Neuro-Oncology Advance Access published February 9, 2011
Neuro-Oncology
doi:10.1093/neuonc/noq203
NEURO-ONCOLOGY

Clinical trial end points for high-grade glioma: the evolving landscape*

David A. Reardon, Evanthia Galanis, John F. DeGroot, Timothy F. Cloughesy, Jeffrey S. Wefel, Kathleen R. Lamborn, Andrew B. Lassman, Mark R. Gilbert, John H. Sampson, Wolfgang Wick, Marc C. Chamberlain, David R. Macdonald, Minesh P. Mehta, Michael A. Vogelbaum, Susan M. Chang, Martin J. Van den Bent, and Patrick Y. Wen

The Preston Robert Tisch Brain Tumor Center, Duke University Medical Center, Durham, North Carolina (D.A.R., J.H.S.); Department of Medical Oncology, Mayo Clinic, Rochester, Minnesota (E.G.); Department of Neuro-Oncology, MD Anderson Cancer Center, Houston, Texas (J.F.D., J.S.W., M.R.G); Neuro-Oncology Program, David Geffen School of Medicine at UCLA, Los Angeles, California (T.F.C.); Division of Neuro-Oncology, Department of Neurological Surgery, University of California–San Francisco (K.R.L, S.M.C); Department of Neurology and Brain Tumor Center, Memorial Sloan Kettering Cancer Center, New York, New York (A.B.L.); Department of Neurooncology, University of Heidelberg, Heidelberg, Germany (W.W.); Department of Neurology, University of Washington, Seattle (M.C.C.); Department of Oncology and Department of Clinical Neurological Sciences, University of Western Ontario and London Regional Cancer Program, London, Ontario, Canada (D.R.M.); Department of Human Oncology, University of Wisconsin–Madison School of Medicine (M.P.M.), Cleveland Clinic Foundation, Brain Tumor and NeuroOncology Center, Cleveland, Ohio (M.A.V.); Neuro-Oncology Unit, Daniel den Hoed Cancer Center/Erasmus University Hospital Rotterdam, the Netherlands (M.J.V.); Center for Neuro-Oncology, Dana Farber/Brigham and Women's Cancer Center and Division of Neurology, Brigham and Women's Hospital, Boston, Massachusetts (P.Y.W.)

Background: To review the strengths and weaknesses of primary and auxiliary end points for clinical trials among patients with high-grade glioma (HGG).

Methods: Recent advances in outcome for patients with newly diagnosed and recurrent HGG, coupled with the development of multiple promising therapeutics with myriad antitumor actions, have led to significant growth in the number of clinical trials for patients with HGG. Appropriate clinical trial design and the incorporation of optimal end points are imperative to efficiently and effectively evaluate such agents and continue to advance outcome.

Findings: Growing recognition of limitations weakening the reliability of traditional clinical trial primary end points has generated increasing uncertainty of how best to evaluate promising therapeutics for patients with HGG. The phenomena of pseudoprogression and pseudoresponse have made imaging-based end points, including overall radiographic response and progression-free survival, problematic. Although overall survival is considered the "gold-standard" end point, recently identified active salvage therapies such as bevacizumab may diminish the association between presalvage therapy and overall survival. Finally, advances in imaging as well as the assessment of patient function and well being have strengthened interest in auxiliary end points assessing these aspects of patient care and outcome.

Interpretation: Better appreciation of the strengths and limitations of primary end points will lead to more effective clinical trial strategies. Technical advances in imaging as well as improved survival for patients with HGG support the further development of auxiliary end points evaluating novel imaging approaches as well as measures of patient function and well being.

Keywords: clinical trials, malignant glioma, end point.

Received September 15, 2010; accepted November 26, 2010.

*A Report of the Response Assessment in Neuro-Oncology Working Group

Correspondence Author: David A. Reardon, MD, The Preston Robert Tisch Brain Tumor Center at Duke, Duke University Medical Center, Box 3624, Durham, NC 27710 (reard003@mc.duke.edu).

© The Author(s) 2011. Published by Oxford University Press on behalf of the Society for Neuro-Oncology.
All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Downloaded from neuro-oncology.oxfordjournals.org at University of California, Los Angeles on February 17, 2011

The identification of optimal end points for high-grade glioma (HGG) clinical trials has become an increasingly complex challenge, in part because of advances in diagnostic techniques, assessment of sequential imaging, and therapy. Traditional radiologic response guidelines defined 20 years ago that use 2-dimensional measurement of abnormal enhancement as a surrogate of tumor size and activity[1] have become less applicable and reliable in the current therapeutic era, in which temozolomide chemoradiotherapy is routinely used for newly diagnosed patients and angiogenic inhibitors, such as bevacizumab, are administered at recurrence. Greater uncertainty in accurately defining progression due to pseudoprogression and pseudoresponse has, in turn, made the use of progression-free end points more problematic. The recently defined Response Assessment for Neuro-Oncology (RANO) criteria[2] and technological advances in neuroimaging are expected to address some of these challenges. Although overall survival has traditionally been "the gold standard" for HGG clinical trials, active salvage therapies, such as bevacizumab, may impact survival and thus weaken the association between presalvage therapy and survival. As overall survival is improving, metrics of patient function and well being are increasingly valued by patients, caregivers, clinicians, and regulatory agencies and thus warrant consideration as end points of importance.

The RANO working group, consisting of leaders from European and North American neuro-oncology organizations and cooperative groups, includes neuro-oncologists, neurosurgeons, radiation oncologists, neuroradiologists, neuropsychologists, and biostatisticians, who are updating guidelines for evolving and complex clinical scenarios affecting patients with brain tumors. In addition to the criteria for response assessment in HGG, updates for response assessment for low-grade gliomas and for neurosurgical interventions are in preparation. The current article provides a review of the literature as well as consensus recommendations from the RANO working group for commonly employed end points in HGG clinical trials, and it discusses their applicability in the design of clinical trials for newly diagnosed and recurrent patients. We also highlight auxiliary end points that may play a growing role in the design and interpretation of future clinical trials. End points are ultimately relevant in the context of their application to specific clinical trial designs. Faulty study architecture may yield misleading or uninterpretable study results, regardless of selected study end points. Therefore, a separate manuscript dedicated to reviewing key controversies and considerations of clinical trial design for patients with malignant glioma is warranted and planned.

