastases during the past 30 years. Phase I and phase III studies, with their respective goals of defining maximal tolerated dose and overall survival, do not rely on neuroimaging as a primary end point. Imaging is crucial in phase II studies, however, because radiographic response, in combination with clinical status, is used to assess therapeutic effect. Phase II studies are usually conducted in patients with progressive tumors, and serial imaging examinations are performed after initiation of treatment and compared with a baseline pretreatment study. Radiographic response for each patient is then assigned according to predetermined criteria.

Neuroradiologists should have a firm understanding of the issues involved in imaging brain tumors in clinical trials, not only with respect to review of quantitative imaging but also to assist in the optimal use of imaging in trial design.

Measurement Techniques
Two major approaches are available for the evaluation of contrast-enhancing tumor size on serial imaging studies during clinical trials: 1) diameter-based measurement on a single-axial section containing the largest diameter of tumor, and 2) computer-assisted volumetric analysis of all sections containing tumor.

Diameter Methods
Diameter-based approaches involve the measurement of either a single largest diameter (1D or Response Evaluation Criteria in Solid Tumors [RECIST; http://ctep.cancer.gov/guidelines/recist.html]) or 2 orthogonal diameters (2D or Macdonald criteria) on a single axial section (Fig 1).2–4

RECIST (1D). Imaging response guidelines called RECIST were published in 2000 to replace the 1981 World Health Organization (WHO) guidelines.5–3 The new criteria were designed as a uniform, simplified, and conservative standard to determine response to therapy for solid tumors. RECIST has not been widely adopted for clinical trials involving brain tumors, mainly because of the historic use of 2D criteria (see below).

RECIST measures the longest single linear enhancing diameter across a lesion in the axial plane (Fig 1). This measurement is repeated with each study, always by using the longest diameter, even if it varies from the original orientation or section. Minimal measurable lesion diameter is 10 mm or at least 2 times the imaging section thickness, to reduce the variability due to volume averaging and variations in section selection between studies. Cystic or necrotic foci of tumor and leptomeningeal lesions are considered nonmeasurable according to the written criteria. When multiple lesions are analyzed, the individual diameter measurements are recorded separately and then summed for response evaluation. RECIST response criteria are described below (see “Response Criteria”).

Macdonald Criteria (2D). Most trials for patients with malignant gliomas use the WHO-based “Macdonald criteria.”4 A measurement is made of the maximal enhancing tumor diameter on a single axial gadolinium-enhanced T1-weighted section, and then the largest perpendicular diameter is measured on the same image (Fig 1). The product of the 2 diameters is calculated, and the measurements are repeated with each scan. Measurements from multiple lesions are summed. The Macdonald criteria do not specifically address the presence of necrotic portions of lesions. Macdonald response criteria are discussed below.

Computer-Aided Volumetric Methods
In the computer-aided volumetric or perimeter approach, semiautomated tumor-segmentation software is used to determine the tumor volume. The computer generates a border between the enhancing and nonenhancing regions on all axial sections that contain enhancing tumor (Fig 1), by using a combination of image processing techniques such as histogram statistics and morphologic filtering. These images are reviewed, the perimeters are adjusted by a neuroradiologist, and the program calculates an enhancing volume, a nonenhancing volume (ie, the centrally necrotic or cystic portion), and total or combined lesion volume in cubic centimeters by adding the lesion volumes from the stack of images.

Response Criteria
Once serial studies become available for lesion measurement during a clinical trial, the following terms are used to describe responses. The specific numeric values for each of these categories are given in the Table.

Complete Response. Complete response (CR) is the disappearance of all enhancing tumor, including measurable and nonmeasurable areas. CR should be confirmed with a repeat study at 4 weeks.
Partial Response. Partial response (PR) is a specified percentage decrease in the tumor measurement compared with the baseline scan (Table). PR should be confirmed with a repeat study at 4 weeks.

Stable Disease. Stable disease (SD) includes changes that do not meet criteria for CR, PR, or progressive disease (PD).

Progressive Disease. PD is a specified percentage increase in tumor measurement compared with the lesion size that defines the nadir, or smallest measurement, in the serial studies. PD should be confirmed with a repeat study at 4 weeks. Any unequivocally new lesion is also a criterion for PD. Criteria for new lesions should be defined by protocol.
**Best Response.** Best response is the largest reduction in tumor measurement during the study (compared with the initial response). Best response may differ from initial response in patients with PR followed by a CR.

