



HHS Public Access

Author manuscript

Nat Clin Pract Oncol. Author manuscript; available in PMC 2016 March 17.

Published in final edited form as:

Nat Clin Pract Oncol. 2008 November ; 5(11): 634–644. doi:10.1038/ncponc1204.

Response criteria for glioma

A Gregory Sorensen^{*}, Tracy T Batchelor, Patrick Y Wen, Wei-Ting Zhang, and Rakesh K Jain

AG Sorensen is Associate Professor of Radiology, Harvard Medical School, and Associate Professor of Health Sciences and Technology, Harvard Medical School and Massachusetts Institute of Technology, and Co-Director, AA Martinos Center for Biomedical Imaging, Department of Radiology, at Massachusetts General Hospital. TT Batchelor is Associate Professor of Neurology, Harvard Medical School, and Executive Director, Stephen E and Catherine Pappas Center for Neuro-Oncology, Massachusetts General Hospital. PY Wen is Associate Professor of Neurology, Center for Neuro-Oncology, Dana-Farber/Brigham and Women's Hospital, and Harvard Medical School. RK Jain is Andrew Werk Cook Professor of Tumor Biology and Director of the Edwin L Steele Laboratory of Tumor Biology, Department of Radiation Oncology at Massachusetts General Hospital, Harvard Medical School, Boston, MA. W Zhang is Research Fellow, Neuro-Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

SUMMARY

The current method for assessing the response to therapy of glial tumors was described by Macdonald *et al.* in 1990. Under this paradigm, response categorization is determined on the basis of changes in the cross-sectional area of a tumor on neuroimaging, coupled with clinical assessment of neurological status and corticosteroid utilization. These categories of response have certain limitations; for example, cross-sectional assessment is not as accurate as volumetric assessment, which is now feasible. Disentangling antitumor effects of therapies from their effects on blood–brain barrier permeability can be challenging. The use of insufficient response criteria might be overestimating the true benefits of drugs in early-stage studies, and, therefore, such therapies could mistakenly move forward into later phases, only to result in disappointment when overall survival is measured. We propose that studies report both radiographic and clinical response rates, use volumetric rather than cross-sectional area to measure lesion size, and incorporate findings from mechanistic imaging and blood biomarker studies more frequently, and also suggest that investigators recognize the limitations of imaging biomarkers as surrogate end points.

^{*}Correspondence: Mailcode 149-2301, Building 149, 13th Street, Boston, MA 02129, USA, sorensen@nmr.mgh.harvard.edu.

Competing interests

AG Sorensen declared associations with the following companies and organizations: ACR ImageMetrix, Amgen, AstraZeneca, Breakaway Imaging, Bayer–Schering, Eli Lilly, EPIX Pharmaceuticals, Exelixis, Genentech, General Electric Healthcare, Mitsubishi Pharma, National Institutes of Health, Novartis, Northwest Biosciences, Pfizer, Schering–Plough, Siemens Medical Solutions, Takeda-Millennium and Thermal Technologies Inc. TT Batchelor declared associations with the following companies: Enzon and Schering–Plough. RK Jain declared associations with the following companies: AstraZeneca, Dyax and SynDevRx. PY Wen declared associations with the following companies: Amgen, AstraZeneca, Exelixis, Genentech, Novartis and Schering–Plough. See the article online for full details of the relationships. W-T Zhang declared no competing interests.

Keywords

blood–brain barrier; criteria; glioma; MRI; RECIST

INTRODUCTION

Biomarkers are increasingly being used in the development of new therapies; however, there are strengths and weaknesses associated with the use of biomarkers to assess such treatments. Imaging is increasingly being used to measure relevant biomarkers both by scientists and regulators. One FDA official recently stated, “Cancer is probably the most promising field right now for biomarkers, and from FDA’s point of view, I think biomarkers are the future of medical therapy, both for diagnostic purposes as well as for cancer therapeutics.”¹ The FDA has identified imaging as “at the forefront of [their] efforts” in the Critical Path Initiative² for identifying new technologies and processes that could speed up the progress of new therapy development and assessment.

Recent data support the major role that imaging has in the assessment of new oncology therapies and the associated regulatory decisions. In a recent study of oncology drug approvals, 53 of 71 FDA approvals were based on end points other than survival, the majority of which were imaging end points.³ The field of neuro-oncology uses advanced imaging techniques, particularly MRI. Neuro-oncology investigators use standardized response criteria for assessment of efficacy of new therapies. The most widely used criteria for assessing the response of glial tumors were developed by Macdonald and co-workers over 18 years ago.⁴ Since this time, however, both imaging technology and therapeutic approaches have advanced substantially; for example, in malignant glioma there has been a profound shift to using MRI rather than X-ray CT to image tumors. A number of groups have described some limitations of these criteria.^{5–12} Also, experience has shown that the Macdonald *et al.* criteria, which are widely regarded as a considerable step forward from previous modes of assessment, are ambiguous in key features such as the appropriate threshold for lesion size and the actual methods for applying the stated criteria. In addition, these response criteria preceded the advent of new antiangiogenic therapies, which might cause pseudoregression of gliomas on MRI via an antivascular effect that diminishes enhancement rather than produce actual regression through an antitumor effect.^{13–15}

A review and update of glioma response criteria is, therefore, both timely and necessary. We identify the strengths and shortcomings of the current approach and also highlight technological advances in both drug therapy and imaging that necessitate this reassessment. Our goal is to raise awareness of the unique challenges in assessing malignant glioma and to propose potential solutions to these problems. We also hope to encourage active discussion of these issues in order to improve the methods used to advance new therapies for this particularly challenging type of cancer.

CURRENT ASSESSMENT METHODS AND MOTIVATION FOR REASSESSMENT

The radiographic response criteria established by Macdonald *et al.* are based on familiar terminology in solid-tumor oncology and comprise the classification of responses into four categories, as described in Table 1.⁴ This categorization is helpful because it uses similar vocabulary and meaning to other fields of oncology; thus, when a new therapy is described as having a particular response profile, this information is meaningful to oncologists. Overall, the terminology and the broad framework of these four categories have proven to be quite successful. Another key benefit of these criteria relates to the objectivity of imaging. Imaging-based categorization techniques are favored over other methods because central reviewers or regulatory auditors can verify them. This authentication strongly reduces unintended local interpreter bias of patient status or tumor response. Gliomas are particularly difficult to treat and, as few drugs are successful, historically the most important aspect of categorization has been progressive disease (PD) status. Of note, most tumor volume assessments are performed using contrast-enhanced imaging (e.g. MRI after the administration of a gadolinium-containing agent). This approach is still appropriate, although we discuss the use of other MRI methods.