Radiographic Response

As with other solid tumors, the assessment of overall radiographic response (ORR) has been an important end point for HGG clinical trials, especially those that use cytotoxic chemotherapy. Reduction in tumor size may correlate with symptomatic improvement and longer survival and therefore may provide a surrogate for clinical benefit. Twenty years ago, Macdonald et al[3] proposed criteria for ORR in patients with supratentorial HGG based on 2-dimensional measurement of contrast enhancement partly resulting from imaging-based responses of oligodendrogliomas to chemotherapy.[4] Similar to the brain tumor response criteria proposed by Levin et al,[5] Macdonald and colleauges' criteria integrated clinical status into the definitions of both response and progression, as well as corticosteroid use, into the definition of response. The latter was based on the observation that corticosteroid use can substantially decrease tumor contrast enhancement on both CT[6] and MRI scans.[7] Macdonald and colleagues' criteria have subsequently remained the adopted, de facto standard for response assessment in HGG clinical trials. Although a number of studies suggest that 1-dimensional tumor measurements using the Response Evaluation Criteria in Solid Tumors (RECIST) criteria[8] may be equally effective, these criteria have not been widely adopted for patients with brain tumor[9,10] Volumetric imaging may potentially be more accurate but is currently neither widely available nor standardized.[11,12] Finally, change in tumor size as a continuous variable has been advocated as a measure of activity,[13] but this approach also requires validation for patients with HGG.

The use of ORR as a primary end point has several advantages. ORR is a direct measure of therapeutic effect and is not affected by disease natural history, which may affect duration-based end points. It can also be determined relatively rapidly. In addition, there are extensive historical ORR data for a variety of salvage therapies, including cytotoxic chemotherapy[14,15] and bevacizumab-based regimens,[16,17] that can serve as comparators to the activity of planned study therapies. Furthermore, ORR is not affected by subsequent salvage therapies. For these reasons, some have argued that ORR is the only reliable end point for single-arm studies. Finally, regulatory agencies such as the US Food and Drug Administration (FDA) continue to place emphasis on ORR as an end point.[18] In fact, durable ORR was the primary criterion cited by the FDA in their recent accelerated approval of bevacizumab for recurrent GBM.

However, the utility of ORR as a primary end point has potential limitations. First, the ability of ORR to consistently predict meaningful clinical benefit is variable.[10–12,19] One possible explanation is that extravasation of contrast material across leaky intratumoral blood vessels, the surrogate used to determine tumor burden and thus ORR, can be influenced by several factors, including use of corticosteroids, inflammation, seizures, surgery, ischemia, and radiation effect on tumor, normal brain, and vascular permeability.[20–22] Variations in radiological techniques can also influence contrast uptake.[11] Furthermore, numeric cutoffs to define ORR were originally derived from the World Health Organization's solid tumor criteria, which may or may not be valid for CNS tumors. Importantly, ORR does not credit prolonged disease stabilization. Many new therapies, including targeted molecular agents, are

Downloaded from neuro-oncology.oxfordjournals.org at University of California, Los Angeles on February 17, 2011

predominantly cytostatic and are likely to produce stable disease rather than tumor regression. Furthermore, most classical cytotoxic agents do not yield significant ORR rates, yet they may show efficacy in terms of progression-free survival (PFS) or overall survical (OS) when given to patients with newly diagnosed conditions (eg, temozolomide). Such tumor control generally represents meaningful clinical benefit among patients with HGG. For cytostatic agents, other end points such as PFS at 6 months (PFS-6), may better reflect antitumor activity. Finally, radiographic assessment can be challenging because of difficulties in measuring irregularly shaped tumors, interobserver variability, lack of guidelines for assessment of multifocal disease, and the inability to measure nonenhancing tumor.[11] A strategy to offset some of these issues is the use of blinded central imaging review to reduce bias and subjectivity in tumor measurement, thereby improving reliability. However, centralized review may not be practical for many phase II studies and may neglect clinical deterioration.

The utility of ORR based on contrast assessment is further limited by 2 clinically relevant scenarios, particularly in the current era. Although it has been noted after radiotherapy alone,[23] an increase in contrast enhancement and edema, with or without accompanying neurologic findings, referred to as pseudoprogression, has been reported in up to 30% of patients with newly diagnosed HGG following completion of temozolomide chemoradiotherapy.[24–26] In addition, pseudoprogression has been observed with other therapies, including biodegradable chemotherapy polymers,[27] immunotoxins administered by convection-enhanced delivery,[28] viral gene therapy,[29] focal irradiation with brachytherapy or stereotactic radiosurgery,[30] and immunotherapies.[31]Although the underlying mechanism is not fully understood, these changes are felt to reflect increased permeability and/or vascular damage. Unfortunately, no currently available imaging modality reliably distinguishes pseudoprogression from true progression with sufficient sensitivity and specificity for routine use. In such situations, progression can only be reliably defined if it occurs at a distant site, is confirmed histopathologically, or worsens on sequential imaging.

Paradoxically, the second scenario limiting current radiographic assessment is the phenomenon of pseudoresponse. Pseudoresponse refers to improved contrast enhancement due to diminished vascular permeability that does not necessarily reflect a true underlying antitumor effect. For example, rapid and marked improvement in contrast enhancement has been noted less than 1-2 days after antivascular endothelial growth factor (VEGF) therapy.[32] In other cases, despite improved enhancement, progressive, infiltrative nonenhancing tumor as assessed by FLAIR/T2 sequences has been noted.[33]

The recently announced RANO criteria[34] build on the foundation established by the criteria of Levin et al[5] and Macdonald et al[1] and were developed to address the radiographic findings associated with pseudoprogression and pseudoresponse. In addition, pending experience and evaluation of the RANO criteria, concerns remain regarding the interpretation of ORR as an end point for trials investigating therapeutic approaches that are associated with either pseudoprogression or pseudoresponse. Future studies that include an evaluation of the potential correlation of ORR, preferably using the RANO criteria, with OS, will be particularly helpful to better define the value of ORR as an end point in the modern era.