With diameter approaches, CR and PR were seen in approximately 10% of patients in clinical trials for newly diagnosed malignant gliomas and in 1%-10% of patients with progressive disease. The criteria for measurement was not well delineated in these studies. Given this low response rate and the fact that cytostatic therapies would not be expected to produce tumor shrinkage, some studies used the duration of SD as a measure of response, by means of either time to progression (TTP) or progression-free survival (PFS).

**TTP.** TTP is the interval between the treatment start date and a subsequent imaging study that shows PD.

**PFS.** PFS is the percentage of patients who have not experienced PD at a specified time point after beginning treatment (eg, 2-month PFS or 6-month PFS).

A common end point for phase II clinical trials is 6-month PFS. Phase II studies compare the percentage of patients with PFS to that of an historical control group. There are significant issues with the use of PFS in a phase II study, however, because it is difficult to identify an appropriate control group or to determine what constitutes a meaningful improvement over the historical comparison group. Most studies seek to demonstrate a response rate (CR + PR) of 20% or a 20% improvement of a measure of response over a historical comparison population.

To our knowledge, correlations between quantitative parameters of response criteria between 1D (RECIST), 2D (Macdonald), and a volumetric approach have not been closely examined. The Table provides a theoretic comparison based on an idealized spheric tumor. Some investigators have used a straight percentage-change approach, such that 50% reduction is used for volume (3D) as in the area (2D or Macdonald) criteria. Another approach would be to use mathematical extrapolation between linear and volumetric measurements, such that a 50% decrease in area (2D) would be equal to a 65% decrease in volume (3D) for a PR. These issues can have an impact on the stringency of criteria for PR and PD. Further research regarding the range of deviation of tumors from the idealized spheric shape will be necessary to determine the best values for most accurate response rate comparisons, and validation of each approach will be important. Nonetheless, several retrospective studies have compared these measurement approaches.

**Comparison of Diameter and Volumetric Approaches**

Diameter approaches are simple to perform and have reasonably low inter-reader variation. However, the use of a linear measurement on a single-axial image ignores the tendency of malignant gliomas to be highly irregular in shape, to progress in a pattern of eccentric nodular growth, and to have cystic and central necrotic areas that are unlikely to be affected by nonsurgical treatment.

Volumetric approaches analyze the entire tumor by using multiple sections and have the ability to exclude nonenhancing areas. Measurement variation is low. In many patients with malignant gliomas, irregular peripheral enhancement and extension along the subependymal regions can make it very challenging to outline an enhancing volume. Two other issues with the volumetric approach are investment of time and the need for specialized technical skills and computer software.

At least 4 retrospective studies have compared diameter and volumetric approaches for the detection of response in neuro-oncology clinical trials. Three of these articles concluded that there was no difference in the detection of response, as defined either by reduction in tumor size by a specified amount or by TTP (see the discussion of response criteria, below). Sorenson et al found that computer-aided volumetric analysis was more sensitive in the early detection of progression, especially with smaller lesions, and, in a subset analysis, concluded that a different response determination was indicated by volumetrics compared with a diameter method in one quarter of cases. In this article, however, the comparison was between volumetric analysis and a 3-dimension measurement approach, and for that reason, the data are difficult to compare with the other 3 studies that compared RECIST and Macdonald approaches. In the 3 other reports, the diameter measurements were made on 5-mm, skip 1-mm axial images (when specified), and the criteria for lesion inclusion were not specified in terms of minimal diameter and whether the lesion could be centrally necrotic. Nonetheless, the computer-assisted volumetric approach yielded a higher response rate than did the single linear (RECIST) measurements in the large retrospective study by Shah et al. Specifically, PR would have been declared in 8% of patients (n = 284 studies) by using the 1D measurement compared with 17% PR rate by using the volumetric approach. The authors did not address the statistical significance of this finding. The studies did not find differences in measures of TTP or PFS. The lack of significant difference between techniques may reflect the inclusion of lesions of substantial nonenhancing components, thereby decreasing the sensitivity for response in the enhancing component.

**Choosing a Measurable Lesion in Clinical Trials**

Definition of a measurable lesion is of major importance in those clinical trials that have an imaging end point. As noted previously, there are many factors that make some lesions ideal for serial measurement and render others less valuable. At present, most clinical trial protocols do not specify criteria for lesion selection.