TECHNICAL ADVANCES DRIVING RESPONSE CRITERIA REASSESSMENT

Although the general classification categories should be maintained, ambiguities in the Macdonald *et al.* criteria and advances in technology indicate the need for clarification and for new criteria. Thanks to technological advances, diagnosis and treatments have markedly evolved. Imaging techniques and methods of analysis have also dramatically expanded and improved over the past two decades. In 1990, CT was still the most common means of assessing brain tumors. MRI has largely supplanted CT as the method of choice for monitoring lesions on the basis of its superior ability to visualize glial tumors. Moreover, there has been a substantial change in the sensitivity, specificity, and overall performance of MRI technology since 1990. Equally relevant to this discussion is the fact that surgical technology has changed substantially in recent years; for example, image-guided surgery has permitted neurosurgeons to resect even the most infiltrative tumor while maximizing efforts to spare eloquent brain tissue. As a result, almost all patients with high-grade glioma and many with low-grade glioma now undergo surgical resection. This practice means that residual or recurrent disease in such individuals is often highly irregular in shape. In addition, new classes of agents that affect vascular permeability and hence tumor contrast enhancement, such as inhibitors of VEGF, might influence classification systems that rely predominantly on contrast enhancement.

How do these changes in imaging and surgical technology affect response criteria? One important criterion is how change in tumor ‘size’ is typically quantified. Macdonald *et al.* noted that, “volume measurements are technically difficult in many glioma patients,” and suggested that “size be considered the tumor’s largest cross-sectional area”.⁴ Cross-sectional area is typically computed using the assumption that the overall lesion can be described by an ellipsoid. The usual approach is to find the ‘slice’ or image on which the tumor area is greatest, then measure the longest single diameter and the longest diameter perpendicular to

that. This technique sounds very plausible as this approach is often used to measure cross-sectional area outside the brain. However, cross-sectional area quickly becomes ambiguous when patients with a brain tumor are studied post surgery. Figure 1 shows a typical post-surgical scan of a patient with glioblastoma, demonstrating enhancing recurrent tumor around a surgical cavity before and after chemotherapy. The volume-based approach to measuring tumor size identifies enough change to warrant classification as a partial response but the cross-sectional area does not. Furthermore, the two-diameter method has been updated and measurement now encompasses the cystic cavity, so a change in the cystic cavity might be included even if the enhancing tumor itself does not change. Even in the absence of a cystic cavity, malignant gliomas can have such an irregular shape that the ellipsoid assumption is erroneous, as shown in Figure 2. In addition, the MRI acquisition plane (e.g. sagittal, coronal, or axial) used to obtain images can vary arbitrarily from scan to scan, thus potentially adding further variability to repeat cross-sectional area measurements unless special software is used.¹⁶ Finally, glial lesions rarely grow in a smooth, spherical shape, which means that the imaging plane with the greatest cross-sectional area can vary from visit to visit, thereby adding further inconsistency.

Volume measurements do not suffer from the problems associated with measuring cross-sectional area. High-resolution volumetric scans of typically 1 mm × 1 mm × 1 mm volume element or voxel size can be obtained on modern MRI scanners. The advantage of this resolution over assessment of thicker (e.g. 5 mm) tissue slabs is not yet fully established, but probably allows for better detection and quantification of small lesions. Another benefit of the volume-based approach over the area-based approach is the capacity to distinguish between an ‘evaluable’ versus a ‘measurable’ lesion. This artifactual dichotomy is evident when using the RECIST criteria¹⁷ and other measurement criteria, but inevitably leads to complexity in glioma measurement because multifocal areas of enhancement or multiple small satellite lesions are common. Complexity in turn leads to increased confusion and higher variability in study design and reporting.

The disadvantages of area-based measurement of glial lesions are not recognized by all practitioners; for example, the February 2000 proposal for new response criteria for solid tumors, known as RECIST,¹⁷ suggested measuring brain tumors by the use of a diameter-based assessment approach. Owing to the high frequency of irregularities (Figures 1 and 2), volumetric approaches have substantially less inter-reader and intra-reader variability than other methods.¹⁸ Fortunately, computer software is now available that can aid in quickly segmenting enhancing tumor from normal tissues and providing a volumetric assessment, without the penalty in accuracy that assumptions of ellipsoid or other shapes inevitably introduce.

Why should the inaccuracy of area-based size determination matter? Some investigators propose that although the area or even single-diameter approach might suffer in reproducibility or precision compared with the volumetric approach, the benefits of a truly effective drug would be powerful enough to be evaluated by even simpler assessment criteria. The RECIST proposal suggests, “It was not thought that increased precision of measurement of tumor volume was an important goal for its own sake.”¹⁷ In retrospect, it is now clear that precision of measurement has a tremendous effect on the powering of clinical

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

trials because power calculations must be based on statistical estimates of the therapeutic benefit of a new treatment, termed effect size, and sample size is critically related to both the effect size of the therapy and the variance of the outcome measurement. A linear change in the variability of the outcome measure can necessitate a geometric change in sample size; for example, doubling the variability requires four times the sample size. Since area measurement techniques have greater variability than do volume approaches, when change in size is used as a primary outcome measure the use of cross-sectional methodology can directly affect the number of patients that need to be recruited and therefore have a major effect on the cost of drug development. Indeed, the RECIST investigators state that there were substantial “concerns regarding the ease with which a patient may be considered mistakenly to have disease progression...primarily because of measurement error.”¹⁷ Events in the 8 years since the RECIST criteria were introduced have only highlighted these concerns to many investigators. Although in certain limited settings (e.g. newly diagnosed low-grade lesions not undergoing surgery) some gliomas might be so nearly approximated by an ellipsoid that an area or unidimensional approach may be valid, general use of the RECIST criteria adds measurement variability and therefore will increase sample size needlessly. Increased sample sizes not only add expense, but, by slowing the progress of a trial, also delay the introduction of potentially useful therapies to clinical practice or unnecessarily expose more patients than needed to an unsuccessful therapy.