OS

OS is considered the definitive primary end point for patients with both newly diagnosed and recurrent HGG, because it reflects a clinically meaningful benefit that can be objectively and unequivocally assessed. Furthermore, the natural history of HGG does not typically include long periods of disease quiescence, as may occur in low-grade gliomas. Therefore, improvement in OS is believed to directly reflect therapeutic efficacy.

However, the use of OS as a primary end point can have drawbacks. Factors unrelated to current treatment (eg, factors related to underlying disease biology) may affect survival time more than PFS or ORR and may therefore introduce greater variability in outcome. Studies based on OS also inherently require longer assessment time than do other measures, such as PFS or ORR. In addition, the impact of study therapy on OS may be less pronounced than its impact on PFS or ORR, resulting in the need for larger sample sizes. These factors may increase study duration and sample size and, hence, expense. Finally, start times may also vary across studies which can impact time to end point measures including OS. This potential variable highlights the need for studies to clearly specify start times and for clinicians to appropriately take study start time into account when comparing outcome across studies and control groups.

Another important potential drawback of OS as a primary end point is the impact of effective salvage therapies administered at recurrence, which may increase subsequent OS and lead to an apparent disconnect between a salvage therapy and OS improvement. Specifically, given the rapidly expanding use of bevacizumab for patients with HGG and its accelerated approval by the FDA in May 2009 for recurrent GBM, the interpretation of ongoing and future GBM studies will likely need to take into account the potential effect of such salvage therapies on OS. Future studies should consider documenting salvage therapy administered after study failure to better assess their potential impact on study OS. A strategy to help deal with the potentially confounding effect of salvage therapy, particularly for definitive or registration studies, is to include a statistical design that is sufficiently powered to evaluate a second end point that is not affected by salvage therapy.

For single-arm phase II studies that include historical controls, differences between the study and control groups, such as eligibility criteria, start times, or salvage therapy use, can impact end point comparisons including but not limited to OS. In addition, OS may improve over time due to general medical advances,

Downloaded from neuro-oncology.oxfordjournals.org at University of California, Los Angeles on February 17, 2011

such as earlier diagnosis, improved care of tumor symptoms and therapy complications, and better neurosurgical and radiation approaches. These factors coupled with the impact of random biases, such as patient selection, argue strongly for the incorporation of a contemporaneous and comparable randomized control group, particularly for advanced phase II studies.[35,36] Nonetheless, there remains a need for study designs for early, small phase II trials to rapidly evaluate potentially advantageous treatments. For such studies, PFS at a defined time point may be a preferable end point. Although such studies may provide a signal of antitumor activity, definitive conclusions, including the decision to proceed to a pivotal phase III study, will be more appropriately based on a randomized study design.

For exploratory phase II trials designed to determine whether proceeding to a randomized phase III study is appropriate, OS rate measured at a defined time point for which there are established historical data, such as OS at 12 months (OS-12), can shorten trial length while still providing a measure of treatment efficacy. However, required accrual may increase for OS at a specified time point compared to overall median OS. Alternatively, clinical trial designs that incorporate a "run-in" phase II component to evaluate a more rapidly measured event (such as OS-12 or PFS-6) can be performed before transitioning into a randomized phase III trial that incorporates a median OS primary end point. Statistical models exist to permit a seamless shift from phase II to phase III.[37]

PFS

Several metrics have been used to gauge the durability of treatment antitumor activity. PFS measures time from treatment initiation to either progression or death from any cause. PFS end points have several potential strengths. First, the duration of time without tumor progression is usually clinically meaningful and, in general, reliably reflects treatment effect, because HGGs do not typically exhibit prolonged disease inactivity (in contrast to low-grade glioma). Second, PFS end points are not affected by subsequent therapies. Third, PFS-based studies are typically completed in a shorter time frame than OS-based studies because the outcome occurs sooner. They may also require smaller sample sizes because treatment effect is often greater on PFS than OS. Quicker, smaller studies are more feasible, economical, and rational in view of the large number of drugs under development.

Another value of PFS end points among patients with recurrent HGG is that they were recently shown to correlate with OS in 2 independent meta-analyses of cooperative group phase II trials.[38,39] Similarly, recent retrospective data suggest that PFS-6 also correlates with OS among patients with newly diagnosed conditions.[40] However, all of these data were generated prior to the routine use of bevacizumab salvage therapy and defined PD based on enhancing tumor measurement according to the criteria of Macdonald et al.[1] The association of PFS-6 and OS for both patients with newly diagnosed GBM and those with recurrent GBM will require reevaluation among patients treated with anti-VEGF therapy and should include parameters to define PD that assess nonenhancing as well as enhancing disease such as the RANO criteria.[34]

There are potential disadvantages of PFS end points.[18] First and foremost, determination of PFS is critically dependent on a reliable, standardized, objective means to define tumor progression. Such a definition must be quantifiable and able to be confirmed by external review and auditing. Current radiographic methods to define progression are problematic due to confounding factors including pseudoprogression and pseudoresponse, as discussed above. Improved guidelines, such as those defined by the RANO criteria, are expected to lessen the impact of these issues.[19,34] The use of blinded independent central review of PFS can lessen the potential bias involved in determining PFS, but it is also not practical in all trials.[41,42]

Another potential drawback of PFS is evaluation time bias due to variably timed planned and unplanned patient assessments within and across clinical trials. One strategy to circumvent this issue is to assess PFS at a fixed time point. PFS-6 is usually preferred for patients with recurrent HGG, because 6 months is believed to represent a clinically meaningful period. In addition, meta-analysis of PFS-6 across comparable trials can provide a useful historical benchmark.[38,39,43]

Auxilliary end points

This section briefly examines alternative imaging assessments and measures of patient function and well being that may be considered as auxiliary or secondary end points in future clinical trials. Unfortunately, an extensive review of such measures, as well as the potential association of tumor biomarkers and outcome, is beyond the scope of this manuscript.