**Enhancement**

Malignant gliomas are histopathologically and radiographically heterogeneous in appearance, with geographically irregular margins, variable enhancement, regions of central necrotic or cystic change, and surrounding T2-weighted hyperintensity (Fig 1). Brain metastases are typically more spheric and lend themselves more easily to analysis. Contrast enhancement provides the best currently available measure of tumor size. A component of the lesion that is as homogeneously enhancing as possible should be selected for response assessment because cystic and necrotic portions of a lesion are unlikely to respond to interventions other than surgery. RECIST specifically excludes measurement of cystic lesions in systemic cancers, and this criterion should be adhered to as much as possible with malignant gliomas and metastases.
Central necrotic or cystic areas within a region of enhancement are less of an issue with computer-assisted volumetric assessment (Fig 1, and see below) because the computer software can segment and quantify enhancing and nonenhancing components. In a malignant glioma with irregular nodular enhancement, it may be better to choose the largest area of solidly enhancing tumor for measurement if the margins of the nodular focus can be clearly defined, rather than to try to outline the entire margin of the lesion. Protocols should define the maximal percentage of an enhancing focus that can be nonenhancing, as assessed visually on the section containing the largest lesion diameter (eg, 80% for volumetric studies).

**Minimal Diameter for Inclusion of Measurable Lesions**

Specification of a minimal lesion diameter for measurable lesions aims to reduce the potential for variation in the measurement of smaller lesions due to section selection and volume averaging. In accordance with RECIST, the minimal lesion diameter should be greater than or equal to 2 times the section thickness. With a routine T1-weighted image in which the sections are 5 mm, skip 1 mm, the minimal diameter would be 12 mm. The use of 3-mm, skip 0-mm gadolinium-enhanced T1-weighted images and specification of a 10-mm minimal lesion diameter would be more in keeping with RECIST criteria.

**Multicentric Lesions**

Approximately one third of malignant gliomas are multicentric at the time of diagnosis, and in half of these cases, there are discrete foci of enhancement. The approach in this situation is to measure and record each separately enhancing lesion that meets inclusion criteria and the sum the measurements.

**Nonmeasurable Lesions**

The concept of nonmeasurable lesions at baseline can be important in clinical trials because tumor progression may occur in these sites. A nonmeasurable lesion might include foci of enhancement that are less than the specified smallest diameter (Fig 1), the region of T2-weighted hyperintensity surrounding enhancing tumor, discrete foci of nonenhancing T2-weighted hyperintensity that represent multicentric tumor, hemorrhagic or predominantly cystic or necrotic lesions, and leptomeningeal tumor. Hemorrhagic lesions often have intrinsic T1-weighted hyperintensity that could be misinterpreted as enhancing tumor, and for this reason, the precontrast T1-weighted image must be examined at baseline to prevent this error. Notes regarding nonmeasurable lesions should be recorded on the imaging case report form for each study.

**Postoperative Changes**

Many patients entering clinical trials have undergone recent resection of tumor, whether at initial diagnosis or at the time of progression. This introduces a special set of considerations in choice of region for measurement. Gadolinium-enhanced MR imaging should be performed within the first 72 hours following surgery to minimize postoperative enhancement along the margin of the surgical cavity.

Hyperintense T1-weighted postoperative blood products along the surgical margin can make it difficult to determine whether there is residual enhancing disease. Surgical cavities should not be included in tumor measurement, but it is often difficult to determine the edge of the operative margin. Comparison of T1-weighted images before and after gadolinium enhancement with fast spin-echo (FSE) T2-weighted images gives the most useful information in this regard.

Areas of restricted diffusion along the margin of the surgical cavity are present on diffusion-weighted images (DWI) in up to two thirds of patients on immediate postoperative scans following surgery for malignant gliomas and usually represent areas of infarction (Fig 2). Enhancement of the infarcted tissue occurs in approximately one half of cases, can be seen within 1–2 weeks of surgery, and is a finding that can easily be misinterpreted as progressive tumor if the presence of the perioperative infarct is not realized. Thus, it is important to obtain and review the immediate postoperative DWI when new enhancement appears within 2–4 weeks of surgery.

Enhancing foci that have been treated with radiosurgery are usually excluded from measurement because of the possibility of radiation necrosis.

**Technical Considerations**

The same imaging technique must be used at every time point (ie, MR imaging and CT measurements must not be mixed and use of differing magnet strengths should be avoided). Measurements are typically made in the axial plane. The acquisition of 3-mm, skip 0-mm T1-weighted contrast-enhanced images would improve resolution, particularly when dealing with smaller enhancing nodules. Higher resolution images increase the acquisition time from 3 minutes to 5–6 minutes for 3-mm, skip 0-mm sections, but the improved resolution is a strong argument in favor of these images.