Another benefit of using volume rather than cross-sectional area is that changes in tumor size (i.e. response) can be determined earlier (Figure 3). Volume is a three-dimensional measure, so as the radius of a sphere increases or decreases there is a larger percentage change in volume than there is in area. This fact means that patients who have shrinking lesions in response to a novel treatment will be identified sooner, and likewise patients who have growing lesions despite a novel treatment will also be recognized earlier. Given the difficulty in identifying promising therapies for glial tumors, such an improvement in temporal efficiency is an attractive prospect. Although it might be argued that altering size thresholds from area to volume precludes comparison of new trials with earlier studies that report area-based progression rates, there is such ambiguity present in the current definitions that true comparability of studies is extremely rare and thus shifting to more-precise terminology can only benefit the field. It might also be argued that current methods of imaging gliomas do not adequately identify tumor boundaries or serve as a useful surrogate for tumor mass or patient survival. A recent meta-analysis showed that changes in the size of a glioma correlates with survival.¹⁹ The data supporting this link are still limited, however, so this assumption should be revisited in the future. Nevertheless, for now these data provide an adequate basis for moving forward.

In 1990, when practitioners were predominantly using CT scans, the use of cross-sectional area as measured with two diameters was a practical approach. In addition, centralized review of imaging scans was rare. Central review is now required by most scientific agencies, such as the Cancer Therapy Evaluation Program or FDA, and, now that powerful computers and software are widely available, this approach cannot be described as ‘technically difficult’. Although no study has yet proven that classification and treatment on the basis of volumetric assessment correlates better with overall survival than does classification and management according to area-based assessment, waiting for the

publication of such a study before changing to volumetric assessment seems unwise. The logic of using area instead of volume is based on two assumptions: that gliomas can be approximated by an ellipsoid, and that volumetric assessment is not routinely feasible. Sorensen *et al.* found that using cross-sectional area instead of volume measurement resulted in different classification of tumor progression in at least 26% of cases,¹⁸ and it seems logical that the volumetric approach is more accurate. Furthermore there are few studies that demonstrate the relationship between progression as determined by area and overall survival.¹⁹ Such qualification needs to be performed in a large population (e.g. through dozens of multicenter trials) because although in some cancers radiographic response is highly correlated with overall survival,²⁰ it is not in other cases.²¹

A challenge to the widespread use of volumetrics is the number of steps necessary to provide such assessments. Tools to perform these assessments need to be easily available in order to lessen the effort inherent in these techniques. Ideally, volumetrics would be performed as part of routine neuroradiological examination, thus providing data that the clinician could use at each patient visit or assessment. Naturally, these methods and criteria should also be subject to prospective validation studies.

AMBIGUITIES DRIVING THE NEED FOR REASSESSMENT OF RESPONSE CRITERIA

The Macdonald *et al.* criteria have ambiguities that can increase the variability of how the measures are applied, which in turn makes comparison of therapies more difficult. There are uncertainties associated with the imaging (even after conversion from cross-sectional area to volumetric) and also the non-imaging components. Here, we describe these issues and propose potential solutions.

The non-imaging components of the assessment criteria are a potentially powerful addition to the evaluation of size changes. Unlike the RECIST criteria used for other tumors, the existing Macdonald *et al.* classification criteria for glial tumors also take into account neurological status of the patient and steroid dosing. This clinical information is of course relevant: a drug that shrinks tumors but causes severe neurological decline would not represent an advance in the field. The first problem with non-imaging criteria is their implementation. In practice, the division between radiographic progression and clinical progression or alteration of steroid dosing is rarely made. We reviewed 46 clinical trials of patients with glioma published in 2006, 30 of which describe progression-based survival data (such as 6 month progression-free survival). Of these studies, none reported progression data that explicitly described change in steroid and/or neurological status as a marker of progression. Four (13%) of these 30 trials mentioned the incorporation of steroid or neurological information into the definition of a progression end point, but none of these reports indicated that neurological decline or a change in steroid dosing alone was a cause for categorizing a patient as having progressive disease. It seems that these other measures are often given less priority than radiographic response, despite being important. In our own recent central review of 877 evaluation visits by 240 patients participating in a phase II study, a third of the categorizations of progression were made on the basis of non-radiographic reasons (142 of 416 determinations of PD for a given visit, half of which were

on the basis of neurological decline; AG Sorensen, unpublished data). This finding suggests that because of the failure to report non-radiographic responses, the true progression rate might be as much as 50% higher than is commonly reported. Thus a drug that has advanced to phase III testing might actually have a much higher likelihood of failure than might have been expected on the basis of the reported results alone. The 90% overall failure rate of oncology drugs²² indicates that the pipeline evaluation process is suboptimal.

One reason for the inadequate nature of this system might be the challenge of centralized review of neurological change or steroid dosage; another reason might be simply the lower profile that these non-imaging metrics have traditionally had. We propose that these non-imaging components of the Macdonald *et al.* criteria be re-emphasized and clarified. We agree with the Macdonald and co-workers that, “The neurologic examination is not a reliable measure of response, but it can be an important and valid measure of progression.” Precise assessment of neurological worsening is not specified in these criteria, only the phrase “unequivocal neurologic deterioration” is used. We too believe that experimental therapies “should be reserved for patients with better function” (e.g. Karnofsky Performance Score [KPS] ≥ 60), and we propose that clinical deterioration should be better assessed. Validated quantitative measures of clinical status, however, are not available. One potential metric that could be employed is decline in a measure such as the KPS; for example, a decline of more than two levels probably represents a meaningful clinical change. Although this is an arbitrary threshold, it has the advantage of being practical and probably clinically meaningful. A change in KPS of more than two levels or a drop below 50 is a reasonable threshold for classifying a patient as having progressive disease according to clinical findings, and specific evaluation of whether such clinical deterioration is caused by the tumor (as opposed to an unrelated event) should be undertaken. Although other thresholds could be considered, a change of more than two levels in the KPS is large enough to overcome the variability noted in this measurement approach.⁵

Knowing steroid dosage is important when assessing tumor response—the document by Macdonald and coauthors states, “By themselves, these drugs improve symptoms and signs, maintain clinical improvement for extended periods even at low or reduced doses, and substantially decrease the size of some malignant gliomas on CT [or MRI] scans.” Furthermore, steroid doses can typically be quantified and confirmed in medical records, thereby reducing potential bias. We propose retaining this measurement as a response criterion. We are concerned, however, that some individuals have interpreted the Macdonald *et al.* criteria as indicating that an increase in steroid dosing alone be cause for declaration of PD status, particularly when there is no clinical or radiographic evidence of progression. The rationale for increases or decreases in steroid dose in a given patient can be subjective and practitioner-dependent, and the true effect of steroids on imaging is dependent on both dose and time. As a result, strictly classifying a tumor into the ‘progressive disease’ category on the basis of a small increase in steroid dose seems inappropriate. Furthermore, a decrease in steroid dosing can cause pseudo-progression on an MRI scan (i.e. an apparent increase in the size of the enhancing area).