Alternative Imaging Approaches

Ideally, alternative imaging approaches should provide either an early predictive marker of an established clinical benefit (ORR, PFS, or OS) or a more accurate measure of tumor burden or extent compared to established techniques. However, before implementation, new techniques must be standardized, validated, and made widely available in terms of both access and cost.

Volumetric tumor assessment—Studies comparing 1-, 2- and 3-dimensional measurements and volumetric assessment suggest that there is good concordance between methods in determining radiologic response in adults with newly diagnosed and recurrent HGGs.[9,10] However, given the geometric complexity of both contrast enhancement and T2/FLAIR changes associated with most glial tumors, volumetric evaluations may improve accuracy, whereas computerization of such measures may lessen interobserver variability.[11]

Downloaded from neuro-oncology.oxfordjournals.org at University of California, Los Angeles on February 17, 2011

Assessment of microvascular changes: blood volume, blood flow, and permeability—MRI can be used to define surrogates of vascular effect, such as cerebral blood volume and flow, as well as vascular permeability using techniques such as dynamic susceptibility contrast imaging, arterial spin labelling, and dynamic contrast enhanced MRI .[32,44] A recent analysis suggests that multivariate algorithms of such indices may provide an accurate strategy to predict outcome following anti-angiogenic therapy.[45]

Magnetic resonance spectroscopy—Evaluating major tissue metabolites such as choline, creatinine, *N*-acetyl-acetate, lipids, and lactic acid using magnetic resonance spectroscopy can help define the complex biology of HGGs.[46] Although these data have yet to be successfully integrated into HGG trials, the ability to noninvasively interrogate tissue metabolites over time is attractive as a method to monitor treatment effects.

Imaging cell density—Diffusion-weighted imaging measures the diffusivity of water within brain tissue and can provide information on cell density and characteristics of tissue architecture through either the creation of apparent diffusion coefficient (ADC) maps[47,48] or white matter tract morphology via fractional anisotropy maps.[32] Preliminary analyses suggest that ADC values correlate with response to therapy.[48–50] A recent report also suggests that diffusion MRI may be able to detect nonenhancing tumor progression in patients receiving anti-VEGF therapy.[51]

Positron emission tomography (PET)—PET[52] can monitor glioma therapy using radiolabeled, biologically relevant ligands such as fluorodeoxyglucose, amino acids and fluoro-L-thymidine.[53,54] Each ligand is associated with potential merits and limitations and such approaches offer promise for a wide array of therapeutics.[53,54] For example, in a recent study of patients with recurrent HGG treated with bevacizumab, a $\geq$ 25% reduction in tumor fluoro-L-thymidine uptake was predictive of PFS and OS.[55]

Assessment of Patient Function and Well Being

Importantly, "established" outcome parameters such as ORR, PFS, or OS are truly of clinical benefit only if they stabilize or improve patient function via tumor control and do not worsen patient function or well being due to treatment toxicity. Patient function and well being are particularly meaningful because they directly reflect the patient perspective. Objective assessment of patient function and well being is generally preferred as it reduces self-report biases and other confounders. For example, neurocognitive testing provides an objective approach to evaluating patient function that reflects neurologic integrity.[56] In contrast, subjective self-reporting of neurocognitive status by patients is not well correlated to objective neurocognitive function and is more often associated with symptoms of depression or fatigue.[57] However, some patient experiences (eg, symptoms such as pain, health related quality of life [HRQOL]) require use of a patient reported outcome (PRO) scale.

Several practical concerns warrant consideration as metrics of patient function and well being gain interest as potential clinical trial end points. Effective assessment tools must take into account the heterogeneity and complexity of neurologic manifestations caused by brain tumors and must also strive to minimize subjectivity, bias, and interobserver variability. Ideally, assessments could distinguish the effects of comorbid events (eg, strokes, seizures, and infections) or concurrent use of medications from those of underlying tumor. Just as imaging end points are often aided by the incorporation of clinical data (eg, information regarding comorbid, events such as strokes and seizures, or use of concurrent medications), inclusion of such data in the interpretation of patient assessments can be valuable.[58] Furthermore, because long-term data regarding these measures among survivors are sparse, it may be difficult to distinguish a treatment effect from that of the natural disease course in uncontrolled, noncomparative studies. Other challenges of such end points include the relatively short period to progression in many patients, and the practical limitations posed by assessment frequency. Compliance and missing data can also be critical issues and may limit data analysis.[59] Finally, because better-functioning patients are more likely to participate with such evaluations, attempts should be made to apply assessments broadly across good and poor performance patients.

Objective assessment of these dimensions may help to establish differences in antitumor activity, treatment-related toxicity, and time to neurological deterioration, the latter being of clinical importance, even in the absence of an objective antitumor response. The value of these alternative end points in HGG trials is partly dependent on the phase of the trial and the mechanism of action of the tested agent. For example, interpreting the impact of therapy on neurocognitive function or quality of life in uncontrolled phase II studies of novel agents with unproven activity may be impractical. However, once evidence of antitumor activity has been established, and a randomized phase II or III trial is developed, data from auxiliary end points assessing objective neurocognitive function and subjective patient reported outcomes become increasingly important.

Corticosteroid use—In patients with HGG, corticosteroids are commonly used to decrease vascular permeability, mitigate peritumoral edema, and improve neurologic function. However, long-term corticosteroid use can contribute substantially to morbidity due to impairment of immune function, diminished wound healing, and the development of proximal myopathy, osteoporosis, mood and sleep disorders, hypertension, weight gain and diabetes. Given the negative effects of corticosteroids and the association of their use with tumor burden, it is reasonable to consider a decrease in corticosteroid dose as a clinical benefit.