Postcontrast axial images should be acquired at a standardized time interval after gadolinium injection, starting no less than 5 minutes after injection. Other sequences not affected by gadolinium (eg, FSE T2-weighted images and DWI performed for a total of 6 minutes) can be routinely acquired after contrast injection to optimize scanner resources and to standardize the postcontrast interval. The same dose of gadolinium should be used for every study.

**Timing of Imaging Studies and Confirmation of Response**

The frequency of imaging should take into account the interval during which a response might be detected as well as an estimate of the likely duration of response. In the large retrospective study of Hess et al, the median time to response (CR + PR) in progressive malignant gliomas in patients was 14 weeks, with 74% of responses occurring by 26 weeks. A baseline study is usually required within 14 days before initiation of protocol treatment. In some cases, this will be the immediate postoperative examination or, in others, a routine follow-up study demonstrating PD on current therapy. Follow-up studies in patients with malignant gliomas are commonly performed after every second cycle of treatment, which is usually every 8 weeks.

Response confirmation at 4 weeks is a key component of both the RECIST and Macdonald criteria and is designed to avoid overestimating the response rate. When these data are not available, the patient does not have a confirmed response and this information should be reported in the results of clinical trials.
Validity of Imaging End Points in Clinical Trials

Radiographic response correlates with survival in newly diagnosed anaplastic oligodendroglioma, in which response, longer survival, and loss of heterozygosity of chromosome 1p are related. For newly diagnosed anaplastic astrocytomas, a study by the Radiation Therapy Oncology Group showed that patients with no progression at 6 months had a median survival of 67 months versus a median survival of 19 months for patients who had shown progression by 6 months. Thus, for anaplastic gliomas, there is evidence that response to initial treatment correlates with survival. A study by Galanis et al of the relationship between response and survival in patients with newly diagnosed “enhancing gliomas” (n = 36; with low-grade glioma = 25%, anaplastic astrocytomas = 14%, and glioblastoma = 61%) found no relationship between the 2 categories of response (PR and SD) and survival but did observe a correlation between PD and shorter survival. However, these were studies of adjuvant therapy soon after diagnosis, not salvage therapy at the time of progression, whereas most phase II trials are conducted in patients with PD in the salvage setting.

Radiographic response is often used as an end point in the phase II setting with the assumption that it is a valid surrogate measure for improved overall survival. Grant et al reviewed the imaging studies of 136 patients with progressive gliomas following 2 cycles of nitrosourea-based therapy and found no correlation between TTP and survival. However, this study also documented a relationship between PD and shorter survival. Shah et al found an association between 2-month PFS and survival but no relationship between 6-month PFS and overall survival in patients entered into a number of different clinical trials for progressive malignant gliomas. The latter fact suggests that 6 months may be too long an interval for response assessment in patients with progressive malignant gliomas. Two other large retrospective studies have demonstrated a positive correlation between response, PFS, and overall survival in patients with progressive high-grade gliomas. Overall, the data support the validity of response as a predictor of survival. Any measurement approach that is more sensitive to response would likely provide the best measure of effect.

Challenges with Novel Therapeutics

This review has focused on the use of gadolinium enhancement as the most useful characteristic for the measurement of brain tumors in clinical trials. New therapies currently in clinical testing and moving into daily clinical practice highlight the shortcomings of enhancement as a measurement tool. A particularly instructive example relates to the fact that inhibitors of vascular endothelial growth factor (VEGF), such as bevacizumab, produce a rapid decrease in the degree of contrast enhancement within malignant gliomas and in the extent of surrounding hyperintense T2-weighted signal intensity (Fig 3). These effects, which are consistent with decrease in the permeability of tumor capillaries, make it exceedingly difficult to use conventional measures of tumor size in the setting of
clinical trials. It is very important to incorporate advanced techniques such as perfusion and permeability imaging and diffusion imaging into clinical trials of these agents to develop better ways of measuring biologic effects.

Conclusions
The measurement of brain tumors in clinical trials should be performed according to a standardized set of criteria. Sensitive volumetric measurement techniques are likely to improve the chance of identifying agents that effect tumor growth. New imaging techniques are needed to measure the effects of novel brain tumor therapies.

References