We therefore propose that radiographic complete response only be declared in the absence of steroid treatment. A classification of partial response requires a stable or decreasing

dosage of steroids. PD cannot be determined by steroid dosing change alone, but only in addition to clinical progression or radiographic progression, as described herein. Furthermore, a classification of radiographic or clinical progression requires stable or increasing doses of steroids in order to exclude pseudoprogression. We propose that a 'stable' dose of steroids comprise at least 3 days and preferably 7 days of an unchanged regimen prior to imaging. Although the exact time course of steroid activity in glioma has not been formally established, 3 days is a practical threshold and steroids are in fact known to have an antiedema effect within a few hours.²³ We propose that investigators identify at the beginning of a study the reference source for determining steroid equivalent dosing.

We also propose resolutions to certain ambiguities involving neuroimaging. Execution of imaging has improved since 1990, leading to better delineation of tumor boundaries in some cases, and raising questions in other cases. Given the highly infiltrative nature of gliomas, separate areas of enhancement might represent the same parent lesion. Similarly, a 'new' lesion might simply represent growth of the pre-existing lesion. An example of this occurrence is shown in Figure 4. We suggest the biological features of glioma and that a given lesion can have 'daughter' lesions that are part of the same abnormality be taken into consideration when assessing neuroimaging results. The daughter lesion must be biologically plausible, however; for example, such a lesion should be in the same hemisphere as the parent lesion or connected and less than 50 mm away. This point is true whether the 'daughter' is present at initial entry into a trial or at follow-up. This designated distance from the parent lesion is arbitrary, but might best reflect the known infiltrative nature of glioma. We propose that investigators identify at the beginning of a study the distance at which parent and daughter lesions are defined as separate entities. Using volumetric rather than area measurements minimizes confounding effects due to unusual geometry.

An issue often raised by neuroradiologists is that of pseudoprogression, in particular the possibility of radiation necrosis mimicking a tumor. At the present time, no neuroimaging technique has been shown to sensitively and specifically distinguish radiation necrosis from tumor, although numerous techniques have been tried without success. We propose that areas of enhancement that are clearly not tumor—such as choroid plexus, vessels, and extra-axial scar tissue—should be excluded from any assessment of tumor size. When there is ambiguity, we suggest avoiding the temptation to interpret whether intra-axial enhancement represents radiation necrosis or scar rather than tumor until an international consensus can be reached on criteria to distinguish between tumor and non-tumor, and propose including all enhancing intra-axial tissue in the tumor burden measurement. A number of groups have concluded that distinguishing radiation necrosis or other forms of pseudoprogression from recurrent or growing tumor is extremely difficult with current technology.^{24–27} This situation might change when a more powerful diagnostic technique arrives that can reliably distinguish enhancement due to tumor from other forms of enhancement and that is robust and usable in a multicenter setting, but no such technique is currently available. On the other hand, we strongly endorse the principle that, whenever possible, true progression be confirmed with a follow-up imaging study, as newer therapies such as immunomodulators or antiangiogenic agents can cause waxing and waning of lesion volume, although the effects of these agents in particular need careful attention. Furthermore, the timing of therapies must

be taken into account when performing neuroimaging: certain therapies might cause transient effects such as inflammation or other reactions. It seems logical to avoid using scans obtained during such therapies without further study of the typical time course of tumor appearance during and shortly after treatment.

Another important issue associated with neuroimaging relates to how comparisons are made between scans. We propose that scans are interpreted by a given reader who has data from all time points available but is blinded to the order. This approach ensures that images are appropriately compared and that meaningless differences in boundaries do not confuse measurements. Reader bias can be decreased by maintaining blinding with respect to temporal sequence.

NOVEL THERAPIES DRIVING RESPONSE CRITERIA REASSESSMENT

With the advent of cell-based techniques, antiangiogenic agents, immune modulators, and other approaches, as well as advanced imaging tools, it might not always be optimum to use volume of enhancing tumor tissue as the final metric. A good example of this point is the effect of antiangiogenic agents on gadolinium enhancement. Decreased enhancement on MRI scans is reported after antiangiogenic treatment,^{14,15,28,29} and, therefore, this effect has increased reports that lesion volume on T2-weighted or fluid-attenuated inversion recovery (FLAIR) MRI are taken into account,^{14,15,28} although clear criteria for these techniques are not yet established. Some investigators have noted an apparent increased tendency for patients to develop infiltrating progression after antiangiogenic treatment—perhaps as a result of cooption of existing blood vessels—that might be visible only on FLAIR or T2-weighted images. We prefer FLAIR over spin echo or ‘fast’ spin echo T2-weighted images when possible, and suggest again that volume rather than area be used as the primary metric. We also propose the exclusion of hyperintense areas that are not likely to represent the tumor (e.g. periventricular changes in the centrum semiovale contralateral to the enhancing tumor). We posit that volumetric criteria apply for FLAIR and T2-weighted images, and that investigators describe in the abstract of reports describing tumor responses whether or not T2-weighted data as well as clinical data (e.g. steroid dosage, clinical deterioration) are included. In certain drug therapies it might be important to ensure that measurement of response includes both T1 and T2 data. In low-grade tumors without enhancement, FLAIR or T2-weighted imaging might be used to measure the boundaries of lesions. This approach is not as well-studied as the measurement of contrast enhancement, and although the concepts we describe could probably be applied to low-grade, non-enhancing lesions, it is not yet clear that implementation would be feasible. Furthermore, many processes besides tumor-induced edema can lead to changes on T2-weighted or FLAIR imaging; for example, post-radiation changes and post-surgical changes, chemotherapy, or tumor infiltration. Therefore, the caveat of pseudoprogression also applies in this approach. Until a reliable method for distinguishing edema from tumor is developed, T2-weighted or FLAIR imaging will remain a secondary rather than primary method for evaluating tumor response.

In addition to monitoring T2-weighted images, we also recommend assessing for other evidence beyond lesion volume, such as a reduction or increase in mass effect (e.g. midline shift). The most robust response conclusions can be drawn when multiple imaging

modalities all point to the same outcome. For example, for antiedema effects, a reduction in the size of a the lesion as seen on FLAIR or T2-weighted imaging, as well as a decrease in the apparent diffusion coefficient and a decrease in mass effect,²⁸ give a greater degree of confidence than any one of these findings alone. Mass effect could be measured by millimeters of midline shift or by displacement of other prospectively defined boundaries. Furthermore, although we have predominantly described enhancing lesions as this type is most common, FLAIR or T2-weighted imaging could be used to evaluate nonenhancing lesions, or certain therapies on the basis of initial human experience; this parameter should be defined prospectively at the start of the study.