Downloaded from neuro-oncology.oxfordjournals.org at University of California, Los Angeles on February 17, 2011

Furthermore, ~ 70% of patients with newly diagnosed GBM[60] and 50% of those with recurrent GBM[16] are receiving corticosteroids. These relatively high percentages make it feasible to power a study incorporating the reduction of corticosteroid dose as an end point. However, there is currently no uniform approach to initiating and adjusting corticosteroid dosing among patients with HGG, making the use of corticosteroid adjustment as a reliably measurable end point problematic. Nevertheless, reduction in steroid requirements was a positive effect noted in the recent FDA accelerated approval of bevacizumab for recurrent GBM.[16]

Neurologic status—Even though the neurologic examination is modularized and neurologic scales exist for a number of neurologic disease categories, there is no clear method to define significant neurologic decline for brain tumor trials. Pretreatment patient performance, a widely evaluated metric across many oncology trials, provides a consistent and statistically significant prognostic factor for OS in brain cancer trials.[61,62] Using change in performance status to indicate clinical progression is attractive given the ease and pertinence of its evaluation. However, HGG trials have yet to identify the performance status scale of choice for consistent use across studies. In addition, changes in performance status due to comorbid events or concurrent use of medications must be distinguished from those attributed to underlying tumor progression.

Neurocognitive assessment—Neurocognitive dysfunction is ubiquitous among patients with HGG, is associated with diminished independence in activities of daily living,[63] and can impact HRQOL along with several other factors.[63,64] Furthermore, neurocognitive function can predict prognosis[58,65,66] and may decline in advance of imaging evidence of progression.[66–68] Previous research has demonstrated that assessment of more specific signs (eg, cognitive function) and symptoms (eg, fatigue) in patients with brain tumors are more prognostic than assessment of global constructs such as overall HRQOL.[69] The assessment of neurocognitive function in clinical trials involves a compromise between the desire for sensitive and extensive testing and practical constraints posed by available resources and patient compliance. The absence of neuropsychological expertise at each center limits the extent to which cognition can be thoroughly tested. Therefore, a practical neurocognitive assessment in clinical trials should be brief, sensitive, and repeatable; have known psychometric properties, such as test-retest reliability and published norms; and be able to be administered by trained nonneuropsychologists. In a large multiple-center trial, neurocognitive assessments have demonstrated value.[56]

Clinical trials may consider implementing a small core set of neurocognitive tests for all participating patients or centers and incorporate more extensive testing at select centers with appropriate neuropsychological expertise. A core neurocognitive test battery that is inexpensive and poses low patient burden developed at the MD Anderson Cancer Center has been widely used in several Radiation Therapy Oncology Group trials.[58] The tests have been translated into a limited number of languages for use in multinational trials.

Although assessments of mental status, such as the Mini-Mental Status Examination (MMSE), can be administered quickly and easily, these measures are most useful for evaluating delirium or dementia and are insensitive measures of neurocognitive function in patients with brain tumors.[70] Nonetheless, MMSE assessment was found to be prognostic for OS in some HGG trials[71] and may provide an early cue of tumor progression.[66] Therefore, baseline MMSE can be considered in future trials as a possible stratification factor; however, the longitudinal evaluation of neurocognitive function is more sensitively achieved by formal testing.

Symptoms and HRQOL assessments—Any assessment of a patient's perceived level of function, symptom experience or HRQOL is a PRO, which is inherently subjective in nature. Although HRQOL and self-reported symptom questionnaires are both subjective measures of an individual's health status that come directly from the patient, they are conceptually distinct forms of patient reported outcome (PROs). In terms of PRO measures, symptom questionnaires have been described as a preferred approach to HRQOL questionnaires to determine the impact of disease and treatment in a clinical trial as symptoms are thought to be closer to the biologic and physiologic factors impacting the patient than HRQOL. Specific symptom questionnaires (eg, Brief Fatigue Inventory[52]) as well as brain tumor specific multisymptom questionnaires (M.D. Anderson Symptom Inventory-Brain Tumor Module[72]) have been developed.

HRQOL has been defined as "the physical, psychological, and social domains of health, seen as distinct areas that are influenced by a person's experiences, beliefs, expectations, and perceptions."[73] Beliefs and expectations regarding health and the ability to cope with limitations and disability can greatly affect a person's perception of health, satisfaction with life and reporting on HRQOL questionnaires. Susceptibility to response shift over time and other biases further limit the use of HRQOL questionnaires to assess clinical benefit in clinical trials. Nonetheless, validated HRQOL questionnaires for patients with brain tumors can assess the physical, psychological, and social impact of the disease and its treatment.[74] The general Functional Assessment of Cancer Therapy questionnaire combined with the Brain Tumor module (FACT-BR) addresses physical, social, emotional, and functional well being to provide a broad view of HRQOL issues with robust psychometric properties. Similarly, the QLQ-C30 developed by the EORTC as a general oncology questionnaire, which interrogates 5 main functional areas and provides symptom scales for fatigue, nausea/vomiting, and pain can be combined with a brain-tumor-specific questionnaire (BN20), which provides additional queries of neurologic and treatment parameters.

Downloaded from neuro-oncology.oxfordjournals.org at University of California, Los Angeles on February 17, 2011

Discussion

Recent clinical trials have identified therapeutic agents and strategies that meaningfully improve outcome for both newly diagnosed and recurrent patients with HGG. On the basis of laboratory advances that have significantly enhanced our understanding of cancer biology, a large number of therapeutic agents with diverse mechanisms of action are in clinical development. It is therefore now more important than ever to recognize the strengths and potential pitfalls associated with various clinical trial end points to evaluate these agents as efficiently as possible. Although OS remains the gold standard for phase III trials and most trials involving newly diagnosed patients, some of these trials may be enhanced by an additional end point that is not affected by subsequent salvage therapy, such as PFS at a defined interval. For most phase II studies, particularly for those involving patients with recurrent conditions, median OS and PFS at a defined time interval represent appropriate primary end points. Imaging response remains of value for selected situations but is potentially confounded by several factors including pseudoprogression and pseudoresponse. Application of the RANO criteria[2] may lessen the impact of these factors and improve assessment of ORR as well as defining progression. Given the limitations of commonly used imaging assessment techniques, the development of alternative imaging approaches as secondary end points for clinical trials is encouraged. Finally, validated metrics of patient function and well being are available and increasingly valued as secondary end points.

Summary of considerations for selection of clinical trial end points for HGG—1. End points of therapeutic clinical trials should be clearly defined and based on a rationale statistical hypothesis.