We also strongly encourage the greater use and acceptance in the clinician community of data from mechanistic imaging techniques. Many of these tools are still experimental and have not yet been used in multicenter settings. Such imaging tools include PET, magnetic resonance spectroscopy, perfusion and diffusion MRI, and many other approaches.³⁰ These techniques are not ready for use as response criteria, but should be included whenever possible because they can add value when attempting to understand responses, especially with novel therapies. As the time course and full mechanism of action of standard therapies (e.g. chemoradiation) are still incompletely understood, such mechanistic imaging approaches could boost our attempts to comprehend not only whether a given therapy succeeds or fails, but also why a therapy is successful or not.

RECOMMENDATIONS

Figure 5 and Table 2 summarize the limitations of the Macdonald *et al.* criteria and provide proposals for developing new standardized response criteria for glial tumors. These proposals describe the problems detailed within this article and the solutions that we believe are most workable. We wish to emphasize that there are a number of excellent features in the Macdonald *et al.* recommendations that we wish to re-state and fully endorse. These are listed in Box 1, and consist of sensible approaches to minimizing variance in assessment.

Box 1

Key recommendations carried forward from the Macdonald *et al.* criteria for the assessment of glial tumor response.⁴

- All protocol patients should undergo central pathology review by an experienced neuropathologist
- Phase II studies of malignant glioma should focus on a single tumor type
- Investigational drugs should be reserved for patients with better function (e.g. KPS \geq 6), as those with severe disability might not live long enough to be assessable for response
- Investigational treatment following major tumor resection should be delayed unless there is unequivocal residual tumor on MRI scans
- A rebiopsy should be performed whenever there is doubt about diagnosis

- Steroid dose should be kept stable for at least 1 week during periods critical for response evaluation
- A uniform scanning technique should be used (i.e. identical scanner, patient position, dose of contrast, injection–scan interval). Whenever possible use of the same scanner and automated repositioning software will aid in minimizing variability¹⁶

Abbreviation: KPS, Karnofsky Performance Score.

CONCLUSIONS

It is important to note that all biomarkers have substantial limitations^{20,31} and surrogate end points such as objective response rates will always have some less-desirable features compared with traditional clinical end points such as overall survival. On the basis of the limitations of biomarkers we suggest that investigators focus on overall survival as an end point in phase III studies and use biomarkers (imaging or otherwise) for decision-making in phase II studies. Gliomas are difficult to treat, and biomarkers— especially mechanistic ones—can provide such valuable information for the future that we also strongly encourage investigators to include biomarkers (blood, imaging, etc.) whenever possible in phase III studies. The final benefit of biomarkers to patients is survival, and we emphasize that a key purpose of these proposed criteria is to improve the link between reports of survival in phase II studies, which are based on progression-free (objective or radiographic) response criteria, and overall survival as documented in phase III studies. Far too many reports have not included this full range of information, thus making direct comparison between new and previous studies difficult. As a result, too many therapies have advanced to phase III trials only to fail, at tremendous cost to patients, investigators, and sponsors. We believe that use of tumor volume rather than area, specifically describing both radiographic and composite (radiographic plus clinical and/or steroid) progression rates, and adherence to a common set of unambiguous guidelines will accelerate the process of developing successful new therapies for glial tumors.

Acknowledgments

This work was funded by the US Public Health Service (grants M01-RR-01066, 1R21CA117079-01, 5T32CA009502-20, 5P41RR014075 and 5P01CA080124) and The MIND Institute. We wish to thank Craig Peterson for technical assistance, and Zariana Nikolova, Richard Parker, Wendy Hayes, and Steven Green for helpful discussions.

References

1. [accessed 28 April 2007] The Online NewsHour (online 28 March 2007) Extended interview: Janet Woodcock discusses cancer biomarkers. http://www.pbs.org/newshour/bb/health/jan-june07/cancerwoodcock_03-28.html
2. Woodcock, J. [accessed 28 April 2007] The Critical Path initiative: one year later. online 5 May 2005 www.fda.gov/cder/regulatory/medlmaging/woodcock.ppt
3. Johnson JR, et al. End points and United States Food and Drug Administration approval of oncology drugs. *J Clin Oncol*. 2003; 21:1404–1411. [PubMed: 12663734]

4. Macdonald DR, et al. Response criteria for phase II studies of supratentorial malignant glioma. *J Clin Oncol.* 1990; 8:1277–1280. [PubMed: 2358840]
5. Brada M, Yung WK. Clinical trial end points in malignant glioma: need for effective trial design strategy. *Semin Oncol.* 2000; 27:11–19. [PubMed: 10866345]
6. Dempsey MF, et al. Measurement of tumor “size” in recurrent malignant glioma: 1D, 2D, or 3D? *AJNR Am J Neuroradiol.* 2005; 26:770–776. [PubMed: 15814919]
7. Galanis E, et al. Validation of neuroradiologic response assessment in gliomas: measurement by RECIST, two-dimensional, computer-assisted tumor area, and computer-assisted tumor volume methods. *Neuro Oncol.* 2006; 8:156–165. [PubMed: 16533757]
8. Grant R, et al. Chemotherapy response criteria in malignant glioma. *Neurology.* 1997; 48:1336–1340. [PubMed: 9153469]
9. Hess KR, et al. Response and progression in recurrent malignant glioma. *Neuro Oncol.* 1999; 1:282–288. [PubMed: 11550320]
10. Kaplan RS. Complexities, pitfalls, and strategies for evaluating brain tumor therapies. *Curr Opin Oncol.* 1998; 10:175–178. [PubMed: 9619351]
11. Shah GD, et al. Comparison of linear and volumetric criteria in assessing tumor response in adult high-grade gliomas. *Neuro Oncol.* 2006; 8:38–46. [PubMed: 16443946]
12. Vos MJ, et al. Interobserver variability in the radiological assessment of response to chemotherapy in glioma. *Neurology.* 2003; 60:826–830. [PubMed: 12629241]
13. Chamberlain MC. MRI in patients with high-grade gliomas treated with bevacizumab and chemotherapy. *Neurology.* 2006; 67:2089. [PubMed: 17159134]
14. Pope WB, et al. MRI in patients with high-grade gliomas treated with bevacizumab and chemotherapy. *Neurology.* 2006; 66:1258–1260. [PubMed: 16636248]
15. Vredenburgh JJ, et al. Phase II trial of bevacizumab and irinotecan in recurrent malignant glioma. *Clin Cancer Res.* 2007; 13:1253–1259. [PubMed: 17317837]
16. Benner T, et al. Comparison of manual and automatic section positioning of brain MR images. *Radiology.* 2006; 239:246–254. [PubMed: 16507753]
17. Therasse P, et al. 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; 92:205–216. [PubMed: 10655437]
18. Sorensen AG, et al. Comparison of diameter and perimeter methods for tumor volume calculation. *J Clin Oncol.* 2001; 19:551–557. [PubMed: 11208850]
19. Ballman KV, et al. The relationship between six-month progression-free survival and 12-month overall survival end points for phase II trials in patients with glioblastoma multiforme. *Neuro Oncol.* 2007; 9:29–38. [PubMed: 17108063]
20. Fleming TR. Surrogate endpoints and FDA’s accelerated approval process. *Health Aff (Millwood).* 2005; 24:67–78. [PubMed: 15647217]
21. Bruzzi P, et al. Objective response to chemotherapy as a potential surrogate end point of survival in metastatic breast cancer patients. *J Clin Oncol.* 2005; 23:5117–5125. [PubMed: 15955906]
22. Kamb A, et al. Why is cancer drug discovery so difficult? *Nat Rev Drug Discov.* 2007; 6:115–120. [PubMed: 17159925]
23. Ostergaard L, et al. Early changes measured by magnetic resonance imaging in cerebral blood flow, blood volume, and blood-brain barrier permeability following dexamethasone treatment in patients with brain tumors. *J Neurosurg.* 1999; 90:300–305. [PubMed: 9950501]
24. Butowski NA, et al. Diagnosis and treatment of recurrent high-grade astrocytoma. *J Clin Oncol.* 2006; 24:1273–1280. [PubMed: 16525182]
25. Kumar AJ, et al. Malignant gliomas: MR imaging spectrum of radiation therapy- and chemotherapy-induced necrosis of the brain after treatment. *Radiology.* 2000; 217:377–384. [PubMed: 11058631]
26. Mullins ME, et al. Radiation necrosis versus glioma recurrence: conventional MR imaging clues to diagnosis. *AJNR Am J Neuroradiol.* 2005; 26:1967–1972. [PubMed: 16155144]

27. Ricci PE, et al. Differentiating recurrent tumor from radiation necrosis: time for re-evaluation of positron emission tomography? *AJNR Am J Neuroradiol.* 1998; 19:407–413. [PubMed: 9541290]
28. Batchelor TT, et al. AZD2171, a pan-VEGF receptor tyrosine kinase inhibitor, normalizes tumor vasculature and alleviates edema in glioblastoma patients. *Cancer Cell.* 2007; 11:83–95. [PubMed: 17222792]
29. Stark-Vance V. Bevacizumab and CPT-11 in the treatment of relapsed malignant glioma [abstract # 369]. *Neuro Oncol.* 2005; 7:369.
30. Jacobs AH, et al. Imaging in neurooncology. *NeuroRx.* 2005; 2:333–347. [PubMed: 15897954]
31. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med.* 1996; 125:605–613. [PubMed: 8815760]
32. Bagley CM Jr. Measurement of brain tumor volumes by the perimeter method. *J Clin Oncol.* 2001; 19:3159–3160. [PubMed: 11408516]

REVIEW CRITERIA

The information for this review was compiled by searching the PubMed and MEDLINE databases for articles published until 1 December 2007. Electronic early-release publications were also included. Only articles published in English were considered. The search terms used included “glioma” and “glioblastoma” in association with the following search terms: “response criteria”, “MRI” “computed tomography” and “RECIST”. When possible, primary sources have been quoted.

KEY POINTS

- Volume rather than area measurements of tumor burden are now feasible and have less inter-observer variability
- Non-volumetric reasons for progression including rate of clinical worsening and steroid dosing should be routinely reported in describing the results of clinical trials
- Mechanistic biomarkers including imaging should be employed wherever possible, especially in early stage trials, to provide more meaning beyond response rates alone
- Resolution is possible for previously ambiguous aspects of response criteria including minimum lesion size, degree of neurological worsening, definition of what 50% change in size means, degree of enhancement, etc.
- The advent of new therapies such as antiangiogenic agents that directly affect tumor vessels and tumor enhancement require particular care in response evaluation. FLAIR and/or T2-weighted imaging should be used in addition to T1-weighted post-contrast images

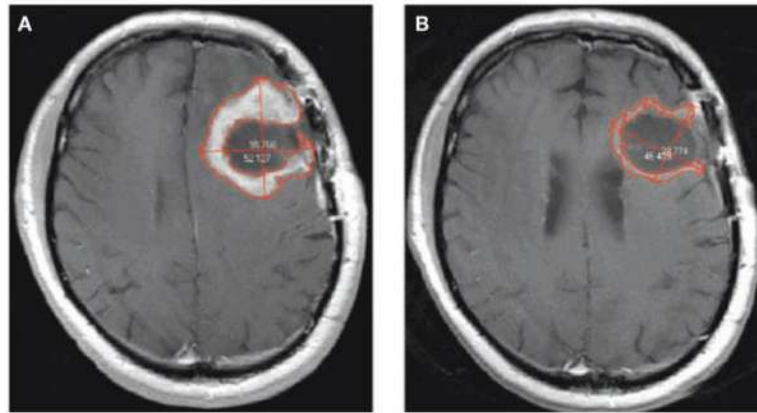


Figure 1.

MRI scans demonstrating shortcomings of the diameter method of measuring a change in glioma size. **(A)** T1 contrast-enhanced MRI scan of a 69-year-old male with recurrent glioblastoma, before initiation of antiangiogenic therapy. Bidimensional measurements are shown and the enhancing tumor is outlined in red as per typical volumetric analysis. In this patient the glioma is 42.912(units) by 42.961 (units), which constitutes a cross-sectional area of 55.768 mm by 52.127 mm, which constitutes a cross sectional area of 22.83cm² **(B)** A scan from the same patient 29 days after initiation of chemotherapy. In this instance the glioma is 46.409mm by 38.774 mm, which constitutes a cross-sectional area of 14.13 cm². If diameter measurements are used to estimate cross-sectional area, the lesion can be said to have shrunk to 62% of its original size, and would be classified as stable disease. According to the volume-based approach (which includes all enhancing tissue, not only on the image shown but of the entire lesion) the glioma has decreased to 28% of its original size, and would be classified as a partial response.