2. Although ORR is an important end point, its reliability is problematic in studies evaluating agents or regimens with cytostatic potential or that are associated with either pseudoprogression or pseudoresponse, particularly anti-VEGF therapies.

3. Metrics of clinical and radiographic evaluation should be standardized and should incorporate the recently defined RANO criteria.

4. Overall survival remains the gold-standard; however, if potentially compromised by salvage therapy, an additional end point of outcome should be considered that is independent of salvage therapy, such as ORR or PFS.

5. FS, including assessment at a fixed time point, is an appropriate end point for most phase II studies, especially among patients with recurrent conditions. Correlation of PFS end points with OS should be further explored in the current era and ideally validated with the RANO criteria. PFS end points can enhance studies with an OS primary end point if subsequent salvage therapy may affect OS.

6. To optimize study end point value, appropriate comparison groups should be used. Although concurrent controls are preferable in general, a historical comparison group may be considered but should reflect patients with comparable eligibility, chronological era of care, study start times, and subsequent salvage therapy.

7. Objective tests of neurocognitive function and subjective patient reported outcome measures are highly valued and should be practical, reliable, and validated in controlled study designs.

Acknowledgements

We gratefully acknowledge Wendy Gentry for the preparation of this manuscript.

Conflict of interest statement. D.A.R. served on advisory boards for Genentech, Merck/S.P., Bayer, AstraZeneca, and Genzyme and consulted for EMD Serono and Genentech. T.F.C consulted for Genentech, AstraZeneca, Exelixis, Imclone, and Roche. A.B.L consulted for Bristol Meyers Squibb, Campus Bio, Cephalon, Eisai, Genentech, Imclone, and Merck/S.P. and served on speaker's bureau for Merck/S.P. M.R.G. consulted for Genentech and Merck/S.P. J.H.S. consulted for BRAINlab. M.C.C. served as an advisor for Genentech, Merck/S.P., Exelixix and consulted for Genentech, and Merck/S.P. M.P.M. has stock in Colby, Pharmacyclins, Procertus, Stemina, and Tomotherapy; served as an advisor for Colby, Stemina, and Procertus; served on the board of directors for Pharmacyclics; served on data safety monitoring board for Apogenix; consulted for Adnexus, Bayer, Merck, Roche, and Tomotherapy; and served as a speaker for Merck. M.A.V. consulted for AstraZeneca, Merck/S.P., BRAINlab, and Genentech; served on the Editorial Board for Neuro-Oncology and Secretary/Treasurer for Society of Neuro-Oncology. All other authors: none declared.

References

1. Macdonald DR, Cascino TL, Schold SC, Jr., Cairncross JG. Response criteria for phase II studies of supratentorial malignant glioma. *J Clin Oncology*. 1990;8(7):1277–1280.
2. Wen PY, Macdonald DR, Reardon DA, et al. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *J Clin Oncol*. 2010;28(11):1963–1972.
3. Macdonald DR, Cascino TL, Schold SC, Jr., Cairncross JG. Response criteria for phase II studies of supratentorial malignant glioma. *J Clin Oncol*. 1990;8(7):1277–1280.
4. Cairncross JG, Macdonald DR. Successful chemotherapy for recurrent malignant oligodendroglioma. *Ann Neurol*. 1988;23(4):360–364.

Downloaded from neuro-oncology.oxfordjournals.org at University of California, Los Angeles on February 17, 2011