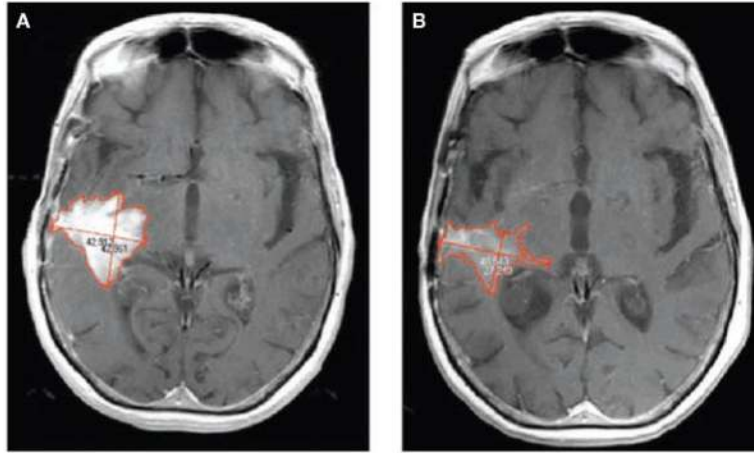


Figure 2. Shortcomings of the ellipsoid assumption used in the diameter method of measuring a change in glioma size. **(A)** T1 contrast-enhanced MRI scan of a 55-year-old female with recurrent glioblastoma, before therapy initiation. **(B)** A scan from the same individual 26 days later. Note that the lesion has not only decreased in size but also in intensity of enhancement. According to the diameter-based cross-sectional area technique, the lesion has decreased to 72% of its original size, and would be classified in the stable disease category. On the basis of the volume of enhancement, the lesion is 35% of its original size, and would be classified as a partial response.

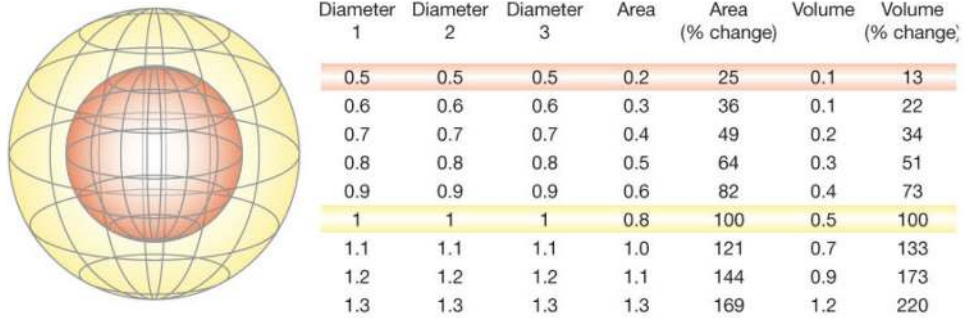


Figure 3. Sensitivity of volume compared with area for the determination of change in tumor size. The inner sphere (red) has half the diameter of the outer sphere (yellow), a quarter of the relative area, and an eighth of the relative volume, as shown in the table on the right that describes diameters, areas, and volumes (with red and yellow sphere values highlighted). Using a threshold of a 50% change in high quality volume measurements would allow patients who are responding to treatment to be identified earlier than would be the case with a 50% change in area measurements.

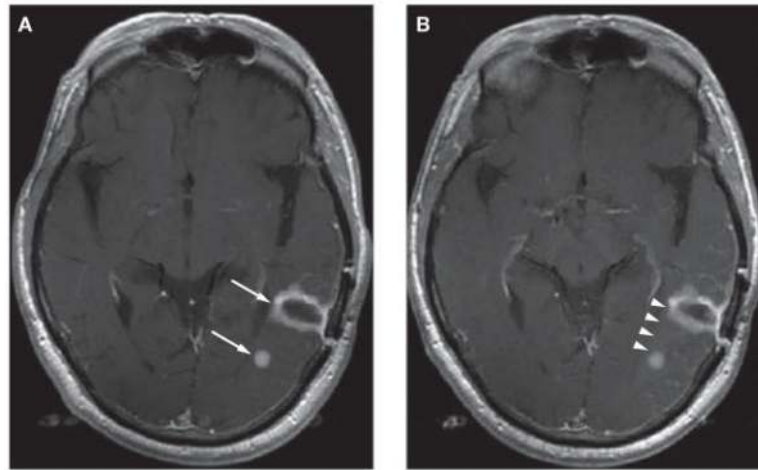


Figure 4.

When neuroimaging gliomas, ‘daughter’ lesions should be counted as part of the parent malignancy. **(A)** An MRI scan of a 69-year-old male with recurrent glioblastoma. Arrows show two areas of enhancement that might be considered separate lesions. **(B)** The same patient on repeat imaging the next day. There is a faint suggestion of enhancement connecting the two affected areas (arrowheads). Due to the infiltrative nature of glioma these are the same lesion with different foci of enhancement. Had the small focus appeared on a follow-up scan, we would propose considering that this region represents growth of the parent lesion rather than a new focus that would lead to classification of the patient as having progressive disease.

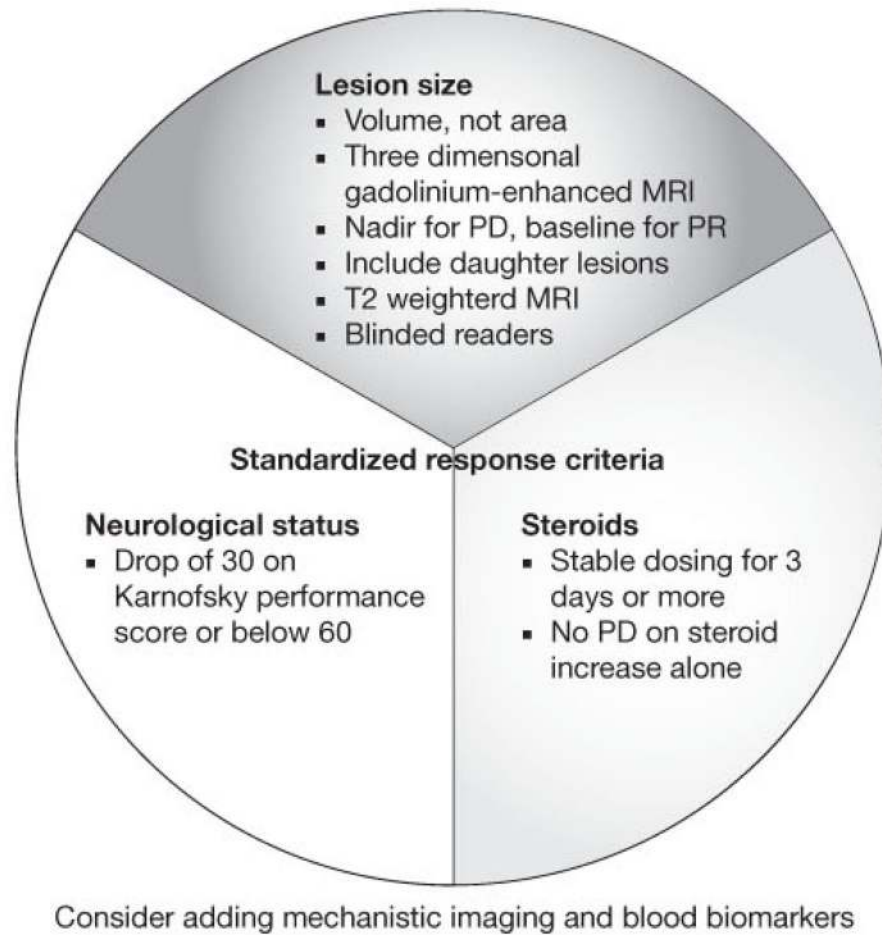


Figure 5.

A schematic showing recommendations for glioma standardized response criteria. The figure identifies three key components: lesion size, neurological status, and steroids. Elaborations on these recommendations are described in Tables 2 and 3. Abbreviations: PD, progressive disease; PR, partial response.

Table 1

Traditional glial tumor response criteria.

Response category	Categorization criteria
Complete response	Disappearance of all enhancing tumor on consecutive CT or MRI scans at least 1 month apart and patient taken off steroid therapy, and neurologically stable or improved
Partial response	Reduction in size of enhancing tumor by $\geq 50\%$ on consecutive CT or MRI scans at least 1 month apart, steroid therapy stable or reduced and patient neurologically stable or improved
Progressive disease	Increase in size of enhancing tumor or any new tumor by $>25\%$ on CT or MRI scans, with steroid therapy stable or increased, or patient neurologically worse
Stable disease	All other situations

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Challenges in glioma response criteria and proposed solutions.

Challenge	Macdonald criterion for progression	Concerns	Proposals
Size	“50% increase in size”	<ul style="list-style-type: none"> Suggests size is equivalent to cross-sectional area Highly irregular tumors are the norm Poor reproducibility Cystic components might affect measurement Presence of multiple lesions might affect measurement 50% threshold is arbitrary 	<ul style="list-style-type: none"> Use volume of enhancing tissue after standard dose of gadolinium on MRI as a marker instead of overall size^{6,18} Do not include cysts^{11,12,17,18} Sum all enhancing intra-axial tumor voxels, contiguous or not^{6,12,18} Define basis for change. We use nadir as basis for PD, baseline as basis for PR^{5,17} Define minimum size a priori on the basis of imaging technique. We use 500 μl as the smallest amount for volumetric approaches (approximately the volume of a sphere 10 mm in diameter) when 1 mm \times 1 mm \times 1 mm three-dimensional MRI data of good quality are available Define what the 50% change in size means. We propose 50% change in volume, not 50% change in area. This approach is more sensitive: patients who are failing treatment will be declared sooner; patients who are succeeding on treatment will also be declared sooner^{7,11,32}
Steroid dosing	“Escalating steroid doses...in the absence of significant CT worsening... are included in the stable category.”	<ul style="list-style-type: none"> Might not be able to determine progression on the basis of steroid increase Affected by number of days that a steroid dose has remained stable 	<ul style="list-style-type: none"> CR or PR not possible if steroid dose is increasing⁴ MRI assessment requires at least 3 days of stable steroid dosing, preferably 7 days.^{4,5,11} PD not called due to steroids alone, but only by radiographic or clinical criteria⁴ PD not determined due to steroids alone, but only by radiographic or clinical criteria⁴
Neurological status	Unequivocal neurological deterioration	<ul style="list-style-type: none"> Scales are not ideal Changes might not be due to the tumor 	<ul style="list-style-type: none"> Define prospectively the degree of change required to make a determination of progression. One option is a drop in Karnofsky Performance Score to below 50 or a decrease of more than two levels⁶

Challenge	Macdonald criterion for progression	Concerns	Proposals
Distinction between a tumor border and a new lesion	"...new areas of tumor"	<ul style="list-style-type: none"> Satellite lesions are common in glioma 	<ul style="list-style-type: none"> Determine the distance within which a second focus of enhancement is considered part of parent lesion. We arbitrarily define a distance of within 50 mm in the same hemisphere or along the corpus callosum as not a new lesion Contralateral hemisphere lesions (e.g. on opposite side of falx) are immediately considered new lesions¹²
Degree of enhancement	No comment	<ul style="list-style-type: none"> Antiangiogenic therapy clearly affects the degree of enhancement Enhancement is not binary 	<ul style="list-style-type: none"> Identify a threshold for enhancement before study initiation. We propose any enhancement visible to the interpreter¹² Include secondary effects, such as mass effect or surrounding edema (see comment on multimodal imaging below)
Tumor mimics or pseudoprogression	"Investigator [must] carefully exclude... pseudoprogression."	<ul style="list-style-type: none"> Numerous concerns; maintain as much control as possible over variables 	<ul style="list-style-type: none"> Use precontrast MRI as well as post-contrast MRI⁴ Continue to attempt to maintain control over variables (Box 1)⁴ Include all enhancing intra-axial tissue as part of the tumor until or unless consensus criteria are developed for distinguishing tumor from pseudoprogression
Multimodal imaging	No comment	<ul style="list-style-type: none"> All imaging modalities need to support the same conclusion 	<ul style="list-style-type: none"> Maintain vigilant awareness of the possibility that all modalities need to supply the same conclusion Include changes T2-weighted or FLAIR images as secondary end point Use only MRI with and without gadolinium unless patient unable to undergo MRI Use standard dosing of gadolinium contrast agent and standardize and record timing of scan after dosing⁴ Consider mechanistic imaging whenever possible⁵

Abbreviations: CR, complete response; FLAIR, fluid-attenuated inversion recovery; PD, progressive disease; PR, partial response.