5. Levin VA, Crafts DC, Norman DM, Hoffer PB, Spire JP, Wilson CB. Criteria for evaluating patients undergoing chemotherapy for malignant brain tumors. *J Neurosurg.* 1977;47(3):329–335.
6. Cairncross JG, Macdonald DR, Pexman JH, Ives FJ. Steroid-induced CT changes in patients with recurrent malignant glioma. *Neurology.* 1988;38(5):724–726.
7. Watling CJ, Lee DH, Macdonald DR, Cairncross JG. Corticosteroid-induced magnetic resonance imaging changes in patients with recurrent malignant glioma. *J Clin Oncol.* 1994;12(9):1886–1889.
8. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer.* 2009;45(2):228–247.
9. Galanis E, Buckner JC, Maurer MJ, 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-oncology.* 2006;8(2):156–165.
10. Shah GD, Kesari S, Xu R, et al. Comparison of linear and volumetric criteria in assessing tumor response in adult high-grade gliomas. *Neuro-oncology.* 2006;8(1):38–46.
11. Sorensen AG, Batchelor TT, Wen PY, Zhang WT, Jain RK. Response criteria for glioma. *Nat Clin Pract Oncol.* 2008;5(11):634–644.
12. Henson JW, Ulmer S, Harris GJ. Brain tumor imaging in clinical trials. *AJNR Am J Neuroradiol.* 2008;29(3):419–424.
13. Karrison TG, Maitland ML, Stadler WM, Ratain MJ. Design of phase II cancer trials using a continuous endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non small-cell lung cancer. *J Natl Cancer Inst.* 2007;99(19):1455–1461.
14. Yung WK, Albright RE, Olson J, et al. A phase II study of temozolomide vs. procarbazine in patients with glioblastoma multiforme at first relapse. *Br J Cancer.* 2000;83(5):588–593.
15. Yung WK, Prados MD, Yaya-Tur R, et al. Multicenter phase II trial of temozolomide in patients with anaplastic astrocytoma or anaplastic oligoastrocytoma at first relapse. Temodal Brain Tumor Group. *J Clin Oncol.* 1999;17(9):2762–2771.
16. Friedman HS, Prados MD, Wen PY, et al. Bevacizumab alone and in combination with irinotecan in recurrent glioblastoma. *J Clin Oncol.* 2009;27(28):4733–4740.
17. Kreisl TN, Kim L, Moore K, et al. Phase II trial of single-agent bevacizumab followed by bevacizumab plus irinotecan at tumor progression in recurrent glioblastoma. *J Clin Oncol.* 2009;27(5): 740–745.
18. FDA. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics: US Department of Health and Human Services, 2007.
19. van den Bent MJ, Vogelbaum MA, Wen PY, Macdonald DR, Chang SM. End point assessment in gliomas: novel treatments limit usefulness of classical Macdonald's Criteria. *J Clin Oncol.* 2009;27(18): 2905–2908.
20. Henegar MM, Moran CJ, Silbergeld DL. Early postoperative magnetic resonance imaging following nonneoplastic cortical resection. *J Neurosurg.* 1996;84(2):174–179.
21. Kumar AJ, Leeds NE, Fuller GN, et al. Malignant gliomas: MR imaging spectrum of radiation therapy- and chemotherapy-induced necrosis of the brain after treatment. *Radiology.* 2000;217(2):377–384.
22. Finn MA, Blumenthal DT, Salzman KL, Jensen RL. Transient postictal MRI changes in patients with brain tumors may mimic disease progression. *Surg Neurol.* 2007;67(3):246–250; discussion 250.
23. de Wit MC, de Bruin HG, Eijkenboom W, Sillevis Smitt PA, van den Bent MJ. Immediate post-radiotherapy changes in malignant glioma can mimic tumor progression. *Neurology.* 2004;63(3):535–537.
24. Brandsma D, Stalpers L, Taal W, Sminia P, van den Bent MJ. Clinical features, mechanisms, and management of pseudoprogression in malignant gliomas. *Lancet Oncol.* 2008;9(5):453–461.
25. Chamberlain MC, Glantz MJ, Chalmers L, Van Horn A, Sloan AE. Early necrosis following concurrent Temodar and radiotherapy in patients with glioblastoma. *J Neurooncol.* 2007;82(1):81–83.
26. Brandes AA, Tosoni A, Spagnolli F, et al. Disease progression or pseudoprogression after concomitant radiochemotherapy treatment: pitfalls in neurooncology. *Neuro-oncology.* 2008;10(3):361–367.
27. Kleinberg LR, Weingart J, Burger P, et al. Clinical course and pathologic findings after Gliadel and radiotherapy for newly diagnosed malignant glioma: implications for patient management. *Cancer Invest.* 2004; 22(1):1–9.
28. Parney IF, Kunwar S, McDermott M, et al. Neuroradiographic changes following convection-enhanced delivery of the recombinant cytotoxin interleukin 13-PE38QQR for recurrent malignant glioma. *J Neurosurgery.* 2005;102(2):267–275.
29. Aghi M, Chiocca EA. Gene therapy for glioblastoma. *Neurosurg Focus.* 2006;20(4):E18.
30. Shrieve DC, Alexander E, 3rd, Wen PY, et al. Comparison of stereotactic radiosurgery and brachytherapy in the treatment of recurrent glioblastoma multiforme. *Neurosurgery.* 1995;36(2):275–282; discussion 282–274.
31. Yu JS, Wheeler CJ, Zeltzer PM, et al. Vaccination of malignant glioma patients with peptide-pulsed dendritic cells elicits systemic cytotoxicity and intracranial T-cell infiltration. *Cancer Res.* 2001;61(3):842–847.
32. Batchelor TT, Sorensen AG, di Tomaso E, et al. AZD2171, a Pan-VEGF receptor tyrosine kinase inhibitor, normalizes tumor vasculature and alleviates edema in glioblastoma patients. *Cancer Cell.* 2007; 11(1):83–95.
33. Norden AD, Young GS, Setayesh K, et al. Bevacizumab for recurrent malignant gliomas: efficacy, toxicity, and patterns of recurrence. *Neurology.* 2008;70(10):779–787.
34. Wen P, Macdonald D, Reardon D, et al. Updated Response Assessment Criteria For High-Grade Gliomas: Response Assessment in Neuro-Oncology (RANO) Working Group. *J Clin Oncol.* 2010;28(11): 1963–1972.
35. Grossman SA, Ye X, Piantadosi S, et al. Current survival statistics for patients with newly diagnosed glioblastoma treated with radiation and temozolomide on research studies in the United States. Paper presented at: *Proc Am Soc Clin Oncol* 2009.
36. Rubinstein L, Crowley J, Ivy P, Leblanc M, Sargent D. Randomized phase II designs. *Clin Cancer Res.* 2009;15(6):1883–1890.
37. Inoue LY, Thall PF, Berry DA. Seamlessly expanding a randomized phase II trial to phase III. *Biometrics.* 2002;58(4):823–831.
38. Ballman KV, Buckner JC, Brown PD, 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-oncology.* 2007;9(1):29–38.
39. Lamborn KR, Yung WK, Chang SM, et al. Progression-free survival: An important end point in evaluating therapy for recurrent high-grade gliomas. *Neuro-oncology.* 2008;10(2):162–170.
40. Polley MY, Lamborn K, Chang S, Butowski N, Clarke J, Prados M. Six-Month Progression-Free Survival as an Alternative Primary Efficacy Endpoint to Overall Survival in Newly-Diagnosed Glioblastoma Patients Receiving Temozolomide. *Neuro-oncology.* 2010.
41. Dodd LE, Korn EL, Freidlin B, et al. Blinded independent central review of progression-free survival in phase III clinical trials: important design element or unnecessary expense? *J Clin Oncol.* 2008;26(22): 3791–3796.

Downloaded from neuro-oncology.oxfordjournals.org at University of California, Los Angeles on February 17, 2011

42. Freidlin B, Korn EL, Hunsberger S, Gray R, Saxman S, Zujewski JA. Proposal for the use of progression-free survival in unblinded randomized trials. *J Clin Oncol.* 20 2007;25(15):2122–2126.
43. Wong ET, Hess KR, Gleason MJ, et al. Outcomes and prognostic factors in recurrent glioma patients enrolled onto phase II clinical trials. *J Clin Oncol.* 1999;17(8):2572–2578.
44. Jackson A, Jayson GC, Li KL, et al. Reproducibility of quantitative dynamic contrast-enhanced MRI in newly presenting glioma. *Br J Radiol.* 2003;76(903):153–162.
45. Sorensen AG, Batchelor TT, Zhang WT, et al. A "vascular normalization index" as potential mechanistic biomarker to predict survival after a single dose of cediranib in recurrent glioblastoma patients. *Cancer Res.* 2009;69(13):5296–5300.
46. Soares DP, Law M. Magnetic resonance spectroscopy of the brain: review of metabolites and clinical applications. *Clin Radiol.* 2009;64(1):12–21.
47. Higano S, Yun X, Kumabe T, et al. Malignant astrocytic tumors: clinical importance of apparent diffusion coefficient in prediction of grade and prognosis. *Radiology.* 2006;241(3):839–846.
48. Moffat BA, Chenevert TL, Lawrence TS, et al. Functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response. *Proc Natl Acad Sci U S A.* 2005;102(15):5524–5529.
49. Pope WB, Kim HJ, Huo J, et al. Recurrent glioblastoma multiforme: ADC histogram analysis predicts response to bevacizumab treatment. *Radiology.* 2009;252(1):182–189.
50. Jain R, Scarpace LM, Ellika S, et al. Imaging response criteria for recurrent gliomas treated with bevacizumab: role of diffusion weighted imaging as an imaging biomarker. *J Neurooncol.* 2009;96(3):423–431.
51. Gerstner ER, Frosch MP, Batchelor TT. Diffusion magnetic resonance imaging detects pathologically confirmed, nonenhancing tumor progression in a patient with recurrent glioblastoma receiving bevacizumab. *J Clin Oncol.* 2009;28(6):e91–93.
52. Mendoza TR, Wang XS, Cleeland CS, et al. The rapid assessment of fatigue severity in cancer patients: use of the Brief Fatigue Inventory. *Cancer.* 1999;85(5):1186–1196.
53. Chawla S, Poptani H, Melhem ER. Anatomic, physiologic and metabolic imaging in neuro-oncology. *Cancer treatment and research.* 2008;143:3–42.
54. Chen W. Clinical applications of PET in brain tumors. *J Nucl Med.* 2007;48(9):1468–1481.
55. Chen W, Delaloye S, Silverman DH, et al. Predicting treatment response of malignant gliomas to bevacizumab and irinotecan by imaging proliferation with [18F] fluorothymidine positron emission tomography: a pilot study. *J Clin Oncol.* 2007;25(30):4714–4721.
56. Meyers CA, Brown PD. Role and relevance of neurocognitive assessment in clinical trials of patients with CNS tumors. *J Clin Oncol.* 10 2006;24(8):1305–1309.
57. Armstrong TS, Gilbert MR. Net clinical benefit: functional endpoints in brain tumor clinical trials. *Current oncology reports.* 2007;9(1):60–65.
58. Meyers CA, Smith JA, Bezjak A, et al. Neurocognitive function and progression in patients with brain metastases treated with whole-brain radiation and motexafin gadolinium: results of a randomized phase III trial. *J Clin Oncol.* 2004;22(1):157–165.
59. Walker M, Brown J, Brown K, Gregor A, Whittle IR, Grant R. Practical problems with the collection and interpretation of serial quality of life assessments in patients with malignant glioma. *J Neurooncol.* 2003;63(2):179–186.
60. Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med.* 2005;352(10):987–996.
61. Buckner JC. Factors influencing survival in high-grade gliomas. *Semin Oncol.* 2003;30(6 Suppl 19):10–14.
62. Compostella A, Tosoni A, Blatt V, Franceschi E, Brandes AA. Prognostic factors for anaplastic astrocytomas. *J Neurooncol.* 2007;81(3):295–303.
63. Li J, Bentzen SM, Li J, Renschler M, Mehta MP. Relationship between neurocognitive function and quality of life after whole-brain radiotherapy in patients with brain metastasis. *Int J Radiat Oncol Biol Phys.* 2008;71(1):64–70.
64. Giovagnoli AR, Silvani A, Colombo E, Boiardi A. Facets and determinants of quality of life in patients with recurrent high grade glioma. *Journal of neurology, neurosurgery, and psychiatry.* 2005;76(4):562–568.
65. Meyers CA, Hess KR, Yung WK, Levin VA. Cognitive function as a predictor of survival in patients with recurrent malignant glioma. *J Clin Oncol.* 2000;18(3):646–650.
66. Brown PD, Jensen AW, Felten SJ, et al. Detrimental effects of tumor progression on cognitive function of patients with high-grade glioma. *J Clin Oncol.* 2006;24(34):5427–5433.
67. Meyers CA, Hess KR. Multifaceted end points in brain tumor clinical trials: cognitive deterioration precedes MRI progression. *Neuro-oncology.* 2003;5(2):89–95.
68. Armstrong CL, Goldstein B, Shera D, Ledakis GE, Tallent EM. The predictive value of longitudinal neuropsychologic assessment in the early detection of brain tumor recurrence. *Cancer.* 2003;97(3):649–656.
69. Gotay CC, Kawamoto CT, Bottomley A, Efficace F. The prognostic significance of patient-reported outcomes in cancer clinical trials. *J Clin Oncol.* 2008;26(8):1355–1363.
70. Meyers CA, Wefel JS. The use of the mini-mental state examination to assess cognitive functioning in cancer trials: no ifs, ands, buts, or sensitivity. *J Clin Oncol.* 2003;21(19):3557–3558.
71. van den Bent MJ, Carpentier AF, Brandes AA, et al. Adjuvant procarbazine, lomustine, and vincristine improves progression-free survival but not overall survival in newly diagnosed anaplastic oligodendrogliomas and oligoastrocytomas: a randomized European Organisation for Research and Treatment of Cancer phase III trial. *J Clin Oncol.* 2006;24(18):2715–2722.
72. Armstrong TS, Mendoza T, Gning I, et al. Validation of the M.D. Anderson Symptom Inventory Brain Tumor Module (MDASI-BT). *J Neurooncol.* 2006;80(1):27–35.
73. Testa MA, Simonson DC. Assesment of quality-of-life outcomes. *N Engl J Med.* 1996;334(13):835–840.
74. Mauer ME, Bottomley A, Taphoorn MJ. Evaluating health-related quality of life and symptom burden in brain tumour patients: instruments for use in experimental trials and clinical practice. *Curr Opin Neurol.* 2008;21(6):745–753.

Downloaded from neuro-oncology.oxfordjournals.org at University of California, Los Angeles on February 17, 2011